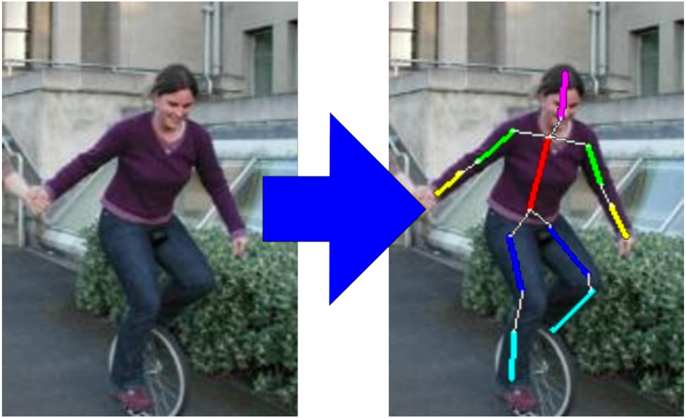
Feature Selection in GPLVM's

Carl Henrik Ek
{chek}@csc.kth.se

Royal Institute of Technology

August 15, 2014



Introduction

Setting

- Observed variables $\mathbf{Y}^{(1)} \in \mathbb{R}^{d_Y^{(1)}}$, $\mathbf{Y}^{(\mathbf{2})} \in \mathbb{R}^{d_Y^{(2)}}$
- Task
  - Infer $\mathbf{y}_i^{(2)}$ from $\mathbf{y}_i^{(1)}$

Introduction

Setting

- Observed variables $\mathbf{Y}^{(1)} \in \mathbb{R}^{d_Y^{(1)}}$, $\mathbf{Y}^{(\mathbf{2})} \in \mathbb{R}^{d_Y^{(2)}}$
- Task
  - Infer $\mathbf{y}_i^{(2)}$ from $\mathbf{y}_i^{(1)}$

Challenge

- $\mathbf{Y}^{(1)}$ is a high-dimensional, noisy, redundant and sometimes ambiguous representation of $\mathbf{Y}^{(2)}$

Modelling paradigm

Generative

$$p(\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)})$$

- Jointly models all data, uncertainty in "input"
- High dimensional, parametrise $\mathbb{R}^{d_{Y^{(1)}}} \times \mathbb{R}^{d_{Y^{(2)}}}$

Discriminative

$$p(\mathbf{Y}^{(2)} | \mathbf{Y}^{(1)})$$

- Only model "decision" boundary
- Low dimensional "model" $\mathbb{R}^{d_{Y^{(2)}}}$

Computer Vision Challenges

- Pascal VOC Challenge [URL]
  - Discriminative methods
  - Lots of feature engineering to achieve generalisation
- ImageNet [URL]
  - Feature learning through Neural Networks
  - Representation learning tweaks and tricks to explain away irrelevant variations
- *little success (nor focus) by actual models of images*

Computer Vision Challenges

- Pascal VOC Challenge [URL]
  - Discriminative methods
  - Lots of feature engineering to achieve generalisation
- ImageNet [URL]
  - Feature learning through Neural Networks
  - Representation learning tweaks and tricks to explain away irrelevant variations
- *little success (nor focus) by actual models of images*

Computer Vision Challenges

- Pascal VOC Challenge [URL]
  - Discriminative methods
  - Lots of feature engineering to achieve generalisation
- ImageNet [URL]
  - Feature learning through Neural Networks
  - Representation learning tweaks and tricks to explain away irrelevant variations
- *little success (nor focus) by actual models of images*

The problem with generative models

Variations

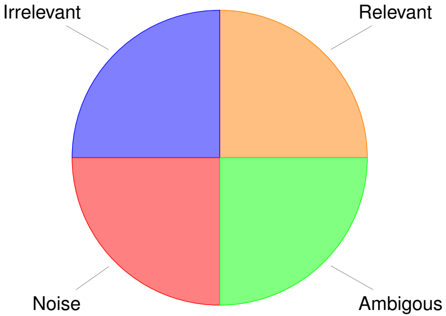
- Signal
  1. $\mathbf{Y}^{(1)}$ informative of $\mathbf{Y}^{(2)}$ (Relevant)
  2. $\mathbf{Y}^{(1)}$ non-informative of $\mathbf{Y}^{(2)}$ (Irrelevant)
  3. $\mathbf{Y}^{(2)}$ non-informative of $\mathbf{Y}^{(1)}$ (Ambiguous)
- Noise in $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$
- *All variations need to be explained in a model of the data*

The problem with generative models

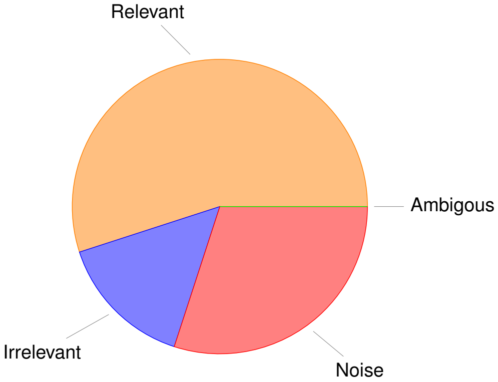
Variations

- Signal
  1. $\mathbf{Y}^{(1)}$ informative of $\mathbf{Y}^{(2)}$ (Relevant)
  2. $\mathbf{Y}^{(1)}$ non-informative of $\mathbf{Y}^{(2)}$ (Irrelevant)
  3. $\mathbf{Y}^{(2)}$ non-informative of $\mathbf{Y}^{(1)}$ (Ambiguous)
- Noise in $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$
- *All variations need to be explained in a model of the data*

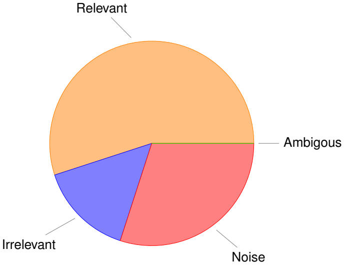
The problem with generative models

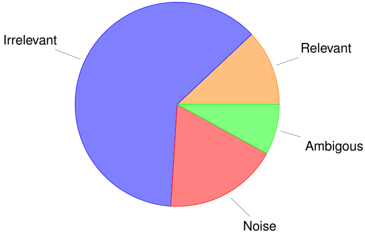


The problem with generative models

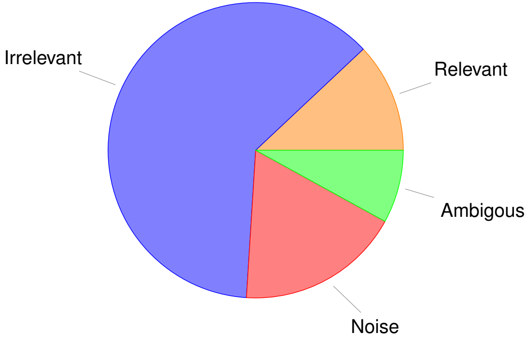


The problem with generative models





The problem with generative models



The problem with generative models

Approaches

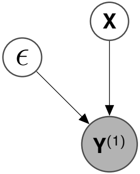
- Heuristics
  - Remove non-informative and noise by "hand" from data (pre-processing)
- Pseudo-heuristics
  - similarity engineering
- Full model
  - Factorise variations

This Talk

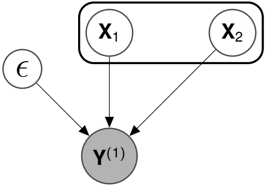
Factorised representation learning as a means of performing *feature selection* in a generative model.

- Factor Analysis
- Multiview learning
- GP formulation

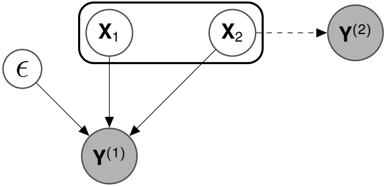
Factorised Representation Learning



Factorised Representation Learning



Factorised Representation Learning



Factor Analysis

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \epsilon$$
$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x}, \Sigma)$$

- **A** - factor loadings
- **X** - latent representation
- Solution not identifiable
- Introduce additional information

Factor Analysis

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \epsilon$$
$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x}, \Sigma)$$

- **A** - factor loadings
- **X** - latent representation
- Solution not identifiable
- Introduce additional information

Factor Analysis

FA according to Carl

- Structure of factor loadings

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & a_{13} & a_{14} & 0 & a_{16} & 0 \\ a_{21} & a_{22} & 0 & 0 & a_{25} & a_{26} & a_{27} \\ a_{31} & 0 & a_{33} & a_{34} & 0 & a_{36} & a_{37} \end{bmatrix}$$

- Column space structure of loadings

Factor Analysis

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \epsilon$$
$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x}, \Sigma)$$

Covariance

- Isotropic covariance implies PCA/MDS
- Full covariance plus diagonal implies “traditional” factor analysis

Factor Analysis

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \epsilon$$
$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x}, \Sigma)$$

Latent Variable

- Gaussian distribution for PCA and FA
- Non Gaussian for ICA

Factor Analysis

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \epsilon$$
$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x}, \Sigma)$$

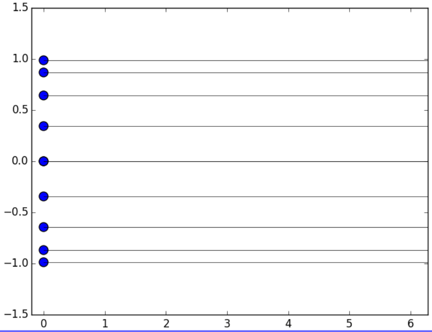
Mapping

- Introduce general mapping f

$$p(\mathbf{y}|\mathbf{f}, \mathbf{x}) = p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\mathbf{x})$$

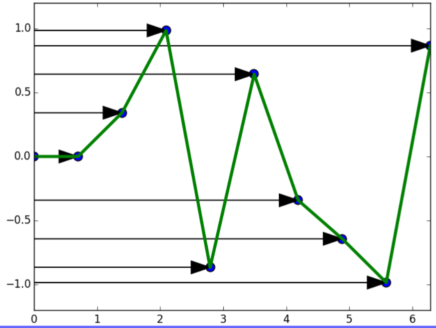
- Gaussian Process prior on mapping

Place a GP-prior over the mapping and get GP-LVM



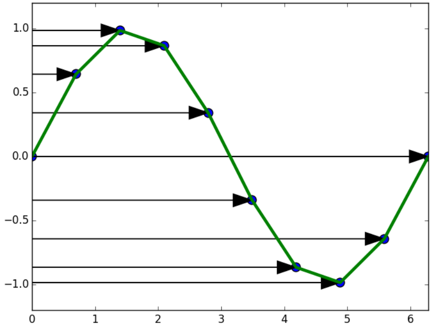
GP-LVM according to Carl

- FA: *given the output* $[\mathbf{y}_1, \ldots, \mathbf{y}_N]$ *how should we associate them with input* $[\mathbf{x}_1, \ldots, \mathbf{x}_N]$*?*
- GP-LVM: *assume functional relationship,* $\mathcal{GP}$ *encodes preference*



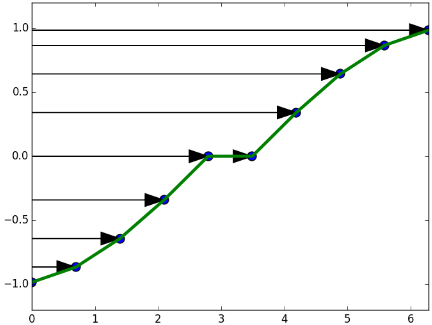
GP-LVM according to Carl

- FA: *given the output* $[\mathbf{y}_1, \ldots, \mathbf{y}_N]$ *how should we associate them with input* $[\mathbf{x}_1, \ldots, \mathbf{x}_N]$*?*
- GP-LVM: *assume functional relationship,* $\mathcal{GP}$ *encodes preference*



GP-LVM according to Carl

- FA: *given the output* $[\mathbf{y}_1, \ldots, \mathbf{y}_N]$ *how should we associate them with input* $[\mathbf{x}_1, \ldots, \mathbf{x}_N]$*?*
- GP-LVM: *assume functional relationship,* $\mathcal{GP}$ *encodes preference*

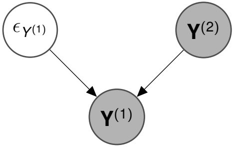


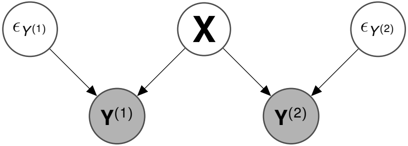
GP-LVM according to Carl

- FA: *given the output* $[\mathbf{y}_1, \ldots, \mathbf{y}_N]$ *how should we associate them with input* $[\mathbf{x}_1, \ldots, \mathbf{x}_N]$*?*
- GP-LVM: *assume functional relationship,* $\mathcal{GP}$ *encodes preference*

Variations

- Signal
  1. $\mathbf{Y}^{(1)}$ informative of $\mathbf{Y}^{(2)}$ (relevant)
  2. $\mathbf{Y}^{(1)}$ non-informative of $\mathbf{Y}^{(2)}$
  3. $\mathbf{Y}^{(2)}$ non-informative of $\mathbf{Y}^{(1)}$
- Noise in $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ (irrelevant)


$\epsilon_{Y^{(1)}}$
$\mathbf{Y}^{(2)}$
$\mathbf{Y}^{(1)}$



$\epsilon_{Y^{(1)}}$
X
$\epsilon_{Y^{(2)}}$
$\mathbf{Y}^{(1)}$
$\mathbf{Y}^{(2)}$

Multiview Factor Analysis

Cannonical Correlation Analysis (Hotelling 1936)

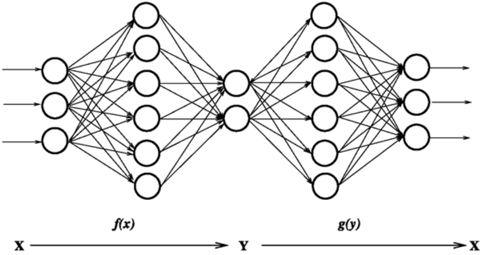
$$\{\hat{\mathbf{u}}, \hat{\mathbf{v}}\} = \underset{\mathbf{u},\mathbf{v}}{\operatorname{argmax}} \rho(\mathbf{u}^{\mathrm{T}}\mathbf{X}, \mathbf{v}^{\mathrm{T}}\mathbf{Y})$$

- Correlation

$$\rho(\mathbf{X}, \mathbf{Y}) = \frac{\mathbb{E}\left[(\mathbf{X} - \mu_X))(\mathbf{Y} - \mu_Y))\right]}{\sqrt{\mathbb{E}\left[\mathbf{X} - \mu_X\right]\mathbb{E}\left[\mathbf{Y} - \mu_Y\right]}}$$

- Learn a project of the data

Multiview Factor Analysis



Hybrid models

- Neuroscale (Lowe and Tipping 1997)
- Bottleneck networks (Hinton and Salakhutdinov 2006)
- De-noising Auto-encoders (Vincent *et al.* 2008)

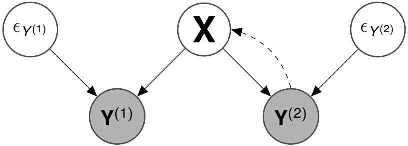
Multiview Factor Analysis

$$p(\mathbf{y}|\mathbf{f},\mathbf{x}) = p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\mathbf{x})$$
$$\mathbf{x} = g(\mathbf{y})$$

BC GP-LVM (Lawrence and Quiñonero-Candela 2006)

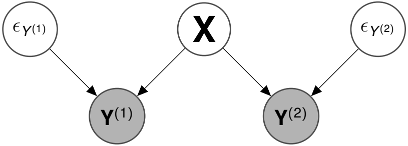
- Constrain latent space to reflect similarity in input
- Multi-view constrained (Ek *et al.* 2007,Snoek *et al.* 2012)
- Constrain latent space to only represent variation in input space that exist in output

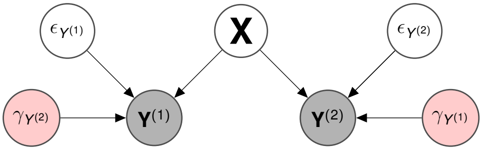
Multiview Factor Analysis

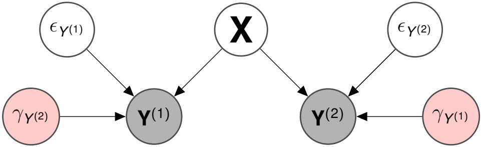


BC GP-LVM (Lawrence and Quiñonero-Candela 2006)

- Constrain latent space to reflect similarity in input
- Multi-view constrained (Ek *et al.* 2007,Snoek *et al.* 2012)
- Constrain latent space to only represent variation in input space that exist in output


$\epsilon_{Y^{(1)}}$
$\mathbf{X}$
$\epsilon_{Y^{(2)}}$
$\mathbf{Y}^{(1)}$
$\mathbf{Y}^{(2)}$


$\epsilon_{Y^{(1)}}$
$\mathbf{X}$
$\epsilon_{Y^{(2)}}$
$\gamma_{Y^{(2)}}$
$\mathbf{Y}^{(1)}$
$\mathbf{Y}^{(2)}$
$\gamma_{Y^{(1)}}$



Variations

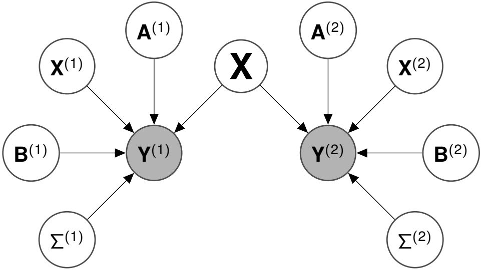
- Signal
  1. $\mathbf{Y}^{(1)}$ informative of $\mathbf{Y}^{(2)}$ (relevant)
  2. $\mathbf{Y}^{(1)}$ non-informative of $\mathbf{Y}^{(2)}$ (structured noise)
  3. $\mathbf{Y}^{(2)}$ non-informative of $\mathbf{Y}^{(1)}$ (ambiguities)
- Noise in $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ (irrelevant)

Inter-Battery Factor Analysis[1]

$$\begin{aligned}\mathbf{y}^{(m)} &\sim \mathcal{N}(\mathbf{A}^{(m)}\mathbf{x} + \mathbf{B}^{(m)}\mathbf{x}^{(m)}, \Sigma^{(m)})\\ \mathbf{x} &\sim \mathcal{N}(\mathbf{0}, \mathbf{I})\\ \mathbf{x}^{(m)} &\sim \mathcal{N}(\mathbf{0}, \mathbf{I})\end{aligned}$$

[1] Tucker 1958.

Inter-Battery Factor Analysis[1]



[1] Tucker 1958.

Inter-Battery Factor Analysis[1]

$$\mathbf{y}^{(m)} \sim \mathcal{N}(\mathbf{A}^{(m)}\mathbf{x} + \mathbf{B}^{(m)}\mathbf{x}^{(m)}, \Sigma^{(m)})$$
$$\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$
$$\mathbf{x}^{(m)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

- Explain away *both* structured and unstructued noise
- Specific model of ambiguities
- Even more unidentifiable
  - Rank preserving transformations
  - Allocations of factors

[1] Tucker 1958.

Inter-Battery Factor Analysis[1]

$$
\begin{aligned}
\mathbf{y}^{(m)} &\sim \mathcal{N}(\mathbf{A}^{(m)}\mathbf{x} + \mathbf{B}^{(m)}\mathbf{x}^{(m)}, \Sigma^{(m)}) \\
\mathbf{x} &\sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \\
\mathbf{x}^{(m)} &\sim \mathcal{N}(\mathbf{0}, \mathbf{I})
\end{aligned}
$$

- Explain away *both* structured and unstructued noise
- Specific model of ambiguities
- Even more unidentifiable
  - Rank preserving transformations
  - Allocations of factors

[1] Tucker 1958.

Inter-Battery Factor Analysis[1]

$$
\begin{aligned}
\mathbf{y}^{(m)} &\sim \mathcal{N}(\mathbf{A}^{(m)}\mathbf{x} + \mathbf{B}^{(m)}\mathbf{x}^{(m)}, \Sigma^{(m)}) \\
\mathbf{x} &\sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \\
\mathbf{x}^{(m)} &\sim \mathcal{N}(\mathbf{0}, \mathbf{I})
\end{aligned}
$$

- Marginalise view dependent latent variable

$$
\mathbf{y}^{(m)} \sim \mathcal{N}(\mathbf{A}^{(m)}\mathbf{x}, \mathbf{B}^{(m)}(\mathbf{B}^{(m)})^{\mathrm{T}} + \Sigma^{(m)})
$$

- Full covariance (Bach and Jordan 2005)

[1] Tucker 1958.

Inter-Battery Factor Analysis[1]

$$\mathbf{y}^{(m)} \sim \mathcal{N}(\mathbf{A}^{(m)}\mathbf{x} + \mathbf{B}^{(m)}\mathbf{x}^{(m)}, \Sigma^{(m)})$$
$$\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$
$$\mathbf{x}^{(m)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

- Do not want to "explain away" the view dependent variations

$$p(\mathbf{x}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{B}^{(1)}, \mathbf{B}^{(2)}, \Sigma^{(1)}, \Sigma^{(2)} | \mathbf{Y}^{(1)}, \mathbf{Y}^{(2)})$$

---

[1] Tucker 1958.

Inter-Battery Factor Analysis[1]

$$\mathbf{X} = [\mathbf{x}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}], \mathbf{Y} = [\mathbf{y}, \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(N)}]$$

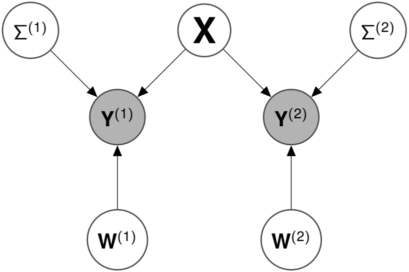
$$\mathbf{W} = \begin{bmatrix} \mathbf{A}^{(1)} & \mathbf{B}^{(1)} & 0 & \dots & 0 \\ \mathbf{A}^{(2)} & 0 & \mathbf{B}^{(2)} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{A}^{(N)} & 0 & 0 & 0 & \mathbf{B}^{(N)} \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \Sigma^{(1)} & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \Sigma^{(N)} \end{bmatrix}$$

$$\mathbf{Y} \sim \mathcal{N}(\mathbf{WX}, \Sigma)$$

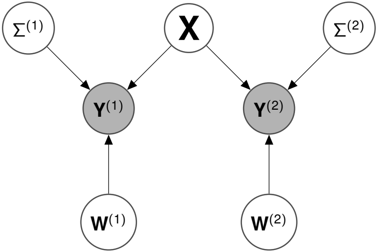
[1] Tucker 1958.

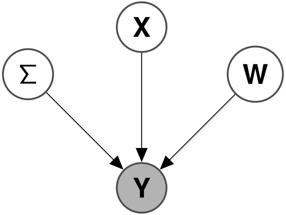
Inter-Battery Factor Analysis[1]



[1] Tucker 1958.

Inter-Battery Factor Analysis[1]





[1] Tucker 1958.

Inter-Battery Factor Analysis[1]

Other models

- Concatenate model reduces to FA with specific structure of $\mathbf{W}$
- Bayesian FA[a]: ignore structure of $\mathbf{W}$
- PPCA[b]: spherical Σ

[a] Ghahramani and Beal 1999.
[b] Tipping and Bishop 1999.

[1] Tucker 1958.

Bayesian IBFA[2]

$$\Sigma \sim IW(\mathbf{S}_0, v_0)$$

$$p(\mathbf{W}) = \prod_{m=1}^{2} p(\mathbf{W}^{(m)}|\alpha_0, \beta_0)$$

$$p(\mathbf{W}^{(m)}|\alpha_0, \beta_0) = \prod_{k=1}^{K} p(w_k^{(m)}|\alpha_k^{(m)}) p(\alpha_k^{(m)}|\alpha_0, \beta_0)$$

$$p(\alpha_k^{(m)}|\alpha_0, \beta_0) \sim \Gamma(\alpha_0, \beta_0)$$

$$p(w_k^{(m)}|\alpha_k^{(m)}) = \mathcal{N}\left(\mathbf{0}, \left(\alpha_k^{(m)}\right)^{-1} \mathbf{I}\right)$$

[2] Klami *et al.* 2013.

Bayesian IBFA[2]

Factorisation

- Prior on **W** induces group row-wise sparsity
- Jointly encourages *shared* representation (columns)
- Variational inference of parameters
- Linear generative mapping

[2] Klami *et al.* 2013.

Bayesian IBFA[2]

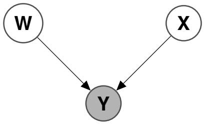
Factorisation

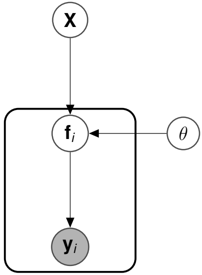
- Prior on **W** induces group row-wise sparsity
- Jointly encourages *shared* representation (columns)
- Variational inference of parameters
- Linear generative mapping

[2] Klami *et al.* 2013.

Non-parametric IBFA[3]

$$\mathbf{X} = \mathbf{W}\mathbf{Y}$$



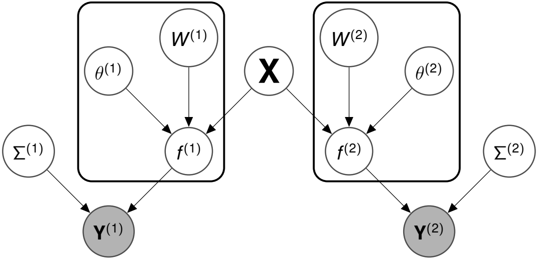


Next step

- History repeats itself
  - MDS/PCA $\Rightarrow$ linear probabilistic $\Rightarrow$ non-linear probabilistic
- IBFA with nonparametric mapping allows for non-linearities

[3] Damianou *et al.* 2012.

Non-parametric IBFA[3]



[3] Damianou *et al.* 2012.

Non-parametric IBFA[3]

Manifold Relevance Determination

- Factorisation inside mapping prior

$$k^{Y}\left(\mathbf{x}_i, \mathbf{x}_j\right) = (\sigma^{Y}_{ard})^2 e^{-\frac{1}{2}\sum_{q=1}^{Q} w_q^Y \left(x_{i,q} - x_{j,q}\right)^2}$$

- Requires bayesian treatment[a]
  - Encourages reduction of (dimensions of) latent space
  - ARD parameters facilitates “turning dimensions off”
- Probabilistic non-linear IBFA

[a] Titsias and Lawrence 2010.

[3] Damianou *et al.* 2012.

Summary

- Feature *learning* in a generative model can be viewed as factor analysis
- Feature *selection* in a generative model can be viewed as multiview factor analysis or inter battery factor analysis
- GP/GP-LVM framework allows for non-parametric formulation of inter battery factor analysis

Introduction

Supervised Factorised Representation Learning

Experiments

Experiments

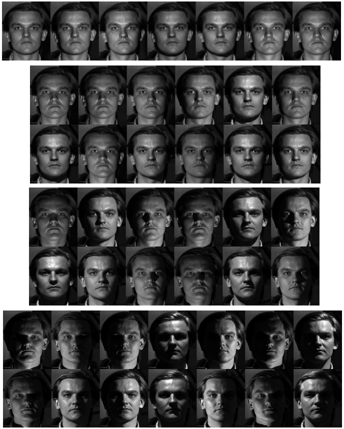
Yale Faces

- Three faces
- 64 illuminations
- $\mathbf{y}_i \in \mathbb{R}^{192\times168}$
- Light alignment

Experiments

Yale Faces

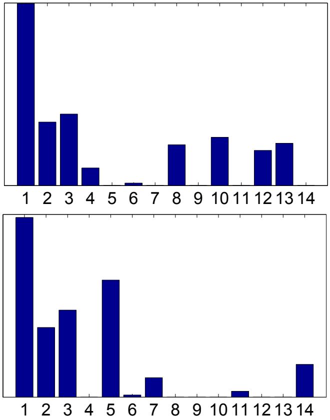
- Three faces
- 64 illuminations
- $\mathbf{y}_i \in \mathbb{R}^{192\times168}$
- Light alignment



Experiments

Yale Faces

- Three faces
- 64 illuminations
- $\mathbf{y}_i \in \mathbb{R}^{192\times168}$
- Light alignment



Experiments

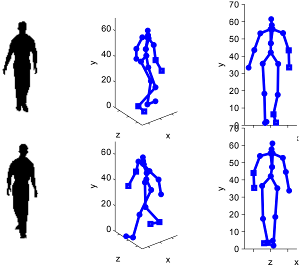
Loading video

Experiments

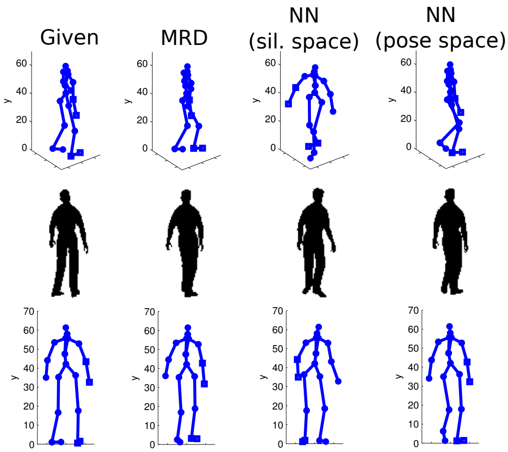
Pose Estimation ([a])

[a] Agarwal and Triggs 2003.

- Silhouette images
- Image features
- Estimate 3D pose
- Highly Ambigous



Experiments



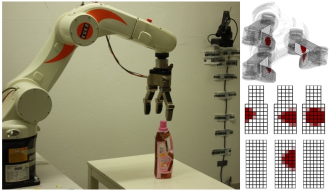
Experiments

	Error
Mean Training Pose	6.16
Linear Regression	5.86
GP Regression	4.27
Nearest Neighbour (sil. space)	4.88
Nearest Neighbour with sequences (sil. space)	4.04
Nearest Neighbour (pose space)	2.08
Shared GP-LVM	5.13
MRD without Dynamics	4.67
MRD with Dynamics	**2.94**

Experiments

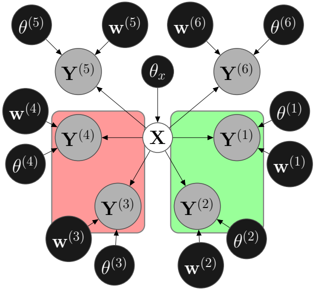
Robotic Grasping

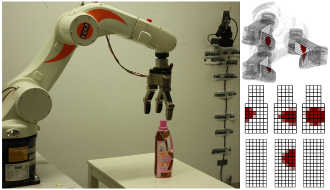
- Gripper pose
- Tactile sensor
- Object pose and identity



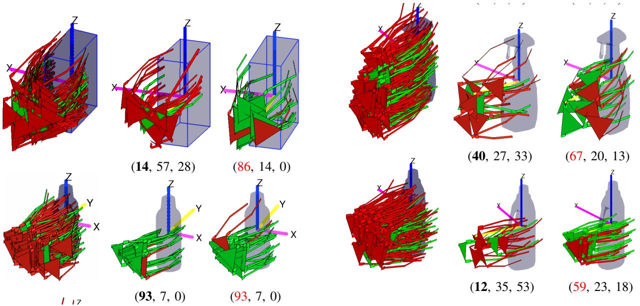
Experiments

Experiments

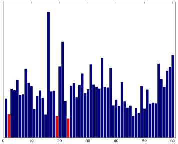


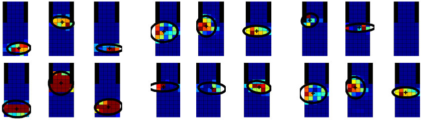


Experiments

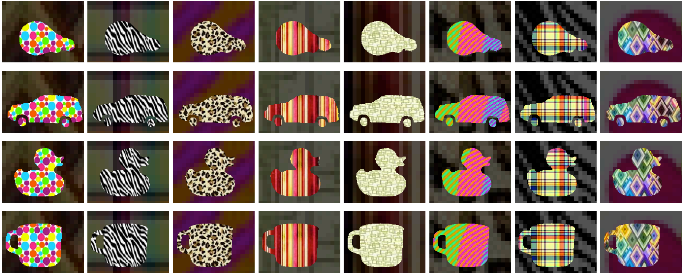


Experiments



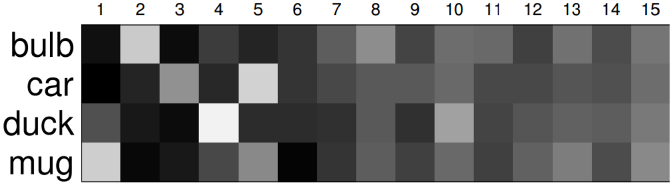


Bonus: Topic Modelling[4]



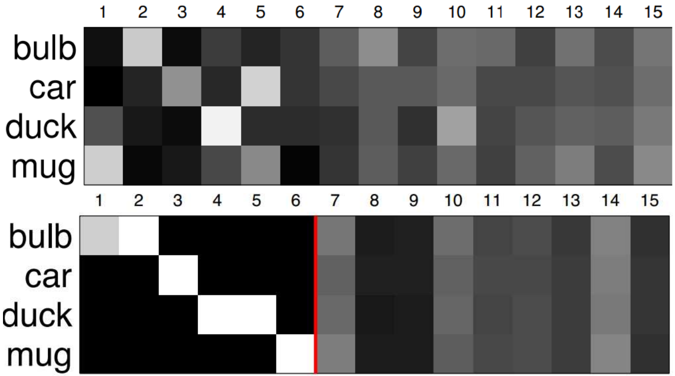
[4] Zhang *et al.* 2013.

Bonus: Topic Modelling[4]



[4] Zhang *et al.* 2013.

Bonus: Topic Modelling[4]



[4] Zhang *et al.* 2013.

Bonus: Topic Modelling[4]

0.25	0.13	0.5	0.13
0.13	0.5	0.25	0.13
0.25	0.13	0.5	0.13
0.13	0.25	0.5	0.13

0.88	0	0.13	0
0.25	0.75	0	0
0	0.13	0.88	0
0.13	0	0.13	0.75

[4] Zhang *et al.* 2013.

Future Work

- Approximate marginalisation of latent space
  - interesting priors
  - auto-encoders
  - deep models
- Bigger data-sets
- Automatic alignment

e.o.f.

References I

- H Hotelling. “Relations between two sets of variates”. In: *Biometrika* 28.3/4 (1936), pp. 321–377.
- D. Lowe and Michael E Tipping. “NeuroScale: Novel topographic feature extraction using RBF networks”. In: *Advances in Neural Information Processing Systems* (1997), pp. 543–549.
- Geoffrey E Hinton and Ruslan R Salakhutdinov. “Reducing the Dimensionality of Data with Neural Networks”. In: *Science* 313.5786 (July 2006), pp. 504–507.

References II

- Pascal Vincent *et al.* “Extracting and composing robust features with denoising autoencoders”. In: *International Conference on Machine Learning*. New York, USA: ACM Press, 2008, pp. 1096–1103.
- Neil D Lawrence and J Quiñonero-Candela. “Local distance preservation in the GP-LVM through back constraints”. In: *Proceedings of the 23rd international conference on Machine learning* (2006), pp. 513–520.
- Carl Henrik Ek *et al.* “Gaussian process latent variable models for human pose estimation”. In: *International conference on Machine learning for multimodal interaction* (2007), pp. 132–143.

References III

- Jasper Snoek *et al.* "Nonparametric guidance of autoencoder representations using label information". In: *Journal of Machine Learning Research* 13 (2012), pp. 2567–2588.
- Ledyard R Tucker. "An Inter-Battery Method of Factory Analysis". In: *Psychometrika* 23 (June 1958).
- Francis R Bach and Michael I Jordan. *A probabilistic interpretation of canonical correlation analysis*. Tech. rep. 2005.

References IV

Zoubin Ghahramani and Matthew J Beal. “Variational Inference for Bayesian Mixtures of Factor Analysers.” In: *Advances in Neural Information Processing Systems*. Dec. 1999, pp. 449–455.

Michael E Tipping and C.M. Bishop. “Probabilistic principal component analysis”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61.3 (1999), pp. 611–622.

Arto Klami *et al.* “Bayesian Canonical Correlation Analysis”. In: *The Journal of Machine Learning Research* 14 (Apr. 2013), pp. 965–1003.

References V

- Andreas C Damianou *et al.* "Manifold Relevance Determination". In: *International Conference on Machine Learning*. June 2012, pp. 145–152.
- Michalis Titsias and Neil D Lawrence. "Bayesian Gaussian Process Latent Variable Model". In: *International Conference on Airtificial Inteligence and Statistical Learning*. 2010, pp. 844–851.
- A Agarwal and B Triggs. "3D human pose from silhouettes by relevance vector regression". In: *IEEE Conference on Computer Vision and Pattern Recognition*. Dec. 2003, pp. II–I2.

References VI

Cheng Zhang *et al.* "Factorized Topic Models". In: *International Conference on Learning Representations*. Apr. 2013.