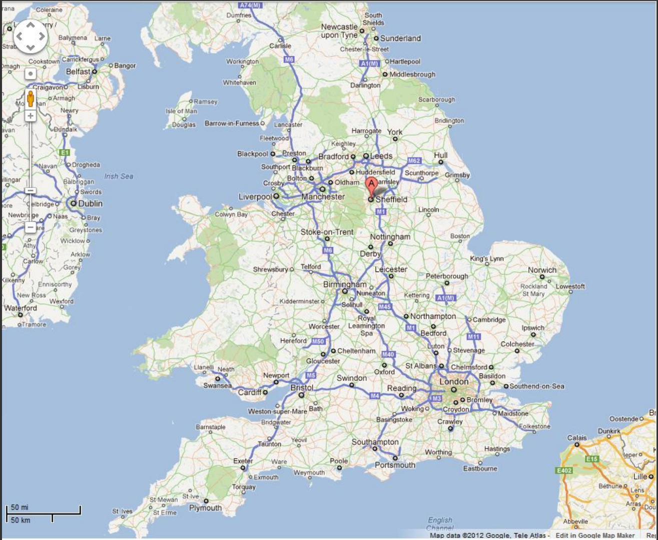
Non-linear probabilistic dimensionality reduction for dynamical and multi-modal vision datasets

Andreas Damianou[1]
joint work with Carl Henrik Ek[2], Michalis Titsias[3] and Neil Lawrence[1]

[1] Department of Neuro- and Computer Science, University of Sheffield, UK
[2] Computer Vision and Active Perception Lab , KTH
[3] Wellcome Trust Centre for Human Genetics, University of Oxford


Coleraine
Ballymena
Larne
Cookstown
Carrickfergus
Bangor
Belfast
Lisburn
Craigavon
Armagh
Newry
Dundalk
Navan
Drogheda
Balbriggan
Swords
Dublin
Celbridge
Naas
Bray
Greystones
Wicklow
Athy
Arklow
Carlow
Gorey
Kilkenny
Enniscorthy
New Ross
Wexford
Waterford
Tramore
Irish Sea
Isle of Man
Ramsey
Douglas
Dumfries
Carlisle
Workington
Whitehaven
Barrow-in-Furness
Lancaster
Fleetwood
Blackpool
Preston
Southport
Blackburn
Crosby
Bolton
Liverpool
Manchester
Oldham
Colwyn Bay
Chester
Newcastle upon Tyne
South Shields
Sunderland
Chester-le-Street
Hartlepool
Middlesbrough
Darlington
Scarborough
Bridlington
Harrogate
York
Keighley
Bradford
Leeds
Huddersfield
Hull
Scunthorpe
Grimsby
Sheffield
Lincoln
Stoke-on-Trent
Nottingham
Derby
Boston
King's Lynn
Norwich
Lowestoft
Shrewsbury
Telford
Leicester
Peterborough
Birmingham
Nuneaton
Kettering
Kidderminster
Solihull
Royal Leamington Spa
Northampton
Cambridge
Worcester
Bedford
Ipswich
Hereford
Luton
Stevenage
Colchester
Cheltenham
Gloucester
St Albans
Chelmsford
Llanelli
Neath
Swansea
Newport
Cardiff
Bristol
Swindon
Oxford
Basildon
London
Southend-on-Sea
Reading
Bromley
Croydon
Woking
Maidstone
Weston-super-Mare
Bath
Bridgwater
Basingstoke
Crawley
Folkestone
Barnstaple
Taunton
Yeovil
Southampton
Worthing
Hastings
Exeter
Poole
Portsmouth
Eastbourne
Exmouth
Weymouth
Torquay
St Ives
St Mewan
St Erme
Plymouth
English Channel
Calais
Dunkirk
Oostende
Lille
Béthune
Lens
Arras
Abbeville
50 mi
50 km
Map data ©2012 Google, Tele Atlas - Edit in Google Map Maker


SITraN


SITraN

Outline

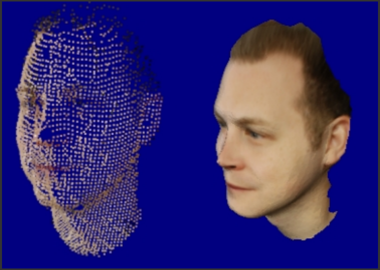
Dimensionality reduction techniques

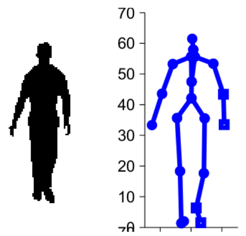
Gaussian process latent variable model (GP-LVM)

Bayesian GP-LVM

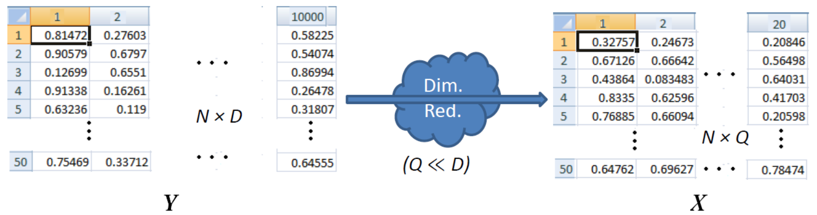
Structure in the latent space
- Modelling dynamics
- Multi-modal modelling

Real-world datasets in computer vision are usually high-dimensional, complex and noisy





Dimensionality reduction



Dimensionality reduction techniques 1/2

Probabilistic vs non-probabilistic

A probabilistic interpretation allows us to:

- Have a model of the data
- Handle incomplete data
- Generate/sample novel data
- Extend the model with prior information or integrate it with other models (e.g. mixtures)

Probabilistic, generative methods

- Observed (high-dimensional) data: $Y \in \mathbb{R}^{N \times D}$
  These contain redundant information
- Actual (low-dimensional) data: $X \in \mathbb{R}^{N \times Q}$, $Q \ll D$
  These are unobserved and (ideally) contain only the minimum amount of information needed to correctly describe the phenomenon
- Work "backwards": learn $f : X \mapsto Y$

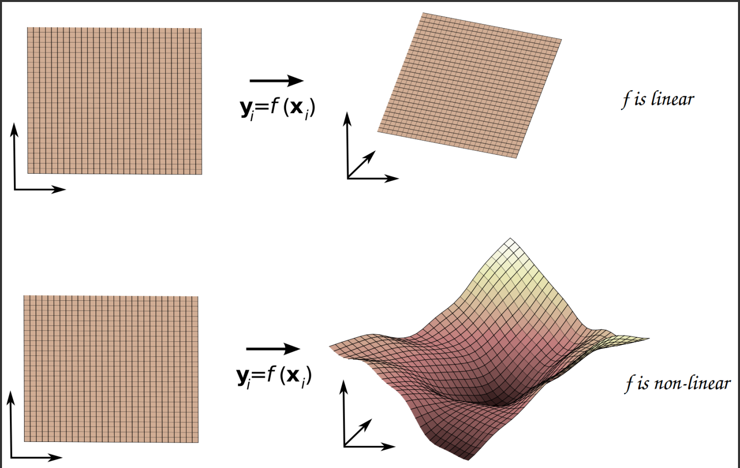
Probabilistic, generative methods

- Observed (high-dimensional) data: $Y \in \mathbb{R}^{N \times D}$
  These contain redundant information
- Actual (low-dimensional) data: $X \in \mathbb{R}^{N \times Q}$, $Q \ll D$
  These are <u>unobserved</u> and (ideally) contain only the minimum amount of information needed to correctly describe the phenomenon
- Work "backwards": learn $f : X \mapsto Y$
- Model:

$$y_{nd} = f_d(\mathbf{x}_n, W) + \epsilon_n \ , \quad \epsilon_n \sim \mathcal{N}(0, \beta^{-1})$$

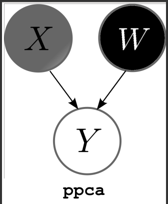
Dimensionality reduction techniques 2/2

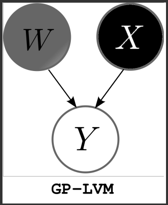
Linear vs non-linear



Gaussian process latent variable model (GP-LVM)

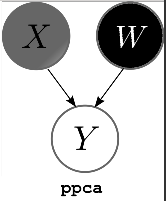
- PPCA places a prior on and marginalises the latent space X and optimises the *linear* mapping's parameters W
- GP-LVM does the opposite: the prior is placed on the mapping.

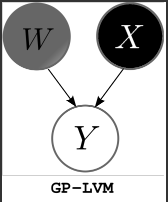




Gaussian process latent variable model (GP-LVM)

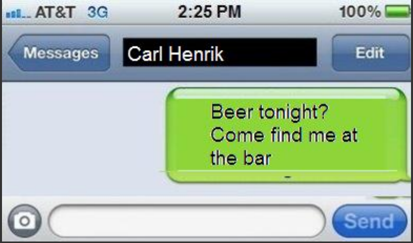
- PPCA places a prior on and marginalises the latent space X and optimises the *linear* mapping's parameters W
- GP-LVM does the opposite: the prior is placed on the mapping.

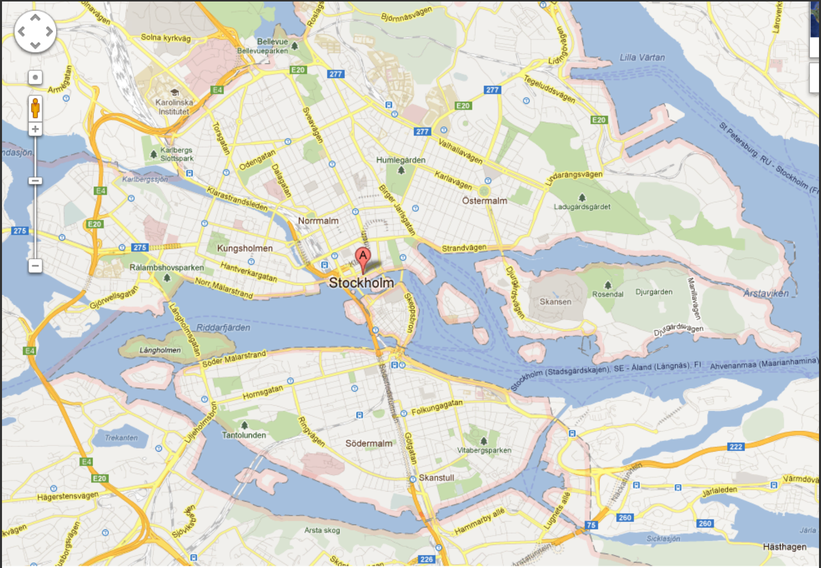




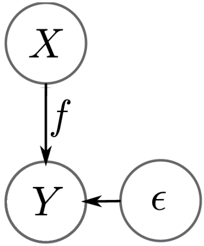
- A GP prior $f \sim \mathcal{GP}(\mathbf{0}, k(x, x'))$ allows for *non-linear mappings* if the kernel k is non-linear. For example:

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


AT&T 3G
2:25 PM
100%
Messages
Carl Henrik
Edit
Beer tonight?
Come find me at
the bar
Send

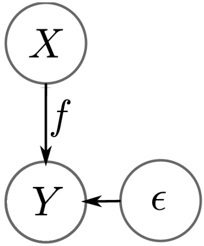

Stockholm
Solna kyrkväg
Bellevue
Bellevueparken
Björnnäsvägen
Lilla Värtan
Karolinska Institutet
Tegeluddsvägen
Valhallavägen
Humlegården
Karlavägen
Östermalm
Lindarängsvägen
Ladugårdsgärdet
St Petersburg, RU - Stockholm (F
Karlbergs Slottspark
Karlbergssjön
Odengatan
Torsgatan
Sveavägen
Dalagatan
Klarastrandsleden
Norrmalm
Birger Jarlsgatan
Strandvägen
Kungsholmen
Hantverkargatan
Rålambshovsparken
Norr Mälarstrand
Gjörwellsgatan
Skeppsbron
Djurgårdsvägen
Skansen
Rosendal
Djurgården
Manillavägen
Årstaviken
Riddarfjärden
Långholmen
Långholmsgatan
Söder Mälarstrand
Stockholm (Stadsgårdskajen), SE - Åland (Långnäs), FI
Ahvenanmaa (Maarianhamina)
Hornsgatan
Folkungagatan
Tantolunden
Ringvägen
Södermalm
Götgatan
Vitabergsparken
Skanstull
Trekanten
Liljeholmsbron
Hägerstensvägen
Årsta skog
Hammarby allé
Lugnets allé
Järlaleden
Värmdöväg
Sicklasjön
Hästhagen
E4
E20
277
275
222
260
75
226

Optimising the GP-LVM



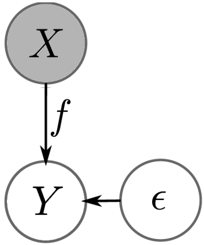
- Objective function for optimisation is $p(Y|X)$

Optimising the GP-LVM



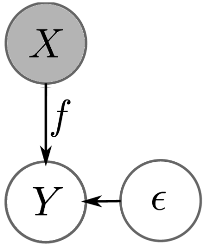
- Objective function for optimisation is $p(Y|X)$
- Problem: this finds a single point *(MAP)* estimate for X

Optimising the GP-LVM



- Objective function for optimisation is $p(Y|X)$
- Problem: this finds a single point *(MAP)* estimate for X
- We would prefer to instead find a *distribution* over $X \Rightarrow$ *Bayesian GP-LVM*

Optimising the GP-LVM



- Objective function for optimisation is $p(Y|X)$
- Problem: this finds a single point *(MAP)* estimate for X
- We would prefer to instead find a *distribution* over $X \Rightarrow$ *Bayesian GP-LVM*
- This allows for:
  - training robust to overfitting
  - automatic detection for the dimensionality of X
  - forcing known structure on the latent space

Bayesian GPLVM

- Marginal likelihood in GPLVM:

$$p(Y|X) = \int p(Y|\mathbf{f})\, p(\mathbf{f}|X) d\mathbf{f} = \mathcal{N}(Y|\mathbf{0}, K_{NN} + \beta^{-1} I_N)$$

  The GPLVM is trained by maximizing $p(Y|X)$ w.r.t the mapping's parameters and X (jointly) $\Rightarrow$ *MAP* estimate,

- Bayesian GPLVM: Also integrate out X's:

$$p(Y) = \int p(Y|X)\, p(X) dX$$

$$p(X) = \prod_{n=1}^{N} N(\mathbf{x}_n|\mathbf{0}, I_Q)$$

Bayesian GPLVM

- Marginal likelihood in GPLVM:

$$p(Y|X) = \int p\,(Y|\mathbf{f})\,p(\mathbf{f}|X)d\mathbf{f} = \mathcal{N}(Y|\mathbf{0}, K_{NN} + \beta^{-1}I_N)$$

  The GPLVM is trained by maximizing $p(Y|X)$ w.r.t the mapping's parameters and X (jointly) $\Rightarrow$ *MAP* estimate,

- Bayesian GPLVM: Also integrate out X's:

$$p(Y) = \int p\,(Y|X)\,p(X)dX$$

$$p(X) = \prod_{n=1}^{N} N(\mathbf{x}_n|\mathbf{0}, I_Q)$$

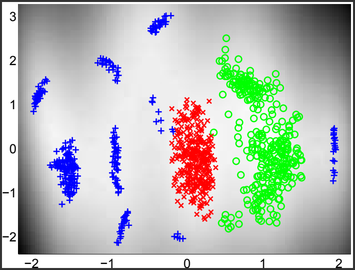
- Problem: The marginal likelihood as well as the posterior $p(X|Y)$ are intractable $\Rightarrow$ the variational framework of [Titsias and Lawrence, 2010] resolves this

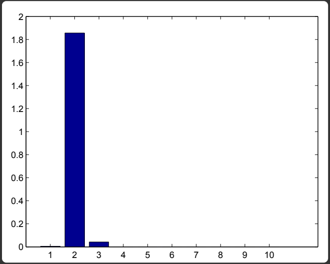
Automatic dimensionality detection

- Achieved by employing *automatic relevance determination (ARD)* priors for the mapping f.
- $f \sim \mathcal{GP}(\mathbf{0}, k_f)$ with:

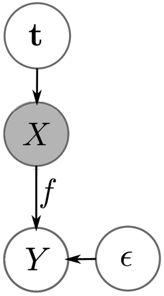
$$k_f\left(\mathbf{x}_i, \mathbf{x}_j\right) = \sigma^2 e^{-\frac{1}{2}\sum_{q=1}^{Q} w_q (x_{i,q} - x_{j,q})^2}$$

- Example:





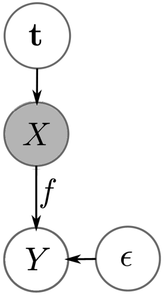
Modelling dynamics



- If Y form is a multivariate time-series, then X also has to be one

[Damianou et al., 2011]

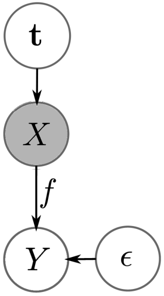
Modelling dynamics



- If Y form is a multivariate time-series, then X also has to be one
- Place a temporal GP prior on the latent space: $\mathbf{x} = x(t) = \mathcal{GP}(\mathbf{0}, k_x)$

[Damianou et al., 2011]

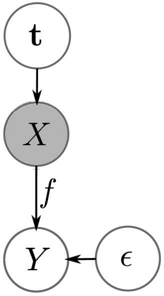
Modelling dynamics



- If Y form is a multivariate time-series, then X also has to be one
- Place a temporal GP prior on the latent space: $\mathbf{x} = x(t) = \mathcal{GP}(\mathbf{0}, k_x)$
- Dynamics are encoded in the covariance matrix $K_x = k_x(\mathbf{t}, \mathbf{t})$, e.g. forcing K_x to be block-diagonal allows to jointly model individual sequences

[Damianou et al., 2011]

Modelling dynamics

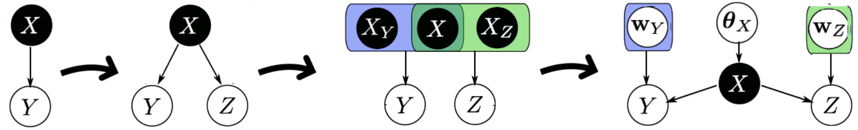


- If Y form is a multivariate time-series, then X also has to be one
- Place a temporal GP prior on the latent space: $\mathbf{x} = x(t) = \mathcal{GP}(\mathbf{0}, k_x)$
- Dynamics are encoded in the covariance matrix $K_x = k_x(\mathbf{t}, \mathbf{t})$, e.g. forcing K_x to be block-diagonal allows to jointly model individual sequences
- *Video examples...*

[Damianou et al., 2011]

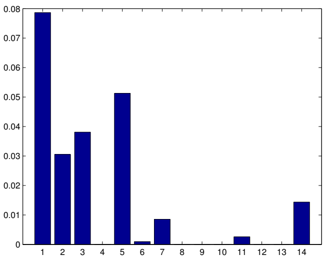
Multi-modal modelling

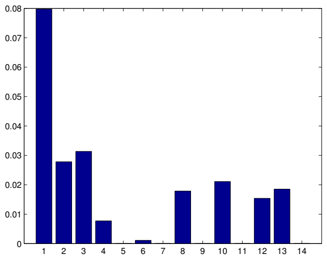
- Several observation modalities for the same underlying phenomenon
- Challenge: factorise the latent space into parts that are either private or shared for all modalities
- Bayesian solution: use a separate set of *ARD* parameters for each modality
- The ARD weights are optimised to learn the responsibility of each latent dimension for generating each of the observation spaces



Example

- Latent space X initialised with 14 dimensions
- Optimisation factorises X as:
  - Shared subspace: $q = \{1, 2, 3\}$
  - Private subspace a: $q = \{5, 7, 11, 14\}$
  - Private subspace b: $q = \{4, 8, 10, 12, 13\}$





- Video...

Summary

- GP-LVM: probabilistic non-linear dimensionality reduction
- Bayesian GP-LVM: placing a prior over and marginalising the latent space
- Dynamical framework: constraining the latent space to be a timeseries
- Multi-modal framework: automatically segment the latent space to shared and private subspaces

Tack!

KTH

Carl Henrik Ek

Univ. of Oxford

Michalis Titsias

Univ. of Sheffield

Neil Lawrence

Funding

- University of Sheffield Moody endowment fund
- Greek State Scholarships Foundation (IKY)