Exercises for the course

Machine Learning 1

Winter semester 2016/17

Abteilung Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

Exercise Sheet 3

Exercise 1: Maximum Likelihood vs. Bayes (10+15+15 P)

An unfair coin is tossed seven times and the event (head or tail) is recorded at each iteration. The observed sequence of events is

$$\mathcal{D} = (x_1, x_2, \dots, x_7) = (\text{head}, \text{head}, \text{tail}, \text{tail}, \text{head}, \text{head}, \text{head}).$$

We assume that all tosses $x_1, x_2, \dots$ have been generated independently following the Bernoulli probability distribution

$$P(x \mid \theta) = \begin{cases} \theta & \text{if} \quad x = \text{head} \\ 1 - \theta & \text{if} \quad x = \text{tail}, \end{cases}$$

where $\theta \in [0, 1]$ is an unknown parameter.

(a) *State* the likelihood function $P(\mathcal{D}|\theta)$, that depends on the parameter θ.

(b) *Compute* the maximum likelihood solution $\hat{\theta}$, and *evaluate* for this parameter the probability that the next two tosses are "head", that is, evaluate

$$P(x_8 = \text{head}\,,\ x_9 = \text{head} \mid \hat{\theta}).$$

(c) We now adopt a Bayesian view on this problem, where we assume a prior distribution for the parameter θ defined as:

$$p(\theta) = \begin{cases} 1 & \text{if} \quad 0 \leq \theta \leq 1 \\ 0 & \text{else.} \end{cases}$$

Compute the posterior distribution $p(\theta|\mathcal{D})$, and *evaluate* the probability that the next two tosses are head, that is,

$$\int P(x_8 = \text{head}\,,\ x_9 = \text{head} \mid \theta)\, p(\theta|\mathcal{D})\, d\theta.$$

Exercise 2: Convergence of Bayes Parameter Estimation (15+15 P)

We consider Section 3.4.1 of Duda et al., where the data is generated according to the univariate probability density

$$p(x|\mu) \sim \mathcal{N}(\mu, \sigma^2),$$

where σ^2 is known and where μ is unknown with prior distribution

$$p(\mu) \sim \mathcal{N}(\mu_0, \sigma_0^2).$$

Having sampled a dataset $\mathcal{D}$ from the data-generating distribution, the posterior probability distribution over the unknown parameter μ becomes

$$p(\mu|\mathcal{D}) \sim \mathcal{N}(\mu_n, \sigma_n^2),$$

where

$$\frac{1}{\sigma_n^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \qquad \frac{\mu_n}{\sigma_n^2} = \frac{n}{\sigma^2}\hat{\mu}_n + \frac{\mu_0}{\sigma_0^2} \qquad \hat{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} x_k.$$

(a) *Show* that the variance of the posterior can be upper-bounded as follows:

$$\sigma_n^2 \leq \min\Big(\frac{\sigma^2}{n}\,,\ \sigma_0^2\Big),$$

that is, the variance of the posterior is contained both by the uncertainty of the data mean and of the prior.

(b) *Show* that the mean of the posterior can be lower- and upper-bounded as follows:

$$\min(\hat{\mu}_n, \mu_0) \leq \mu_n \leq \max(\hat{\mu}_n, \mu_0),$$

that is, the mean of the posterior distribution lies somewhere on the segment between the mean of the prior distribution and the sample mean.

Exercise 3: Programming (30 P)

Download the programming files on ISIS and follow the instructions.