Expectation Propagation for the Clutter Problem

Daniel Korzekwa

11th March 2013

Problem definition

Clutter Problem is a Gaussian density estimation task, presented by Thomas Minka [1] to illustrate Expectation Propagation [1] algorithm in practice. Imagine that we observe data points $\{\theta_1, ..., \theta_n\}$, generated from a Gaussian distribution. With a probability $1 - w$, we get noisy observations represented by linear Gaussian model $\mathcal{N}(x|\theta, 1)$, while with a probability w, we receive some clutter characterised by Gaussian distribution $\mathcal{N}(x|0, a)$. Additionally, we express prior belief $\mathcal{N}(\theta|m, v)$ around the true distribution, from which data points are generated.

Let's setup a probabilistic graphical model, with the prior and likelihood variables defined as X and $\{Y_1, ..., Y_n\}$ respectively, and the following conditional probability distributions.

$$p(X) \sim \mathcal{N}(\theta|m, v)$$

$$p(Y|X) \sim (1 - w)\mathcal{N}(x|\theta, 1) + w\mathcal{N}(0, a)$$

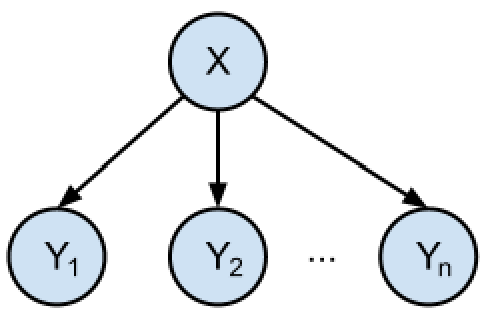

Figure 1: Probabilistic graphical model for the Clutter Problem. X - prior belief of the true distribution $\sim \mathcal{N}(\theta|m, v)$, $\{Y_1, ..., Y_n\}$ - observed data points $\sim (1 - w)\mathcal{N}(x|\theta, 1) + w\mathcal{N}(0, a)$

The posterior probability over variable X given observations $\{Y_1, ..., Y_n\}$, is defined as

$$p(X|Y_1, ..., Y_n) \propto p(X) \prod_{i=1}^{n} p(Y_i|X)$$

Posterior inference with Expectation Propagation

In this section, we infer posterior distribution $p(X|Y)$ using Expectation Propagation algorithm, given the following setting

- Observed data points $\{3, 5\}$

- Prior distribution $p(X) = \mathcal{N}(\theta | m = 15, v = 100)$
- Likelihood parameters, $w = 0.4$, and $a = 10$

First, draw a factor graph [2]

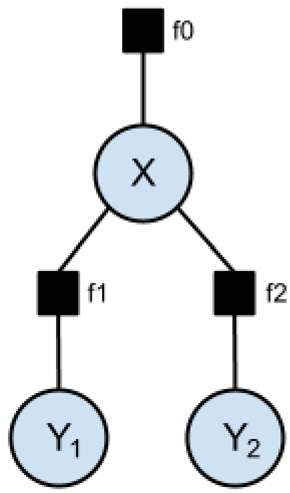

Figure 2: Factor graph for the Clutter Problem with observation points $\{3, 5\}$

Factors:

- $f_0 \sim \mathcal{N}(\theta | m = 15, v = 100)$
- $f_1, f_2 \sim (1 - w)\mathcal{N}(x|\theta, 1) + w\mathcal{N}(0, a)$

Next, create a messaging passing schedule and execute it for a given number of iterations.

$$m_{f_0 \to X} = (f_o m_{f_1} m_{f_2})/(m_{f_1} m_{f_2}) = f_0$$

$$m_{f_1 \to X} = proj(f_1 m_{f_0} m_{f_2})/(m_{f_0} m_{f_2})$$

$$m_{f_2 \to X} = proj(f_2 m_{f_0} m_{f_1})/(m_{f_0} m_{f_1})$$

The $proj(q)$ operator [1, 3] approximates distribution q with a Gaussian distribution by matching the mean and the variance moments. To compute posterior distribution, multiply all incoming messages for a variable X.

$$p(X|Y) = m_{f_0} m_{f_1} m_{f_2}$$

The following chart shows the mean for the posterior distribution $p(X|Y)$ as a function of current iteration. It takes about 6 iterations of the message passing routine, till the posterior mean for $p(X|Y)$ gets very close to the stationary point of ~ 4.34.

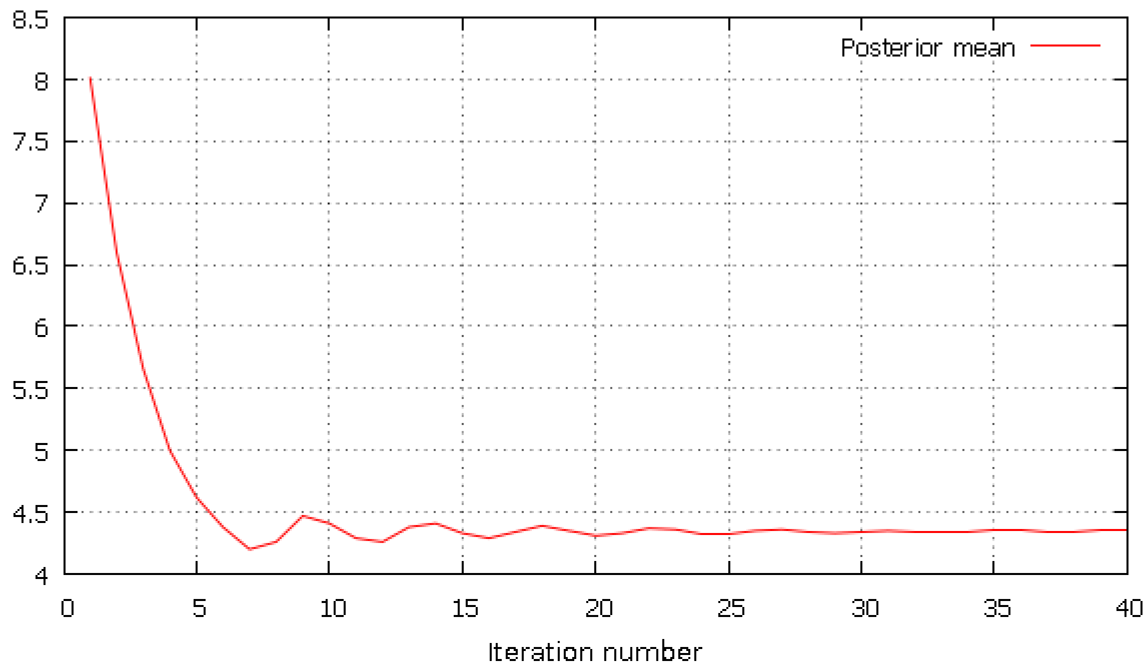

Figure 3: The mean for the posterior distribution $p(X|Y)$ as a function of current iteration

Appendix A

Scala implementation for the Clutter Problem - Bayes-Scala toolbox

References

[1] Thomas P Minka. A family of algorithms for approximate Bayesian inference, 2001

[2] Christopher M. Bishop. Pattern Recognition and Machine Learning (Information Science and Statistics), 2009

[3] Daniel Korzekwa. Gaussian approximation with moment matching, aka proj() operator in Expectation Propagation, 2013