Machine Learning

Project

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

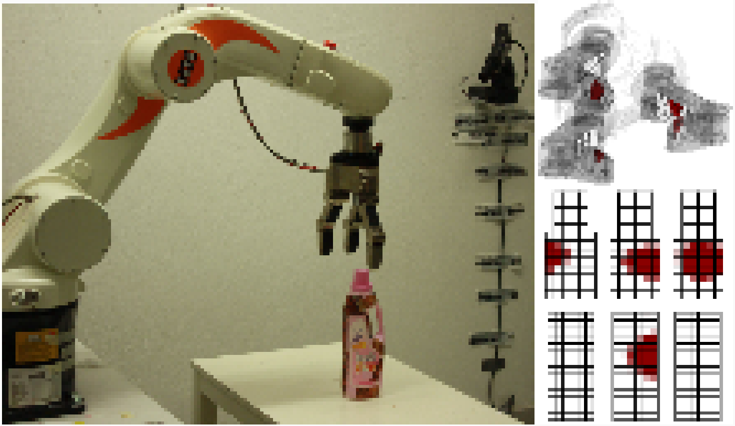
November 21, 2017

http://www.carlhenrik.com

Introduction

Motivation

Robots

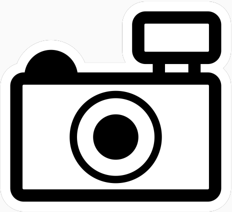


Challenges

Grasping

- how to grasp and pick-up objects?
- where to approach the object?
- how to place fingers?

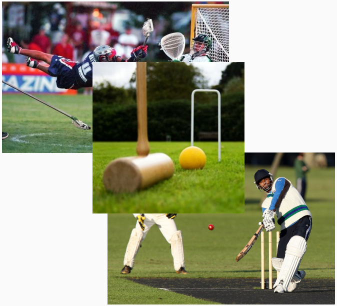
Sensors



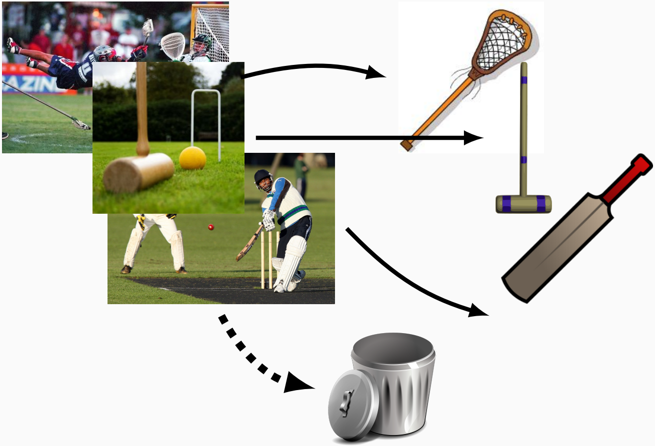
Idea

1. It is easy to get failed grasps
2. Different modalities are often availible and useful at different times
3. Can we learn how to fuse the different modalities?
4. Can we learn how to correct failed grasps rather than learning from scratch?

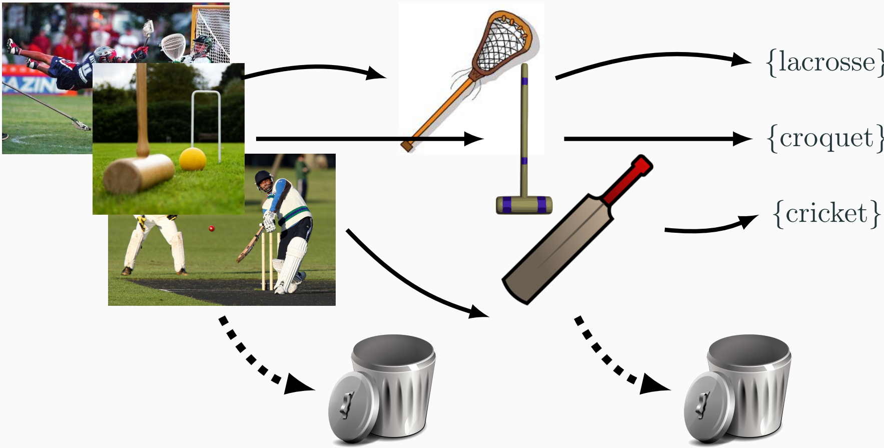
Latent Structure



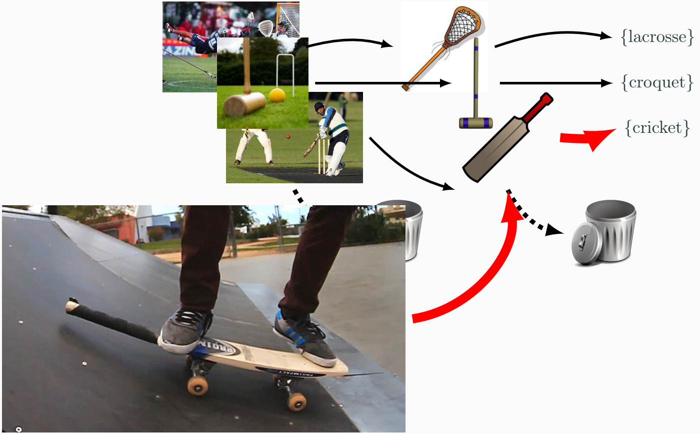
Latent Structure



Latent Structure

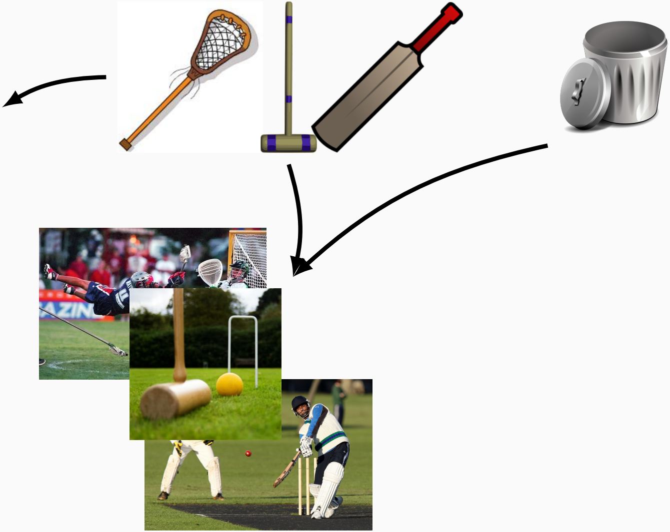


Latent Structure



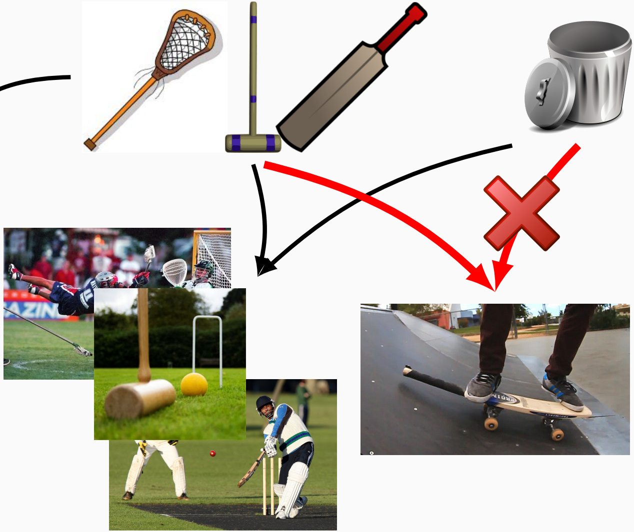
Latent Structure

{lacrosse}
{croquet}
{cricket}



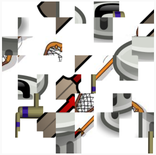
Latent Structure

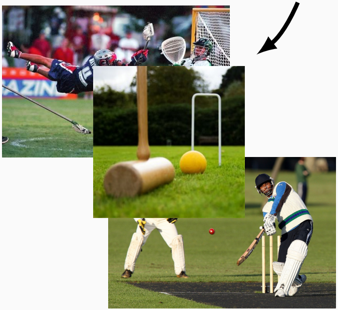
{lacrosse}
{croquet}
{cricket}



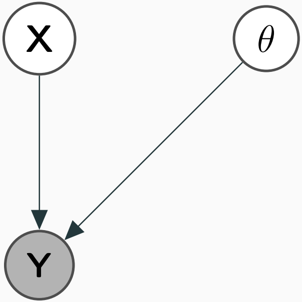
Latent Structure

{lacrosse}
{croquet}
{cricket}

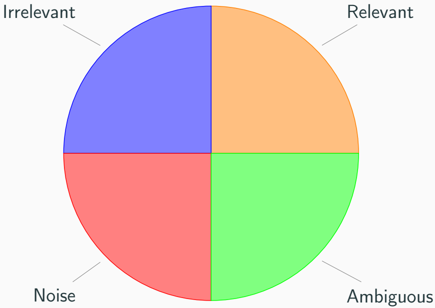




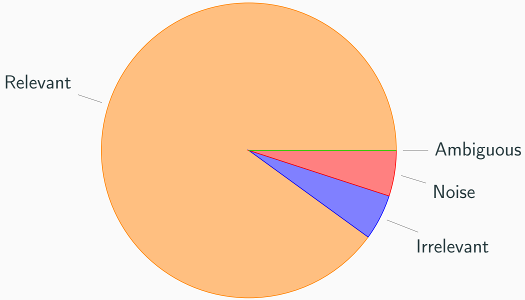
Latent Structure



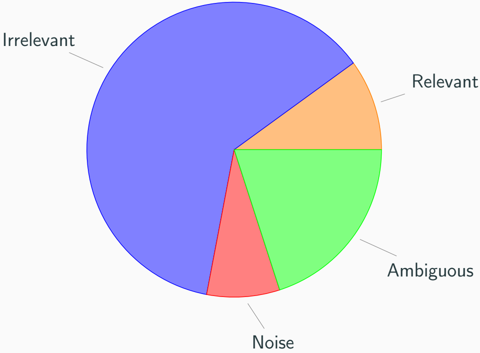
Latent Structure



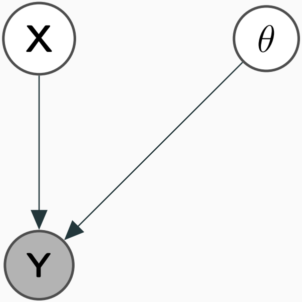
Latent Structure



Latent Structure

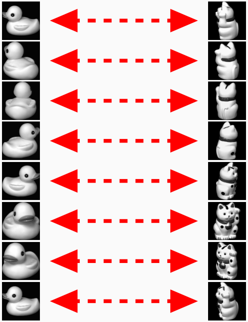


Latent Structure

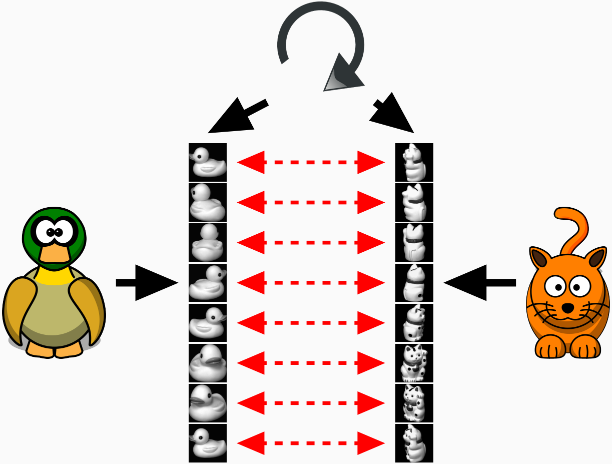


Alignments

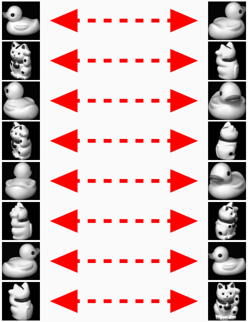
Alignments



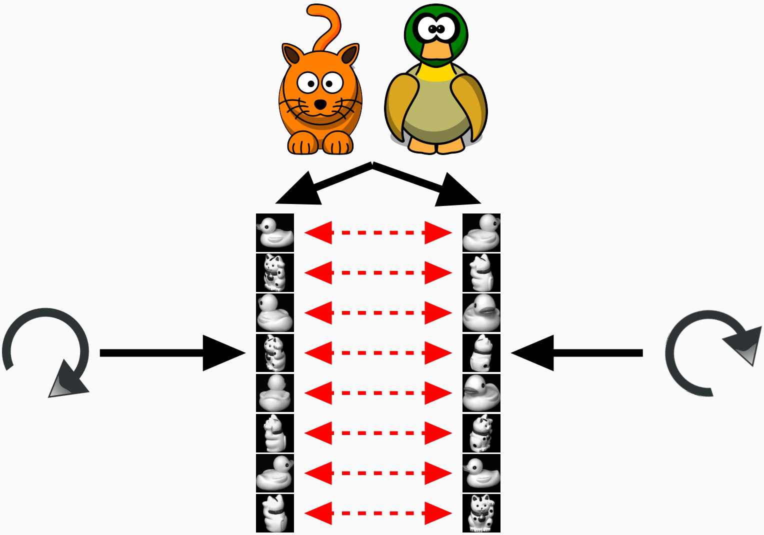
Alignments



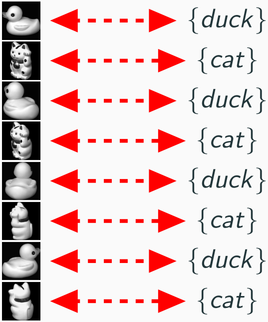
Alignments



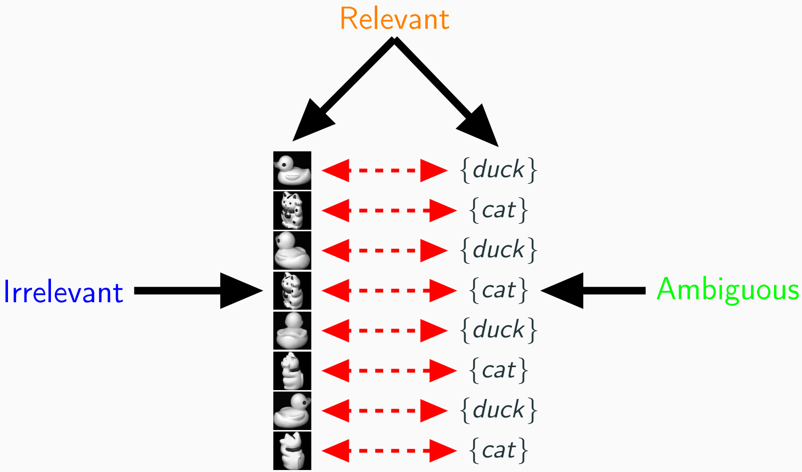
Alignments



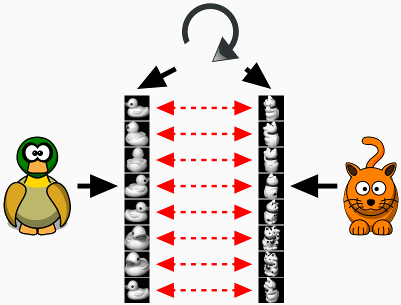
Alignments



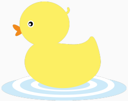
Alignments

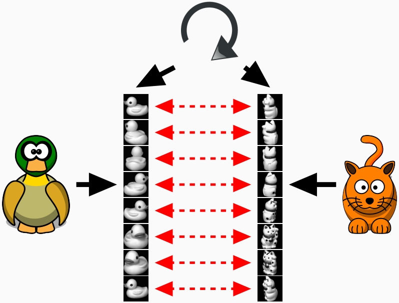

Relevant
Irrelevant
Ambiguous
{duck}
{cat}
{duck}
{cat}
{duck}
{cat}
{duck}
{cat}

Inference in a factorised model

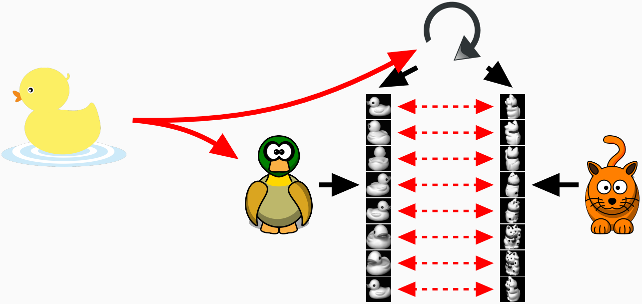


Inference in a factorised model

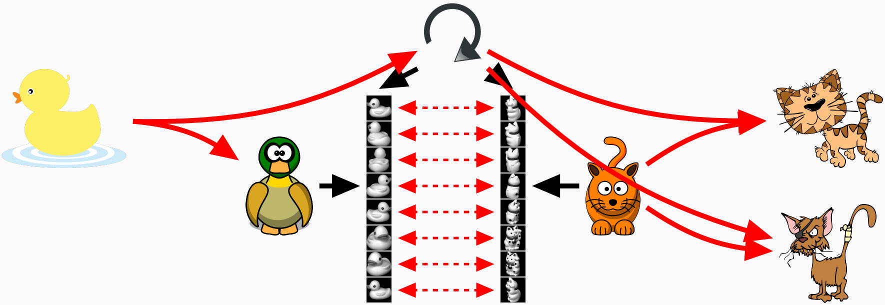




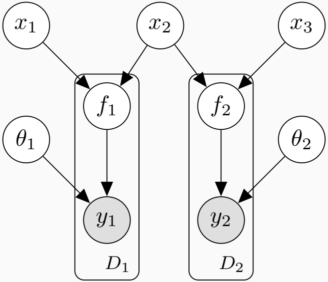
Inference in a factorised model



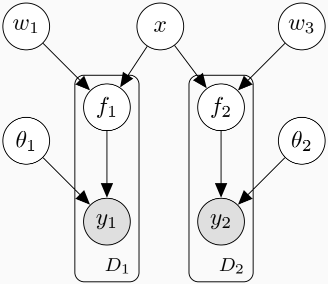
Inference in a factorised model



Explaining Away cont.

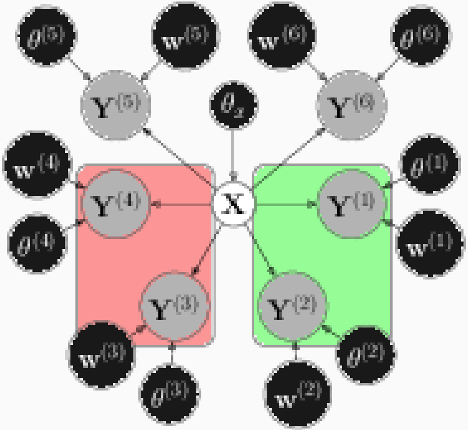


IBFA with GP-LVM

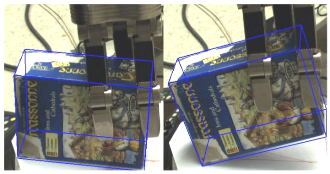


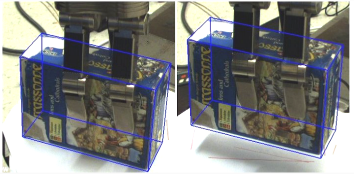
$$y_1 = f\left(w_1^{\mathrm{T}} x\right) \quad y_2 = f\left(w_2^{\mathrm{T}} x\right)$$

Probabilistic Consolidation of Grasp Experience [1]

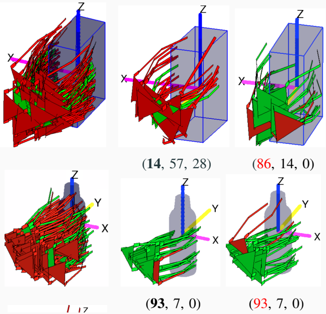


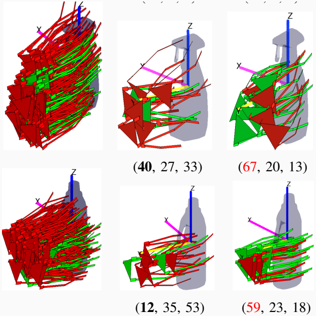
Alignments





Results

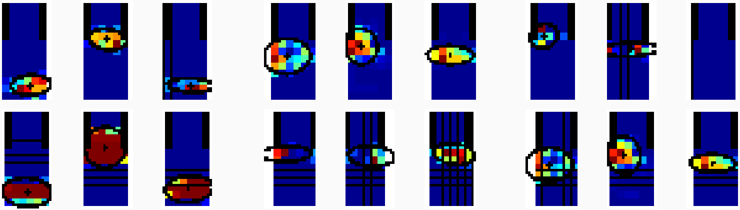




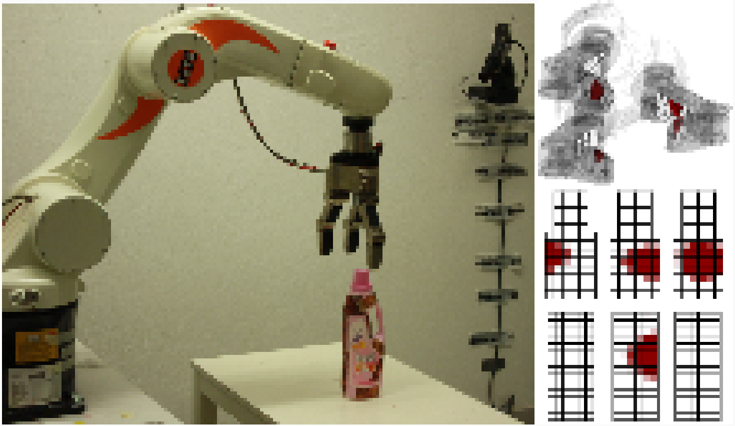
Models

Box	79	21	0	0		100	0	0	0		93	0	7	0
Oval	0	100	0	0		0	100	0	0		6	82	6	6
Spray	0	13	67	20		7	47	47	0		0	73	27	0
Cyl.	0	0	0	100		0	0	0	100		0	0	0	100

Models



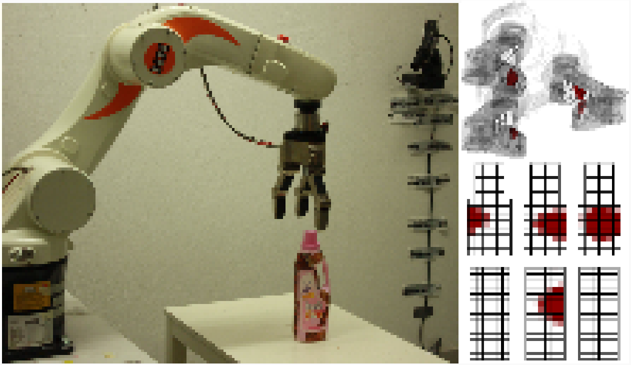
Robot



The project

Specifying Model

Getting data



Developing inference

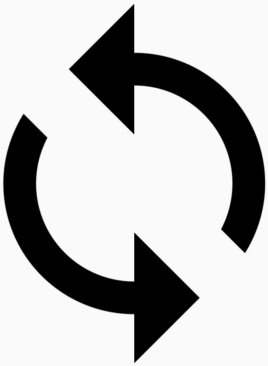
$$\tilde{\mathcal{L}} = \int_{\mathbf{X},\mathbf{F},\mathbf{U}} \prod_{j=1}^{d} p(\mathbf{f}_{:,j}|\mathbf{u}_{:,j}, \mathbf{X}, \mathbf{Z}) q(\mathbf{u}_{:,j}) q(\mathbf{X})$$

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

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

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

Iterate



Most important lesson

Most important lesson II

Clearly separate

- Model
    - how is the data generated
- Algorithms
    - how do I merge model with data
- Prediction
    - how do I get output from the model

-*be aware of your assumptions*

eof

Inference

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

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

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

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

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

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

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

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

ELBO

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

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

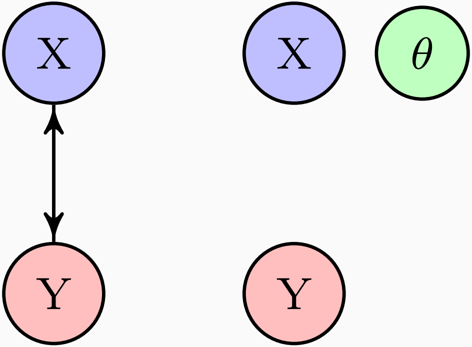
ELBO

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

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

- 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



Lower Bound

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

Lower Bound

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

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

Lower Bound

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

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

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

Lower Bound

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

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

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* $\mathbf{U}$ that completely represents $\mathbf{F}$, i.e. $\mathbf{U}$ is sufficient statistics of $\mathbf{F}$,

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

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

$$
\begin{aligned}
\tilde{\mathcal{L}} &= \int_{\mathbf{X},\mathbf{F},\mathbf{U}} q(\mathbf{F})q(\mathbf{U})q(\mathbf{X}) \log \frac{p(\mathbf{Y},\mathbf{F},\mathbf{U}|\mathbf{X},\mathbf{Z})}{q(\mathbf{F})q(\mathbf{U})} \\
&= \int_{\mathbf{X},\mathbf{F},\mathbf{U}} q(\mathbf{F})q(\mathbf{U})q(\mathbf{X}) \log \frac{\prod_{j=1}^{d} p(\mathbf{y}_{:,j}|\mathbf{f}_{:,j})p(\mathbf{f}_{:,j}|\mathbf{u}_{:,j},\mathbf{X},\mathbf{Z})p(\mathbf{u}_{:,j}|\mathbf{Z})}{q(\mathbf{F})q(\mathbf{U})}
\end{aligned}
$$

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

References

Yasemin Bekiroğlu, Andreas C Damianou, Renaud Detry, A Johannes Stork, Danica Kragic, and Carl Henrik Ek.
Probabilistic Consolidation of Grasp Experience.
In *IEEE International Conference on Robotics and Automation*, Stockholm, May 2016.