Expectation Propagation & Variational Message Passing
A comparison using Infer.NET

John Winn and Tom Minka
Machine Learning and Perception Group

7th Dec 2007
Overview

- Deterministic approximate inference
  - Divergence minimisation
  - Message passing

- Comparing EP and VMP
  - Beta-Bernoulli
  - Fixed-precision Gaussian
  - Learned precision Gaussian
  - Product of Gaussians

- Summary
Divergence minimisation

- Pick family of distributions for approximate posterior \( q \)
- Optimize \( q \) to minimize \( D(p \parallel q) \) for some divergence \( D \)

For example, alpha divergence:

\[
D_\alpha(p \parallel q) = \int x \alpha p(x) + (1 - \alpha)q(x) - p(x)^\alpha q(x)^{1-\alpha} dx \\
\alpha (1 - \alpha)
\]

\( \alpha = -\infty \)  \( \alpha = 0 \)  \( \alpha = 0.5 \)  \( \alpha = 1 \)  \( \alpha = \infty \)

VMP  EP, BP  TRW  Power EP

Minka [2005]
Divergence minimisation

Seeks: Largest mode

Zeroes: Exclusive (zero-forcing)

Mass: Under estimate

Upper bound

Inclusive

Correct

Over estimate

Minka [2005]
Message passing

\[ m_{A \rightarrow f} \rightarrow f \rightarrow m_{f \rightarrow B} \]

**EP**

\[
m_{f \rightarrow B} = \text{proj} \left[ m_{B \rightarrow f} \sum_A f(A, B) m_{A \rightarrow f} \right] / m_{B \rightarrow f}
\]

**BP**

\[
m_{f \rightarrow B} = \sum_A f(A, B) m_{A \rightarrow f}
\]

**VMP**

\[
\log m_{f \rightarrow B} = \sum_A \log f(A, B) m_{A \rightarrow f}
\]
Overview

- Deterministic approximate inference
  - Divergence minimisation
  - Message passing

- Comparing EP and VMP
  - Beta-Bernoulli
  - Fixed-precision Gaussian
  - Learned precision Gaussian
  - Product of Gaussians

- Summary
Beta-Bernoulli

\[ f(A, B) = \text{Bernoulli}(A \mid B) = B^A (1 - B)^{1-A} \]

**VMP**

Bernoulli\((p)\)  \quad \text{Beta}(1 + p, 1 + (1 - p))

Concentration of message \(m_{f \rightarrow B}\) is 3, for any input message \(m_{A \rightarrow f}\).

**EP**

Bernoulli\((0.5)\)  \quad \text{Beta}(1,1)

Bernoulli\((1)\)  \quad \text{Beta}(1,2)

Concentration of \(m_{f \rightarrow B}\) goes from 2 to 3 as certainty in \(m_{A \rightarrow f}\) increases.
Gaussian with fixed variance

\[ f(A, B) = \mathcal{N}(A | B, \sigma^2) \]

VMP

\[ \mathcal{N}(\mu, \tau^2) \quad \mathcal{N}(\mu, \sigma^2) \]

Variance of message \( m_{f \rightarrow B} \) is fixed, for any input message \( m_{A \rightarrow f} \)

EP

\[ \mathcal{N}(\mu, \tau^2) \quad \mathcal{N}(\mu, \sigma^2 + \tau^2) \]

Variance of message \( m_{f \rightarrow B} \) increases as the variance of \( m_{A \rightarrow f} \) increases
Surprising conclusion

VMP does not propagate message uncertainty!

EP does propagate message uncertainty.
Demo #1

Learning the mean of a fixed-precision Gaussian.
Learning mean and precision

\[ \text{Gaussian prior} \quad \mu \quad \text{mean} \quad \beta \quad \text{precision} \quad \Gamma \quad \text{prior} \]

\[ f \quad \text{Gaussian} \quad X_1 \quad \text{data points} \quad g \quad \text{Gaussian} \quad X_2 \]
Learning mean and precision

EP messages

prior
mean
prior
prior
prior

data points

$\mu$

$\beta$

$x_1$

$x_2$

fg
Learning mean and precision

\[ N(x_1, \tau^2) \]

Data points

\( f \)
\( g \)

Mean

Precision
Learning mean and precision

$\sim \mathcal{N}(x_1, \tau^2)$

$\mu$ prior

$\beta$ prior

$\tau$ too small relative to $|x_1 - x_2|$.

$x_1$ data points

$x_2$ data points
So...

EP can have *improper* messages.

VMP messages are always *proper*.
Demo #2

Learning the mean and precision of a Gaussian.
Improper messages are necessary

$$\mathcal{N}(x_1, \tau^2)$$

mean

$\mu$

$\beta$

precision

prior

g

data point

$N(2, \tau^2)$

negative variance if $\tau$ too small relative to $|x_1 - x_2|$.

EP is exact
**Improper message ‘fixes’**

- Recipient treats as uniform
  - never converges
- Recipient treats as point mass
  - leads to further improper messages
- Recipient re-sends last message
  - never converges
- Recipient sends less confident version of last message (?)
- Other suggestions? VMP?
Product of Gaussians

\[ f(A, B, X) = \mathcal{N}(X \mid AB, 1) \]

\[ \mathcal{N}(0, 1000) \quad \mathcal{N}(5, 1) \]

A \quad B

X

f

\[ \mathcal{N}(5, 1) \]

10
Product of Gaussians

\[ f(A, B, X) = \mathcal{N}(X \mid AB, 1) \]

\[ \mathcal{N}(0,1000) \]

\[ \mathcal{N}(5,1) \]

\[ \mathcal{N} \left( \frac{50}{26}, \frac{1}{26} \right) \]

\[ \mathcal{N} \left( \frac{\langle X \rangle \langle B \rangle}{\langle B^2 \rangle}, \frac{1}{\langle B^2 \rangle} \right) \]
Product of Gaussians

\[ f(A, B, X) = \mathcal{N}(X \mid AB, 1) \]

EP: likelihood is Cauchy-like distribution. So does posterior have valid moments?
Exact posterior for A

$P(A)$
Posterior for A vs. variance of B

Mean of P(A)

Variance of P(A)

Variance of P(B)
Posterior for positive $A$ vs. variance of $B$

- **Mean of $P(A)$**
- **Variance of $P(A)$**
- **Prior on $A$ times $1/A$**

![Graph showing the relationship between the variance of $P(B)$ and the mean and variance of $P(A)$, with a line labeled "exact".]
Demo #3

Product of Gaussians.
Summary

- Variational Message Passing
  - Estimates uncertainty locally, rather than propagating it => can be overconfident.
  - Stable (no improper messages)

- Expectation Propagation
  - Propagates uncertainty
  - Can be unstable

- Product factor
  - Remove unnecessary bimodality e.g constrain one argument to be positive (or split out the positive A and negative A cases + tie weights)
Thanks!

*Infer.NET* will be publicly available in Spring 2008.