Technische Universität Berlin
Fakultät IV – Elektrotechnik und Informatik

Probabilistic and Bayesian Modelling in Machine Learning and Artificial Intelligence

Manfred Opper and Théo Galy–Fajou

Summer Term 2018

Problem Sheet 5

Solutions to be discussed in the tutorial on Tuesday, July 3.

Problem 1 – Variational inference

Assume we have n observations $D = (x_1, \ldots, x_n)$ generated independently from a Gaussian density $\mathcal{N}(x|\mu, 1/\tau)$, i.e.

$$p(D|\mu,\tau) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left[-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

We also assume prior densities $p(\mu|\tau) = \mathcal{N}(\mu|\mu_0, (\lambda_0\tau)^{-1})$ and $p(\tau) = \mathrm{Gamma}(\tau|a_0, b_0)$. λ_0 and μ_0 as well as a_0, b_0 are given hyper parameters.

Our goal is to approximate the posterior density $p(\mu,\tau|D)$ by a **factorising density** $q(\mu,\tau) = q_1(\mu)q_2(\tau)$ which minimises the variational free energy

$$F[q] = \int q(\mu,\tau)\ln\frac{q(\mu,\tau)}{p(\mu,\tau,D)}\,d\mu\,d\tau$$

(a) Show that the optimal $q_1(\mu)$ is a **Gaussian density** and give expressions for the mean and variance in terms of expectations with respect to q_2.

(b) Show that the optimal $q_2(\tau)$ is a **Gamma density** and give expressions for the parameters in terms of expectations with respect to q_1.

You can use the following results which follow from the derivations given in the lecture

$$q_1(\mu) \propto \exp\left[E_\tau \ln p(\mu,\tau,D)\right]$$
$$q_2(\tau) \propto \exp\left[E_\mu \ln p(\mu,\tau,D)\right]$$