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 4

Solutions to be discussed in the tutorial on Tuesday, June 26.

Problem 1 – Evidence for Gaussian process (GP) regression

For the GP regression problem, we assume that data are generated as

$$y_i = f(x_i) + \nu_i \qquad i = 1, \ldots, n \tag{1}$$

where the ν_i are independent, zero mean Gaussian noise variables within $E[\nu_i^2] = \sigma^2$ and $f(\cdot)$ has a GP prior with kernel $K(x, x')$. Show that the **Bayesian evidence** is given by

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{n/2} |\det(\mathbf{K} + \sigma^2 \mathbf{I})|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\mathbf{y}^T (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{y}\right] \tag{2}$$

where $\mathbf{y} = (y_1, \ldots, y_n)$ and the kernel matrix is defined by $\mathbf{K}_{ij} = K(x_i, x_j)$.

Hint: Calculate the joint density of $\mathbf{y}$ and use the fact that $f(x_j)$ and ν_i are independent Gaussian random variables. Hence you can add the respective covariance matrices.

Problem 2 – Gibbs sampler for outlier detection

The file `outlier.dat` on the web page of the course contains a data set $D = (y_1, \ldots, y_N)$. Most of the observations have been drawn from a Gaussian probability distribution $\mathcal{N}(y_i; \mu, \sigma^2)$ with mean μ and variance σ^2. However, D contains some *outliers*, which occur with probability ϵ and are displaced by a random offset A_i. For the purpose of *outlier detection* the model is augmented with an indicator variable

$$\delta_i = \begin{cases} 1 & \text{if } y_i \text{ is an outlier,} \\ 0 & \text{if } y_i \text{ is a normal data point,} \end{cases}$$

for each observation. Assuming conjugate priors for the parameters yields the full stochastic model

$$\begin{array}{lllllllll} \mu & \sim & \mathcal{N}(\theta, v^2), & \sigma^{-2} & \sim & \mathrm{Gamma}(\kappa, \lambda), & \epsilon & \sim & \mathrm{Beta}(\alpha, \beta), \\ y_i & \sim & \mathcal{N}(\mu + \delta_i\, A_i, \sigma^2), & \delta_i & \sim & \mathrm{Bernoulli}(\epsilon), & A_i & \sim & \mathcal{N}(0, \tau^2). \end{array}$$

We want to use a Gibbs sampler in order to draw samples from the posterior $p(\mu, \sigma^2, \epsilon, \boldsymbol{\delta}, \mathbf{A}|D)$ with $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_N)$ and $\mathbf{A} = (A_1, \ldots, A_N)$. Some conditional posteriors are given by

$$\begin{aligned} \mu &\sim \mathcal{N}\left(\frac{\sigma^2\theta + v^2 \sum_{i=1}^N (y_i - \delta_i\, A_i)}{\sigma^2 + N v^2}, \frac{\sigma^2 v^2}{\sigma^2 + N v^2}\right), \\ \sigma^{-2} &\sim \mathrm{Gamma}\left(\kappa + \frac{N}{2}, \frac{2\lambda}{2 + \lambda \sum_{i=1}^N (y_i - \delta_i\, A_i - \mu)^2}\right). \end{aligned}$$

(a) Show that the remaining conditional posteriors are given by

$$\begin{aligned} \delta_i &\sim \mathrm{Bernoulli}\left(\frac{\epsilon}{\epsilon + (1-\epsilon)\exp(-A_i(y_i - A_i - \mu)/(2\sigma^2))}\right), \\ A_i &\sim \mathcal{N}\left(\frac{\tau^2 \delta_i (y_i - \mu)}{\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2 \delta_i}\right), \\ \epsilon &\sim \mathrm{Beta}\left(\alpha + \sum_{i=1}^N \delta_i, \beta + \sum_{i=1}^N (1 - \delta_i)\right). \end{aligned}$$

(b) Write a program that implements the *Gibbs sampler*. Generate 10^3 samples from the posterior using the hyperparameters $\theta = 0$, $v^2 = 100$, $\kappa = 2$, $\lambda = 2$, $\alpha = 2$, $\beta = 20$, $\tau^2 = 100$. Plot histograms showing the marginal posteriors $p(\mu|D)$ and $p(\epsilon|D)$.

(c) Which data points in the file `outlier.dat` are outliers? Use the samples generated in part (b) and the condition $p(\delta_i|D) \geq 0.02$ in order to identify them.

Probability distributions

- Beta distribution for $0 \leq z \leq 1$

$$z \sim \mathrm{Beta}(a, b) \quad \Longleftrightarrow \quad \mathrm{Beta}(z; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} z^{a-1}(1-z)^{b-1}$$

  Octave functions: `betapdf`, `betarnd`
  R functions: `dbeta`, `rbeta`

- Bernoulli distribution for $z \in \{0, 1\}$

$$z \sim \text{Bernoulli}(a) \quad \Longleftrightarrow \quad \text{Bernoulli}(z; a) = a^z (1-a)^{1-z}$$

Octave functions: `binopdf`, `binornd`
R functions: `dbinom`, `rbinom`

- Gamma distribution for $z \geq 0$

$$z \sim \text{Gamma}(a, b) \quad \Longleftrightarrow \quad \text{Gamma}(z; a, b) = \frac{1}{\Gamma(a) b^a} z^{a-1} e^{-z/b}$$

Octave functions: `gampdf`, `gamrnd`
R functions: `dgamma`, `rgamma`

- Normal (Gaussian) distribution

$$z \sim \mathcal{N}(a, b) \quad \Longleftrightarrow \quad \mathcal{N}(z; a, b) = \frac{1}{\sqrt{2\pi b}} \exp\left(-\frac{(z-a)^2}{2b}\right)$$

Octave functions: `normpdf`, `normrnd`
R functions: `dnorm`, `rnorm`