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

Problem 1 – Variational inference

Assume we have n observations $D = (x_1, \dots, x_n)$ 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

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

We have the representation of the joint density

$$p(\mu,\tau,D) = p(D|\mu,\tau)p(\mu|\tau)p(\tau)$$

with

$$
\begin{aligned}
p(\mu|\tau) =& \frac{(\lambda_0\tau)^{1/2}}{2\pi} \exp\left(-\frac{(\mu-\mu_0)^2\lambda_0\tau}{2}\right) \\
p(\tau) \propto& \tau^{a_0-1}e^{-b_o\tau}
\end{aligned}
$$

(a) Hence

$$
\begin{aligned}
E_\tau\left[\ln p(\mu,\tau,D)\right] &= -\frac{E_\tau[\tau]}{2}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{\lambda_0 E_\tau[\tau]}{2}(\mu-\mu_0)^2 + \text{const} = \\
&-\frac{1}{2}\left(E_\tau[\tau](n+\lambda_0)\right)\mu^2 + \mu E_\tau[\tau]\left(\sum_i x_i + \lambda_0\mu_0\right) + \text{const}
\end{aligned}
$$

Note, that the second constant differs from the first. We get a Gaussian density for $q_1(\mu)$ with

$$
\begin{aligned}
E[\mu] &= \frac{\sum_i x_i + \lambda_0\mu_0}{n+\lambda_0} \\
\text{VAR}[\mu] &= \frac{1}{E_\tau[\tau](n+\lambda_0)}
\end{aligned}
$$

(b) for the density of $q_2(\tau)$, we use

$$
E_\tau\left[\ln p(\mu,\tau,D)\right] = \ln\left(\tau^{a_0+(n+1)/2-1}e^{-b_0\tau}\right) - \frac{\tau}{2}\sum_{i=1}^{n}E_\mu[(x_i-\mu)^2] - \frac{\lambda_0\tau}{2}E_\mu[(\mu-\mu_0)^2] + \text{const}
$$

We get a Gamma density

$$
q_2(\tau) \propto \tau^{a_n-1}e^{-b_n\tau}
$$

with parameters

$$
\begin{aligned}
a_n =& a_0 + (n+1)/2 \\
b_n =& b_0 + \frac{1}{2}\sum_{i=1}^{n}E_\mu[(x_i-\mu)^2] + \frac{\lambda_0}{2}E_\mu[(\mu-\mu_0)^2]
\end{aligned}
$$

Knowing the form of both variational distributions we can also compute closed form solution of the expectations :

$$
\begin{aligned}
E_\tau\left[\tau\right] =& \frac{a_n}{b_n} \\
E_\mu\left[(x-\mu)^2\right] =& \text{Var}_\mu\left[\mu\right] + \left(x - E_\mu\left[\mu\right]\right)^2
\end{aligned}
$$

And proceed to coordinate ascent updates to converge to the optimal distribution.