Technische Universität Berlin
Fakultät IV – Elektrotechnik und Informatik

Probabilistic and Bayesian Modelling in Machine Learning and Artificial Intelligence

Manfred Opper, Théo Galy–Fajou

Summer Term 2018

Problem Sheet 3

Solutions to be discussed in the tutorial on Tuesday, June 12, 2018

Problem 1 – Bayes inference for the variance of a Gaussian

Use a Bayesian approach to estimate the inverse variance λ of a univariate Gaussian distribution

$$p(x|\lambda) = \sqrt{\frac{\lambda}{2\pi}} \exp\left[-\frac{\lambda x^2}{2}\right].$$

Here we have assumed for simplicity that the data has zero mean $\mu = 0$. To apply Bayesian inference we specify a *Gamma* prior distribution for λ,

$$p(\lambda) = \frac{\lambda^{\alpha-1} \exp\left[-\lambda/\beta\right]}{\Gamma(\alpha)\beta^\alpha}$$

where the positive numbers α and β, the *hyperparameters* of the model are assumed to be known and $\Gamma(\alpha)$ is Euler's *gamma* function (`gamma` in Octave and R). We then observe a dataset $D = (x_1, x_2, \ldots, x_N)$ comprising N independent random samples from $p(x|\lambda)$.

(a) Show that the posterior probability $p(\lambda|D)$ of the inverse variance is also a *gamma* distribution with parameters

$$\alpha_p = \alpha + \frac{N}{2}, \qquad \frac{1}{\beta_p} = \frac{1}{\beta} + \frac{1}{2}\sum_{i=1}^{N} x_i^2.$$

(b) Compute the mean of the posterior distribution of λ. Compare the result with the result from the *maximum-likelihood* estimation, $\lambda_{\mathrm{ML}} = 1/\sigma^2_{\mathrm{ML}}$ and explain what happens if $N \to \infty$.

(c) Show that the variance of the posterior distribution $\mathrm{Var}(\lambda_{\mathrm{post}}) = \langle\lambda^2\rangle - \langle\lambda\rangle^2$ shrinks to zero as $N \to \infty$. Here we have used the notation $\langle \ldots \rangle$ for posterior expectations.

(d) Show that the predictive distribution is

$$p(x|D) = \frac{1}{\sqrt{2\pi}} \frac{\Gamma(\alpha_p + 1/2)}{\Gamma(\alpha_p)} \sqrt{\beta_p} \left(1 + \frac{x^2 \beta_p}{2}\right)^{-\alpha_p - 1/2}$$

where α_p and β_p were defined above. Note, this is **not a Gaussian**!

Problem 2 – Hyperparameter estimation for a generalised linear model

Consider a model for a set of data $D = (y_1, \ldots, y_n)$ defined by

$$p(D|\mathbf{w}, \beta) = \left(\frac{\beta}{2\pi}\right)^{N/2} \exp\left[-\sum_{i=1}^{N} \frac{\beta}{2} \left(y_i - \sum_{j=1}^{K} w_j \Phi_j(x_i)\right)^2\right]$$

with a fixed set $\{\Phi_1(x), \ldots, \Phi_k(x)\}$ of K basis functions. The prior distribution on the weights is given by

$$p(\mathbf{w}|\alpha) = \left(\frac{\alpha}{2\pi}\right)^{K/2} \exp\left[-\frac{\alpha}{2} \sum_{j=1}^{K} w_j^2\right].$$

This *generalised linear model* assumes that the observations are generated from a weighted linear combination of the basis functions with additive Gaussian noise.

(a) The posterior distribution $p(\mathbf{w}|D, \alpha, \beta)$ of the vector of weights is a Gaussian. Compute the posterior mean vector $\mathrm{E}[\mathbf{w}]$ and the posterior covariance in terms of the matrix $\mathbf{X}$ where $X_{lk} = \Phi_k(x_l)$ (this is just a repetition of the calculations done in the lecture).

(b) Derive an EM algorithm for optimising the hyperparameters α and β by maximising the log-evidence

$$p(D|\alpha, \beta) = \int p(D|\mathbf{w}, \beta) p(\mathbf{w}|\alpha)\, d\mathbf{w}$$

Hint: Treat the weights $\mathbf{w}$ as a set of latent variables. Express your result in terms of the posterior mean and variance.