Simple Beamer Class

Sascha Frank

February 12, 2014

Table of contents

Overview of Variational Inference

- Turns a complicated inference problems into an optimization problem
- Minimizes the KL divergence from the variational distribution to the posterior distribution
- EM is a special case

Model Image

put fig 2 here
give gloss of variables: relate them to the input/output/nuisance ones Jason defines

Evidence Lower Bound (ELBO)

- Key insight is the use of Jensen's inequality
- We choose q from some family of distributions $\mathcal{Q}$

$$\log(p(x \mid \alpha)) \quad = \quad \log \int p(x, z, \beta \mid \alpha) dz \; d\beta$$

Evidence Lower Bound (ELBO)

- Key insight is the use of Jensen's inequality
- We choose q from some family of distributions $\mathcal{Q}$

$$
\begin{aligned}
\log(p(x \mid \alpha)) &= \log \int p(x, z, \beta \mid \alpha) dz \; d\beta \\
&= \log \int p(x, z, \beta \mid \alpha) \frac{q(z, \beta)}{q(z, \beta)} dz d\beta
\end{aligned}
$$

Evidence Lower Bound (ELBO)

- Key insight is the use of Jensen's inequality
- We choose q from some family of distributions $\mathcal{Q}$

$$
\begin{aligned}
\log(p(x \mid \alpha)) &= \log \int p(x, z, \beta \mid \alpha) dz \; d\beta \\
&= \log \int p(x, z, \beta \mid \alpha) \frac{q(z, \beta)}{q(z, \beta)} dz d\beta \\
&= \log \left(\mathbb{E}_q \left[\frac{p(x, z, \beta \mid \alpha)}{q(z, \beta)} \right] \right)
\end{aligned}
$$

Evidence Lower Bound (ELBO)

- Key insight is the use of Jensen's inequality
- We choose q from some family of distributions $\mathcal{Q}$

$$
\begin{aligned}
\log(p(x \mid \alpha)) &= \log \int p(x, z, \beta \mid \alpha) dz \; d\beta \\
&= \log \int p(x, z, \beta \mid \alpha) \frac{q(z, \beta)}{q(z, \beta)} dz d\beta \\
&= \log \left(\mathbb{E}_q \left[\frac{p(x, z, \beta \mid \alpha)}{q(z, \beta)} \right] \right) \\
&\geq \mathbb{E}_q \left[\log p(x, z, \beta \mid \alpha) \right] - \mathbb{E}_q \left[q(z, \beta) \right]
\end{aligned}
$$

Evidence Lower Bound (ELBO)

- Key insight is the use of Jensen's inequality
- We choose q from some family of distributions $\mathcal{Q}$

$$
\begin{aligned}
\log(p(x \mid \alpha)) &= \log \int p(x, z, \beta \mid \alpha) dz \; d\beta \\
&= \log \int p(x, z, \beta \mid \alpha) \frac{q(z, \beta)}{q(z, \beta)} dz d\beta \\
&= \log \left(\mathbb{E}_q \left[\frac{p(x, z, \beta \mid \alpha)}{q(z, \beta)} \right] \right) \\
&\geq \mathbb{E}_q \left[\log p(x, z, \beta \mid \alpha) \right] - \mathbb{E}_q \left[q(z, \beta) \right] \\
&= \mathcal{L}(q)
\end{aligned}
$$

Why does maximizes the ELBO minimize the KL-Divergence?

$$\begin{aligned}
\mathrm{KL}(q(z,\beta)||p(z,\beta|x) &= \mathbb{E}_q\left[\log(q(z,\beta)\right] - \mathbb{E}_q\left[\log p(z,\beta|x)\right] \\
&= \mathbb{E}_q\left[\log(q(z,\beta)\right] - \mathbb{E}_q\left[\log p(z,\beta,x)\right] + \log p(x
\end{aligned}$$

$$= -\mathcal{L}(q) + const.$$

- Maximizing $\mathcal{L}(q)$ is just minimizing $-\mathcal{L}(q)$

How is EM a special case?

- How do we choose $\mathcal{Q}$?
- We want it to be easily computable!
- What if $p(z, \beta|x) \in \mathcal{Q}$?
- Then computing expectations under q is just inference in our model!
- Consider a standard HMM, our E-step involves direct computation of the marginals through Forward-Backwards
- Let $p_\theta(z|x)$ be our model and θ_1 be a setting of parameters where z is a latent variable (tags in an HMM) and x is an observed variable (words in an HMM)
- Since $p_\theta \in \mathcal{Q}$, $\forall \theta \in \Theta$, we simply need to maximize $\sum_z p(z|x;\theta_1) \log\left(p(z|x;\theta_2)\right)$ with respect to θ_2
- blah

Variational $+ x, \forall x$

- Variational EM - the process described above
- Variational Bayes - approximate inference (no M-step)
- Variational Decoding - approximation decoding

unnumbered lists

- Introduction to LaTeX
- Course 2
- Termpapers and presentations with LaTeX
- Beamer class

lists with pause

- Introduction to $\LaTeX$

lists with pause

- Introduction to $\LaTeX$
- Course 2

lists with pause

- Introduction to LaTeX
- Course 2
- Termpapers and presentations with LaTeX

lists with pause

- Introduction to LaTeX
- Course 2
- Termpapers and presentations with LaTeX
- Beamer class

numbered lists

1. Introduction to LaTeX
2. Course 2
3. Termpapers and presentations with LaTeX
4. Beamer class

numbered lists with pause

1. Introduction to $\LaTeX$

numbered lists with pause

1. Introduction to LaTeX
2. Course 2

numbered lists with pause

1. Introduction to LaTeX
2. Course 2
3. Termpapers and presentations with LaTeX

numbered lists with pause

1. Introduction to LaTeX
2. Course 2
3. Termpapers and presentations with LaTeX
4. Beamer class

Tables

Date	Instructor	Title
WS 04/05	Sascha Frank	First steps with LaTeX
SS 05	Sascha Frank	LaTeX Course serial

Tables with pause

A	B	C

Tables with pause

A	B	C
1	2	3

Tables with pause

A	B	C
1	2	3
A	B	C

blocs

title of the bloc

bloc text

title of the bloc

bloc text

title of the bloc

bloc text