Gaussian Processes

Neil D. Lawrence

GPRS
25th–27th February 2015

Outline

Gaussian Processes

Multiple Output Processes

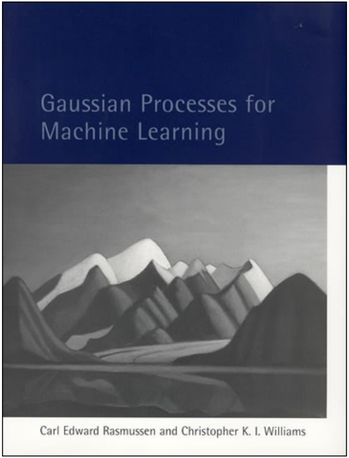
Approximations

Dimensionality Reduction

Latent Force Models

Outline

Book



Rasmussen and Williams (2006)

What is Machine Learning?

data

- data: observations, could be actively or passively acquired (meta-data).

What is Machine Learning?

data +

- data: observations, could be actively or passively acquired (meta-data).

What is Machine Learning?

data + model

- data: observations, could be actively or passively acquired (meta-data).
- model: assumptions, based on previous experience (other data! transfer learning etc), or beliefs about the regularities of the universe. Inductive bias.

What is Machine Learning?

data + model =

- data: observations, could be actively or passively acquired (meta-data).
- model: assumptions, based on previous experience (other data! transfer learning etc), or beliefs about the regularities of the universe. Inductive bias.

What is Machine Learning?

data + model = prediction

- data: observations, could be actively or passively acquired (meta-data).
- model: assumptions, based on previous experience (other data! transfer learning etc), or beliefs about the regularities of the universe. Inductive bias.
- prediction: an action to be taken or a categorization or a quality score.

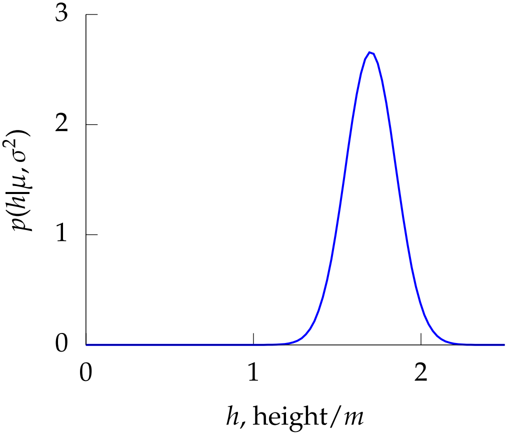
The Gaussian Density

- Perhaps the most common probability density.

$$
\begin{aligned}
p(y|\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)\\
&\triangleq \mathcal{N}\left(y|\mu,\sigma^2\right)
\end{aligned}
$$

- The Gaussian density.

Gaussian Density



The Gaussian PDF with $\mu = 1.7$ and variance $\sigma^2 = 0.0225$. Mean shown as red line. It could represent the heights of a population of students.

Gaussian Density

$$\mathcal{N}\left(y|\mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

σ^2 is the variance of the density and μ is the mean.

Two Important Gaussian Properties

Sum of Gaussians

- Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

Two Important Gaussian Properties

Sum of Gaussians

- Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

And the sum is distributed as

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Two Important Gaussian Properties

Sum of Gaussians

- Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

And the sum is distributed as

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

(*Aside*: As sum increases, sum of non-Gaussian, finite variance variables is also Gaussian [central limit theorem].)

Two Important Gaussian Properties

Sum of Gaussians

- Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

And the sum is distributed as

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

(*Aside*: As sum increases, sum of non-Gaussian, finite variance variables is also Gaussian [central limit theorem].)

Two Important Gaussian Properties

Scaling a Gaussian

- Scaling a Gaussian leads to a Gaussian.

Two Important Gaussian Properties

Scaling a Gaussian

- Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

Two Important Gaussian Properties

Scaling a Gaussian

- Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

And the scaled density is distributed as

$$wy \sim \mathcal{N}\left(w\mu, w^2\sigma^2\right)$$

Two Dimensional Gaussian

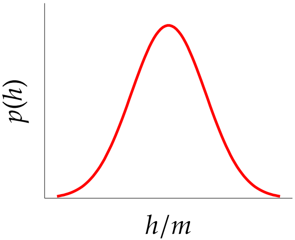
- Consider height, h/m and weight, w/kg.
- Could sample height from a distribution:

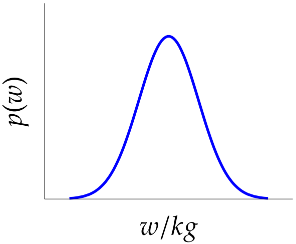
$$p(h) \sim \mathcal{N}(1.7, 0.0225)$$

- And similarly weight:

$$p(w) \sim \mathcal{N}(75, 36)$$

Height and Weight Models

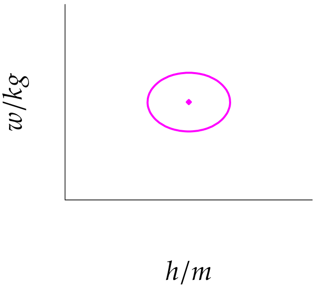




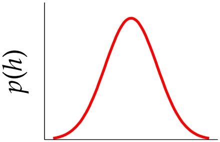
Gaussian distributions for height and weight.

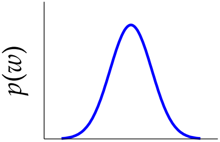
Sampling Two Dimensional Variables

Joint Distribution



Marginal Distributions

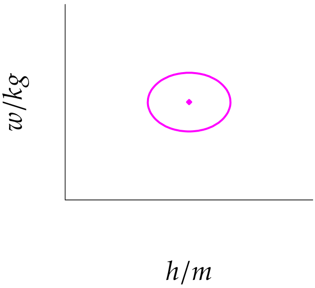




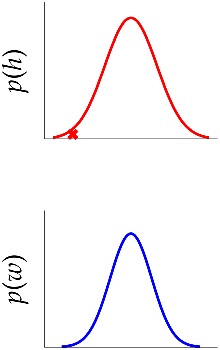
Samples of height and weight

Sampling Two Dimensional Variables

Joint Distribution



Marginal Distributions

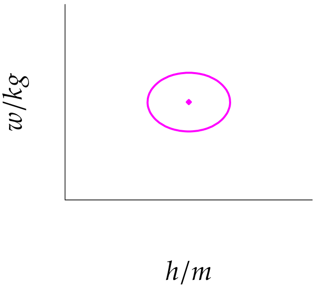


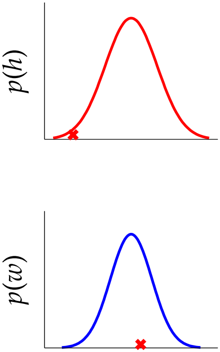
Samples of height and weight

Sampling Two Dimensional Variables

Marginal Distributions

Joint Distribution



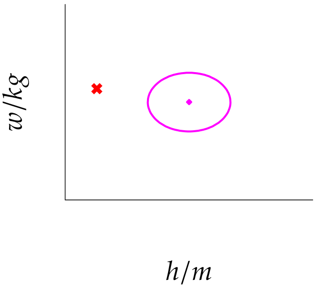


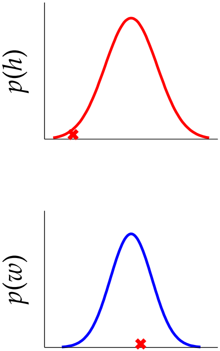
Samples of height and weight

Sampling Two Dimensional Variables

Marginal Distributions

Joint Distribution



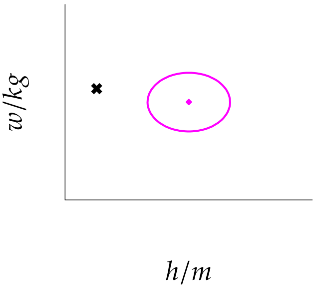


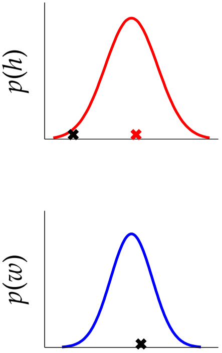
Samples of height and weight

Sampling Two Dimensional Variables

Marginal Distributions

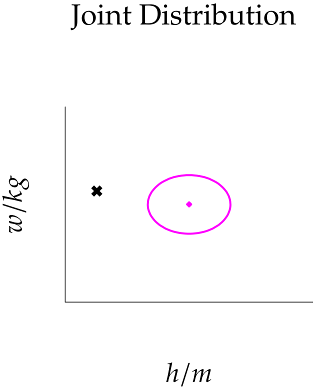
Joint Distribution

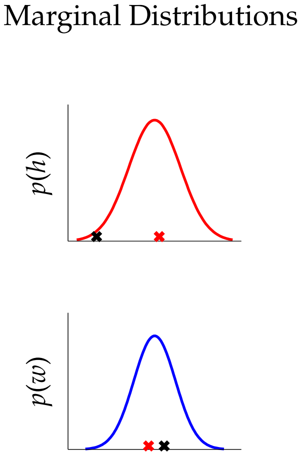




Samples of height and weight

Sampling Two Dimensional Variables

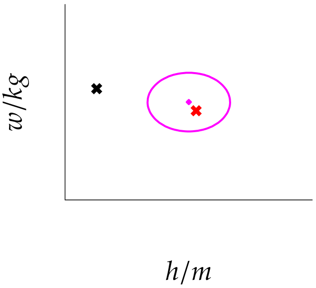




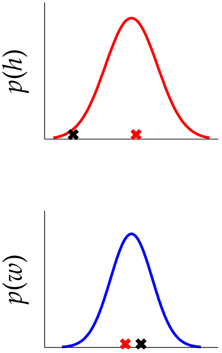
Samples of height and weight

Sampling Two Dimensional Variables

Joint Distribution



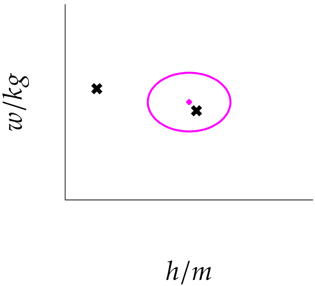
Marginal Distributions



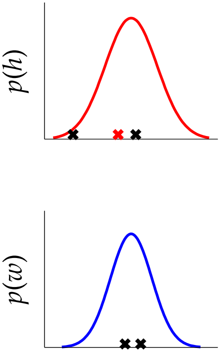
Samples of height and weight

Sampling Two Dimensional Variables

Joint Distribution

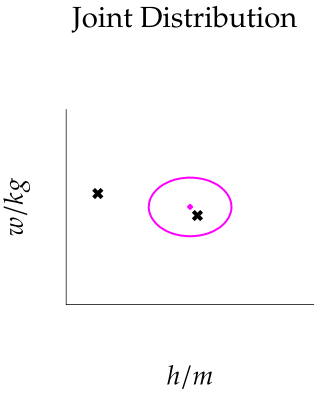


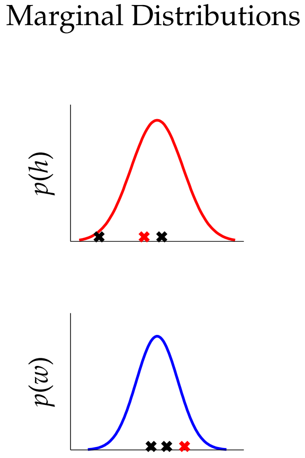
Marginal Distributions



Samples of height and weight

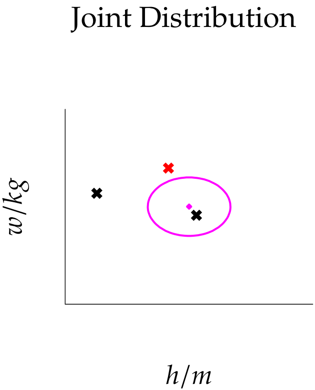
Sampling Two Dimensional Variables

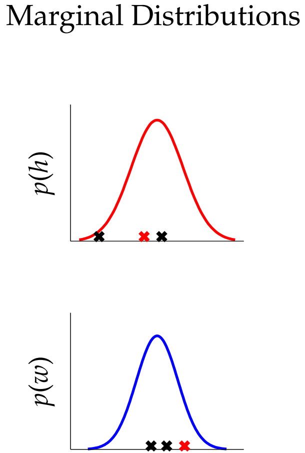




Samples of height and weight

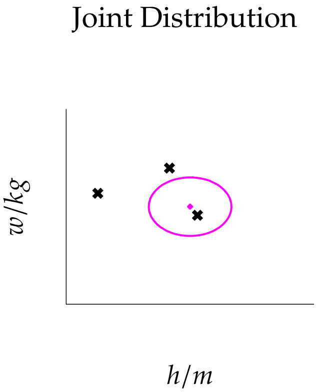
Sampling Two Dimensional Variables

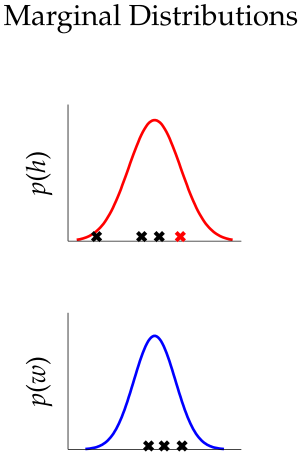




Samples of height and weight

Sampling Two Dimensional Variables

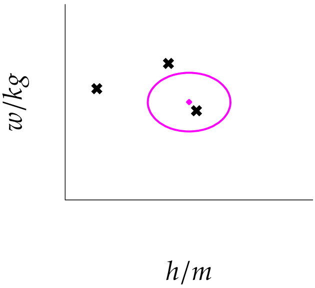




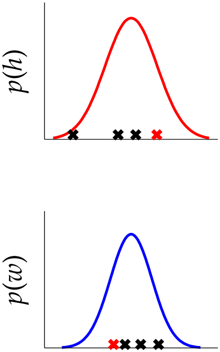
Samples of height and weight

Sampling Two Dimensional Variables

Joint Distribution

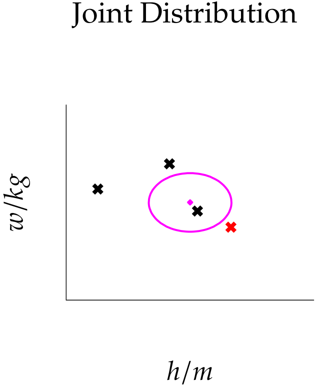


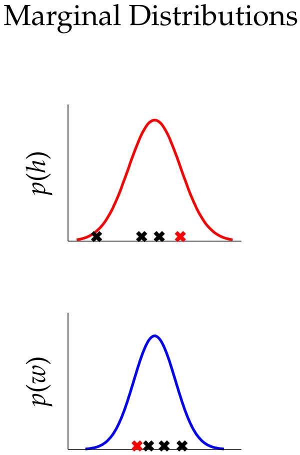
Marginal Distributions



Samples of height and weight

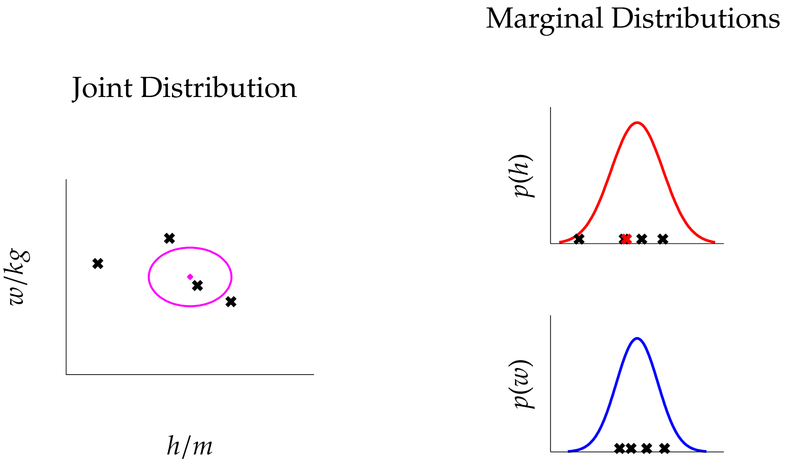
Sampling Two Dimensional Variables





Samples of height and weight

Sampling Two Dimensional Variables

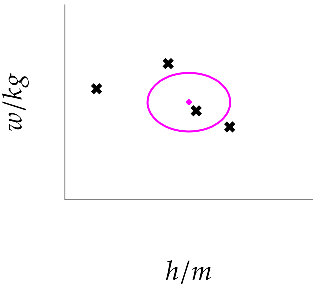


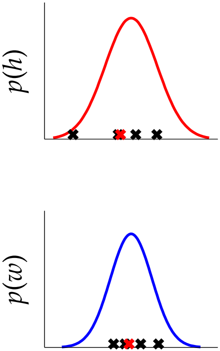
Samples of height and weight

Sampling Two Dimensional Variables

Marginal Distributions

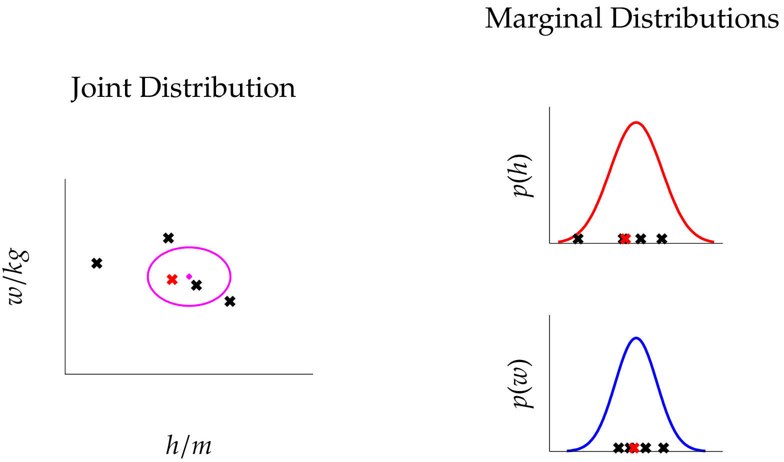
Joint Distribution





Samples of height and weight

Sampling Two Dimensional Variables

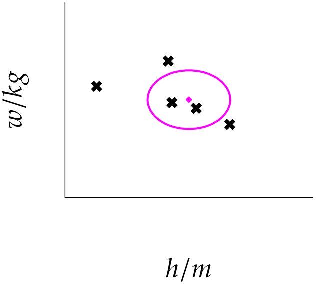


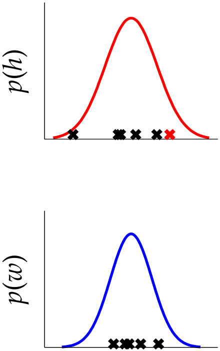
Samples of height and weight

Sampling Two Dimensional Variables

Marginal Distributions

Joint Distribution



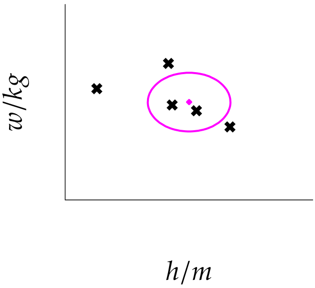


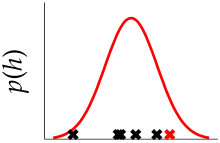
Samples of height and weight

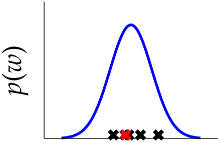
Sampling Two Dimensional Variables

Marginal Distributions

Joint Distribution



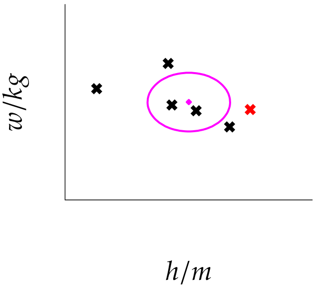




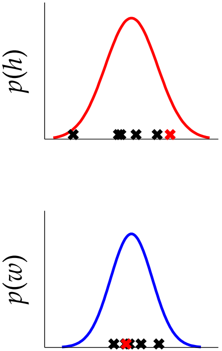
Samples of height and weight

Sampling Two Dimensional Variables

Joint Distribution

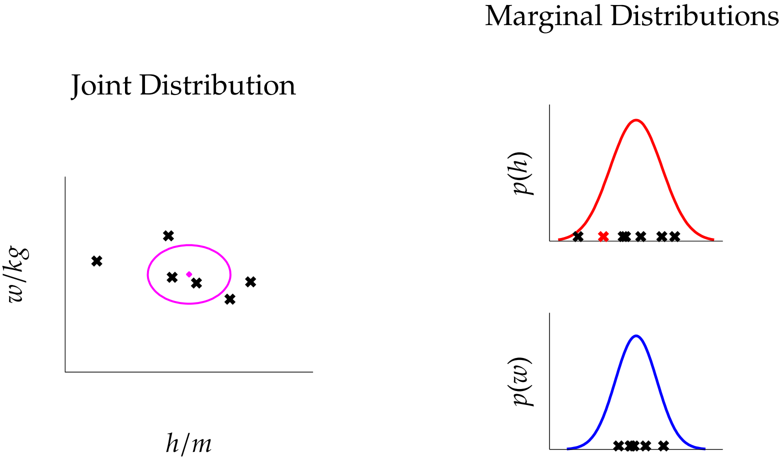


Marginal Distributions



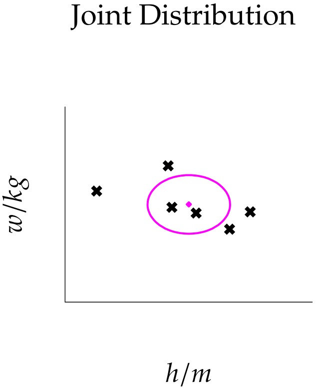
Samples of height and weight

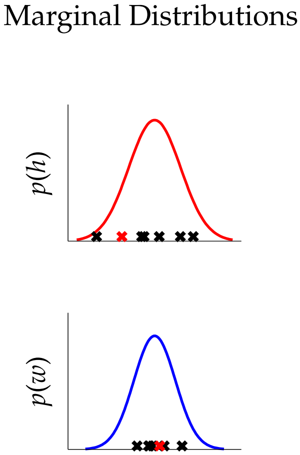
Sampling Two Dimensional Variables



Samples of height and weight

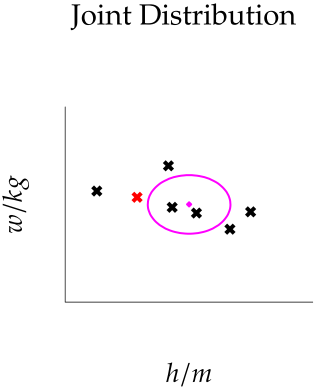
Sampling Two Dimensional Variables

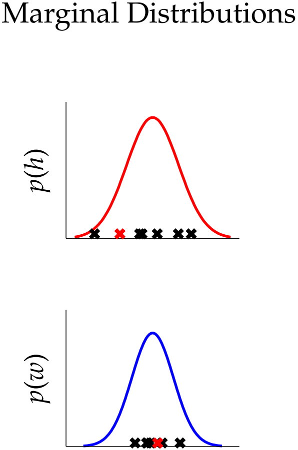




Samples of height and weight

Sampling Two Dimensional Variables

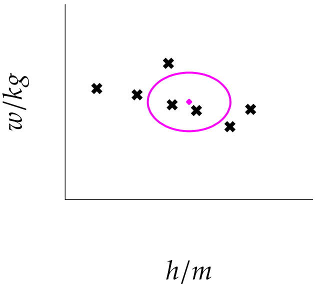




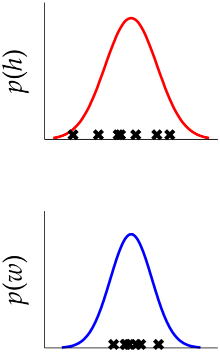
Samples of height and weight

Sampling Two Dimensional Variables

Joint Distribution



Marginal Distributions



Samples of height and weight

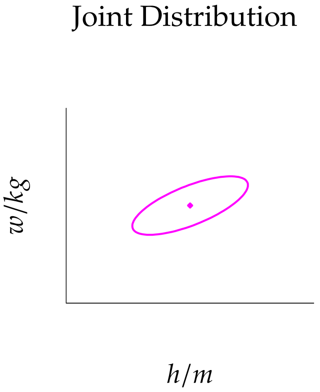
Independence Assumption

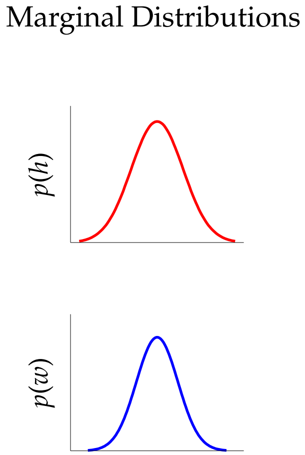
- This assumes height and weight are independent.

$$p(h, w) = p(h)p(w)$$

- In reality they are dependent (body mass index) $= \frac{w}{h^2}$.

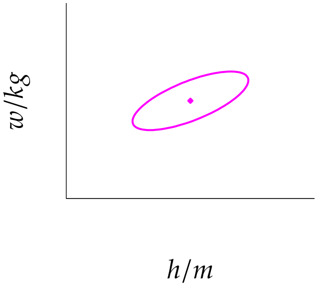
Sampling Two Dimensional Variables



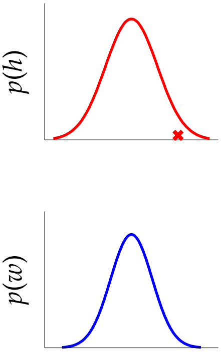


Sampling Two Dimensional Variables

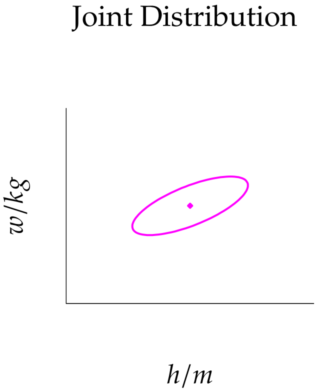
Joint Distribution

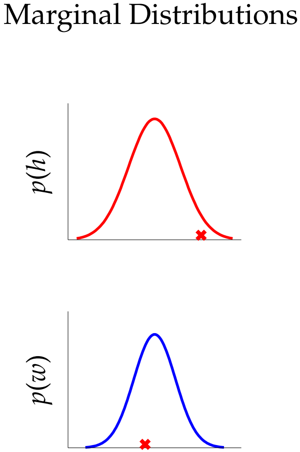


Marginal Distributions



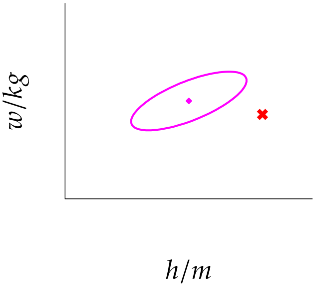
Sampling Two Dimensional Variables



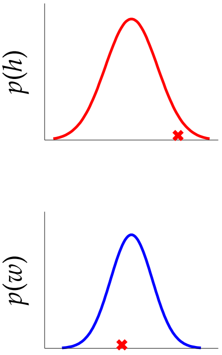


Sampling Two Dimensional Variables

Joint Distribution

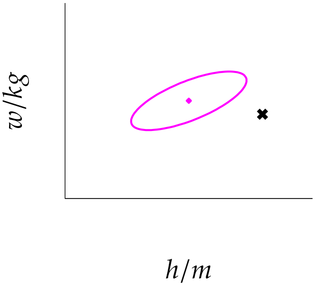


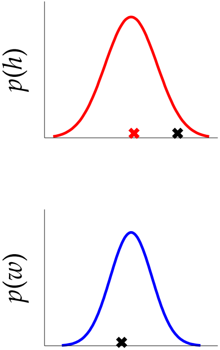
Marginal Distributions



Marginal Distributions

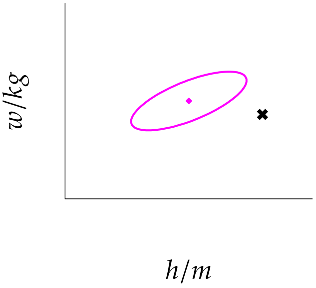
Joint Distribution



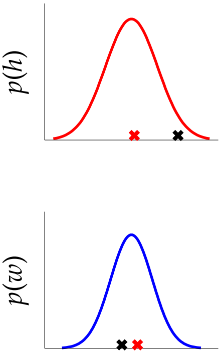


Sampling Two Dimensional Variables

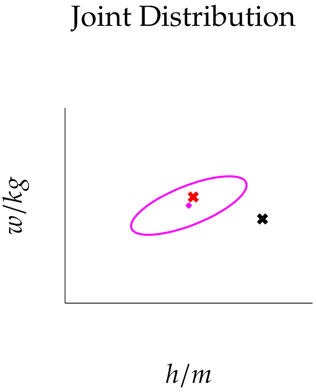
Joint Distribution

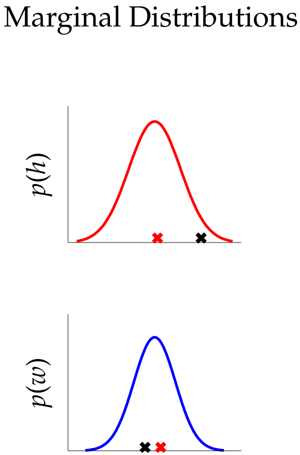


Marginal Distributions



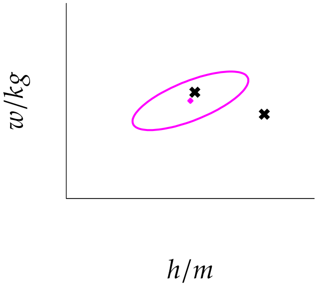
Sampling Two Dimensional Variables



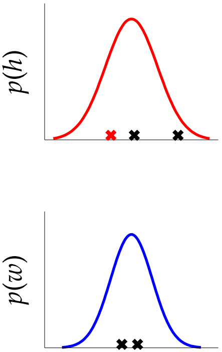


Sampling Two Dimensional Variables

Joint Distribution

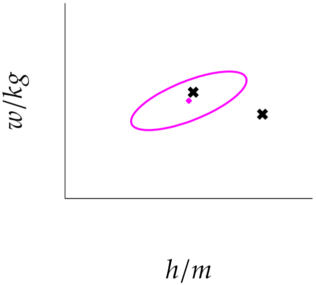


Marginal Distributions

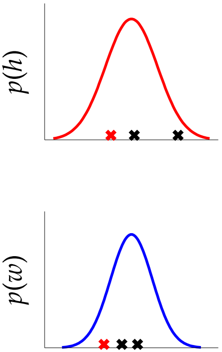


Sampling Two Dimensional Variables

Joint Distribution

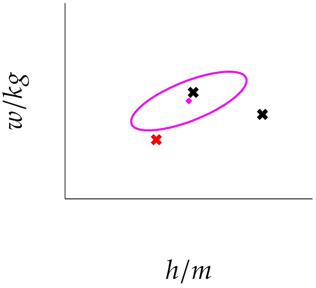


Marginal Distributions

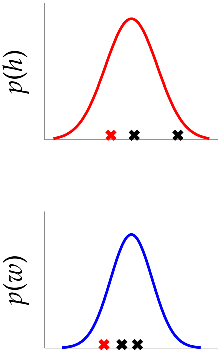


Sampling Two Dimensional Variables

Joint Distribution

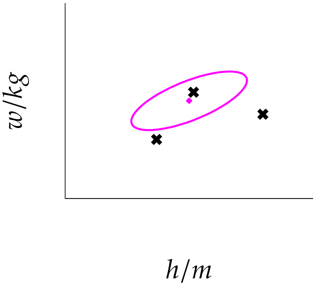


Marginal Distributions

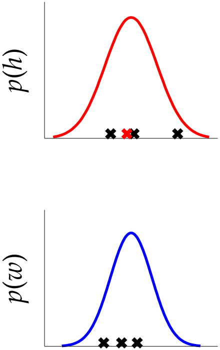


Sampling Two Dimensional Variables

Joint Distribution

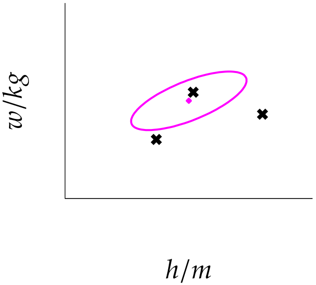


Marginal Distributions

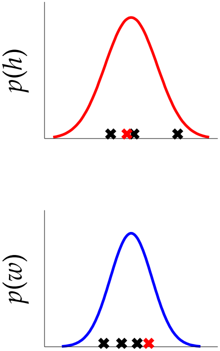


Sampling Two Dimensional Variables

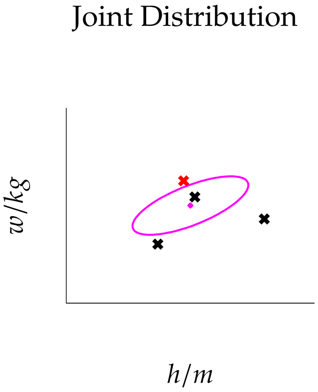
Joint Distribution

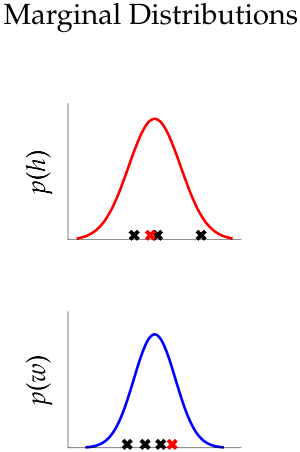


Marginal Distributions



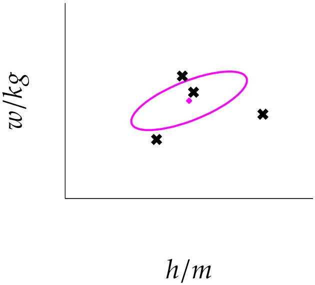
Sampling Two Dimensional Variables



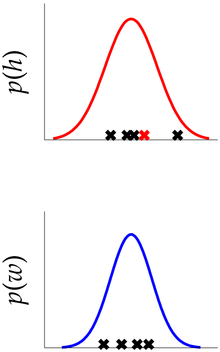


Sampling Two Dimensional Variables

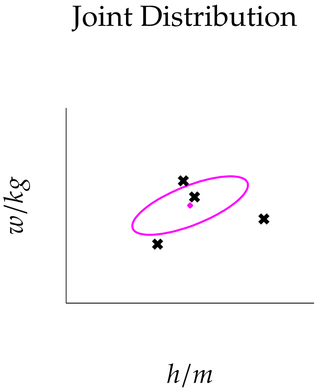
Joint Distribution

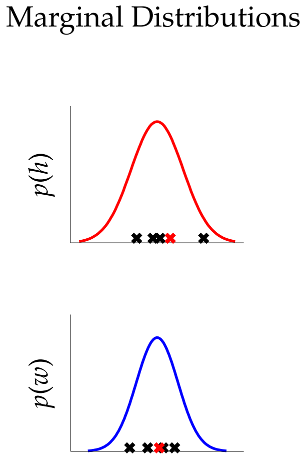


Marginal Distributions



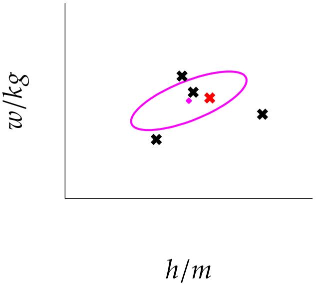
Sampling Two Dimensional Variables



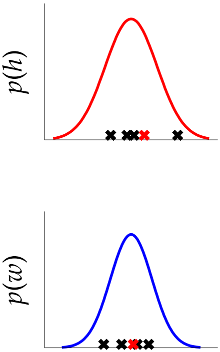


Sampling Two Dimensional Variables

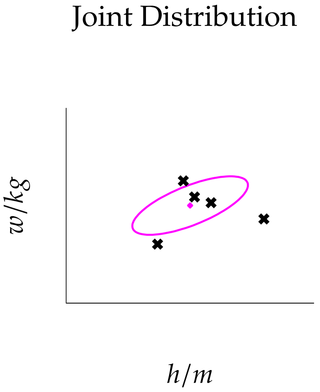
Joint Distribution

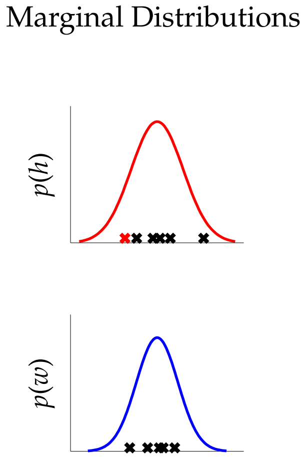


Marginal Distributions



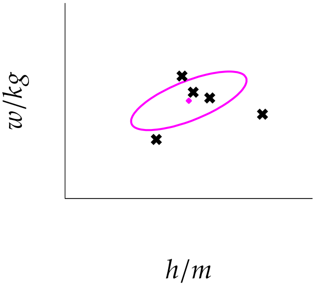
Sampling Two Dimensional Variables



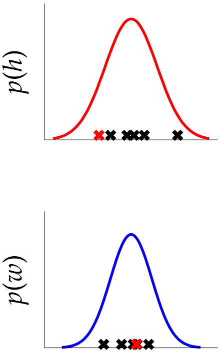


Sampling Two Dimensional Variables

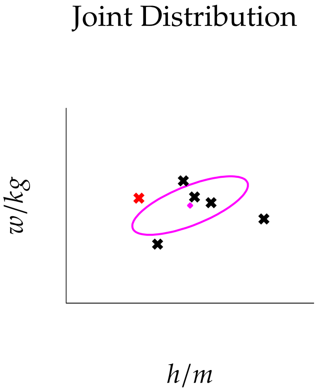
Joint Distribution

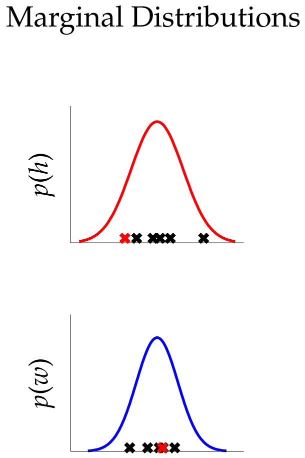


Marginal Distributions



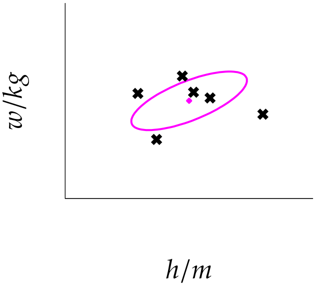
Sampling Two Dimensional Variables



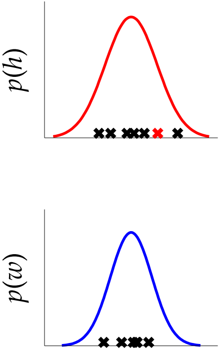


Sampling Two Dimensional Variables

Joint Distribution

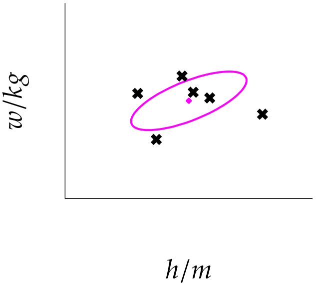


Marginal Distributions

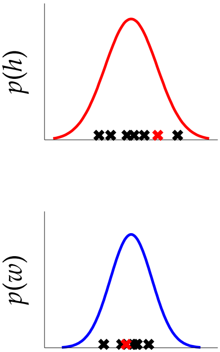


Sampling Two Dimensional Variables

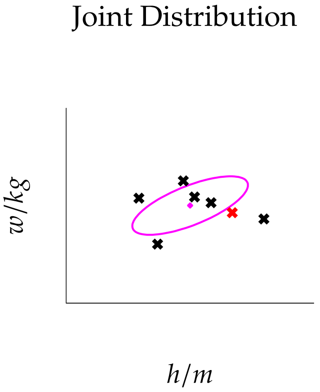
Joint Distribution

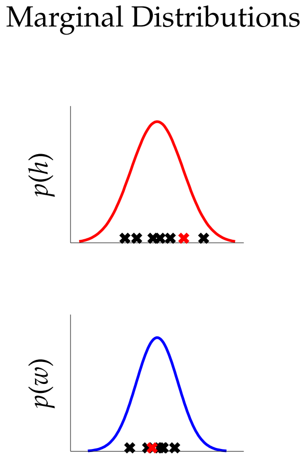


Marginal Distributions



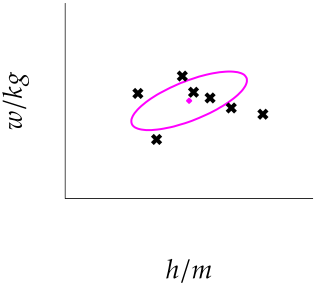
Sampling Two Dimensional Variables



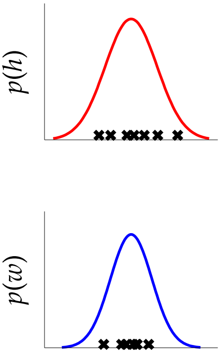


Sampling Two Dimensional Variables

Joint Distribution



Marginal Distributions



Independent Gaussians

$$p(w, h) = p(w)p(h)$$

Independent Gaussians

$$p(w,h) = \frac{1}{\sqrt{2\pi\sigma_1^2}\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\left(\frac{(w-\mu_1)^2}{\sigma_1^2} + \frac{(h-\mu_2)^2}{\sigma_2^2}\right)\right)$$

Independent Gaussians

$$p(w,h) = \frac{1}{\sqrt{2\pi\sigma_1^2 2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix} w \\ h \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\right)^\top \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}^{-1} \left(\begin{bmatrix} w \\ h \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\right)\right)$$

Independent Gaussians

$$p(\mathbf{y}) = \frac{1}{|2\pi\mathbf{D}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \mathbf{D}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)$$

Correlated Gaussian

Form correlated from original by rotating the data space using matrix $\mathbf{R}$.

$$p(\mathbf{y}) = \frac{1}{\left|2\pi\mathbf{D}\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{D}^{-1}(\mathbf{y} - \boldsymbol{\mu})\right)$$

Correlated Gaussian

Form correlated from original by rotating the data space using matrix **R**.

$$p(\mathbf{y}) = \frac{1}{|2\pi\mathbf{D}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{R}^\top\mathbf{y} - \mathbf{R}^\top\boldsymbol{\mu})^\top\mathbf{D}^{-1}(\mathbf{R}^\top\mathbf{y} - \mathbf{R}^\top\boldsymbol{\mu})\right)$$

Correlated Gaussian

Form correlated from original by rotating the data space using matrix $\mathbf{R}$.

$$p(\mathbf{y}) = \frac{1}{|2\pi\mathbf{D}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \mathbf{R}\mathbf{D}^{-1}\mathbf{R}^\top(\mathbf{y}-\boldsymbol{\mu})\right)$$

this gives a covariance matrix:

$$\mathbf{C}^{-1} = \mathbf{R}\mathbf{D}^{-1}\mathbf{R}^\top$$

Correlated Gaussian

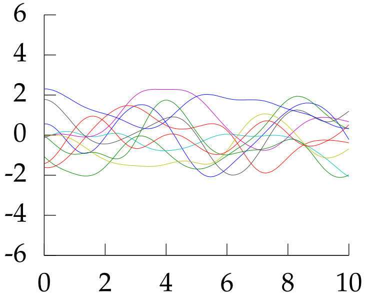
Form correlated from original by rotating the data space using matrix $\mathbf{R}$.

$$p(\mathbf{y}) = \frac{1}{|2\pi\mathbf{C}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \mathbf{C}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)$$

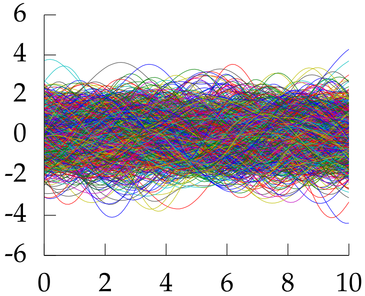
this gives a covariance matrix:

$$\mathbf{C} = \mathbf{R}\mathbf{D}\mathbf{R}^\top$$

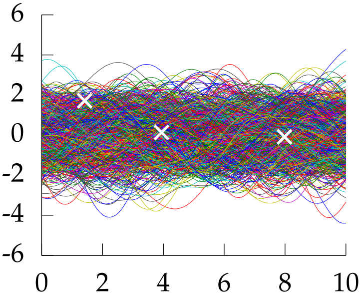
Gaussian Processes: Extremely Short Overview



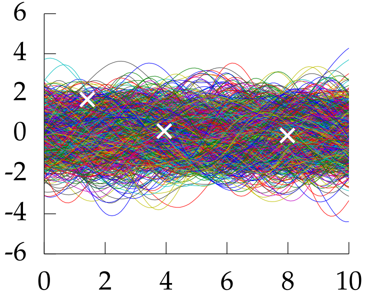
Gaussian Processes: Extremely Short Overview



Gaussian Processes: Extremely Short Overview



Gaussian Processes: Extremely Short Overview

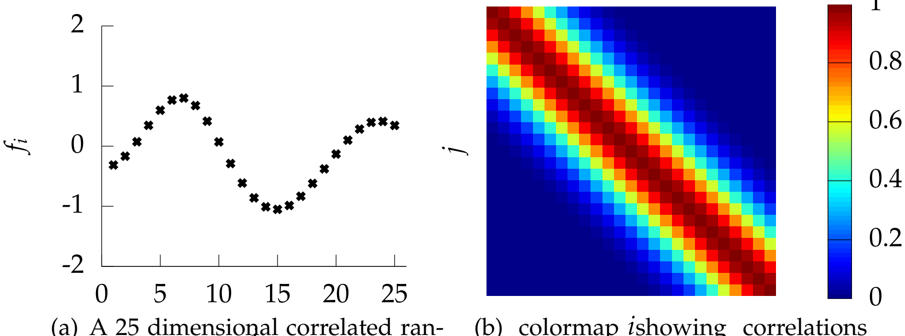


Sampling a Function

Multi-variate Gaussians

- We will consider a Gaussian with a particular structure of covariance matrix.
- Generate a single sample from this 25 dimensional Gaussian distribution, $\mathbf{f} = [f_1, f_2 \dots f_{25}]$.
- We will plot these points against their index.

Gaussian Distribution Sample

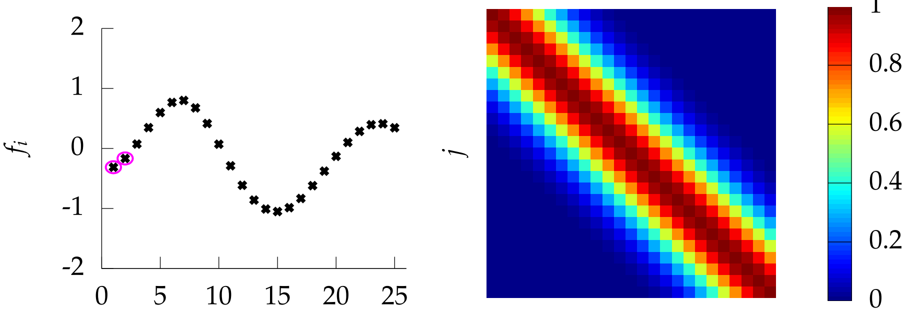


(a) A 25 dimensional correlated random variable (values ploted against index)

(b) colormap showing correlations between dimensions.

Figure : A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample

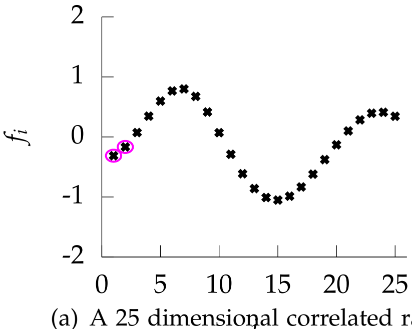


(a) A 25 dimensional correlated random variable (values ploted against index)

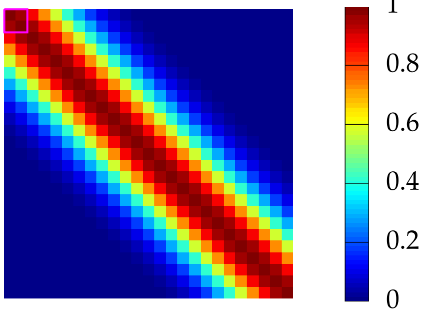
(b) colormap showing correlations between dimensions.

Figure : A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample



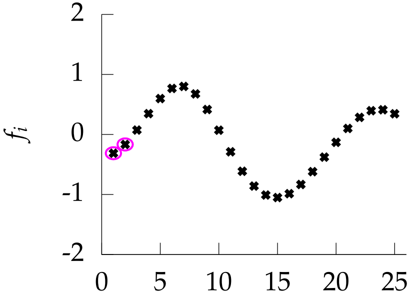
(a) A 25 dimensional correlated random variable (values ploted against index)



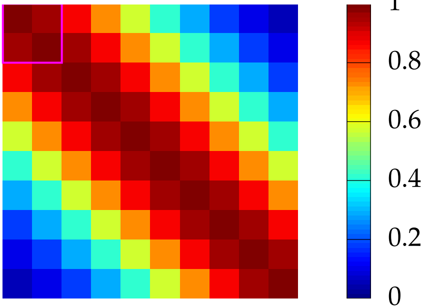
(b) colormap showing correlations between dimensions.

Figure : A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample



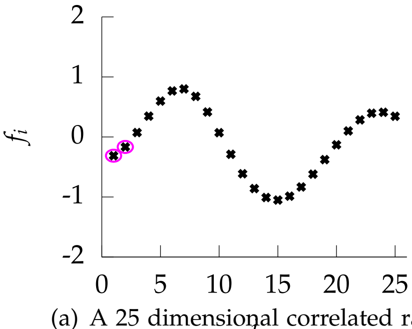
(a) A 25 dimensional correlated random variable (values ploted against index)



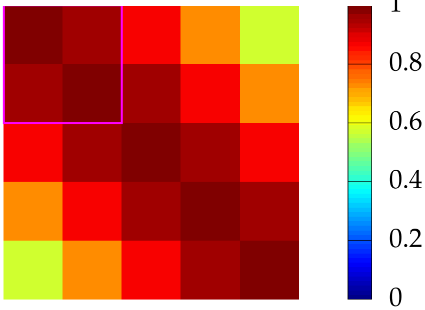
(b) colormap showing correlations between dimensions.

Figure : A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample



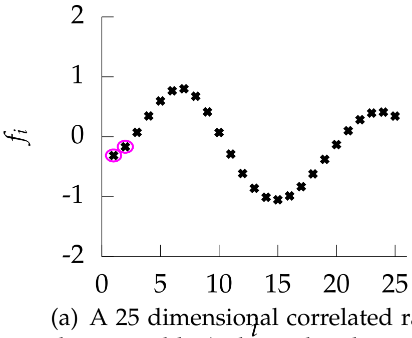
(a) A 25 dimensional correlated random variable (values ploted against index)



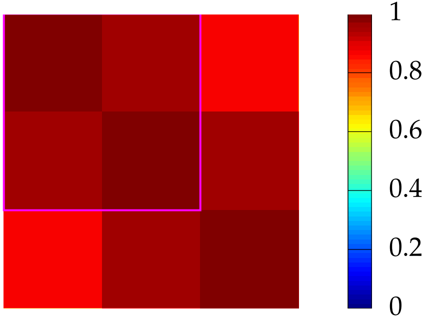
(b) colormap showing correlations between dimensions.

Figure : A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample



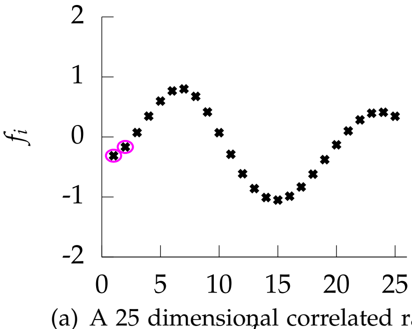
(a) A 25 dimensional correlated random variable (values ploted against index)



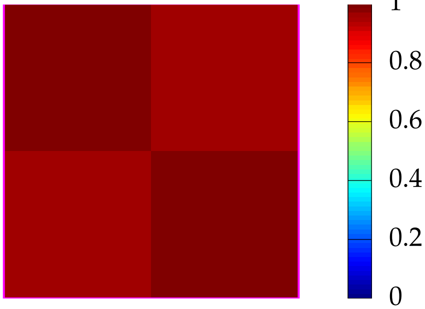
(b) colormap showing correlations between dimensions.

Figure : A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample



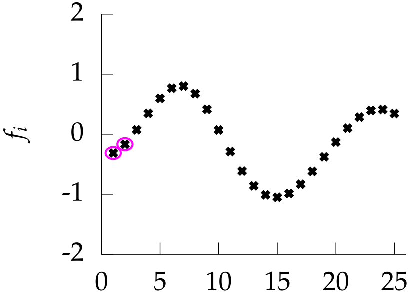
(a) A 25 dimensional correlated random variable (values ploted against index)



(b) colormap showing correlations between dimensions.

Figure : A sample from a 25 dimensional Gaussian distribution.

Gaussian Distribution Sample



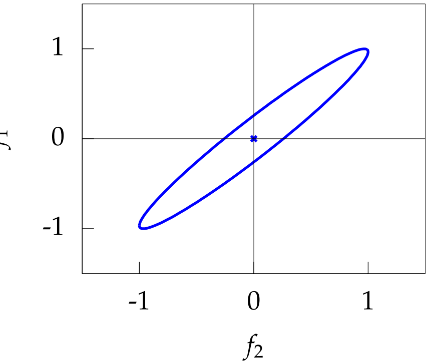
(a) A 25 dimensional correlated random variable (values ploted against index)

$$\begin{bmatrix} 1 & 0.96587 \\ 0.96587 & 1 \end{bmatrix}$$

(b) correlation between f_1 and f_2.

Figure : A sample from a 25 dimensional Gaussian distribution.

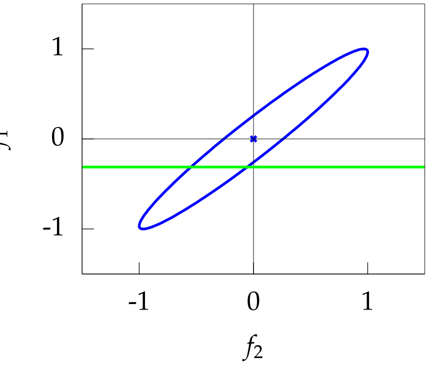
Prediction of f_2 from f_1



$$\begin{bmatrix} 1 & 0.96587 \\ 0.96587 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_2)$.

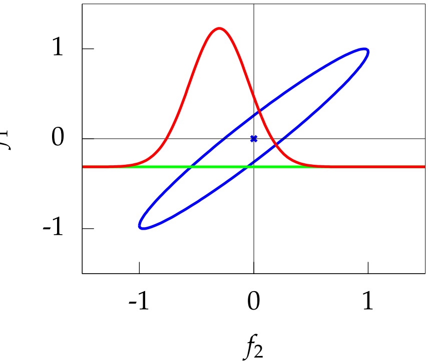
Prediction of f_2 from f_1



$$\begin{bmatrix} 1 & 0.96587 \\ 0.96587 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_2)$.
- We observe that $f_1 = -0.313$.

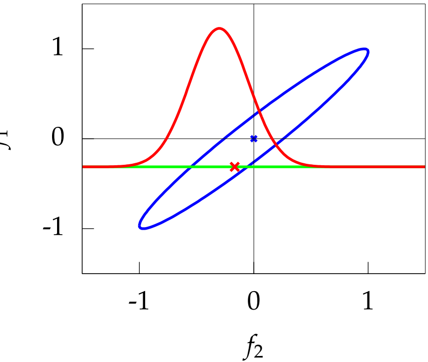
Prediction of f_2 from f_1



$$\begin{bmatrix} 1 & 0.96587 \\ 0.96587 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_2)$.
- We observe that $f_1 = -0.313$.
- Conditional density: $p(f_2|f_1 = -0.313)$.

Prediction of f_2 from f_1



$$\begin{bmatrix} 1 & 0.96587 \\ 0.96587 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_2)$.
- We observe that $f_1 = -0.313$.
- Conditional density: $p(f_2 | f_1 = -0.313)$.

Prediction with Correlated Gaussians

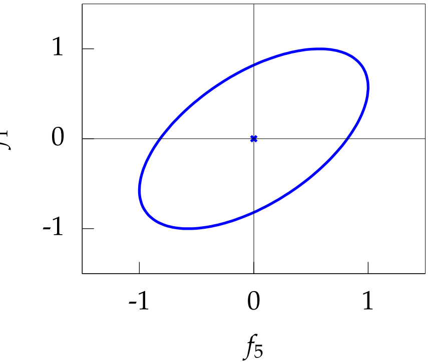
- Prediction of f_2 from f_1 requires *conditional density*.
- Conditional density is *also* Gaussian.

$$p(f_2|f_1) = \mathcal{N}\left(f_2|\frac{k_{1,2}}{k_{1,1}}f_1, k_{2,2} - \frac{k_{1,2}^2}{k_{1,1}}\right)$$

where covariance of joint density is given by

$$\mathbf{K} = \begin{bmatrix} k_{1,1} & k_{1,2} \\ k_{2,1} & k_{2,2} \end{bmatrix}$$

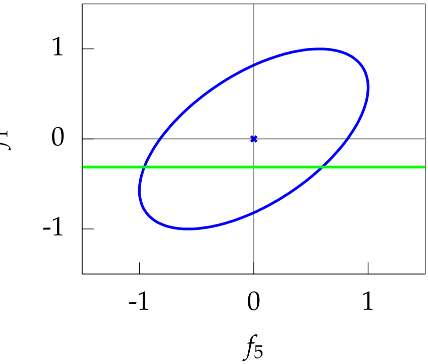
Prediction of f_5 from f_1



$$\begin{bmatrix} 1 & 0.57375 \\ 0.57375 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_5)$.

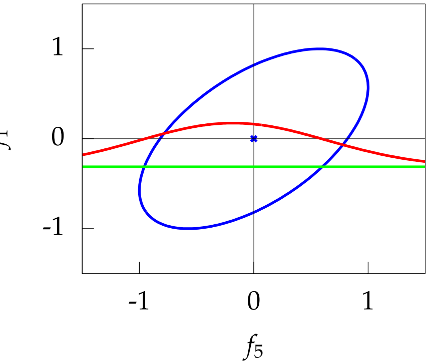
Prediction of f_5 from f_1



$$\begin{bmatrix} 1 & 0.57375 \\ 0.57375 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_5)$.
- We observe that $f_1 = -0.313$.

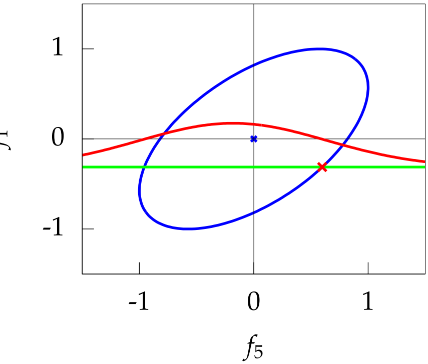
Prediction of f_5 from f_1



$$\begin{bmatrix} 1 & 0.57375 \\ 0.57375 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_5)$.
- We observe that $f_1 = -0.313$.
- Conditional density: $p(f_5|f_1 = -0.313)$.

Prediction of f_5 from f_1



$$\begin{bmatrix} 1 & 0.57375 \\ 0.57375 & 1 \end{bmatrix}$$

- The single contour of the Gaussian density represents the joint distribution, $p(f_1, f_5)$.
- We observe that $f_1 = -0.313$.
- Conditional density: $p(f_5 | f_1 = -0.313)$.

Prediction with Correlated Gaussians

- Prediction of $\mathbf{f}_*$ from $\mathbf{f}$ requires multivariate *conditional density*.
- Multivariate conditional density is *also* Gaussian.

$$p(\mathbf{f}_*|\mathbf{f}) = \mathcal{N}\left(\mathbf{f}_*|\mathbf{K}_{*,\mathbf{f}}\mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1}\mathbf{f}, \mathbf{K}_{*,*} - \mathbf{K}_{*,\mathbf{f}}\mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1}\mathbf{K}_{\mathbf{f},*}\right)$$

- Here covariance of joint density is given by

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{\mathbf{f},\mathbf{f}} & \mathbf{K}_{*,\mathbf{f}} \\ \mathbf{K}_{\mathbf{f},*} & \mathbf{K}_{*,*} \end{bmatrix}$$

Prediction with Correlated Gaussians

- Prediction of $\mathbf{f}_*$ from $\mathbf{f}$ requires multivariate *conditional density*.
- Multivariate conditional density is *also* Gaussian.

$$p(\mathbf{f}_*|\mathbf{f}) = \mathcal{N}\left(\mathbf{f}_*|\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$

$$\boldsymbol{\mu} = \mathbf{K}_{*,\mathbf{f}}\mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1}\mathbf{f}$$

$$\boldsymbol{\Sigma} = \mathbf{K}_{*,*} - \mathbf{K}_{*,\mathbf{f}}\mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1}\mathbf{K}_{\mathbf{f},*}$$

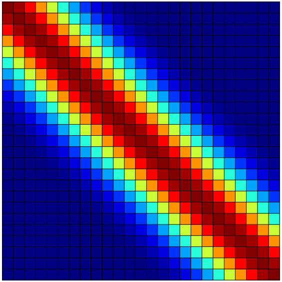
- Here covariance of joint density is given by

$$\mathbf{K} = \begin{bmatrix}\mathbf{K}_{\mathbf{f},\mathbf{f}} & \mathbf{K}_{*,\mathbf{f}}\\ \mathbf{K}_{\mathbf{f},*} & \mathbf{K}_{*,*}\end{bmatrix}$$

Exponentiated Quadratic Kernel Function (RBF, Squared Exponential, Gaussian)

$$k(\mathbf{x}, \mathbf{x}') = \alpha \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|_2^2}{2\ell^2}\right)$$

- Covariance matrix is built using the *inputs* to the function $\mathbf{x}$.
- For the example above it was based on Euclidean distance.
- The covariance function is also know as a kernel.



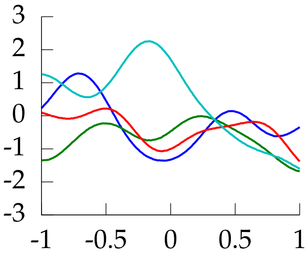
Covariance Functions

Where did this covariance matrix come from?

Exponentiated Quadratic Kernel Function (RBF, Squared Exponential, Gaussian)

$$k(\mathbf{x}, \mathbf{x}') = \alpha \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|_2^2}{2\ell^2}\right)$$

- Covariance matrix is built using the *inputs* to the function $\mathbf{x}$.
- For the example above it was based on Euclidean distance.
- The covariance function is also know as a kernel.



Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_1 = -3.0,\, x_1 = -3.0$$

$$k_{1,1} = 1.00 \times \exp\left(-\frac{(-3.0 - -3.0)^2}{2 \times 2.00^2}\right)$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_1 = -3.0, x_1 = -3.0$$

$$k_{1,1} = 1.00 \times \exp\left(-\frac{(-3.0 - -3.0)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} \boxed{1.00} & & \\ & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.20, x_1 = -3.0$$

$$k_{2,1} = 1.00 \times \exp\left(-\frac{(1.20 - -3.0)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & & \\ & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.20,\, x_1 = -3.0$$

$$k_{2,1} = 1.00 \times \exp\left(-\frac{(1.20 - -3.0)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & & \\ \boxed{0.110} & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.20, x_1 = -3.0$$

$$k_{2,1} = 1.00 \times \exp\left(-\frac{(1.20 - -3.0)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & \\ 0.110 & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.20,\; x_2 = 1.20$$

$$k_{2,2} = 1.00 \times \exp\left(-\frac{(1.20 - 1.20)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & \\ 0.110 & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.20, x_2 = 1.20$$

$$k_{2,2} = 1.00 \times \exp\left(-\frac{(1.20-1.20)^2}{2\times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & \\ 0.110 & \boxed{1.00} & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40,\ x_1 = -3.0$$

$$k_{3,1} = 1.00 \times \exp\left(-\frac{(1.40 - -3.0)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & \\ 0.110 & 1.00 & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40,\ x_1 = -3.0$$

$$k_{3,1} = 1.00 \times \exp\left(-\frac{(1.40 - -3.0)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & \\ 0.110 & 1.00 & \\ \boxed{0.0889} & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40,\ x_1 = -3.0$$

$$k_{3,1} = 1.00 \times \exp\left(-\frac{(1.40 - -3.0)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & \\ 0.0889 & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_2 = 1.20$$

$$k_{3,2} = 1.00 \times \exp\left(-\frac{(1.40-1.20)^2}{2\times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & \\ 0.0889 & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_2 = 1.20$$

$$k_{3,2} = 1.00 \times \exp\left(-\frac{(1.40-1.20)^2}{2\times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & \\ 0.0889 & \boxed{0.995} & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_2 = 1.20$$

$$k_{3,2} = 1.00 \times \exp\left(-\frac{(1.40 - 1.20)^2}{2 \times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & 0.995 \\ 0.0889 & 0.995 & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_3 = 1.40$$

$$k_{3,3} = 1.00 \times \exp\left(-\frac{(1.40-1.40)^2}{2\times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & 0.995 \\ 0.0889 & 0.995 & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_3 = 1.40$$

$$k_{3,3} = 1.00 \times \exp\left(-\frac{(1.40-1.40)^2}{2\times 2.00^2}\right)$$

$$\begin{bmatrix} 1.00 & 0.110 & 0.0889 \\ 0.110 & 1.00 & 0.995 \\ 0.0889 & 0.995 & \boxed{1.00} \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

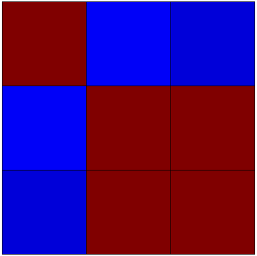
Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$x_3 = 1.40$, $x_3 = 1.40$

$$k_{3,3} = 1.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 2.00^2}\right)$$



$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 2.00$ and $\alpha = 1.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_1 = -3, x_1 = -3$$

$$k_{1,1} = 1.0 \times \exp\left(-\frac{(-3 - -3)^2}{2 \times 2.0^2}\right)$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_1 = -3, x_1 = -3$$

$$k_{1,1} = 1.0 \times \exp\left(-\frac{(-3 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} \boxed{1.0} & & & \\ & & & \\ & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.2, x_1 = -3$$

$$k_{2,1} = 1.0 \times \exp\left(-\frac{(1.2 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & & & \\ & & & \\ & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.2,\ x_1 = -3$$

$$\begin{bmatrix} 1.0 & & & \\ \boxed{0.11} & & & \\ & & & \\ & & & \end{bmatrix}$$

$$k_{2,1} = 1.0 \times \exp\left(-\frac{(1.2 - -3)^2}{2 \times 2.0^2}\right)$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.2,\; x_1 = -3$$

$$k_{2,1} = 1.0 \times \exp\left(-\frac{(1.2 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & & \\ 0.11 & & & \\ & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.2, x_2 = 1.2$$

$$k_{2,2} = 1.0 \times \exp\left(-\frac{(1.2-1.2)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & & \\ 0.11 & & & \\ & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.2,\, x_2 = 1.2$$

$$k_{2,2} = 1.0 \times \exp\left(-\frac{(1.2-1.2)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & & \\ 0.11 & \boxed{1.0} & & \\ & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.4, x_1 = -3$$

$$k_{3,1} = 1.0 \times \exp\left(-\frac{(1.4 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & & \\ 0.11 & 1.0 & & \\ & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.4,\ x_1 = -3$$

$$k_{3,1} = 1.0 \times \exp\left(-\frac{(1.4 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & & \\ 0.11 & 1.0 & & \\ \boxed{0.089} & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.4, x_1 = -3$$

$$k_{3,1} = 1.0 \times \exp\left(-\frac{(1.4 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & & \\ 0.089 & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.4, x_2 = 1.2$$

$$k_{3,2} = 1.0 \times \exp\left(-\frac{(1.4-1.2)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & & \\ 0.089 & & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.4,\ x_2 = 1.2$$

$$k_{3,2} = 1.0 \times \exp\left(-\frac{(1.4-1.2)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & & \\ 0.089 & \boxed{1.0} & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.4, x_2 = 1.2$$

$$k_{3,2} = 1.0 \times \exp\left(-\frac{(1.4-1.2)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.4,\ x_3 = 1.4$$

$$k_{3,3} = 1.0 \times \exp\left(-\frac{(1.4-1.4)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.4, x_3 = 1.4$$

$$k_{3,3} = 1.0 \times \exp\left(-\frac{(1.4-1.4)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 \\ 0.11 & 1.0 & 1.0 \\ 0.089 & 1.0 & \boxed{1.0} \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0, x_1 = -3$$

$$k_{4,1} = 1.0 \times \exp\left(-\frac{(2.0 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & 1.0 & \\ & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0, x_1 = -3$$

$$k_{4,1} = 1.0 \times \exp\left(-\frac{(2.0 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & 1.0 & \\ \boxed{0.044} & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0, x_1 = -3$$

$$k_{4,1} = 1.0 \times \exp\left(-\frac{(2.0 - -3)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0, x_2 = 1.2$$

$$k_{4,2} = 1.0 \times \exp\left(-\frac{(2.0-1.2)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0,\, x_2 = 1.2$$

$$k_{4,2} = 1.0 \times \exp\left(-\frac{(2.0-1.2)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & \boxed{0.92} & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0, x_2 = 1.2$$

$$k_{4,2} = 1.0 \times \exp\left(-\frac{(2.0-1.2)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & 0.92 & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0, x_3 = 1.4$$

$$k_{4,3} = 1.0 \times \exp\left(-\frac{(2.0-1.4)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & 0.92 & & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0, x_3 = 1.4$$

$$k_{4,3} = 1.0 \times \exp\left(-\frac{(2.0-1.4)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & \\ 0.044 & 0.92 & \boxed{0.96} & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0,\ x_3 = 1.4$$

$$k_{4,3} = 1.0 \times \exp\left(-\frac{(2.0 - 1.4)^2}{2 \times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & 0.96 \\ 0.044 & 0.92 & 0.96 & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0, x_4 = 2.0$$

$$k_{4,4} = 1.0 \times \exp\left(-\frac{(2.0-2.0)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & 0.96 \\ 0.044 & 0.92 & 0.96 & \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_4 = 2.0, x_4 = 2.0$$

$$k_{4,4} = 1.0 \times \exp\left(-\frac{(2.0-2.0)^2}{2\times 2.0^2}\right)$$

$$\begin{bmatrix} 1.0 & 0.11 & 0.089 & 0.044 \\ 0.11 & 1.0 & 1.0 & 0.92 \\ 0.089 & 1.0 & 1.0 & 0.96 \\ 0.044 & 0.92 & 0.96 & \boxed{1.0} \end{bmatrix}$$

$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

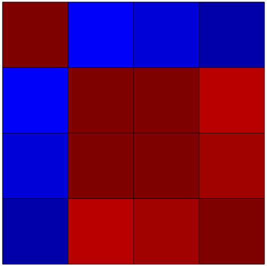
Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$x_4 = 2.0$, $x_4 = 2.0$

$$k_{4,4} = 1.0 \times \exp\left(-\frac{(2.0-2.0)^2}{2\times 2.0^2}\right)$$



$x_1 = -3$, $x_2 = 1.2$, $x_3 = 1.4$, and $x_4 = 2.0$ with $\ell = 2.0$ and $\alpha = 1.0$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_1 = -3.0, x_1 = -3.0$$

$$k_{1,1} = 4.00 \times \exp\left(-\frac{(-3.0 - -3.0)^2}{2 \times 5.00^2}\right)$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_1 = -3.0, x_1 = -3.0$$

$$k_{1,1} = 4.00 \times \exp\left(-\frac{(-3.0 - -3.0)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} \boxed{4.00} & & \\ & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.20, x_1 = -3.0$$

$$k_{2,1} = 4.00 \times \exp\left(-\frac{(1.20 - -3.0)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & & \\ & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$x_2 = 1.20, x_1 = -3.0$

$k_{2,1} = 4.00 \times \exp\left(-\frac{(1.20 - -3.0)^2}{2 \times 5.00^2}\right)$

$$\begin{bmatrix} 4.00 & & \\ \boxed{2.81} & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.20, x_1 = -3.0$$

$$k_{2,1} = 4.00 \times \exp\left(-\frac{(1.20 - -3.0)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & \\ 2.81 & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.20,\, x_2 = 1.20$$

$$k_{2,2} = 4.00 \times \exp\left(-\frac{(1.20-1.20)^2}{2\times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & \\ 2.81 & & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_2 = 1.20,\, x_2 = 1.20$$

$$k_{2,2} = 4.00 \times \exp\left(-\frac{(1.20 - 1.20)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & \\ 2.81 & \boxed{4.00} & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_1 = -3.0$$

$$k_{3,1} = 4.00 \times \exp\left(-\frac{(1.40 - -3.0)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & \\ 2.81 & 4.00 & \\ & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_1 = -3.0$$

$$k_{3,1} = 4.00 \times \exp\left(-\frac{(1.40 - -3.0)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & \\ 2.81 & 4.00 & \\ \boxed{2.72} & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40,\ x_1 = -3.0$$

$$k_{3,1} = 4.00 \times \exp\left(-\frac{(1.40 - -3.0)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & 2.72 \\ 2.81 & 4.00 & \\ 2.72 & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40,\ x_2 = 1.20$$

$$k_{3,2} = 4.00 \times \exp\left(-\frac{(1.40-1.20)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & 2.72 \\ 2.81 & 4.00 & \\ 2.72 & & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40,\; x_2 = 1.20$$

$$k_{3,2} = 4.00 \times \exp\left(-\frac{(1.40 - 1.20)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & 2.72 \\ 2.81 & 4.00 & \\ 2.72 & \boxed{4.00} & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40, x_2 = 1.20$$

$$k_{3,2} = 4.00 \times \exp\left(-\frac{(1.40-1.20)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & 2.72 \\ 2.81 & 4.00 & 4.00 \\ 2.72 & 4.00 & \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & 2.72 \\ 2.81 & 4.00 & 4.00 \\ 2.72 & 4.00 & \end{bmatrix}$$

$$x_3 = 1.40, x_3 = 1.40$$

$$k_{3,3} = 4.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 5.00^2}\right)$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40,\, x_3 = 1.40$$

$$k_{3,3} = 4.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 5.00^2}\right)$$

$$\begin{bmatrix} 4.00 & 2.81 & 2.72 \\ 2.81 & 4.00 & 4.00 \\ 2.72 & 4.00 & \boxed{4.00} \end{bmatrix}$$

$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

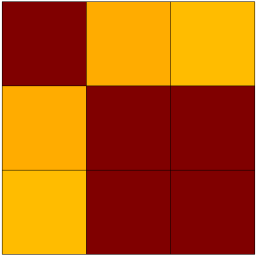
Covariance Functions

Where did this covariance matrix come from?

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\ell^2}\right)$$

$$x_3 = 1.40,\, x_3 = 1.40$$

$$k_{3,3} = 4.00 \times \exp\left(-\frac{(1.40 - 1.40)^2}{2 \times 5.00^2}\right)$$



$x_1 = -3.0$, $x_2 = 1.20$, and $x_3 = 1.40$ with $\ell = 5.00$ and $\alpha = 4.00$.

Gaussian Process Interpolation

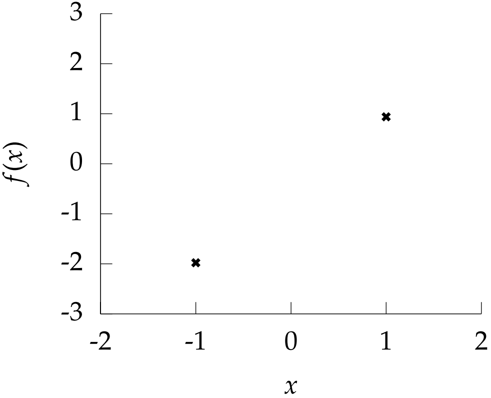


Figure : Real example: BACCO (see *e.g.* (Oakley and O'Hagan, 2002)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

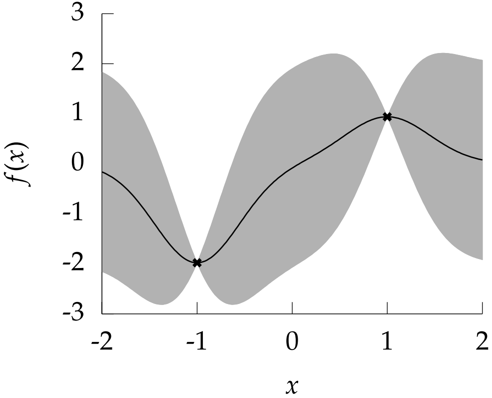


Figure : Real example: BACCO (see *e.g.* (Oakley and O'Hagan, 2002)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

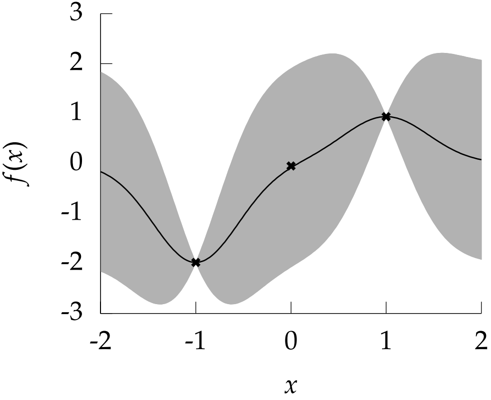


Figure : Real example: BACCO (see *e.g.* (Oakley and O'Hagan, 2002)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

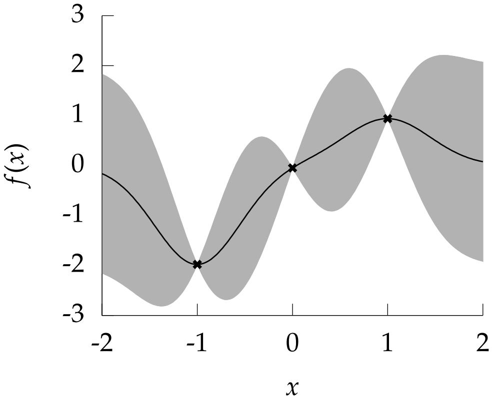


Figure : Real example: BACCO (see *e.g.* (Oakley and O'Hagan, 2002)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

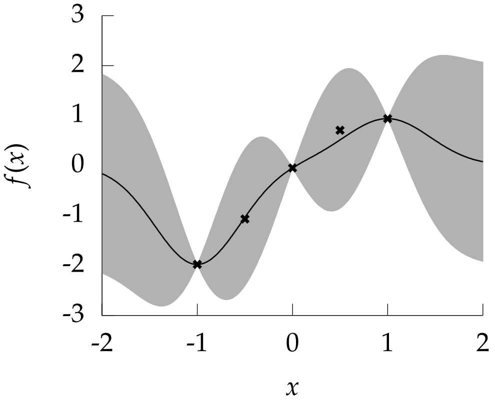


Figure : Real example: BACCO (see *e.g.* (Oakley and O'Hagan, 2002)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

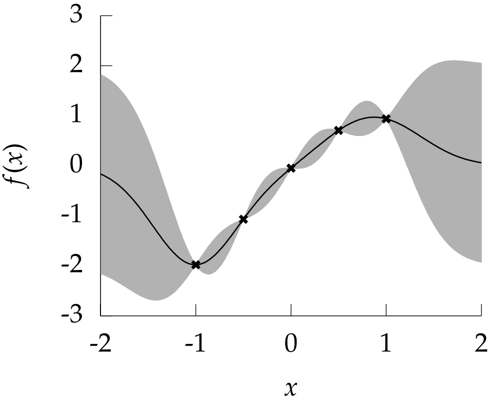


Figure : Real example: BACCO (see *e.g.* (Oakley and O'Hagan, 2002)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

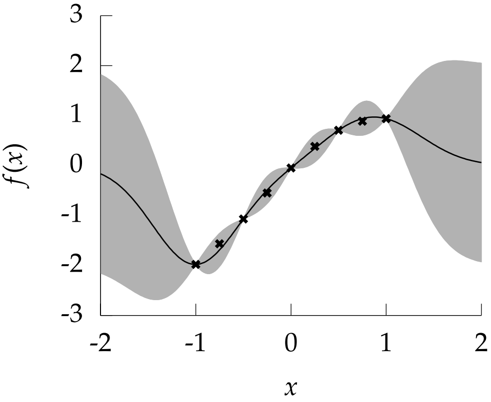


Figure : Real example: BACCO (see *e.g.* (Oakley and O'Hagan, 2002)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Interpolation

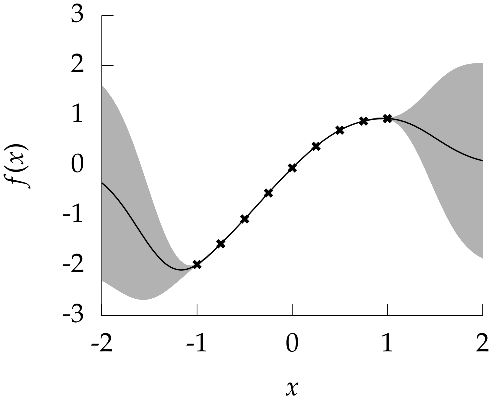


Figure : Real example: BACCO (see *e.g.* (Oakley and O'Hagan, 2002)). Interpolation through outputs from slow computer simulations (*e.g.* atmospheric carbon levels).

Gaussian Process Regression

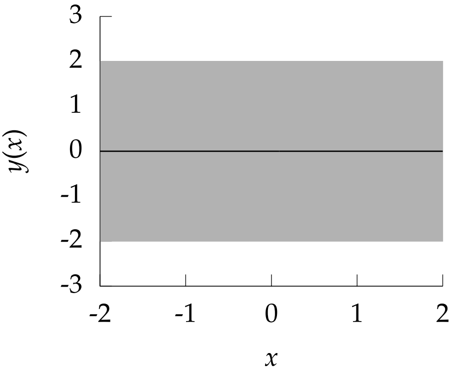


Figure : Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

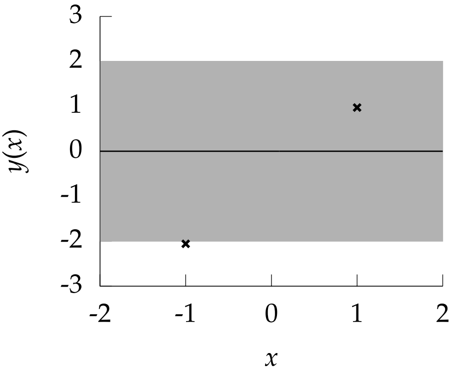


Figure : Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

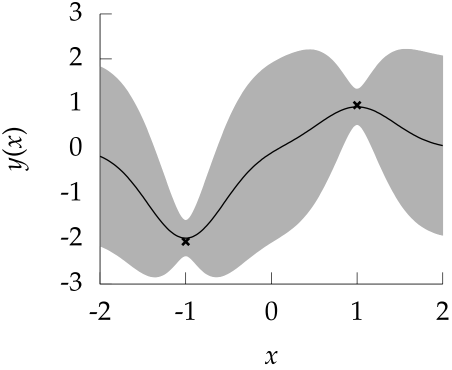


Figure : Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

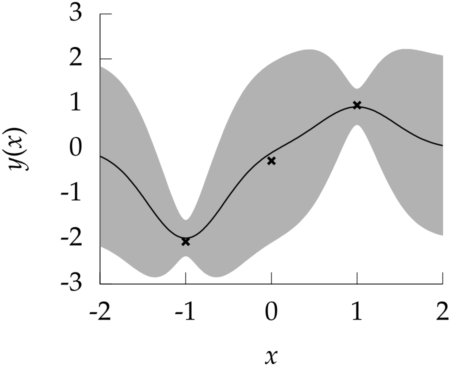


Figure : Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

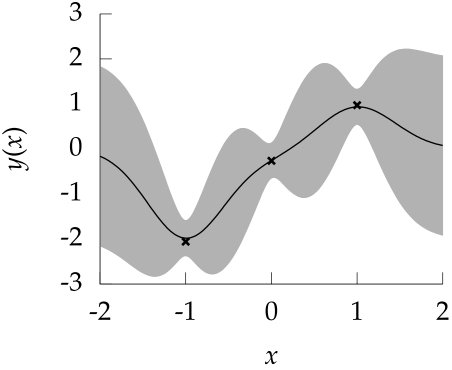


Figure : Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

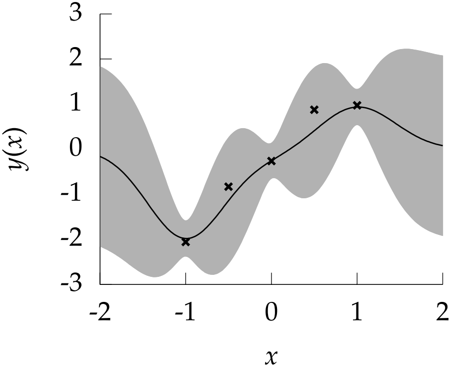


Figure : Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

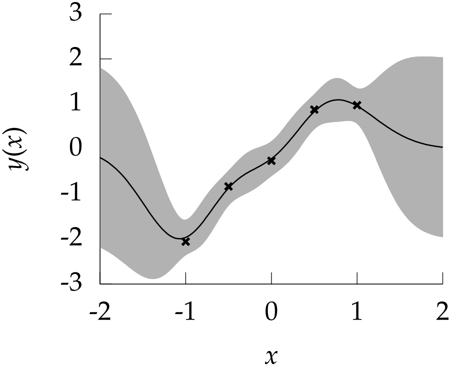


Figure : Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

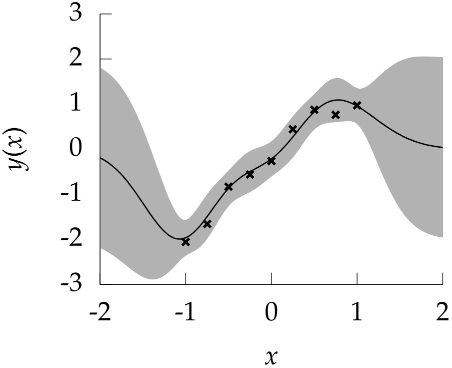


Figure : Examples include WiFi localization, C14 callibration curve.

Gaussian Process Regression

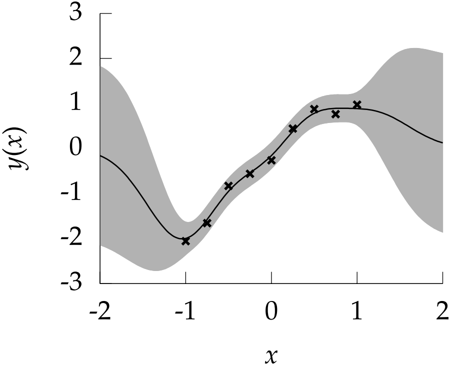


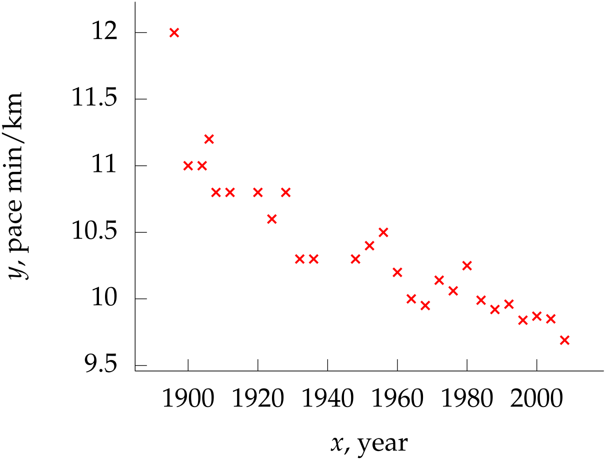
Figure : Examples include WiFi localization, C14 callibration curve.

Olympic 100m Data

- Gold medal times for Olympic 100 m runners since 1896.

Image from Wikimedia Commons
`http://bit.ly/191adDC`

Olympic 100m Data



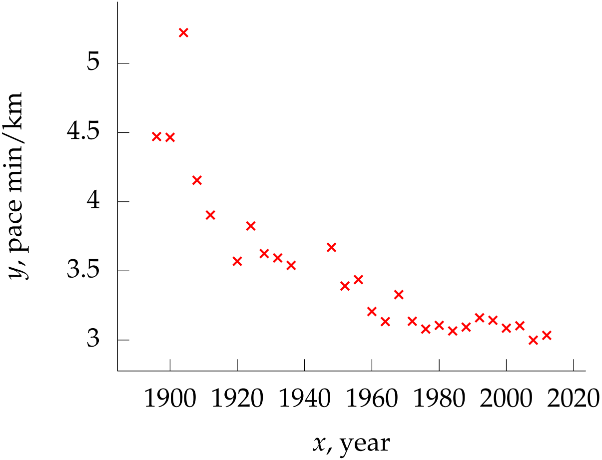
Olympic 100 m Data.

Olympic Marathon Data

- Gold medal times for Olympic Marathon since 1896.
- Marathons before 1924 didn't have a standardised distance.
- Present results using pace per km.
- In 1904 Marathon was badly organised leading to very slow times.

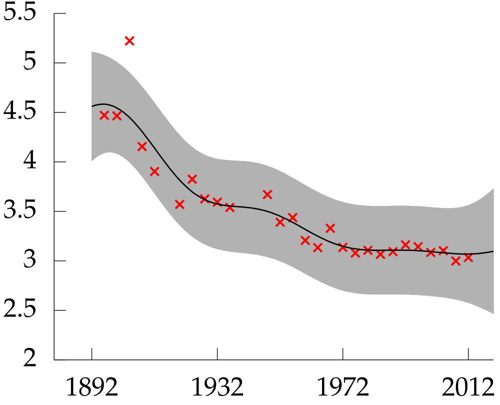
Image from Wikimedia Commons
`http://bit.ly/16kMKHQ`

Olympic Marathon Data



Olympic Marathon Data.

Gaussian Process Fit to Olympic Marathon Data



Learning Covariance Parameters

Can we determine covariance parameters from the data?

$$\mathcal{N}\left(\mathbf{y}|\mathbf{0},\mathbf{K}\right) = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{K}|^{\frac{1}{2}}}\exp\left(-\frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}\right)$$

The parameters are *inside* the covariance function (matrix).

$$k_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\theta})$$

Learning Covariance Parameters

Can we determine covariance parameters from the data?

$$\mathcal{N}\left(\mathbf{y}|\mathbf{0},\mathbf{K}\right) = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{K}|^{\frac{1}{2}}}\exp\left(-\frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}\right)$$

The parameters are *inside* the covariance function (matrix).

$$k_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\theta})$$

Learning Covariance Parameters

Can we determine covariance parameters from the data?

$$\log \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K}) = -\frac{1}{2}\log|\mathbf{K}| - \frac{\mathbf{y}^\top \mathbf{K}^{-1}\mathbf{y}}{2} - \frac{n}{2}\log 2\pi$$

The parameters are *inside* the covariance function (matrix).

$$k_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\theta})$$

Learning Covariance Parameters

Can we determine covariance parameters from the data?

$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^\top\mathbf{K}^{-1}\mathbf{y}}{2}$$

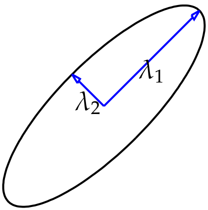
The parameters are *inside* the covariance function (matrix).

$$k_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\theta})$$

Eigendecomposition of Covariance

A useful decomposition for understanding the objective function.

$$\mathbf{K} = \mathbf{R}\boldsymbol{\Lambda}^2\mathbf{R}^\top$$

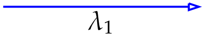


Diagonal of $\boldsymbol{\Lambda}$ represents distance along axes.
$\mathbf{R}$ gives a rotation of these axes.

where $\boldsymbol{\Lambda}$ is a *diagonal* matrix and $\mathbf{R}^\top\mathbf{R} = \mathbf{I}$.

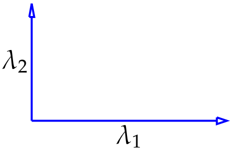
Capacity control: $\log |\mathbf{K}|$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



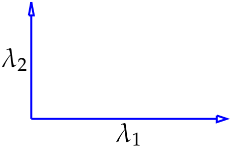
Capacity control: $\log |\mathbf{K}|$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



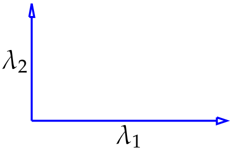
Capacity control: $\log |\mathbf{K}|$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



Capacity control: $\log |\mathbf{K}|$

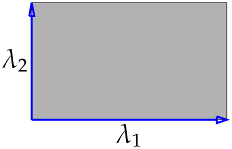
$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



$$|\boldsymbol{\Lambda}| = \lambda_1 \lambda_2$$

Capacity control: $\log |\mathbf{K}|$

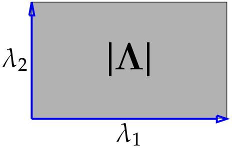
$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



$$|\mathbf{\Lambda}| = \lambda_1 \lambda_2$$

Capacity control: $\log |\mathbf{K}|$

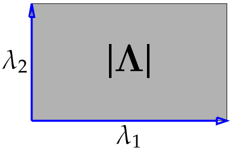
$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



$$|\mathbf{\Lambda}| = \lambda_1 \lambda_2$$

Capacity control: $\log |\mathbf{K}|$

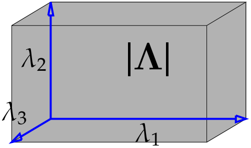
$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$



$$|\boldsymbol{\Lambda}| = \lambda_1 \lambda_2$$

Capacity control: $\log |\mathbf{K}|$

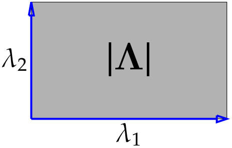
$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ \boxed{0 \quad 0 \quad \lambda_3} \end{bmatrix}$$



$$|\boldsymbol{\Lambda}| = \lambda_1 \lambda_2 \lambda_3$$

Capacity control: $\log |\mathbf{K}|$

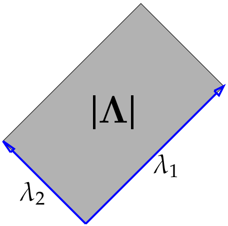
$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



$$|\boldsymbol{\Lambda}| = \lambda_1 \lambda_2$$

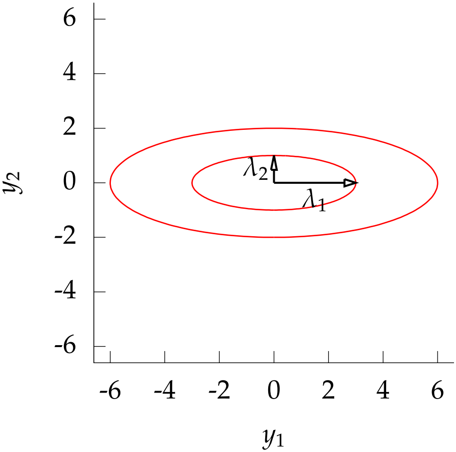
Capacity control: $\log |\mathbf{K}|$

$$\mathbf{R}\boldsymbol{\Lambda} = \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{bmatrix}$$

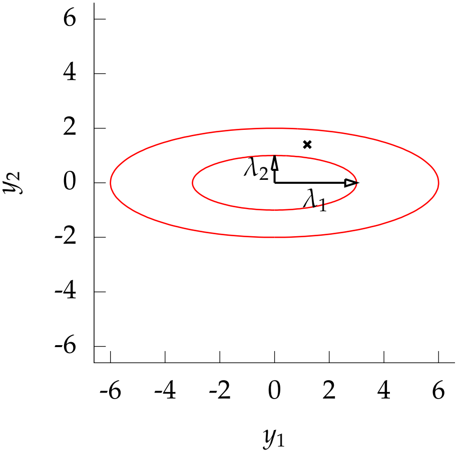


$$|\mathbf{R}\boldsymbol{\Lambda}| = \lambda_1 \lambda_2$$

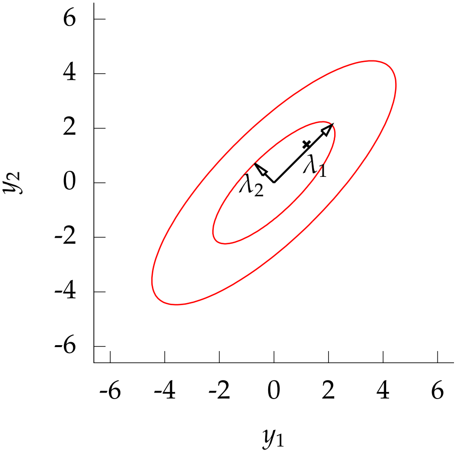
Data Fit: $\frac{\mathbf{y}^\top \mathbf{K}^{-1} \mathbf{y}}{2}$



Data Fit: $\frac{\mathbf{y}^\top \mathbf{K}^{-1}\mathbf{y}}{2}$

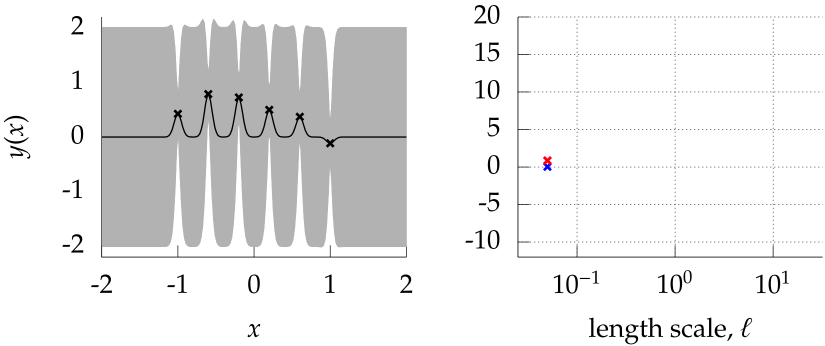


Data Fit: $\frac{\mathbf{y}^\top \mathbf{K}^{-1}\mathbf{y}}{2}$



Learning Covariance Parameters

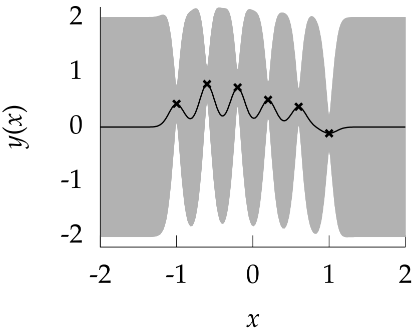
Can we determine length scales and noise levels from the data?

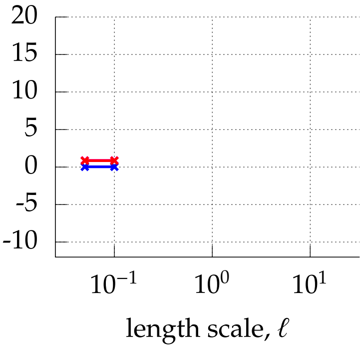


$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

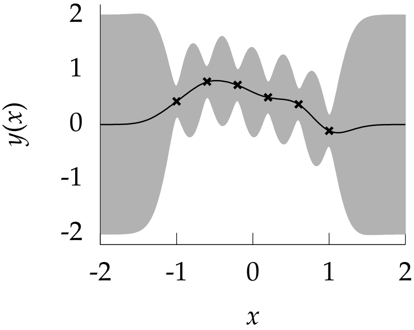


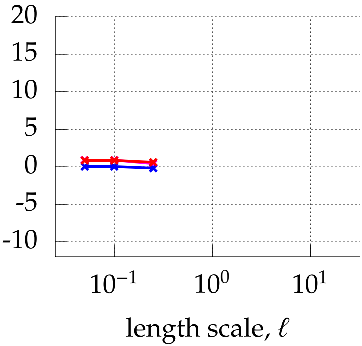


$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

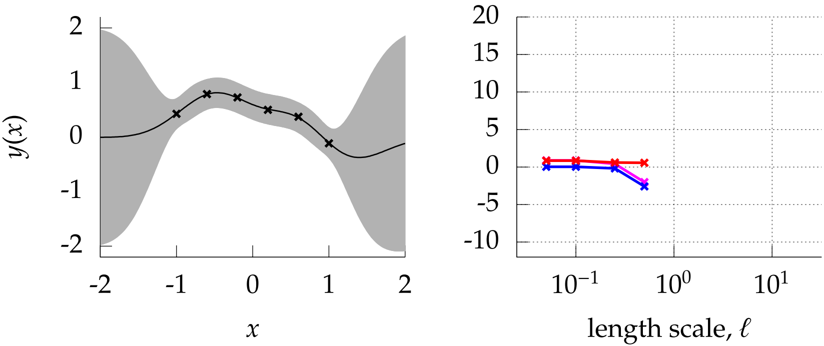




$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^\top \mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

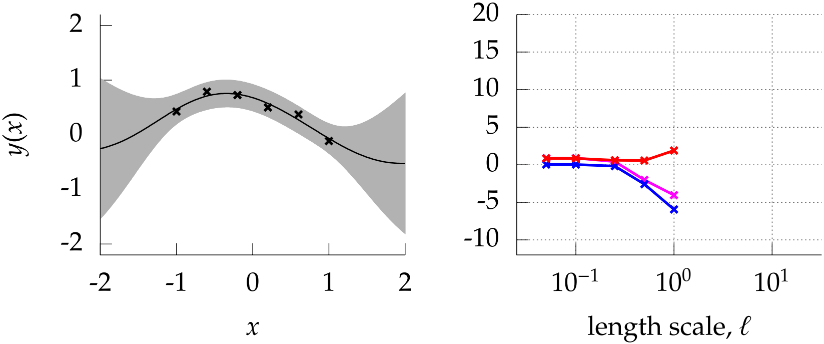
Can we determine length scales and noise levels from the data?



$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^{\top}\mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

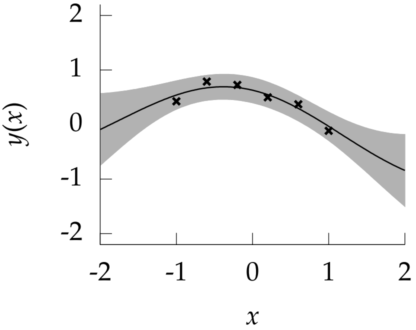
Can we determine length scales and noise levels from the data?

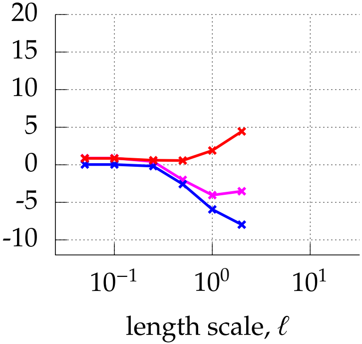


$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^\top\mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

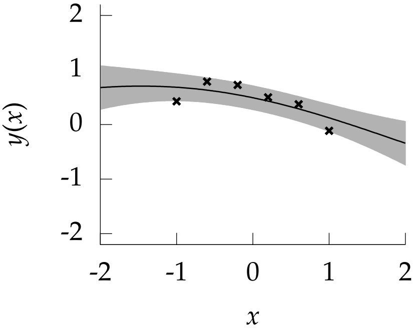


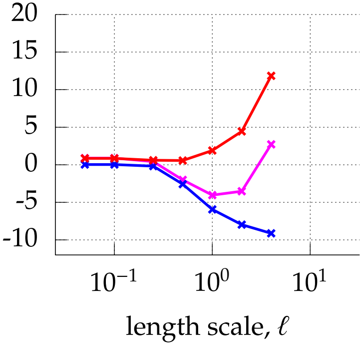


$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^\top \mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

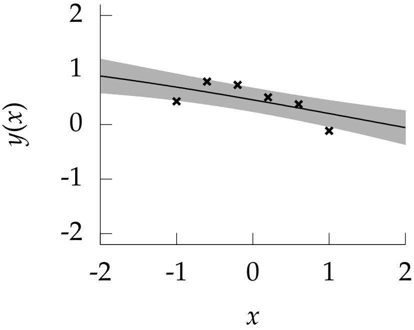


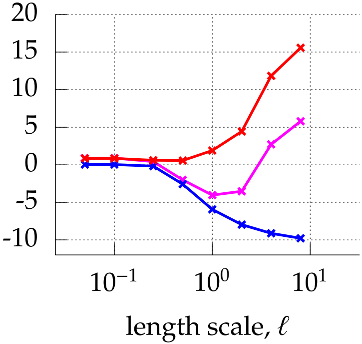


$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^\top \mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?

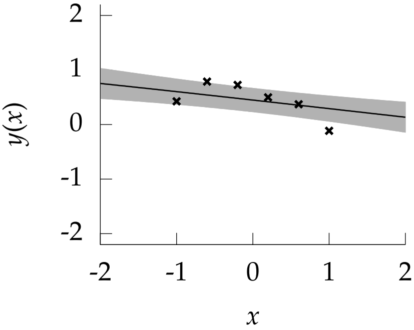


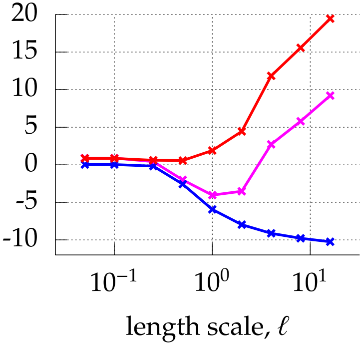


$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^\top\mathbf{K}^{-1}\mathbf{y}}{2}$$

Learning Covariance Parameters

Can we determine length scales and noise levels from the data?





$$E(\boldsymbol{\theta}) = \frac{1}{2}\log|\mathbf{K}| + \frac{\mathbf{y}^\top \mathbf{K}^{-1}\mathbf{y}}{2}$$

Gene Expression Example

- Given given expression levels in the form of a time series from Della Gatta et al. (2008).
- Want to detect if a gene is expressed or not, fit a GP to each gene (Kalaitzis and Lawrence, 2011).

Kalaitzis and Lawrence *BMC Bioinformatics* 2011, **12**:180
http://www.biomedcentral.com/1471-2105/12/180

BMC Bioinformatics

RESEARCH ARTICLE — Open Access

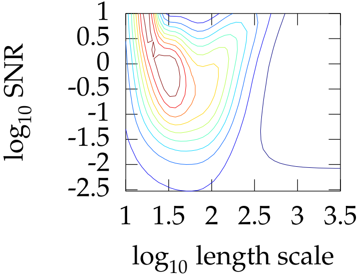
A Simple Approach to Ranking Differentially Expressed Gene Expression Time Courses through Gaussian Process Regression

Alfredo A Kalaitzis* and Neil D Lawrence*

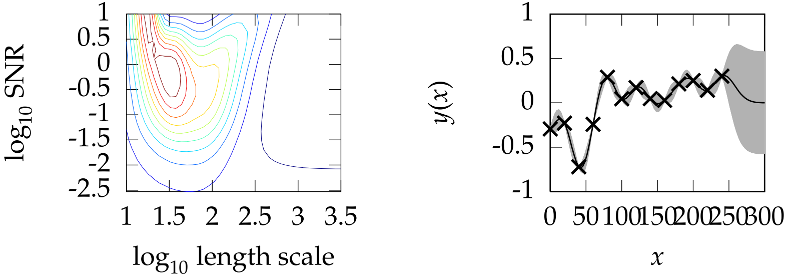
Abstract

Background: The analysis of gene expression from time series underpins many biological studies. Two basic forms of analysis recur for data of this type: removing inactive (quiet) genes from the study and determining which genes are differentially expressed. Often these analysis stages are applied disregarding the fact that the data is drawn from a time series. In this paper we propose a simple model for accounting for the underlying temporal nature of the data based on a Gaussian process.

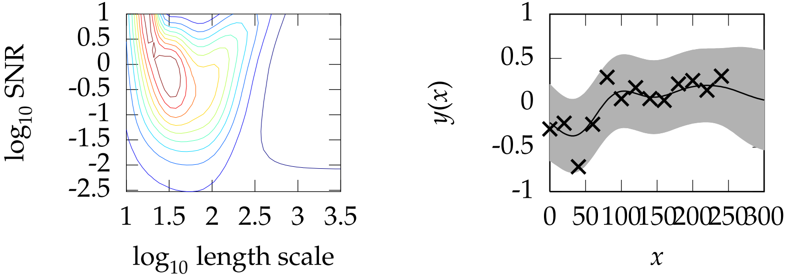
Results: We review Gaussian process (GP) regression for estimating the continuous trajectories underlying in gene expression time-series. We present a simple approach which can be used to filter quiet genes, or for the case of time series in the form of expression ratios, quantify differential expression. We assess via ROC curves the rankings produced by our regression framework and compare them to a recently proposed hierarchical Bayesian model for the analysis of gene expression time-series (BATS). We compare on both simulated and experimental data showing that the proposed approach considerably outperforms the current state of the art.



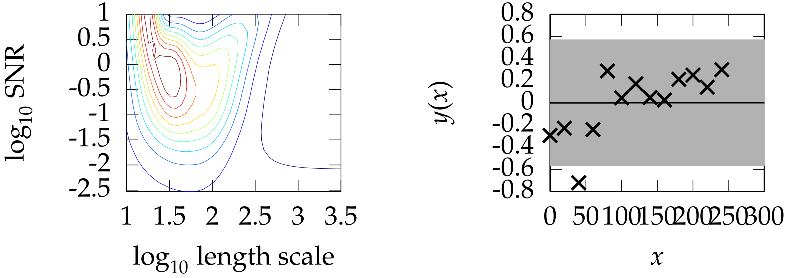
Contour plot of Gaussian process likelihood.



Optima: length scale of 1.2221 and $\log_{10}$ SNR of 1.9654 log likelihood is -0.22317.



Optima: length scale of 1.5162 and $\log_{10}$ SNR of 0.21306 log likelihood is -0.23604.



Optima: length scale of 2.9886 and $\log_{10}$ SNR of -4.506 log likelihood is -2.1056.

Basis Function Form

Radial basis functions commonly have the form

$$\phi_k(\mathbf{x}_i) = \exp\left(-\frac{|\mathbf{x}_i - \boldsymbol{\mu}_k|^2}{2\ell^2}\right).$$

- Basis function maps data into a "feature space" in which a linear sum is a non linear function.

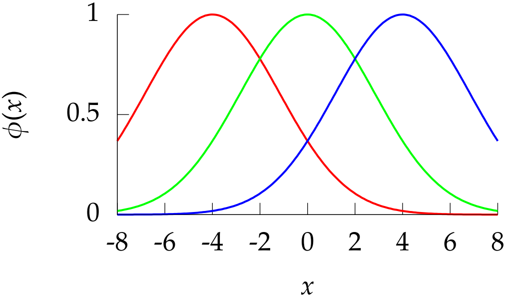


Figure : A set of radial basis functions with width $\ell = 2$ and location parameters $\boldsymbol{\mu} = [-4\ \ 0\ \ 4]^\top$.

Basis Function Representations

- Represent a function by a linear sum over a basis,

$$f(\mathbf{x}_{i,:}; \mathbf{w}) = \sum_{k=1}^{m} w_k \phi_k(\mathbf{x}_{i,:}), \tag{1}$$

- Here: m basis functions and $\phi_k(\cdot)$ is kth basis function and

$$\mathbf{w} = [w_1, \ldots, w_m]^\top .$$

- For standard linear model: $\phi_k(\mathbf{x}_{i,:}) = x_{i,k}$.

Random Functions

Functions derived using:

$$f(x) = \sum_{k=1}^{m} w_k \phi_k(x),$$

where elements of $\mathbf{w}$ are independently sampled from a Gaussian density,

$$w_k \sim \mathcal{N}(0, \alpha).$$

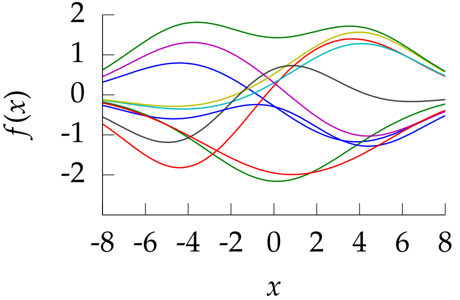


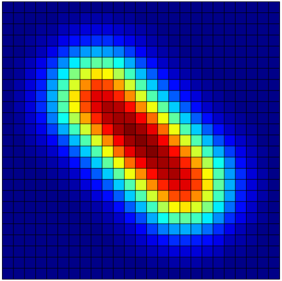
Figure : Functions sampled using the basis set from figure 4. Each line is a separate sample, generated by a weighted sum of the basis set. The weights, $\mathbf{w}$ are sampled from a Gaussian density with variance $\alpha = 1$.

RBF Basis Functions

$$k(\mathbf{x}, \mathbf{x}') = \alpha \boldsymbol{\phi}(\mathbf{x})^\top \boldsymbol{\phi}(\mathbf{x}')$$

$$\phi_k(x) = \exp\left(-\frac{\left\|x - \mu_k\right\|_2^2}{\ell^2}\right)$$

$$\boldsymbol{\mu} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$



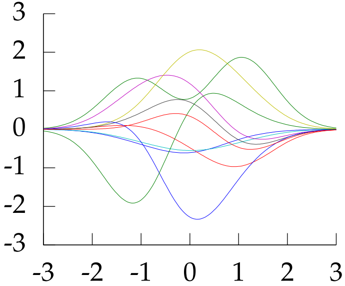
Covariance Functions

RBF Basis Functions

$$k(\mathbf{x}, \mathbf{x}') = \alpha \boldsymbol{\phi}(\mathbf{x})^\top \boldsymbol{\phi}(\mathbf{x}')$$

$$\phi_k(x) = \exp\left(-\frac{\left\|x - \mu_k\right\|_2^2}{\ell^2}\right)$$

$$\boldsymbol{\mu} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$



Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

2. Scaling a Gaussian leads to a Gaussian.

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

2. Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

Recall Univariate Gaussian Properties

1. Sum of Gaussian variables is also Gaussian.

$$y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$$

$$\sum_{i=1}^{n} y_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

2. Scaling a Gaussian leads to a Gaussian.

$$y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$wy \sim \mathcal{N}\left(w\mu, w^2\sigma^2\right)$$

Multivariate Consequence

- If

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Multivariate Consequence

- If

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- And

$$\mathbf{y} = \mathbf{W}\mathbf{x}$$

Multivariate Consequence

- If

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- And

$$\mathbf{y} = \mathbf{W}\mathbf{x}$$

- Then

$$\mathbf{y} \sim \mathcal{N}\left(\mathbf{W}\boldsymbol{\mu}, \mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^\top\right)$$

Basis Function Models

- If

$$f(\mathbf{x};\mathbf{w}) = \sum_{k=1}^{m} w_k \phi_k(\mathbf{x})$$

Basis Function Models

- If

$$f(\mathbf{x}; \mathbf{w}) = \mathbf{w}^\top \phi(\mathbf{x})$$

Basis Function Models

- If

$$f(\mathbf{x}; \mathbf{w}) = \mathbf{w}^\top \phi(\mathbf{x})$$

$$\mathbf{f} = \mathbf{\Phi} \mathbf{w}$$

Basis Function Models

- If
$$f(\mathbf{x}; \mathbf{w}) = \mathbf{w}^\top \phi(\mathbf{x})$$
$$\mathbf{f} = \mathbf{\Phi} \mathbf{w}$$
- If
$$\mathbf{w} \sim \mathcal{N}\left(\mathbf{0}, \alpha \mathbf{I}\right)$$
Then
$$\mathbf{f} \sim \mathcal{N}\left(\mathbf{0}, \alpha \mathbf{\Phi} \mathbf{\Phi}^\top\right)$$

Selecting Number and Location of Basis

- Need to choose
  1. location of centers
  2. number of basis functions

  Restrict analysis to 1-D input, x.
- Consider uniform spacing over a region:

$$k\left(x_i, x_j\right) = \alpha \phi_k(x_i)^\top \phi_k(x_j)$$

Selecting Number and Location of Basis

- Need to choose
  1. location of centers
  2. number of basis functions

  Restrict analysis to 1-D input, x.
- Consider uniform spacing over a region:

$$k\left(x_i, x_j\right) = \alpha \sum_{k=1}^{m} \phi_k(x_i)\phi_k(x_j)$$

Selecting Number and Location of Basis

- Need to choose
  1. location of centers
  2. number of basis functions

  Restrict analysis to 1-D input, x.
- Consider uniform spacing over a region:

$$k\left(x_i, x_j\right) = \alpha \sum_{k=1}^{m} \exp\left(-\frac{(x_i - \mu_k)^2}{2\ell^2}\right) \exp\left(-\frac{(x_j - \mu_k)^2}{2\ell^2}\right)$$

Selecting Number and Location of Basis

- Need to choose
  1. location of centers
  2. number of basis functions

  Restrict analysis to 1-D input, x.
- Consider uniform spacing over a region:

$$k\left(x_i, x_j\right) = \alpha \sum_{k=1}^{m} \exp\left(-\frac{(x_i - \mu_k)^2}{2\ell^2} - \frac{(x_j - \mu_k)^2}{2\ell^2}\right)$$

Selecting Number and Location of Basis

- Need to choose
  1. location of centers
  2. number of basis functions

  Restrict analysis to 1-D input, x.
- Consider uniform spacing over a region:

$$k\left(x_i, x_j\right) = \alpha \sum_{k=1}^{m} \exp\left(-\frac{x_i^2 + x_j^2 - 2\mu_k\left(x_i + x_j\right) + 2\mu_k^2}{2\ell^2}\right),$$

Uniform Basis Functions

- Set each center location to

$$\mu_k = a + \Delta\mu \cdot (k - 1).$$

Uniform Basis Functions

- Set each center location to

$$\mu_k = a + \Delta\mu \cdot (k - 1).$$

- Specify the basis functions in terms of their indices,

$$\begin{aligned} k\left(x_i, x_j\right) =& \alpha' \Delta\mu \sum_{k=1}^{m} \exp\Bigg(- \frac{x_i^2 + x_j^2}{2\ell^2} \\ & - \frac{2\left(a + \Delta\mu \cdot (k-1)\right)\left(x_i + x_j\right) + 2\left(a + \Delta\mu \cdot (k-1)\right)^2}{2\ell^2} \Bigg). \end{aligned}$$

Uniform Basis Functions

- Set each center location to

$$\mu_k = a + \Delta\mu \cdot (k-1).$$

- Specify the basis functions in terms of their indices,

$$\begin{aligned} k\left(x_i, x_j\right) =& \alpha' \Delta\mu \sum_{k=1}^{m} \exp\Bigg(- \frac{x_i^2 + x_j^2}{2\ell^2} \\ & - \frac{2\left(a + \Delta\mu \cdot (k-1)\right)\left(x_i + x_j\right) + 2\left(a + \Delta\mu \cdot (k-1)\right)^2}{2\ell^2}\Bigg). \end{aligned}$$

- Here we've scaled variance of process by $\Delta\mu$.

Infinite Basis Functions

- Take
$$\mu_1 = a \text{ and } \mu_m = b \text{ so } b = a + \Delta\mu \cdot (m-1)$$

Infinite Basis Functions

- Take
$$\mu_1 = a \text{ and } \mu_m = b \text{ so } b = a + \Delta\mu \cdot (m-1)$$

- This implies
$$b - a = \Delta\mu(m-1)$$

Infinite Basis Functions

- Take
$$\mu_1 = a \text{ and } \mu_m = b \text{ so } b = a + \Delta\mu \cdot (m-1)$$

- This implies
$$b - a = \Delta\mu(m-1)$$
and therefore
$$m = \frac{b-a}{\Delta\mu} + 1$$

Infinite Basis Functions

- Take
$$\mu_1 = a \text{ and } \mu_m = b \text{ so } b = a + \Delta\mu \cdot (m-1)$$

- This implies
$$b - a = \Delta\mu(m-1)$$
and therefore
$$m = \frac{b-a}{\Delta\mu} + 1$$

- Take limit as $\Delta\mu \to 0$ so $m \to \infty$

Infinite Basis Functions

- Take

$$\mu_1 = a \text{ and } \mu_m = b \text{ so } b = a + \Delta\mu \cdot (m-1)$$

- This implies

$$b - a = \Delta\mu(m-1)$$

and therefore

$$m = \frac{b-a}{\Delta\mu} + 1$$

- Take limit as $\Delta\mu \to 0$ so $m \to \infty$

$$k(x_i, x_j) = \alpha' \int_a^b \exp\left(-\frac{x_i^2 + x_j^2}{2\ell^2} + \frac{2\left(\mu - \frac{1}{2}\left(x_i + x_j\right)\right)^2 - \frac{1}{2}\left(x_i + x_j\right)^2}{2\ell^2}\right) \mathrm{d}\mu,$$

where we have used $a + k \cdot \Delta\mu \to \mu$.

Result

- Performing the integration leads to

$$k(x_i, x_j) = \alpha' \sqrt{\pi \ell^2} \exp\left(-\frac{\left(x_i - x_j\right)^2}{4\ell^2} \right)$$

$$\times \frac{1}{2} \left[\operatorname{erf}\left(\frac{\left(b - \frac{1}{2}\left(x_i + x_j\right)\right)}{\ell} \right) - \operatorname{erf}\left(\frac{\left(a - \frac{1}{2}\left(x_i + x_j\right)\right)}{\ell} \right) \right],$$

Result

- Performing the integration leads to

$$k(x_i,x_j) = \alpha' \sqrt{\pi\ell^2} \exp\left(-\frac{\left(x_i - x_j\right)^2}{4\ell^2}\right)$$
$$\times \frac{1}{2}\left[\operatorname{erf}\left(\frac{\left(b - \frac{1}{2}\left(x_i + x_j\right)\right)}{\ell}\right) - \operatorname{erf}\left(\frac{\left(a - \frac{1}{2}\left(x_i + x_j\right)\right)}{\ell}\right)\right],$$

- Now take limit as $a \to -\infty$ and $b \to \infty$

Result

- Performing the integration leads to

$$k(x_i,x_j) = \alpha' \sqrt{\pi\ell^2} \exp\left(-\frac{\left(x_i - x_j\right)^2}{4\ell^2}\right)$$
$$\times \frac{1}{2}\left[\operatorname{erf}\left(\frac{\left(b - \frac{1}{2}\left(x_i + x_j\right)\right)}{\ell}\right) - \operatorname{erf}\left(\frac{\left(a - \frac{1}{2}\left(x_i + x_j\right)\right)}{\ell}\right)\right],$$

- Now take limit as $a \to -\infty$ and $b \to \infty$

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left(x_i - x_j\right)^2}{4\ell^2}\right).$$

where $\alpha = \alpha' \sqrt{\pi\ell^2}$.

Infinite Feature Space

- An RBF model with infinite basis functions is a Gaussian process.

Infinite Feature Space

- An RBF model with infinite basis functions is a Gaussian process.
- The covariance function is given by the exponentiated quadratic covariance function.

$$k\left(x_i, x_j\right) = \alpha \exp\left(-\frac{\left(x_i - x_j\right)^2}{4\ell^2}\right).$$

Infinite Feature Space

- An RBF model with infinite basis functions is a Gaussian process.
- The covariance function is the exponentiated quadratic.
- **Note:** The functional form for the covariance function and basis functions are similar.
  - this is a special case,
  - in general they are very different

Similar results can obtained for multi-dimensional input models Williams (1998); Neal (1996).

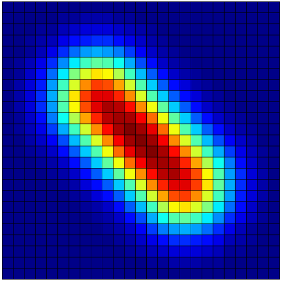
Covariance Functions

RBF Basis Functions

$$k\left(\mathbf{x}, \mathbf{x}'\right) = \alpha \boldsymbol{\phi}(\mathbf{x})^\top \boldsymbol{\phi}(\mathbf{x}')$$

$$\phi_k(x) = \exp\left(-\frac{\left\|x - \mu_k\right\|_2^2}{\ell^2}\right)$$

$$\boldsymbol{\mu} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$



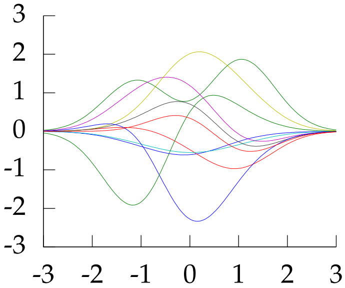
Covariance Functions

RBF Basis Functions

$$k(\mathbf{x}, \mathbf{x}') = \alpha \boldsymbol{\phi}(\mathbf{x})^\top \boldsymbol{\phi}(\mathbf{x}')$$

$$\phi_k(x) = \exp\left(-\frac{\left\|x - \mu_k\right\|_2^2}{\ell^2}\right)$$

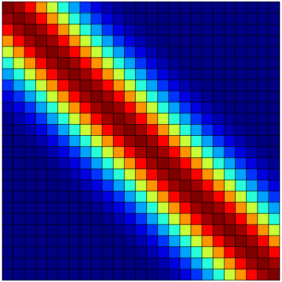
$$\boldsymbol{\mu} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$



Exponentiated Quadratic Kernel Function (RBF, Squared Exponential, Gaussian)

$$k(\mathbf{x}, \mathbf{x}') = \alpha \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|_2^2}{2\ell^2}\right)$$

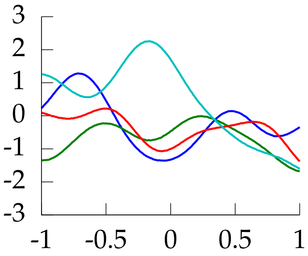
- Covariance matrix is built using the *inputs* to the function $\mathbf{x}$.
- For the example above it was based on Euclidean distance.
- The covariance function is also know as a kernel.



Exponentiated Quadratic Kernel Function (RBF, Squared Exponential, Gaussian)

$$k(\mathbf{x}, \mathbf{x}') = \alpha \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|_2^2}{2\ell^2}\right)$$

- Covariance matrix is built using the *inputs* to the function $\mathbf{x}$.
- For the example above it was based on Euclidean distance.
- The covariance function is also know as a kernel.



Covariance Functions

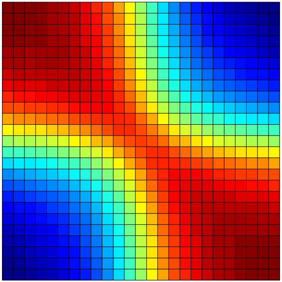
MLP Covariance Function

$$k(\mathbf{x}, \mathbf{x}') = \alpha \operatorname{asin}\left(\frac{w\mathbf{x}^\top \mathbf{x}' + b}{\sqrt{w\mathbf{x}^\top \mathbf{x} + b + 1}\sqrt{w\mathbf{x}'^\top \mathbf{x}' + b + 1}}\right)$$

- Based on infinite neural network model.

$$w = 40$$

$$b = 4$$



Covariance Functions

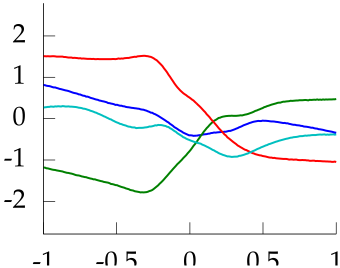
MLP Covariance Function

$$k(\mathbf{x}, \mathbf{x}') = \alpha \operatorname{asin}\left(\frac{w\mathbf{x}^\top\mathbf{x}' + b}{\sqrt{w\mathbf{x}^\top\mathbf{x} + b + 1}\sqrt{w\mathbf{x}'^\top\mathbf{x}' + b + 1}}\right)$$

- Based on infinite neural network model.

$$w = 40$$

$$b = 4$$



Constructing Covariance Functions

- Sum of two covariances is also a covariance function.

$$k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}') + k_2(\mathbf{x}, \mathbf{x}')$$

Constructing Covariance Functions

- Product of two covariances is also a covariance function.

$$k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}')k_2(\mathbf{x}, \mathbf{x}')$$

Multiply by Deterministic Function

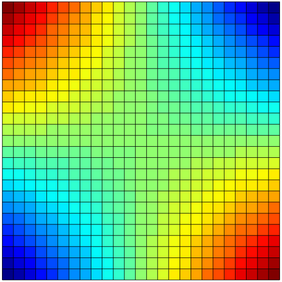
- If $f(\mathbf{x})$ is a Gaussian process.
- $g(\mathbf{x})$ is a deterministic function.
- $h(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x})$
- Then
$$k_h(\mathbf{x}, \mathbf{x}') = g(\mathbf{x})k_f(\mathbf{x}, \mathbf{x}')g(\mathbf{x}')$$
where k_h is covariance for $h(\cdot)$ and k_f is covariance for $f(\cdot)$.

Linear Covariance Function

$$k(\mathbf{x}, \mathbf{x}') = \alpha \mathbf{x}^\top \mathbf{x}'$$

- Bayesian linear regression.

$$\alpha = 1$$

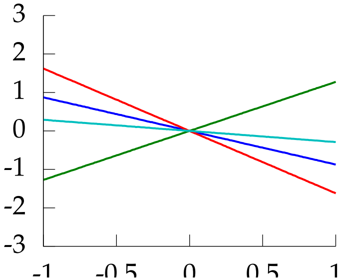


Linear Covariance Function

$$k(\mathbf{x}, \mathbf{x}') = \alpha \mathbf{x}^\top \mathbf{x}'$$

- Bayesian linear regression.

$$\alpha = 1$$



Bochner’s Theorem

Given a positive finite Borel measure μ on the real line $\mathbb{R}$, the Fourier transform Q of μ is the continuous function

$$Q(t) = \int_{\mathbb{R}} e^{-itx} \mathrm{d}\mu(x).$$

Q is continuous since for a fixed x, the function e^{-itx} is continuous and periodic. The function Q is a positive definite function, i.e. the kernel $k(x, x') = Q(x' - x)$ is positive definite.

Bochner’s theorem says the converse is true, i.e. every positive definite function Q is the Fourier transform of a positive finite Borel measure. A proof can be sketched as follows.

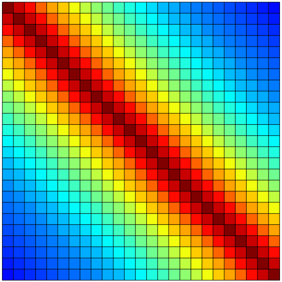
Covariance Functions

Where did this covariance matrix come from?

Ornstein-Uhlenbeck (stationary Gauss-Markov) covariance function

$$k(\mathbf{x}, \mathbf{x}') = \alpha \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|}{2\ell^2}\right)$$

- In one dimension arises from a stochastic differential equation. Brownian motion in a parabolic tube.
- In higher dimension a Fourier filter of the form $\frac{1}{\pi(1+x^2)}$.



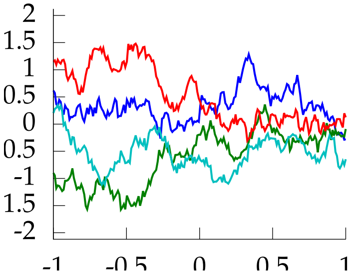
Covariance Functions

Where did this covariance matrix come from?

Ornstein-Uhlenbeck (stationary Gauss-Markov) covariance function

$$k(\mathbf{x}, \mathbf{x}') = \alpha \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|}{2\ell^2}\right)$$

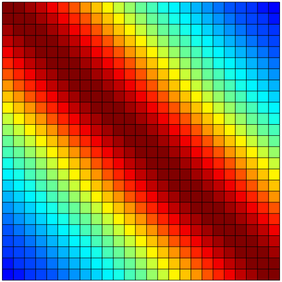
- In one dimension arises from a stochastic differential equation. Brownian motion in a parabolic tube.
- In higher dimension a Fourier filter of the form $\frac{1}{\pi(1+x^2)}$.



Matern 3/2 Covariance Function

$$k(\mathbf{x}, \mathbf{x}') = \alpha \left(1 + \sqrt{3}r\right) \exp\left(-\sqrt{3}r\right) \quad \text{where} \quad r = \frac{\|\mathbf{x} - \mathbf{x}'\|_2}{\ell}$$

- Matern 3/2 is a once differentiable covariance.
- Matern family constructed with Student-t filters in Fourier space.



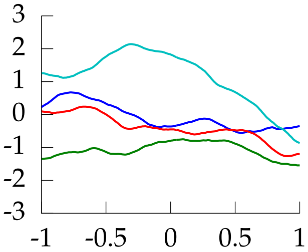
Covariance Functions

Where did this covariance matrix come from?

Matern 3/2 Covariance Function

$$k(\mathbf{x}, \mathbf{x}') = \alpha\left(1 + \sqrt{3}r\right)\exp\left(-\sqrt{3}r\right) \quad \text{where} \quad r = \frac{\|\mathbf{x} - \mathbf{x}'\|_2}{\ell}$$

- Matern 3/2 is a once differentiable covariance.
- Matern family constructed with Student-*t* filters in Fourier space.



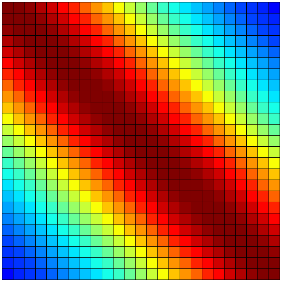
Covariance Functions

Where did this covariance matrix come from?

Matern 5/2 Covariance Function

$$k(\mathbf{x}, \mathbf{x}') = \alpha\left(1 + \sqrt{5}r + \frac{5}{3}r^2\right)\exp\left(-\sqrt{5}r\right) \quad \text{where} \quad r = \frac{\|\mathbf{x} - \mathbf{x}'\|_2}{\ell}$$

- Matern 5/2 is a twice differentiable covariance.
- Matern family constructed with Student-*t* filters in Fourier space.



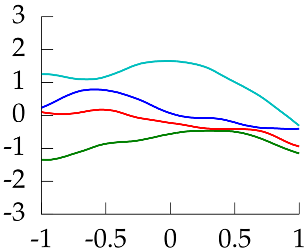
Covariance Functions

Where did this covariance matrix come from?

Matern 5/2 Covariance Function

$$k(\mathbf{x}, \mathbf{x}') = \alpha \left(1 + \sqrt{5} r + \frac{5}{3} r^2\right) \exp\left(-\sqrt{5} r\right) \quad \text{where} \quad r = \frac{\|\mathbf{x} - \mathbf{x}'\|_2}{\ell}$$

- Matern 5/2 is a twice differentiable covariance.
- Matern family constructed with Student-*t* filters in Fourier space.



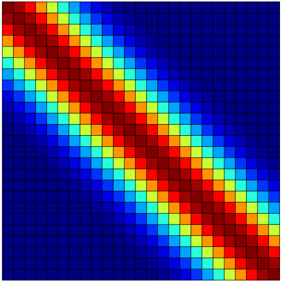
Covariance Functions

Where did this covariance matrix come from?

Exponentiated Quadratic Kernel Function (RBF, Squared Exponential, Gaussian)

$$k(\mathbf{x}, \mathbf{x}') = \alpha \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|_2^2}{2\ell^2}\right)$$

- Covariance matrix is built using the *inputs* to the function $\mathbf{x}$.
- For the example above it was based on Euclidean distance.
- The covariance function is also know as a kernel.



Covariance Functions

Where did this covariance matrix come from?

Exponentiated Quadratic Kernel Function (RBF, Squared Exponential, Gaussian)

$$k(\mathbf{x}, \mathbf{x}') = \alpha \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|_2^2}{2\ell^2}\right)$$

- Covariance matrix is built using the *inputs* to the function $\mathbf{x}$.
- For the example above it was based on Euclidean distance.
- The covariance function is also know as a kernel.