Unsupervised Learning with Gaussian Processes

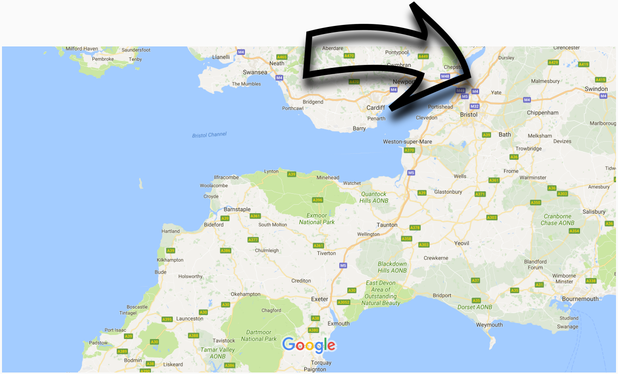
Carl Henrik Ek - carlhenrik.ek@bristol.ac.uk

September 12, 2017

http://www.carlhenrik.com

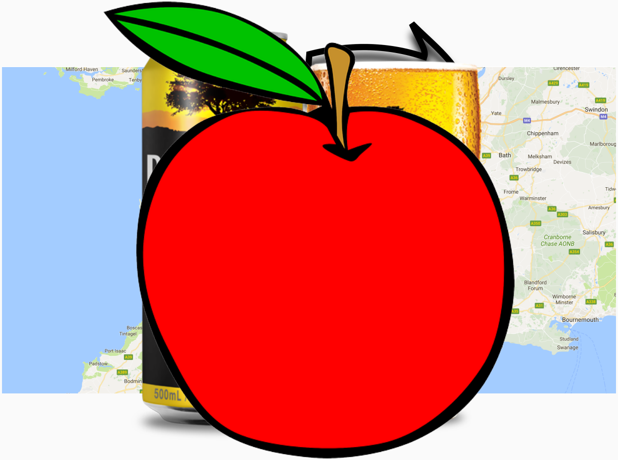
Introductions

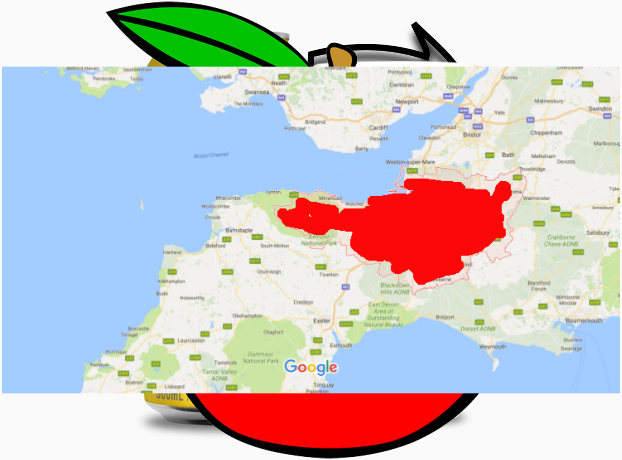
Me

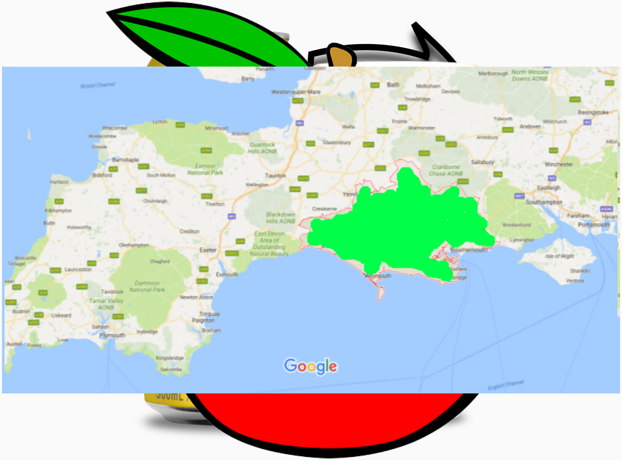

Milford Haven
Saundersfoot
Pembroke
Tenby
Llanelli
Neath
Swansea
The Mumbles
Aberdare
Pontypool
Cwmbran
Bridgend
Porthcawl
Cardiff
Penarth
Barry
Portishead
Clevedon
Bristol
Dursley
Cirencester
Malmesbury
Swindon
Yate
Chippenham
Marlborough
Bath
Melksham
Devizes
Trowbridge
Frome
Warminster
Amesbury
Salisbury
Bristol Channel
Weston-super-Mare
Wells
Glastonbury
Lynton
Ilfracombe
Woolacombe
Croyde
Minehead
Watchet
Quantock Hills AONB
Barnstaple
Exmoor National Park
Bideford
South Molton
Taunton
Wellington
Hartland
Chulmleigh
Tiverton
Yeovil
Crewkerne
Kilkhampton
Bude
Holsworthy
Blackdown Hills AONB
Cranborne Chase AONB
Blandford Forum
Wimborne Minster
Crediton
East Devon Area of Outstanding Natural Beauty
Okehampton
Exeter
Bridport
Dorset AONB
Bournemouth
Boscastle
Tintagel
Launceston
Chagford
Exmouth
Weymouth
Studland
Swanage
Port Isaac
Padstow
Dartmoor National Park
Tavistock
Tamar Valley AONB
Bodmin
Liskeard
Torquay
Paignton
Google

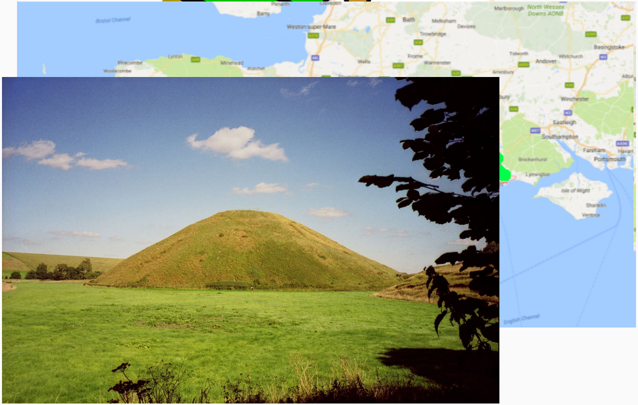

Milford Haven
Saundersfoot
Pembroke
Tenby
Llanelli
Neath
Swansea
The Mumbles
Aberdare
Pontypool
Newport
Cardiff
Penarth
Barry
Portishead
Clevedon
Bristol
Yate
Dursley
Cirencester
Malmesbury
Swindon
Chippenham
Marlborough
Bath
Melksham
Devizes
Trowbridge
Weston-super-Mare
Bristol Channel
Frome
Warminster
Amesbury
Salisbury
Cranborne Chase AONB
Ilfracombe
Woolacombe
Croyde
Hartland
Kilkhampton
Bude
Yeovil
Blandford Forum
Wimborne Minster
Bournemouth
Dorset AONB
Weymouth
Studland
Swanage
Boscastle
Tintagel
Launceston
Port Isaac
Padstow
Tavistock
Tamar Valley AONB
Bodmin
Liskeard
Google

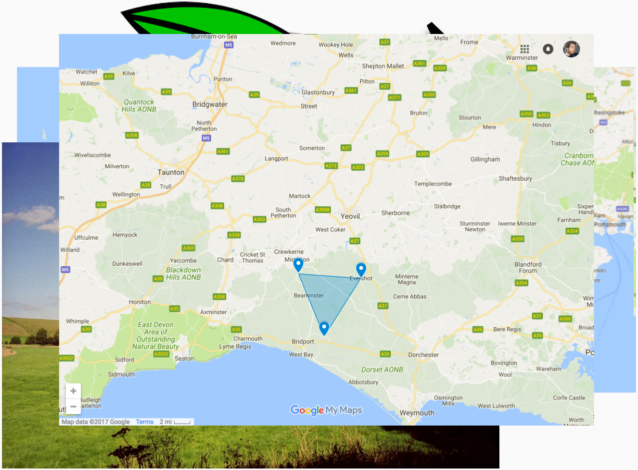

BLACKTHORN
A DISTINCTIVE CRISP, DRY TASTE
FROM ENGLISH CIDER APPLES
IMPORTED
Fermented Hard Cider
PRODUCT OF IRELAND
500mL / 16.9 fl. oz. 6% ALC./VOL.
BLACKTHORN
Milford Haven
Pembroke
Tenby
Tintagel
Port Isaac
Padstow
Bodmin
Cirencester
Dursley
Malmesbury
Swindon
Yate
Chippenham
Bath
Melksham
Devizes
Trowbridge
Frome
Warminster
Amesbury
Salisbury
Cranborne Chase AONB
Blandford Forum
Wimborne Minster
Bournemouth
Weymouth
Studland
Swanage

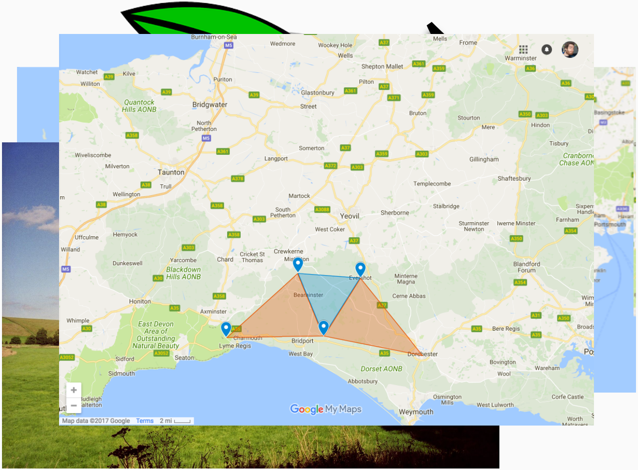

Milford Haven
Pembroke
Tenby
Tintagel
Port Isaac
Padstow
Bodmin
500mL
Cirencester
Dursley
Malmesbury
Swindon
Yate
Chippenham
Bath
Melksham
Devizes
Trowbridge
Frome
Warminster
Amesbury
Salisbury
Cranborne
Chase AONB
Blandford
Forum
Wimborne
Minster
Bournemouth
Studland
Swanage


Swansea
Neath
Bridgend
Cardiff
Newport
Cwmbran
Bristol
Bath
Chippenham
Swindon
Weston-super-Mare
Minehead
Lynton
Ilfracombe
Barnstaple
Bideford
Tiverton
Crediton
Exeter
Exmouth
Okehampton
Launceston
Tavistock
Dartmoor National Park
Exmoor National Park
Blackdown Hills AONB
East Devon Area of Outstanding Natural Beauty
Bridport
Dorset AONB
Weymouth
Bournemouth
Salisbury
Cranborne Chase AONB
Torquay
Google

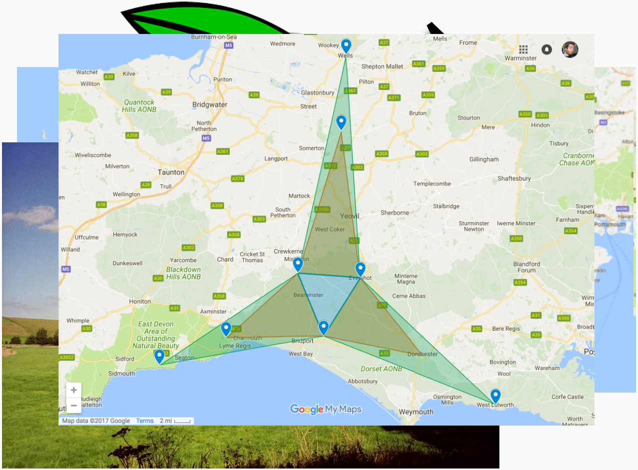

Bath
Weston-super-Mare
Taunton
Exeter
Exmouth
Plymouth
Barnstaple
Salisbury
Winchester
Southampton
Portsmouth
Bournemouth
Weymouth
Andover
Basingstoke
Dartmoor National Park
Exmoor National Park
Google

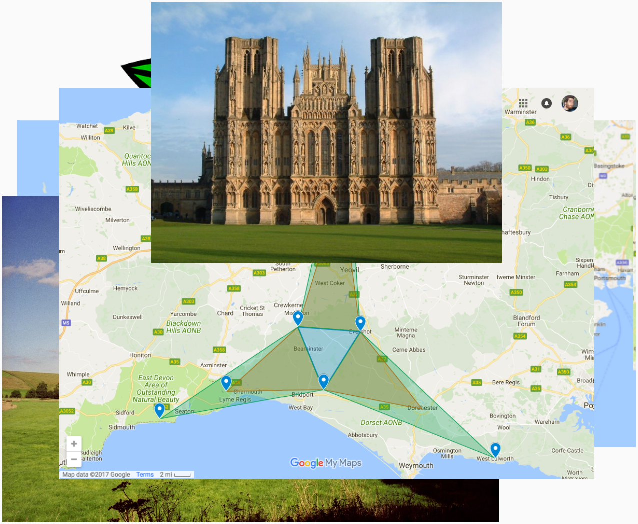

Barry
Weston-super-Mare
Bath
Melksham
Trowbridge
Devizes
Marlborough
North Wessex Downs AONB
Basingstoke
Tidworth
Andover
Frome
Warminster
Wells
Lynton
Ilfracombe
Minehead
Winchester
Eastleigh
Southampton
Fareham
Portsmouth
Lymington
Isle of Wight
Shanklin
Ventnor

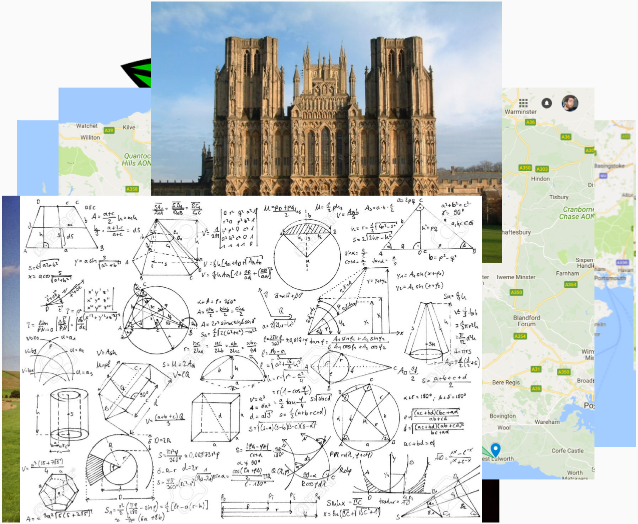

Burnham-on-Sea
Wedmore
Wookey Hole
Wells
Mells
Frome
Warminster
Watchet
Kilve
Williton
Puriton
Shepton Mallet
Pilton
Glastonbury
Street
Quantock Hills AONB
Bridgwater
North Petherton
Bruton
Stourton
Mere
Hindon
Tisbury
Cranborn Chase AON
Wiveliscombe
Milverton
Taunton
Trull
Somerton
Langport
Gillingham
Shaftesbury
Templecombe
Wellington
Martock
South Petherton
Yeovil
Sherborne
Stalbridge
Sturminster Newton
Iwerne Minster
Farnham
West Coker
Uffculme
Hemyock
Willand
Dunkeswell
Yarcombe
Blackdown Hills AONB
Chard
Cricket St Thomas
Crewkerne
Evershot
Minterne Magna
Blandford Forum
Beaminster
Cerne Abbas
Honiton
Axminster
Whimple
East Devon Area of Outstanding Natural Beauty
Charmouth
Lyme Regis
Bridport
West Bay
Bere Regis
Dorchester
Sidford
Seaton
Sidmouth
Dorset AONB
Bovington
Wool
Wareham
Abbotsbury
Osmington Mills
West Lulworth
Corfe Castle
Weymouth
Worth
Google My Maps
Map data ©2017 Google
Terms
2 mi

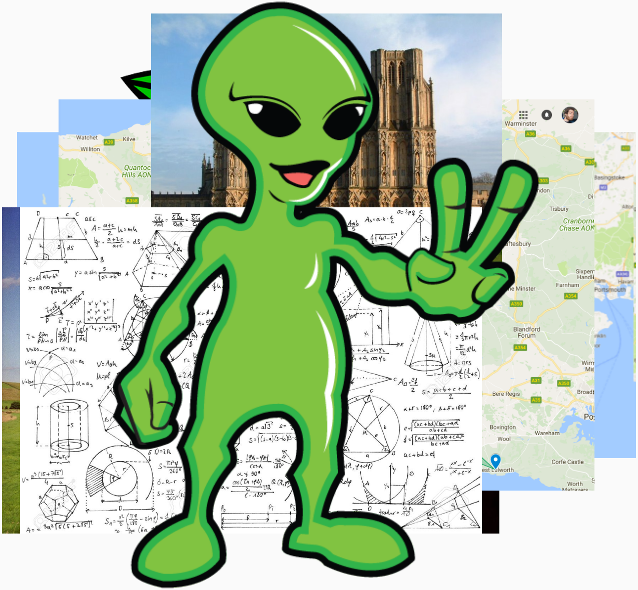

Burnham-on-Sea
Wedmore
Wookey Hole
Wells
Mells
Frome
Warminster
Watchet
Kilve
Williton
Puriton
Shepton Mallet
Pilton
Glastonbury
Street
Quantock Hills AONB
Bridgwater
North Petherton
Bruton
Stourton
Mere
Hindon
Tisbury
Cranborne Chase AON
Wiveliscombe
Milverton
Taunton
Trull
Somerton
Langport
Gillingham
Shaftesbury
Templecombe
Wellington
Martock
South Petherton
Yeovil
Sherborne
Stalbridge
Sturminster Newton
Iwerne Minster
Farnham
West Coker
Uffculme
Hemyock
Willand
Crewkerne
Chard
Cricket St Thomas
Yarcombe
Dunkeswell
Blackdown Hills AONB
Evershot
Minterne Magna
Cerne Abbas
Blandford Forum
Beaminster
Honiton
Axminster
Whimple
East Devon Area of Outstanding Natural Beauty
Charmouth
Lyme Regis
Bridport
West Bay
Bere Regis
Dorchester
Seaton
Sidford
Sidmouth
Dorset AONB
Bovington
Wool
Wareham
Abbotsbury
Osmington Mills
West Lulworth
Corfe Castle
Weymouth
Google My Maps
Map data ©2017 Google
Terms
2 mi


Burnham-on-Sea
Wedmore
Wookey Hole
Wells
Mells
Frome
Warminster
Watchet
Kilve
Williton
Puriton
Shepton Mallet
Pilton
Glastonbury
Street
Quantock Hills AONB
Bridgwater
North Petherton
Bruton
Stourton
Mere
Hindon
Tisbury
Wiveliscombe
Milverton
Somerton
Langport
Gillingham
Cranborn Chase AON
Taunton
Trull
Templecombe
Shaftesbury
Wellington
Martock
South Petherton
Yeovil
Sherborne
Stalbridge
Sturminster Newton
Iwerne Minster
Farnham
West Coker
Uffculme
Hemyock
Willand
Crewkerne
Cricket St Thomas
Chard
Yarcombe
Dunkeswell
Blackdown Hills AONB
Misterton
Evershot
Minterne Magna
Blandford Forum
Beaminster
Cerne Abbas
Honiton
Axminster
Whimple
East Devon Area of Outstanding Natural Beauty
Charmouth
Lyme Regis
Bridport
Bere Regis
Seaton
West Bay
Dorchester
Sidford
Sidmouth
Dorset AONB
Bovington
Wool
Wareham
Abbotsbury
Osmington Mills
West Lulworth
Corfe Castle
Weymouth
Google My Maps
Map data ©2017 Google
Terms
2 mi

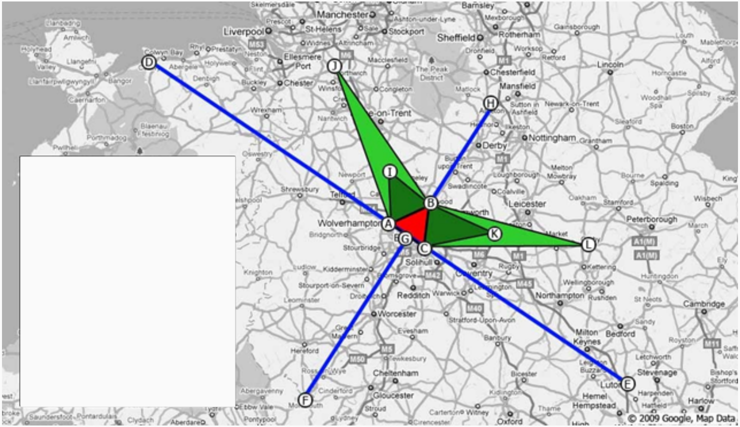

Burnham-on-Sea
Wedmore
Wookey
Wells
Mells
Frome
Warminster
Watchet
Kilve
Williton
Puriton
Shepton Mallet
Pilton
Glastonbury
Street
Quantock Hills AONB
Bridgwater
North Petherton
Bruton
Stourton
Mere
Hindon
Tisbury
Wiveliscombe
Milverton
Taunton
Trull
Somerton
Langport
Gillingham
Cranborne Chase AONB
Shaftesbury
Templecombe
Wellington
Martock
South Petherton
Yeovil
Sherborne
Stalbridge
West Coker
Sturminster Newton
Iwerne Minster
Farnham
Uffculme
Hemyock
Willand
Dunkeswell
Yarcombe
Chard
Cricket St Thomas
Crewkerne
Blackdown Hills AONB
Minterne Magna
Blandford Forum
Beaminster
Cerne Abbas
Honiton
Axminster
Whimple
East Devon Area of Outstanding Natural Beauty
Charmouth
Lyme Regis
Bridport
West Bay
Bere Regis
Seaton
Sidford
Sidmouth
Dorchester
Dorset AONB
Bovington
Wareham
Wool
Abbotsbury
Osmington Mills
West Lulworth
Corfe Castle
Weymouth
Google My Maps
Map data ©2017 Google
Terms
2 mi


Watchet
Kilve
Williton
Wiveliscombe
Milverton
Wellington
Uffculme
Hemyock
Willand
Dunkeswell
Yarcombe
Blackdown Hills AONB
Chard
Cricket St Thomas
Crewkerne
Honiton
Axminster
Whimple
East Devon Area of Outstanding Natural Beauty
Sidford
Sidmouth
Seaton
Lyme Regis
Charmouth
Bridport
West Bay
Beaminster
Yeovil
West Coker
Sherborne
Minterne Magna
Cerne Abbas
Dorchester
Dorset AONB
Abbotsbury
Osmington Mills
Weymouth
West Lulworth
Bovington
Wool
Wareham
Corfe Castle
Bere Regis
Blandford Forum
Sturminster Newton
Iwerne Minster
Farnham
Warminster
Hindon
Tisbury
Shaftesbury
Google My Maps
Map data ©2017 Google
Terms
2 mi


Watchet
Kilve
Williton
A39
Quantock Hills AONB
A358
Warminster
A36
A350
A303
Hindon
Tisbury
Cranborne Chase AONB
Shaftesbury
Iwerne Minster
Farnham
Blandford Forum
A354
A31
Bere Regis
A35
Bovington
Wool
Wareham
Corfe Castle
West Lulworth
Worth Matravers
Basingstoke
Portsmouth


Watchet
Kilve
Williton
Warminster
Tisbury
Farnham
Blandford Forum
Bere Regis
Bovington
Wool
Wareham
Corfe Castle

"Brooks has proved, he explains, that there were keen mathematicians here 5,000 years ago, millennia before the Greeks invented geometry: "Such is the mathematical precision, it is inconceivable that this work could have been carried out by the primitive indigenous culture we have always associated with such structures . . . all this suggests a culture existing in these islands in the past quite outside our expectation and experience today." He does not rule out extraterrestrial help." – The Guardian



[1]Bad Science Blog

"We know so little about the ancient Woolworths stores," he explains, "but we do still know their locations. I thought that if we analysed the sites we could learn more about what life was like in 2008 and how these people went about buying cheap kitchen accessories and discount CDs" – Matt Parker interviewed in The Guardian[1]

[1] Bad Science Blog


Lord Nelson
PALE ALE

Napoleon *"You have written this huge book on the system of the world without once mentioning the author of the universe."*

Napoleon *"You have written this huge book on the system of the world without once mentioning the author of the universe."*

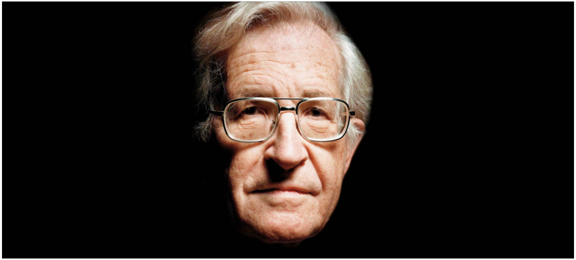
Laplace *"I had no need for that assumption"*

Napoleon *"You have written this huge book on the system of the world without once mentioning the author of the universe."*

Laplace *"I had no need for that assumption"*

Laplace *"Ah, but that is a fine hypothesis. It explains so many things"*

Inductivist Fallacy



2

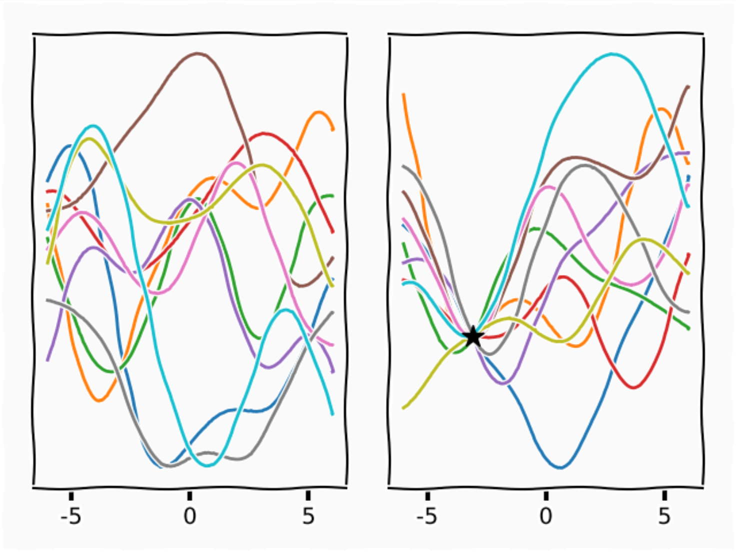
[2]Chomsky, N. A., & Fodor, J. A. (1980). The inductivist fallacy. Language and Learning: The Debate between Jean Piaget and Noam Chomsky, (), .

GPSS

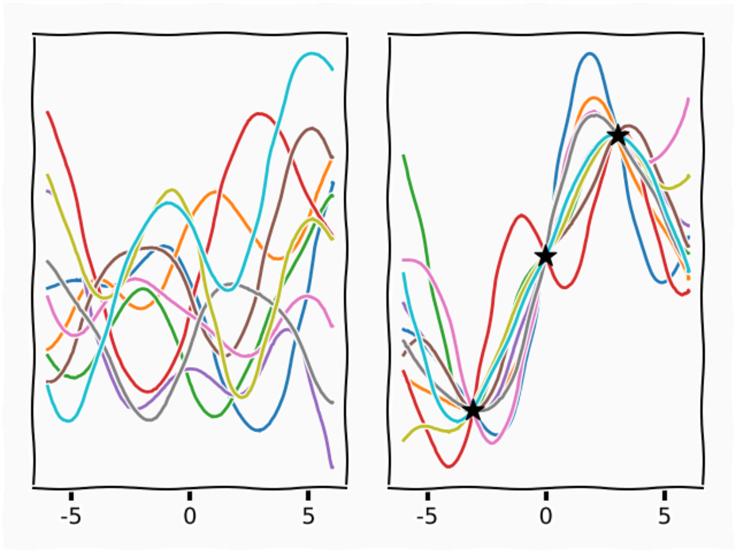
IUDICIUM POSTERIUM DISCIPULUS EST PRIORIS

Unsupervised Learning

Gaussian Processes



Gaussian Processes

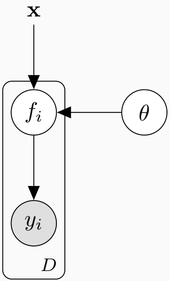


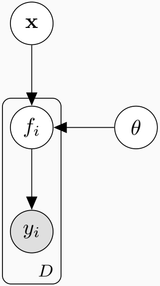
$$p(\theta|y) = p(y|\theta)\frac{p(\theta)}{p(y)}$$

"Scientific modelling is a scientific activity, the aim of which is to make a particular part or feature of the world easier to understand, define, quantify, visualize, or simulate by referencing it to existing and usually commonly accepted knowledge." [3]

[3] Wikipedia

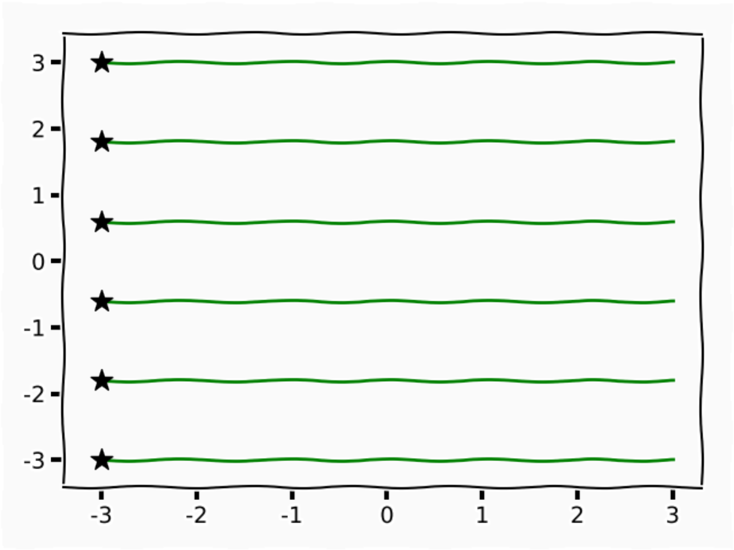
Unsupervised Learning



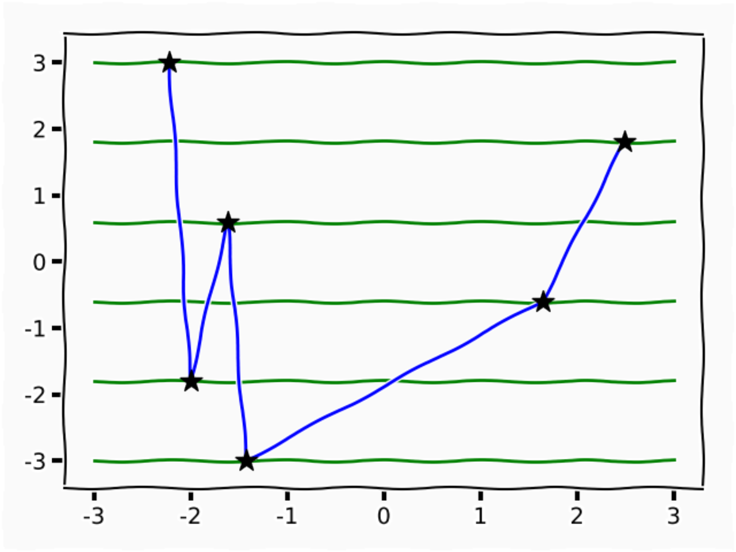


$p(y|x)$ $p(y)$

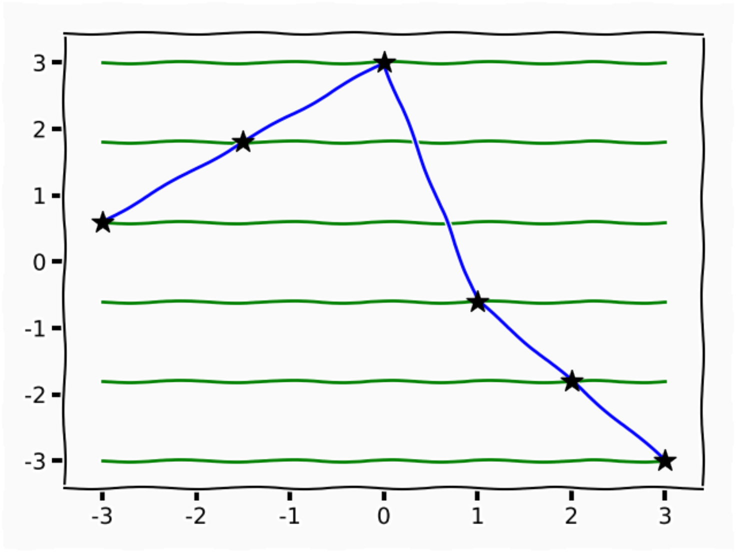
Unsupervised Learning



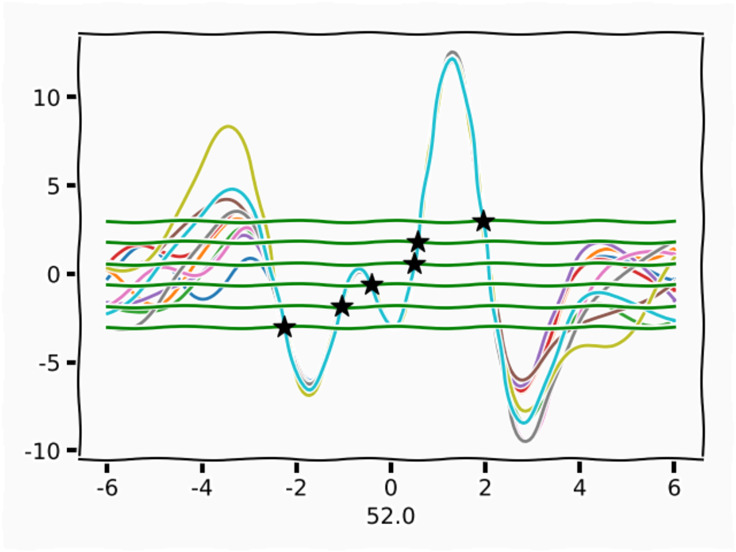
Unsupervised Learning



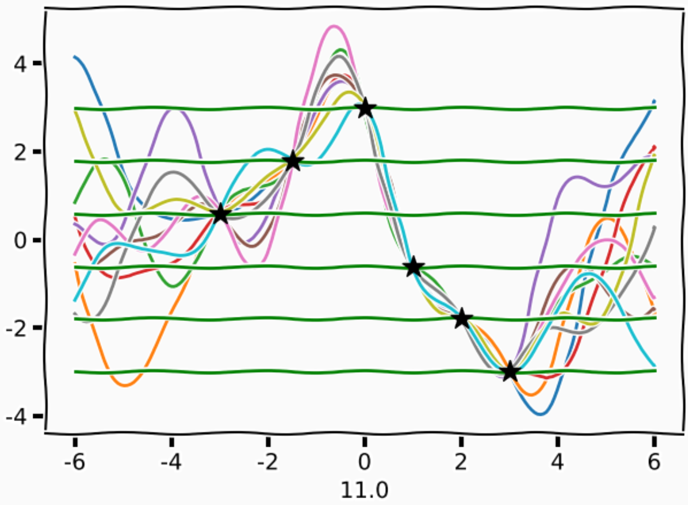
Unsupervised Learning



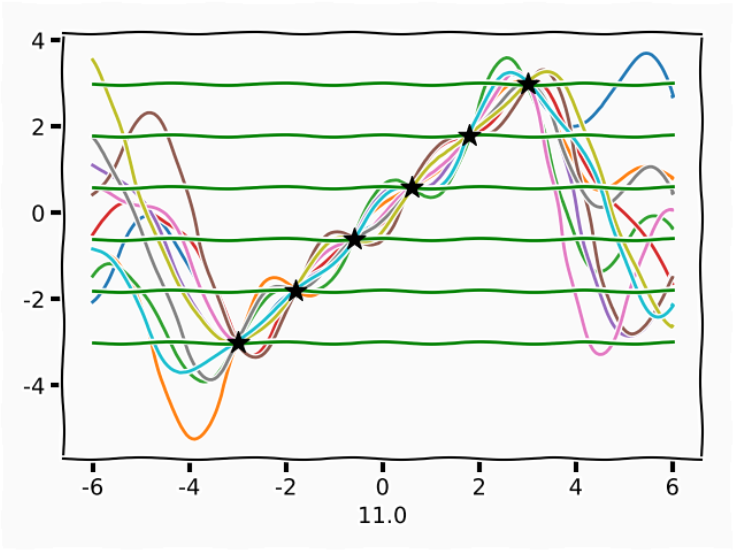
Unsupervised Learning



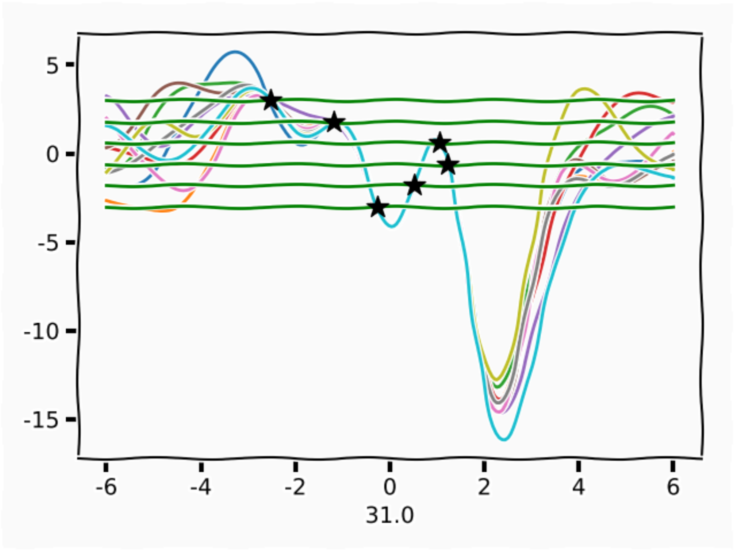
Unsupervised Learning



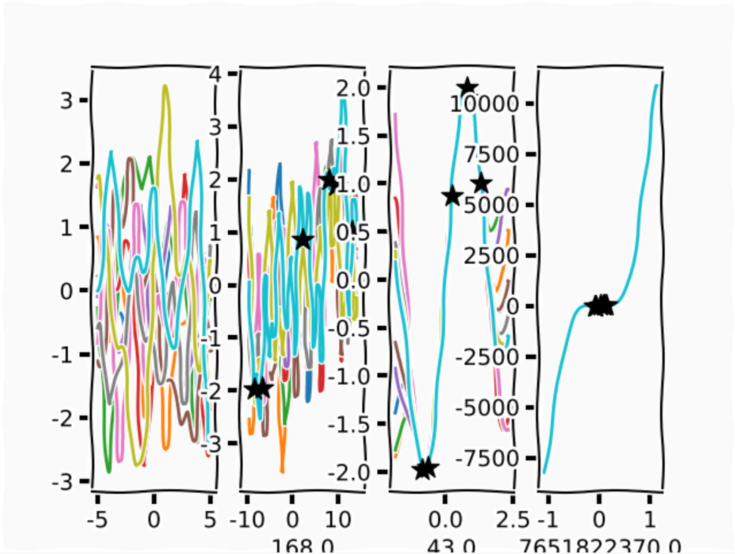
Unsupervised Learning



Unsupervised Learning



Unsupervised Learning



Being Bayesian[4]

[4] By Dieric Bouts (circa 1420-1475) - The Yorck Project: 10.000 Meisterwerke der Malerei, Public Domain, URL

Priors

$$p(y) = \int p(y|f)p(f|x)p(x)\mathrm{d}f\mathrm{d}x$$

$$p(x|y) = p(y|x)\frac{p(x)}{p(y)}$$

1. Priors that makes sense

   p(f) describes our belief/assumptions and defines our notion of complexity in the function

   p(x) expresses our belief/assumptions and defines our notion of complexity in the latent space

2. The priors are *"balanced"*
3. Now lets churn the handle

Relationship between x and data

$$p(y) = \int p(y|f)p(f|x)p(x)\mathrm{d}f\,\mathrm{d}x$$

- GP prior

$$p(f|x) \sim \mathcal{N}(0, K) \propto e^{-\frac{1}{2}(f^{\mathrm{T}}K^{-1}f)}$$

$$K_{ij} = e^{-(x_i - x_j)^{\mathrm{T}}M^{\mathrm{T}}M(x_i - x_j)}$$

Relationship between x and data

$$p(y) = \int p(y|f)p(f|x)p(x)\mathrm{d}f\mathrm{d}x$$

- GP prior

$$p(f|x) \sim \mathcal{N}(0, K) \propto e^{-\frac{1}{2}(f^{\mathrm{T}}K^{-1}f)}$$

$$K_{ij} = e^{-(x_i - x_j)^{\mathrm{T}}M^{\mathrm{T}}M(x_i - x_j)}$$

- Likelihood

$$p(y|f) \sim N(y|f, \beta) \propto e^{-\frac{1}{2\beta}\mathrm{tr}(y-f)^{\mathrm{T}}(y-f)}$$

Relationship between x and data

$$p(y) = \int p(y|f)p(f|x)p(x)\mathrm{d}f\mathrm{d}x$$

- GP prior

$$p(f|x) \sim \mathcal{N}(0, K) \propto e^{-\frac{1}{2}(f^{\mathrm{T}}K^{-1}f)}$$

$$K_{ij} = e^{-(x_i - x_j)^{\mathrm{T}}M^{\mathrm{T}}M(x_i - x_j)}$$

- Likelihood

$$p(y|f) \sim N(y|f, \beta) \propto e^{-\frac{1}{2\beta}\mathrm{tr}(y-f)^{\mathrm{T}}(y-f)}$$

- Analytically intractable (`Non Elementary Integral`) and infinitely differientiable

Laplace Integration

"Nature laughs at the difficulties of integrations"
– Simon Laplace

Unsupervised Learning with GPs

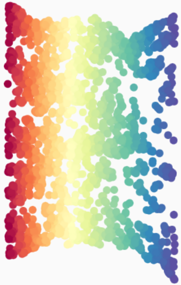
ML/MAP[5]

$$\hat{x} = \text{argmax}_x \int p(y|f)p(f|x)\mathrm{d}f p(x)$$

$$= \text{argmin}_x \frac{1}{2} y^{\mathrm{T}} \mathbf{K}^{-1} y + \frac{1}{2}|\mathbf{K}| - \log\, p(x)$$

[5] Lawrence, N. D. (2005). Probabilistic non-linear principal component analysis with Gaussian process latent variable models.

Dimensionality Reduction



$$y \in \mathbb{R}^d \quad x \in \mathbb{R}^q \quad d > q$$

GP-LVM

- Li, W., Viola, F., Starck, J., Brostow, G. J., & Campbell, N. D. (2016). Roto++: accelerating professional rotoscoping using shape manifolds. (In proceeding of ACM SIGGRAPH'16)
- Grochow, K., Martin, S. L., Hertzmann, A., & Popovi\'c, Zoran (2004). Style-based inverse kinematics. SIGGRAPH '04: SIGGRAPH 2004
- Urtasun, R., Fleet, D. J., & Fua, P. (2006). 3D people tracking with Gaussian process dynamical models. Computer Vision and Pattern Recognition, 2006

Bayesian GP-LVM[7]

- Challenges with ML estimation
  - How to initialise x?
  - What is the dimensionality q?
- *Our assumption on the latent space does not reach the data*

[6] Titsias, M. (2009). Variational learning of inducing variables in sparse Gaussian processes.
[7] Titsias, M., & Lawrence, N. D. (2010). Bayesian Gaussian Process Latent Variable Model

Bayesian GP-LVM[7]

- Challenges with ML estimation
  - How to initialise x?
  - What is the dimensionality q?
- *Our assumption on the latent space does not reach the data*
- Approximate integration![6]

[6] Titsias, M. (2009). Variational learning of inducing variables in sparse Gaussian processes.

[7] Titsias, M., & Lawrence, N. D. (2010). Bayesian Gaussian Process Latent Variable Model

Variational Bayes

$$p(\mathbf{Y})$$

Variational Bayes

$$\log p(\mathbf{Y})$$

Variational Bayes

$$\log p(\mathbf{Y}) = \log \int p(\mathbf{Y}, \mathbf{X}) \mathrm{d}\mathbf{X}$$

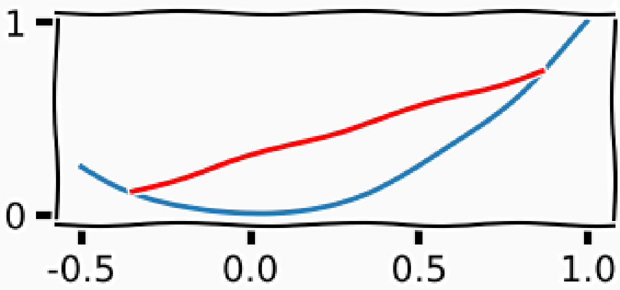
Variational Bayes

$$\log\, p(\mathbf{Y}) = \log \int p(\mathbf{Y}, \mathbf{X}) \mathrm{d}\mathbf{X} = \log \int p(\mathbf{X}|\mathbf{Y}) p(\mathbf{Y}) \mathrm{d}\mathbf{X}$$

Variational Bayes

$$\begin{aligned}\log\, p(\mathbf{Y}) &= \log \int p(\mathbf{Y},\mathbf{X})\mathrm{d}\mathbf{X} = \log \int p(\mathbf{X}|\mathbf{Y})p(\mathbf{Y})\mathrm{d}\mathbf{X} \\ &= \log \int \frac{q(\mathbf{X})}{q(\mathbf{X})} p(\mathbf{X}|\mathbf{Y})p(\mathbf{Y})\mathrm{d}\mathbf{X}\end{aligned}$$

Jensen Inequality



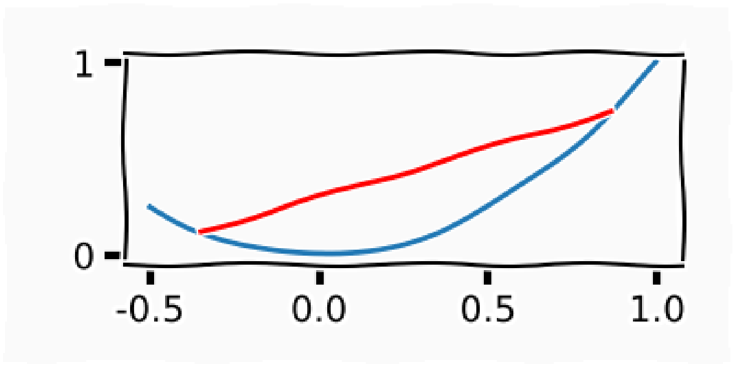
Convex Function

$$\lambda f(x_0) + (1-\lambda) f(x_1) \geq f(\lambda x_0 + (1-\lambda) x_1)$$

$$x \in [x_{min}, x_{max}]$$

$$\lambda \in [0, 1]]$$

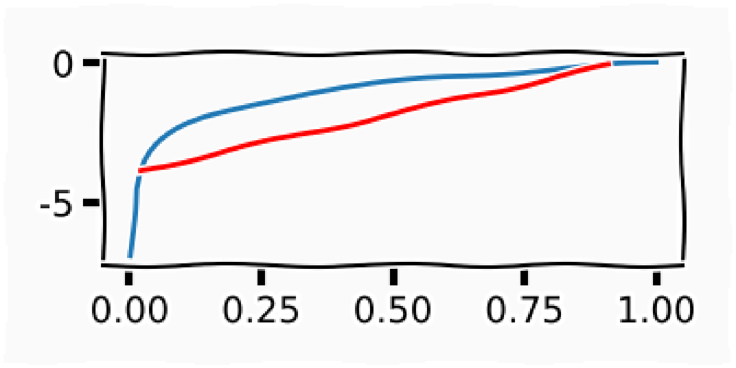
Jensen Inequality



$$\mathbb{E}[f(x)] \geq f(\mathbb{E}[x])$$

$$\int f(x)p(x)\mathrm{d}x \geq f\left(\int xp(x)\mathrm{d}x\right)$$

Jensen Inequality in Variational Bayes



$$\int \log(x)p(x)\mathrm{d}x \leq \log\left(\int xp(x)\mathrm{d}x\right)$$

moving the log inside the the integral is a lower-bound on the integral

Variational Bayes cont.

$$\log p(\mathbf{Y}) = \log \int \frac{q(\mathbf{X})}{q(\mathbf{X})} p(\mathbf{X}|\mathbf{Y}) p(\mathbf{Y}) \mathrm{d}\mathbf{X} =$$

Variational Bayes cont.

$$\begin{aligned}\log p(\mathbf{Y}) &= \log \int \frac{q(\mathbf{X})}{q(\mathbf{X})} p(\mathbf{X}|\mathbf{Y}) p(\mathbf{Y}) \mathrm{d}\mathbf{X} = \\ &\geq \int q(\mathbf{X}) \log \frac{p(\mathbf{X}|\mathbf{Y}) p(\mathbf{Y})}{q(\mathbf{X})} \mathrm{d}\mathbf{X}\end{aligned}$$

Variational Bayes cont.

$$
\begin{aligned}
\log p(\mathbf{Y}) &= \log \int \frac{q(\mathbf{X})}{q(\mathbf{X})} p(\mathbf{X}|\mathbf{Y}) p(\mathbf{Y}) \mathrm{d}\mathbf{X} = \\
&\geq \int q(\mathbf{X}) \log \frac{p(\mathbf{X}|\mathbf{Y}) p(\mathbf{Y})}{q(\mathbf{X})} \mathrm{d}\mathbf{X} \\
&= \int q(\mathbf{X}) \log \frac{p(\mathbf{X}|\mathbf{Y})}{q(\mathbf{X})} \mathrm{d}\mathbf{X} + \int q(\mathbf{X}) \mathrm{d}\mathbf{X} \log\, p(\mathbf{Y})
\end{aligned}
$$

Variational Bayes cont.

$$
\begin{aligned}
\log p(\mathbf{Y}) &= \log \int \frac{q(\mathbf{X})}{q(\mathbf{X})} p(\mathbf{X}|\mathbf{Y}) p(\mathbf{Y}) \mathrm{d}\mathbf{X} = \\
&\geq \int q(\mathbf{X}) \log \frac{p(\mathbf{X}|\mathbf{Y}) p(\mathbf{Y})}{q(\mathbf{X})} \mathrm{d}\mathbf{X} \\
&= \int q(\mathbf{X}) \log \frac{p(\mathbf{X}|\mathbf{Y})}{q(\mathbf{X})} \mathrm{d}\mathbf{X} + \int q(\mathbf{X}) \mathrm{d}\mathbf{X} \log\, p(\mathbf{Y}) \\
&= -\mathrm{KL}\left(q(\mathbf{X}) || p(\mathbf{X}|\mathbf{Y})\right) + \log\, p(\mathbf{Y})
\end{aligned}
$$

Variational Bayes cont.

$$
\begin{aligned}
\log p(\mathbf{Y}) &= \log \int \frac{q(\mathbf{X})}{q(\mathbf{X})} p(\mathbf{X}|\mathbf{Y}) p(\mathbf{Y}) \mathrm{d}\mathbf{X} = \\
&\geq \int q(\mathbf{X}) \log \frac{p(\mathbf{X}|\mathbf{Y}) p(\mathbf{Y})}{q(\mathbf{X})} \mathrm{d}\mathbf{X} \\
&= \int q(\mathbf{X}) \log \frac{p(\mathbf{X}|\mathbf{Y})}{q(\mathbf{X})} \mathrm{d}\mathbf{X} + \int q(\mathbf{X}) \mathrm{d}\mathbf{X} \log\, p(\mathbf{Y}) \\
&= -\mathrm{KL}\left(q(\mathbf{X}) || p(\mathbf{X}|\mathbf{Y})\right) + \log\, p(\mathbf{Y})
\end{aligned}
$$

- if $q(\mathbf{X})$ is the true posterior we have an equality, therefore match the distributions

Variational Bayes cont.

$$\begin{aligned}
\log p(\mathbf{Y}) &= \log \int \frac{q(\mathbf{X})}{q(\mathbf{X})} p(\mathbf{X}|\mathbf{Y})p(\mathbf{Y})\mathrm{d}\mathbf{X} = \\
&\geq \int q(\mathbf{X}) \log \frac{p(\mathbf{X}|\mathbf{Y})p(\mathbf{Y})}{q(\mathbf{X})} \mathrm{d}\mathbf{X} \\
&= \int q(\mathbf{X}) \log \frac{p(\mathbf{X}|\mathbf{Y})}{q(\mathbf{X})} \mathrm{d}\mathbf{X} + \int q(\mathbf{X}) \mathrm{d}\mathbf{X} \log\, p(\mathbf{Y}) \\
&= -\mathrm{KL}\left(q(\mathbf{X}) || p(\mathbf{X}|\mathbf{Y})\right) + \log\, p(\mathbf{Y})
\end{aligned}$$

- if $q(\mathbf{X})$ is the true posterior we have an equality, therefore match the distributions
- i.e. $\mathrm{argmin}_q \; \mathrm{KL}\left(q(\mathbf{X}) || p(\mathbf{X}|\mathbf{Y})\right)$

  $\Rightarrow$ variational distributions are approximations to intractable posteriors

ELBO

$$\mathrm{KL}(q(\mathbf{X})||p(\mathbf{X}|\mathbf{Y}))$$

ELBO

$$\mathrm{KL}(q(\mathbf{X})||p(\mathbf{X}|\mathbf{Y})) = \int q(\mathbf{X}) log \frac{q(\mathbf{X})}{p(\mathbf{X}|\mathbf{Y})} \mathrm{d}\mathbf{X}$$

ELBO

$$
\begin{aligned}
\mathrm{KL}(q(\mathsf{X})||p(\mathsf{X}|\mathsf{Y})) &= \int q(\mathsf{X}) log \frac{q(\mathsf{X})}{p(\mathsf{X}|\mathsf{Y})} \mathrm{d}\mathsf{X} \\
&= \int q(\mathsf{X}) log \frac{q(\mathsf{X})}{p(\mathsf{X},\mathsf{Y})} \mathrm{d}\mathsf{X} + log\ \ p(\mathsf{Y})
\end{aligned}
$$

ELBO

$$
\begin{aligned}
\mathrm{KL}(q(\mathbf{X})||p(\mathbf{X}|\mathbf{Y})) &= \int q(\mathbf{X}) log \frac{q(\mathbf{X})}{p(\mathbf{X}|\mathbf{Y})} \mathrm{d}\mathbf{X} \\
&= \int q(\mathbf{X}) log \frac{q(\mathbf{X})}{p(\mathbf{X},\mathbf{Y})} \mathrm{d}\mathbf{X} + log\ p(\mathbf{Y}) \\
&= H(q(\mathbf{X})) - \mathbb{E}_{q(\mathbf{X})}\left[log\ p(\mathbf{X},\mathbf{Y})\right] + log\ p(\mathbf{Y})
\end{aligned}
$$

ELBO

$$log\ \ p(\mathbf{Y}) = \mathrm{KL}(q(\mathbf{X})||p(\mathbf{X}|\mathbf{Y})) + \underbrace{\mathbb{E}_{q(\mathbf{X})}\left[log\ \ p(\mathbf{X},\mathbf{Y})\right] - H(q(\mathbf{X}))}_{\mathrm{ELBO}}$$

ELBO

$$log\ p(\mathbf{Y}) = \mathrm{KL}(q(\mathbf{X})||p(\mathbf{X}|\mathbf{Y})) + \underbrace{\mathbb{E}_{q(\mathbf{X})}\left[log\ p(\mathbf{X},\mathbf{Y})\right] - H(q(\mathbf{X}))}_{\mathrm{ELBO}}$$

$$\geq \mathbb{E}_{q(\mathbf{X})}\left[log\ p(\mathbf{X},\mathbf{Y})\right] - H(q(\mathbf{X})) = \mathcal{L}(q(\mathbf{X}))$$

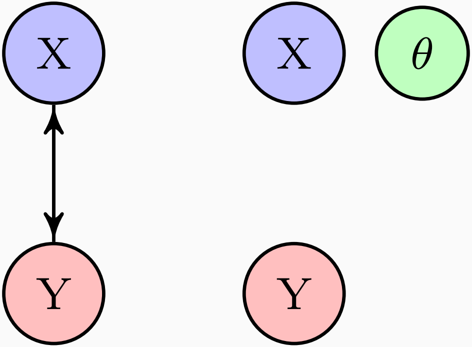
ELBO

$$log\ p(\mathbf{Y}) = \mathrm{KL}(q(\mathbf{X})||p(\mathbf{X}|\mathbf{Y})) + \underbrace{\mathbb{E}_{q(\mathbf{X})}\left[log\ p(\mathbf{X},\mathbf{Y})\right] - H(q(\mathbf{X}))}_{\mathrm{ELBO}}$$

$$\geq \mathbb{E}_{q(\mathbf{X})}\left[log\ p(\mathbf{X},\mathbf{Y})\right] - H(q(\mathbf{X})) = \mathcal{L}(q(\mathbf{X}))$$

- if we maximise the ELBO we,
    - find an approximate posterior
    - get an approximation to the marginal likelihood
- *maximising* $p(\mathbf{Y})$ is learning
- finding $p(\mathbf{X}|\mathbf{Y}) \approx q(\mathbf{X})$ is prediction

ELBO



Why is this useful?

Why is this a sensible thing to do?

- If we can't formulate the joint distribution there isn't much we can do

– Ryan Adams[8]

[8] `Talking Machines Season 2, Episode 5`

Why is this useful?

Why is this a sensible thing to do?

- If we can't formulate the joint distribution there isn't much we can do
- Taking the expectation of a log is usually easier than the expectation

– Ryan Adams[8]

[8] Talking Machines Season 2, Episode 5

Why is this useful?

Why is this a sensible thing to do?

- If we can't formulate the joint distribution there isn't much we can do
- Taking the expectation of a log is usually easier than the expectation
- We are allowed to choose the distribution to take the expectation over

– Ryan Adams[8]

[8] `Talking Machines Season 2, Episode 5`

Lower Bound[9]

$$\mathcal{L} = \int_{\mathbf{X},\mathbf{F}} q(\mathbf{X}) \log \left(\frac{p(\mathbf{Y}, \mathbf{F}, \mathbf{X})}{q(\mathbf{X})} \right)$$

[9] Damianou, A. C. (2015). Deep Gaussian Processes and Variational Propagation of Uncertainty (Doctoral dissertation)

Lower Bound[9]

$$\mathcal{L} = \int_{\mathbf{X},\mathbf{F}} q(\mathbf{X}) \log\left(\frac{p(\mathbf{Y},\mathbf{F},\mathbf{X})}{q(\mathbf{X})}\right)$$

$$\int_{\mathbf{X},\mathbf{F}} q(\mathbf{X}) \log\left(\frac{p(\mathbf{Y}|\mathbf{F})p(\mathbf{F}|\mathbf{X})p(\mathbf{X})}{q(\mathbf{X})}\right)$$

[9]Damianou, A. C. (2015). Deep Gaussian Processes and Variational Propagation of Uncertainty (Doctoral dissertation)

Lower Bound[9]

$$\mathcal{L} = \int_{\mathbf{X},\mathbf{F}} q(\mathbf{X}) \log \left(\frac{p(\mathbf{Y}, \mathbf{F}, \mathbf{X})}{q(\mathbf{X})} \right)$$

$$\int_{\mathbf{X},\mathbf{F}} q(\mathbf{X}) \log \left(\frac{p(\mathbf{Y}|\mathbf{F}) p(\mathbf{F}|\mathbf{X}) p(\mathbf{X})}{q(\mathbf{X})} \right)$$

$$= \int_{\mathbf{F},\mathbf{X}} q(\mathbf{X}) \log p(\mathbf{Y}|\mathbf{F}) p(\mathbf{F}|\mathbf{X}) - \int_{\mathbf{X}} q(\mathbf{X}) \log \frac{q(\mathbf{X})}{p(\mathbf{X})}$$

[9] Damianou, A. C. (2015). Deep Gaussian Processes and Variational Propagation of Uncertainty (Doctoral dissertation)

Lower Bound[9]

$$
\begin{aligned}
\mathcal{L} = & \int_{\mathbf{X},\mathbf{F}} q(\mathbf{X}) \log \left(\frac{p(\mathbf{Y}, \mathbf{F}, \mathbf{X})}{q(\mathbf{X})} \right) \\
& \int_{\mathbf{X},\mathbf{F}} q(\mathbf{X}) \log \left(\frac{p(\mathbf{Y}|\mathbf{F}) p(\mathbf{F}|\mathbf{X}) p(\mathbf{X})}{q(\mathbf{X})} \right) \\
& = \int_{\mathbf{F},\mathbf{X}} q(\mathbf{X}) \log p(\mathbf{Y}|\mathbf{F}) p(\mathbf{F}|\mathbf{X}) - \int_{\mathbf{X}} q(\mathbf{X}) \log \frac{q(\mathbf{X})}{p(\mathbf{X})} \\
& = \tilde{\mathcal{L}} - \mathrm{KL}\left(q(\mathbf{X}) \parallel p(\mathbf{X})\right)
\end{aligned}
$$

[9] Damianou, A. C. (2015). Deep Gaussian Processes and Variational Propagation of Uncertainty (Doctoral dissertation)

Lower Bound

$$\tilde{\mathcal{L}} = \int_{\mathbf{F},\mathbf{X}} q(\mathbf{X}) \log p(\mathbf{Y}|\mathbf{F})p(\mathbf{F}|\mathbf{X})$$

- Has not eliviate the problem at all, X still needs to go through F to reach the data
- Idea of sparse approximations[10]

[10]Quinonero-Candela, Joaquin, & Rasmussen, C. E. (2005). A unifying view of sparse approximate Gaussian process regression & Snelson, E., & Ghahramani, Z. (2006). Sparse Gaussian processes using pseudo-inputs

Lower Bound

- Add another set of samples from the same prior

$$p(\mathbf{U}|\mathbf{Z}) = \prod_{j=1}^{d} \mathcal{N}(\mathbf{u}_{:,j}|\mathbf{0}, \mathbf{K})$$

Lower Bound

- Add another set of samples from the same prior

$$p(\mathbf{U}|\mathbf{Z}) = \prod_{j=1}^{d} \mathcal{N}(\mathbf{u}_{:,j}|\mathbf{0}, \mathbf{K})$$

- Conditional distribution

$$p(\mathbf{f}_{:,j}, \mathbf{u}_{:,j}|\mathbf{X}, \mathbf{Z}) = p(\mathbf{f}_{:,j}|\mathbf{u}_{:,j}, \mathbf{X}, \mathbf{Z})p(\mathbf{u}_{:,j}|\mathbf{Z})$$

$$= \mathcal{N}\left(\mathbf{f}_{:,j}|\mathbf{K}_{fu}(\mathbf{K}_{uu})^{-1}\mathbf{u}_{:,j}, \mathbf{K}_{ff} - \mathbf{K}_{fu}(\mathbf{K}_{uu})^{-1}\mathbf{K}_{uf}\right)\mathcal{N}\left(\mathbf{u}_{:,j}|\mathbf{0}, \mathbf{K}_{uu}\right),$$

Lower Bound

$$p(\mathbf{Y}, \mathbf{F}, \mathbf{U}, \mathbf{X}|\mathbf{Z}) = p(\mathbf{X}) \prod_{j=1}^{d} p(\mathbf{y}_{:,j}|\mathbf{f}_{:,j}) p(\mathbf{f}_{:,j}|\mathbf{u}_{:,j}, \mathbf{X}) p(\mathbf{u}_{:,j}|\mathbf{Z})$$

- we have done nothing to the model, just added *halucinated* observations

Lower Bound

$$p(\mathbf{Y}, \mathbf{F}, \mathbf{U}, \mathbf{X}|\mathbf{Z}) = p(\mathbf{X}) \prod_{j=1}^{d} p(\mathbf{y}_{:,j}|\mathbf{f}_{:,j}) p(\mathbf{f}_{:,j}|\mathbf{u}_{:,j}, \mathbf{X}) p(\mathbf{u}_{:,j}|\mathbf{Z})$$

- we have done nothing to the model, just added *halucinated* observations
- however, we will now interpret $\mathbf{U}$ and $\mathbf{X}_u$ not as random variables but variational parameters

Lower Bound

$$p(\mathbf{Y}, \mathbf{F}, \mathbf{U}, \mathbf{X}|\mathbf{Z}) = p(\mathbf{X}) \prod_{j=1}^{d} p(\mathbf{y}_{:,j}|\mathbf{f}_{:,j}) p(\mathbf{f}_{:,j}|\mathbf{u}_{:,j}, \mathbf{X}) p(\mathbf{u}_{:,j}|\mathbf{Z})$$

- we have done nothing to the model, just added *halucinated* observations
- however, we will now interpret $\mathbf{U}$ and $\mathbf{X}_u$ not as random variables but variational parameters
- i.e. parametrise approximate posterior using these parameters (remember sparse motivation)

Lower Bound

- Variational distributions are approximations to intractable posteriors,

$$
\begin{aligned}
q(\mathbf{U}) &\approx p(\mathbf{U}|\mathbf{Y}, \mathbf{X}, \mathbf{Z}, \mathbf{F}) \\
q(\mathbf{F}) &\approx p(\mathbf{F}|\mathbf{U}, \mathbf{X}, \mathbf{Z}, \mathbf{Y}) \\
q(\mathbf{X}) &\approx p(\mathbf{X}|\mathbf{Y})
\end{aligned}
$$

Lower Bound

- Variational distributions are approximations to intractable posteriors,

$$q(\mathbf{U}) \approx p(\mathbf{U}|\mathbf{Y}, \mathbf{X}, \mathbf{Z}, \mathbf{F})$$
$$q(\mathbf{F}) \approx p(\mathbf{F}|\mathbf{U}, \mathbf{X}, \mathbf{Z}, \mathbf{Y})$$
$$q(\mathbf{X}) \approx p(\mathbf{X}|\mathbf{Y})$$

- Assume that we can *find* **U** that completely represents **F**, i.e. **U** is sufficient statistics of **F**,

$$q(\mathbf{F}) \approx p(\mathbf{F}|\mathbf{U}, \mathbf{X}, \mathbf{Z}, \mathbf{Y}) = p(\mathbf{F}|\mathbf{U}, \mathbf{X}, \mathbf{Z})$$

Lower Bound

$$\tilde{\mathcal{L}} = \int_{\mathbf{X},\mathbf{F},\mathbf{U}} q(\mathbf{F})q(\mathbf{U})q(\mathbf{X}) \log \frac{p(\mathbf{Y},\mathbf{F},\mathbf{U}|\mathbf{X},\mathbf{Z})}{q(\mathbf{F})q(\mathbf{U})}$$

Lower Bound

$$\tilde{\mathcal{L}} = \int_{\mathbf{X},\mathbf{F},\mathbf{U}} q(\mathbf{F})q(\mathbf{U})q(\mathbf{X}) \log \frac{p(\mathbf{Y},\mathbf{F},\mathbf{U}|\mathbf{X},\mathbf{Z})}{q(\mathbf{F})q(\mathbf{U})}$$

$$= \int_{\mathbf{X},\mathbf{F},\mathbf{U}} q(\mathbf{F})q(\mathbf{U})q(\mathbf{X}) \log \frac{\prod_{j=1}^{d} p(\mathbf{y}_{:,j}|\mathbf{f}_{:,j})p(\mathbf{f}_{:,j}|\mathbf{u}_{:,j},\mathbf{X},\mathbf{Z})p(\mathbf{u}_{:,j}|\mathbf{Z})}{q(\mathbf{F})q(\mathbf{U})}$$

Lower Bound

$$\tilde{\mathcal{L}} = \int_{\mathbf{X},\mathbf{F},\mathbf{U}} q(\mathbf{F})q(\mathbf{U})q(\mathbf{X}) \log \frac{p(\mathbf{Y},\mathbf{F},\mathbf{U}|\mathbf{X},\mathbf{Z})}{q(\mathbf{F})q(\mathbf{U})}$$

$$= \int_{\mathbf{X},\mathbf{F},\mathbf{U}} q(\mathbf{F})q(\mathbf{U})q(\mathbf{X}) \log \frac{\prod_{j=1}^{d} p(\mathbf{y}_{:,j}|\mathbf{f}_{:,j})p(\mathbf{f}_{:,j}|\mathbf{u}_{:,j},\mathbf{X},\mathbf{Z})p(\mathbf{u}_{:,j}|\mathbf{Z})}{q(\mathbf{F})q(\mathbf{U})}$$

- Assume that $\mathbf{U}$ is sufficient statistics for $\mathbf{F}$

$$q(\mathbf{F})q(\mathbf{U})q(\mathbf{X}) = p(\mathbf{F}|\mathbf{U},\mathbf{X},\mathbf{Z})q(\mathbf{U})q(\mathbf{X})$$

Lower Bound

$$\tilde{\mathcal{L}} = \int_{\mathbf{X},\mathbf{F},\mathbf{U}} \prod_{j=1}^{d} p(\mathbf{f}_{:,j}|\mathbf{u}_{:,j}, \mathbf{X}, \mathbf{Z}) q(\mathbf{u}_{:,j}) q(\mathbf{X})$$

$$\log \frac{\prod_{j=1}^{d} p(\mathbf{y}_{:,j}|\mathbf{f}_{:,j}) p(\mathbf{f}_{:,j}|\mathbf{u}_{:,j}, \mathbf{X}, \mathbf{Z}) p(\mathbf{u}_{:,j}|\mathbf{Z})}{\prod_{j=1}^{d} p(\mathbf{f}_{:,j}|\mathbf{u}_{:,j}, \mathbf{X}, \mathbf{Z}) q(\mathbf{u}_{:,j})} =$$

Lower Bound

$$\tilde{\mathcal{L}} = \int_{\mathbf{X},\mathbf{F},\mathbf{U}} \prod_{j=1}^{d} p(\mathbf{f}_{:,j}|\mathbf{u}_{:,j}, \mathbf{X}, \mathbf{Z}) q(\mathbf{u}_{:,j}) q(\mathbf{X})$$

$$\log \frac{\prod_{j=1}^{d} p(\mathbf{y}_{:,j}|\mathbf{f}_{:,j}) p(\mathbf{f}_{:,j}|\mathbf{u}_{:,j}, \mathbf{X}, \mathbf{Z}) p(\mathbf{u}_{:,j}|\mathbf{Z})}{\prod_{j=1}^{d} p(\mathbf{f}_{:,j}|\mathbf{u}_{:,j}, \mathbf{X}, \mathbf{Z}) q(\mathbf{u}_{:,j})} =$$

$$= \int_{\mathbf{X},\mathbf{F},\mathbf{U}} \prod_{j=1}^{p} p(\mathbf{f}_{:,j}|\mathbf{u}_{:,j}, \mathbf{X}, \mathbf{Z}) q(\mathbf{u}_{:,j}) q(\mathbf{X}) \log \frac{\prod_{j=1}^{p} p(\mathbf{y}_{:,j}|\mathbf{f}_{:,j}) p(\mathbf{u}_{:,j}|\mathbf{Z})}{\prod_{j=1}^{p} q(\mathbf{u}_{:,j})}$$

$$= \mathbb{E}_{q(\mathbf{F}),q(\mathbf{X}),q(\mathbf{U})} \left[p(\mathbf{Y}|\mathbf{F}) \right] - \mathrm{KL}\left(q(\mathbf{U}) || p(\mathbf{U}|\mathbf{Z}) \right)$$

Summary

$$\mathbb{E}_{q(\mathbf{F}),q(\mathbf{X}),q(\mathbf{U})}\left[p(\mathbf{Y}|\mathbf{F})\right] - \mathrm{KL}\left(q(\mathbf{U})||p(\mathbf{U}|\mathbf{Z})\right) - \mathrm{KL}\left(q(\mathbf{X})||p(\mathbf{X})\right)$$

- Expectation tractable (for some co-variances)
- Reduces to expectations over co-variance functions know as Ψ statistics
- Allows us to place priors and not "regularisers" over the latent representation

Latent space priors

Latent space priors[11]

$$\mathbb{E}_{q(\mathbf{F}),q(\mathbf{X}),q(\mathbf{U})}\left[p(\mathbf{Y}|\mathbf{F})\right] - \mathrm{KL}\left(q(\mathbf{U})||p(\mathbf{U}|\mathbf{Z})\right) - \mathrm{KL}\left(q(\mathbf{X})||p(\mathbf{X})\right)$$

- Importantly $p(\mathbf{X})$ appears only in KL term
- Allows us to express stronger assumptions about the model

[11] Damianou, A. C., Titsias, M., & Lawrence, Neil D, Variational Inference for Uncertainty on the Inputs of Gaussian Process Models (2014)

Non-Gaussian Data

Theorem (Change-of-variable)

$$p_y(y) = p_x(\rho(y))|\nabla\rho(y)|$$

$$x \in \mathcal{X} \subseteq R^{D_x} \quad y \in \mathcal{Y} \subseteq R^{D_y}$$

- $\rho : \mathcal{Y} \to \mathcal{X}$
- ρ *is a bijective function*

Change of Variables

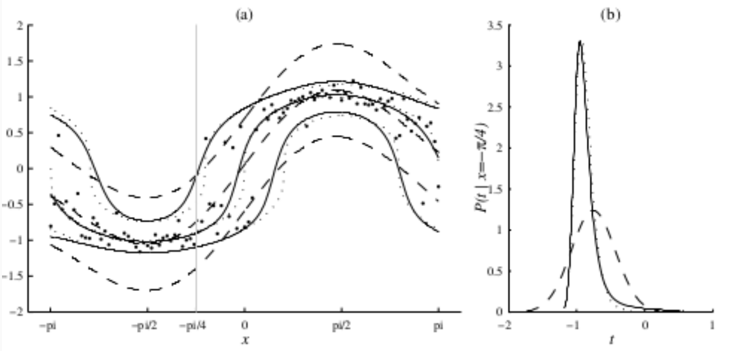
$$p(y) = \mathcal{N}(\rho(y)\mu, \Sigma)|\nabla\rho(y)| \quad p(y_*|y) = \mathcal{N}(\rho(y)|\mu(x_*|x), \Sigma(x_*|x))|\nabla\rho(y)|$$

$$p(x) \sim \mathcal{N}(\mu, \Sigma) \quad p(x_*|x) = \mathcal{N}(\rho(y)|\mu(x_*|x), \Sigma(x_*|x))$$

Change of Variable

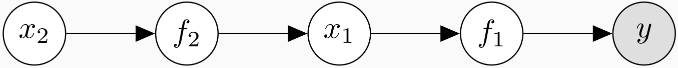
- We can model non-gaussian data y using a Gaussian variable x if we have a transformation ρ that makes it "Gaussian"

Warped Gaussian Processes[12, 13]



[12]Snelson, E., & Ghahramani, Z. (2004). Warped Gaussian Processes
[13]Lazaro-Gredilla, Miguel (2012). Bayesian Warped Gaussian Processes. In , Advances in Neural Information Processing Systems

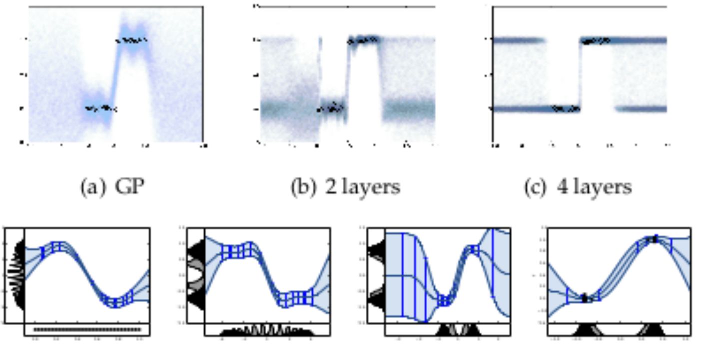
Deep Gaussian Processes[14]



- Place a GP as a warping function, that is warped, ...

[14] Damianou, A. C., & Lawrence, N. D. (2013). Deep Gaussian Processes

Deep Gaussian Processes[15]

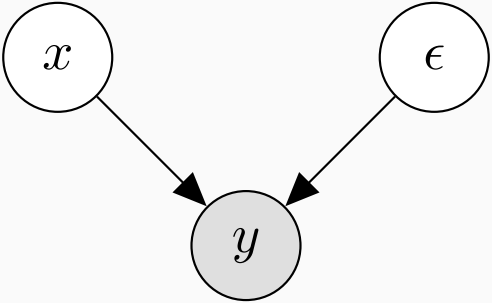


(a) GP (b) 2 layers (c) 4 layers

(d) Hidden spaces for 4 layer model

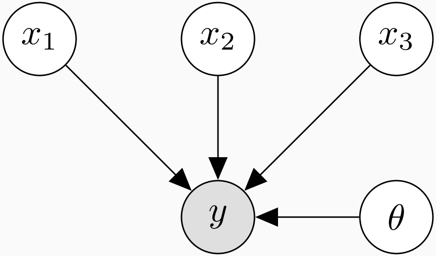
[15] *Stolen from Neil, who borrowed it from James, who we believe genereated the plot*

Explaining Away



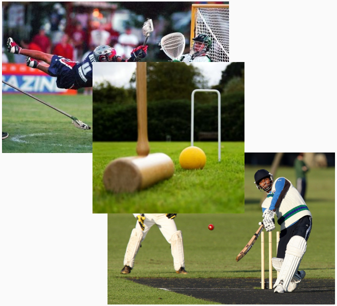
$$y = f(x) + \epsilon$$

Factor Analysis

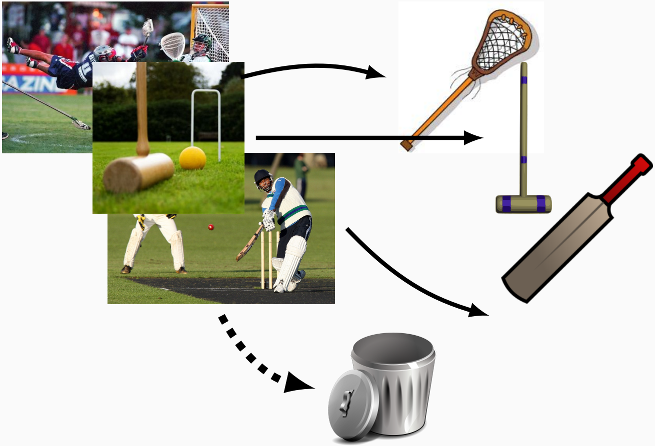


$$y = f(x_1, x_2, x_3) + \epsilon$$

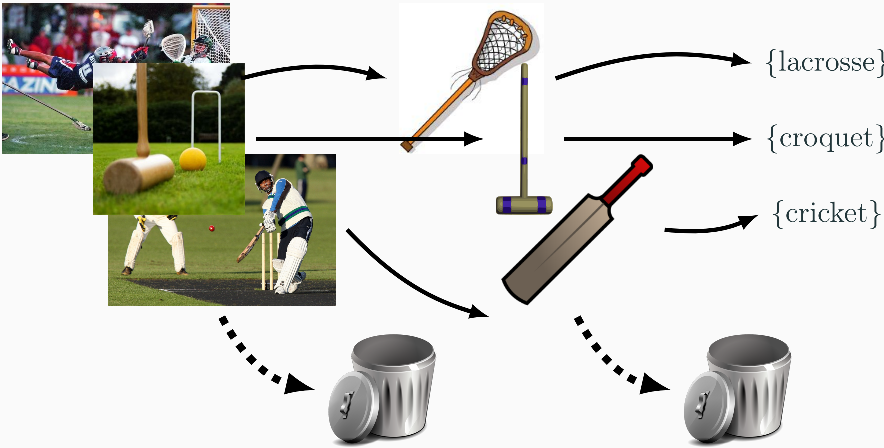
Latent Structure



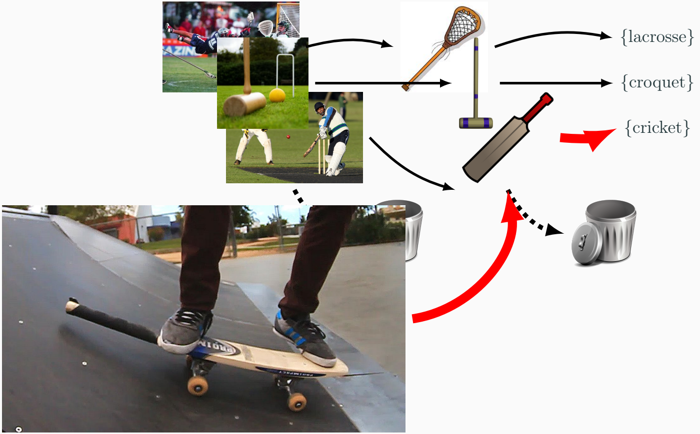
Latent Structure



Latent Structure

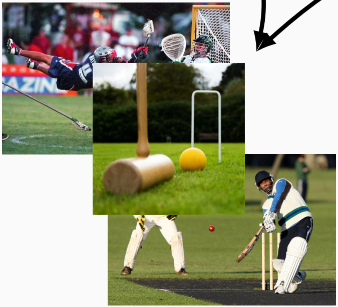


Latent Structure

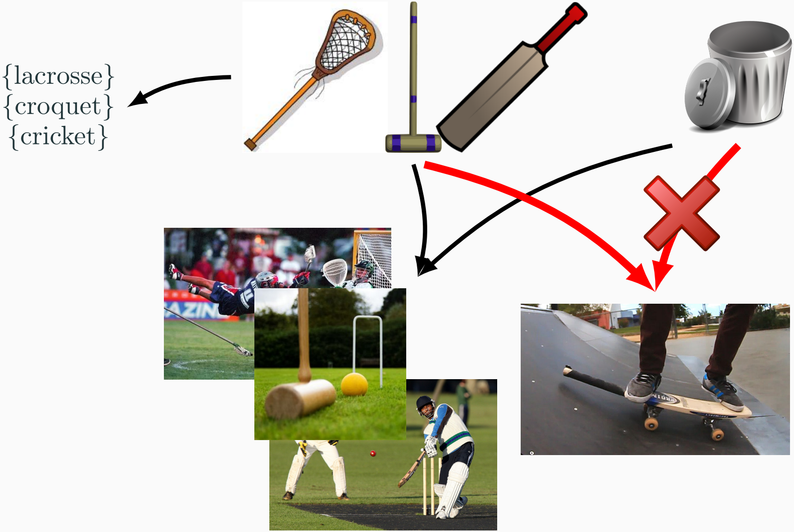


Latent Structure

$\{\text{lacrosse}\}$
$\{\text{croquet}\}$
$\{\text{cricket}\}$

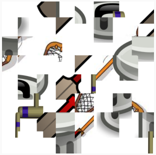


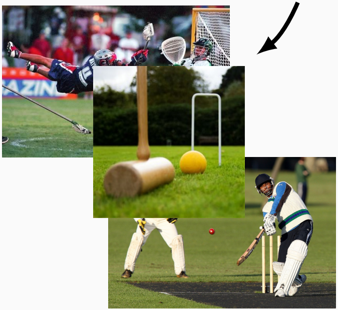
Latent Structure



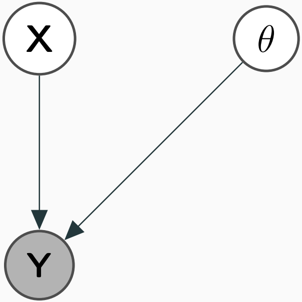
Latent Structure

{lacrosse}
{croquet}
{cricket}

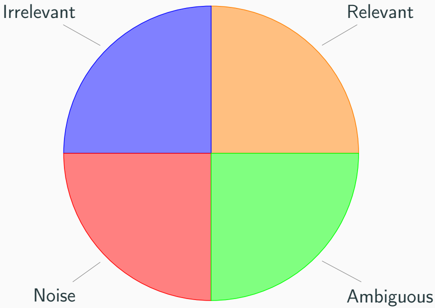




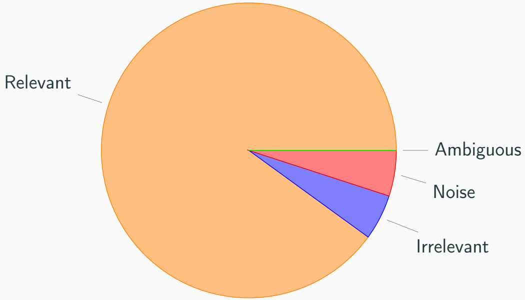
Latent Structure



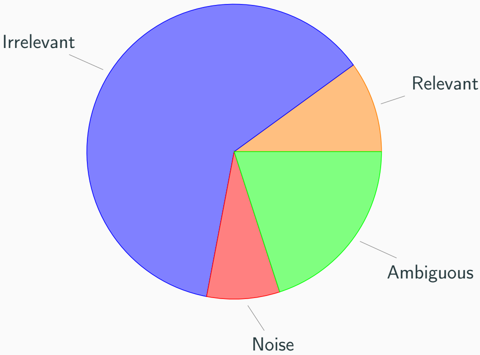
Latent Structure



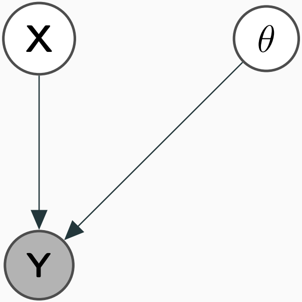
Latent Structure



Latent Structure

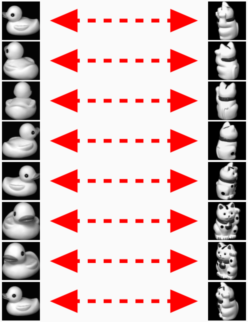


Latent Structure

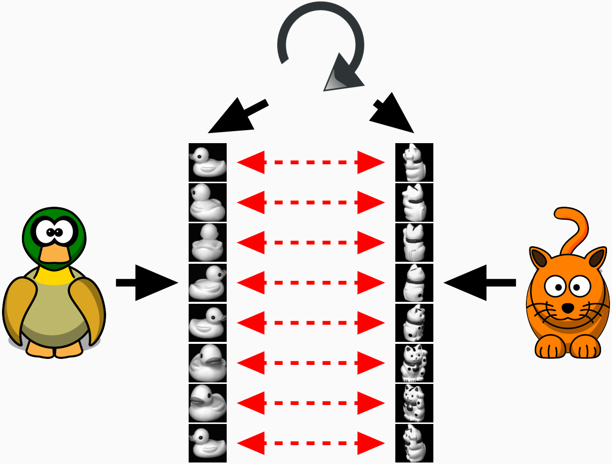


Alignments

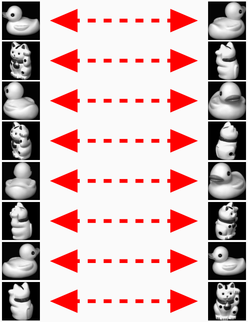
Alignments



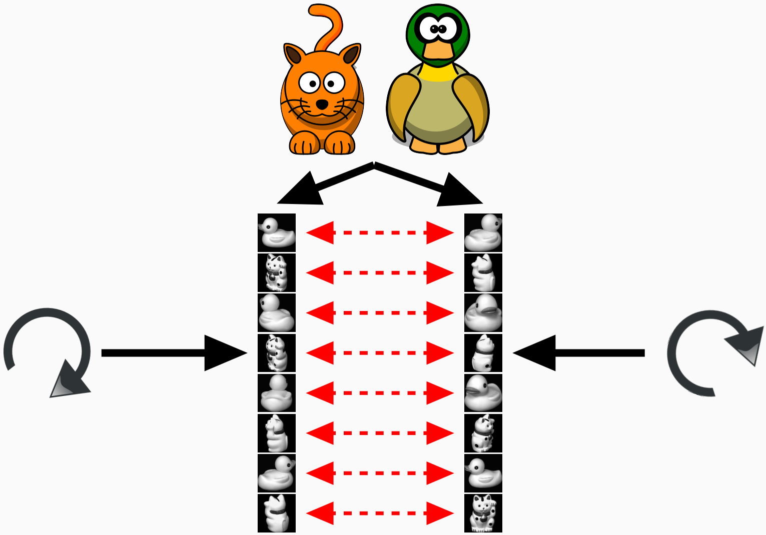
Alignments



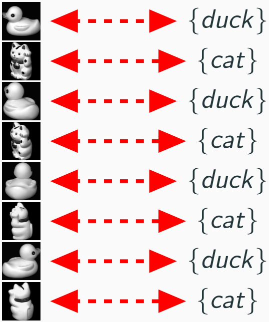
Alignments



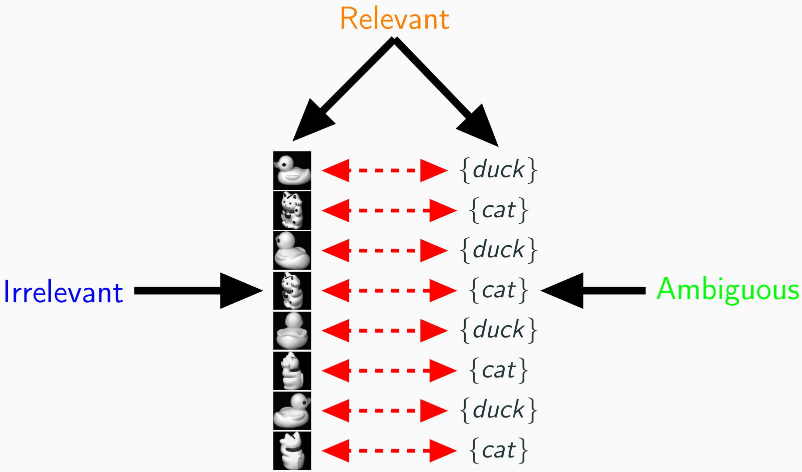
Alignments



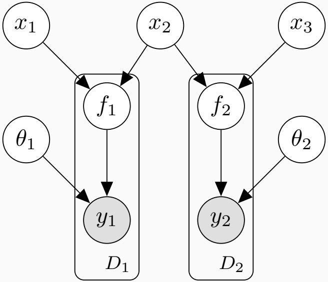
Alignments



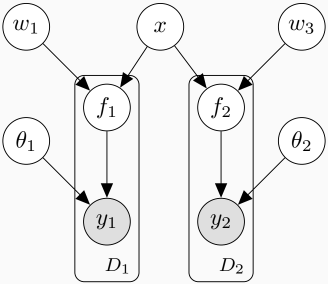
Alignments


Relevant
Irrelevant
Ambiguous
$\{duck\}$
$\{cat\}$
$\{duck\}$
$\{cat\}$
$\{duck\}$
$\{cat\}$
$\{duck\}$
$\{cat\}$

Explaining Away cont.



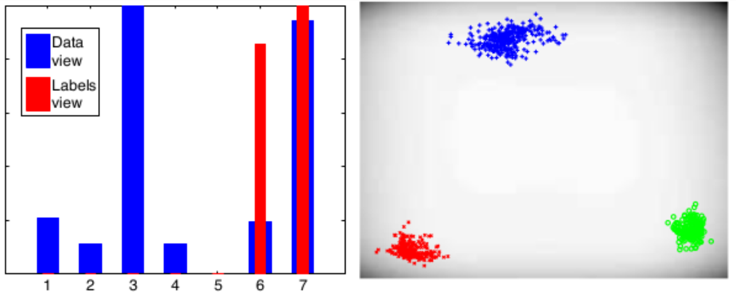
IBFA with GP-LVM[16]



$$y_1 = f(w_1^{\mathrm{T}} x) \quad y_2 = f(w_2^{\mathrm{T}} x)$$

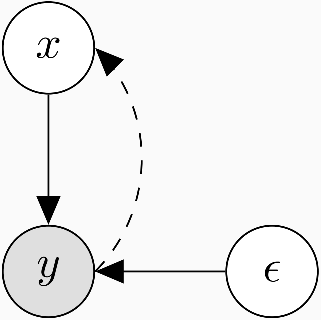
[16]Damianou, A., Lawrence, N. D., & Ek, C. H. (2016). Multi-view learning as a nonparametric nonlinear inter-battery factor analysis

IBFA with GP-LVM



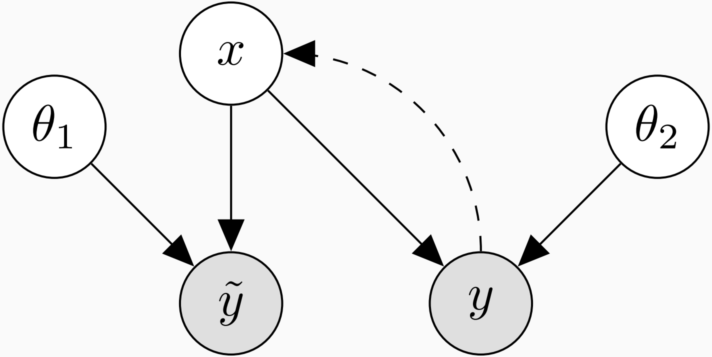
IBFA with GP-LVM

Constrained Latent Space



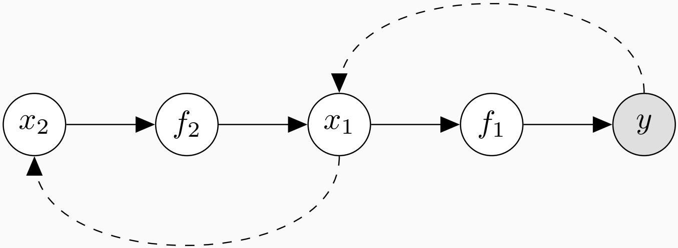
$$y = f(g(y)) + \epsilon$$

Denoising Auto-encoder[17, 18]



[17]Ek, C. H., Torr, P. H. S., & Lawrence, N. D., Gaussian process latent variable models for human pose estimation, International conference on Machine learning for multimodal interaction, (), 132–143 (2007).
[18]Snoek, J., Adams, R. P., & Larochelle, H., Nonparametric guidance of autoencoder representations using label information, Journal of Machine

VAE[19]



$$q(x_l) = g(x_{l-1})$$

[19]Dai, Z., Damianou, A., Gonz\'alez, Javier, & Lawrence, N., Variational auto-encoded deep Gaussian processes, International Conference on Learning Representations (ICLR), (2016).

Summary

Summary

- Unsupervised learning is very hard

Summary

- Unsupervised learning is very hard
    - *Its actually not, its really really easy.*

Summary

- Unsupervised learning is very hard
  - *Its actually not, its really really easy.*
- Relevant assumptions needed to learn anything useful

Summary

- Unsupervised learning is very hard
    - *Its actually not, its really really easy.*
- Relevant assumptions needed to learn anything useful
- Strong assumptions needed to learn anything from "sensible" amounts of data

Summary

- Unsupervised learning is very hard
  - *Its actually not, its really really easy.*
- Relevant assumptions needed to learn anything useful
- Strong assumptions needed to learn anything from "sensible" amounts of data
- GPs provide strong, interpretative assumptions that aligns well to our intuitions allowing us to make relevant assumptions

eof