Energy function of Expectation Propagation

Energy function of Expectation Propagation

This note contains some comments about the energy function proposed by Minka for expectation propagation (EP). For simplicity we consider discrete variables $\mathbf{x}$. The goal is to obtain an approximation to the posterior:

$$p\left(\mathbf{x}|D\right) = p(\mathbf{x}) \prod_i t_i(\mathbf{x}) \approx p(\mathbf{x}) \prod_i \tilde{t}_i(\mathbf{x})$$

using functions $\tilde{t}_i(\mathbf{x})$ from an exponential family

$$\tilde{t}_i(\mathbf{x}) = \exp\left(\sum_j \tau_{ij} f_j(\mathbf{x})\right)$$

for some parameters τ_{ij}.

The primal energy function

Minka defines a *primal energy function*,

$$\begin{aligned} \mathcal{J}(\hat{p}, q) &= \sum_i \sum_{\mathbf{x}} \hat{p}_i(\mathbf{x}) \ln \frac{\hat{p}_i(\mathbf{x})}{t_i(\mathbf{x}) p(\mathbf{x})} - (n-1) \sum_{\mathbf{x}} q(\mathbf{x}) \ln \frac{q(\mathbf{x})}{p(\mathbf{x})} \\ &= \sum_i \mathrm{KL}(\hat{p}_i(\mathbf{x})||t_i(\mathbf{x})p(\mathbf{x})) - (n-1)\mathrm{KL}(q(\mathbf{x})||p(\mathbf{x})) \end{aligned}$$

By the convexity of the Kullback-Leibler (KL) divergence, $\mathcal{J}(\hat{p}, q)$ is convex in $\hat{p}$ and concave in q.

Minka proposes the following minimax optimization to obtain $\hat{p}, q$:

$$\min_{\hat{p}_i} \max_q \mathcal{J}(\hat{p}, q)$$

subject to the following constrains on $\hat{p}_i, q$:

$$p_i(\mathbf{x}) \geq 0, \qquad q(\mathbf{x}) \geq 0$$

$$\sum_{\mathbf{x}} \hat{p}_i(\mathbf{x}) = 1, \qquad \sum_{\mathbf{x}} q\left(\mathbf{x}\right) = 1$$

$$\sum_{\mathbf{x}} f_j(\mathbf{x}) \hat{p}_i(\mathbf{x}) = \sum_{\mathbf{x}} f_j(\mathbf{x}) q(\mathbf{x})$$

We assume that the optimization reaches a point where $p_i(\mathbf{x}) > 0$ and $q(\mathbf{x}) > 0$. Since $\mathcal{J}(\hat{p}, q)$ is convex in $\hat{p}$ and concave in q and the constrains are linear, it follows that $\min_{\hat{p}_i} \max_q$ can be exchanged with $\max_q \min_{\hat{p}_i}$ in this objective.

KL duality

Minka notes the following variational formula for the KL divergence:

$$\mathrm{KL}(p||q) = \max_{\nu} \left\{ \sum_{\mathbf{x}} p(\mathbf{x}) \nu(\mathbf{x}) - \ln \sum_{\mathbf{x}} q(\mathbf{x}) \mathrm{e}^{\nu(\mathbf{x})} \right\}$$

This can be shown directly by differentiation w.r.t. $\nu(\mathbf{x})$. Alternatively, we observe that this equation is equivalent to

$$\mathrm{KL}(p||q) = \max_{\nu} \left\{ \sum_{\mathbf{x}} p(\mathbf{x}) \nu(\mathbf{x}) - \mathrm{KL}^*(\nu||q) \right\}$$

where

$$\mathrm{KL}^*(\xi||v) = 1 + \ln \sum_i v_i \mathrm{e}^{\xi_i - 1}$$

is the Legendre transform of $\mathrm{KL}(p||q)$ w.r.t. p. Therefore it is a direct consequence of the duality of Legendre transforms of convex differentiable functions.

References

1. Thomas P. Minka, "The EP Energy Function and Minimization Schemes" (Technical report, 2001) (link).

2. Minka, Thomas P. "Expectation Propagation for Approximate Bayesian Inference." In Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, 362–369. UAI'01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001. (link).