Inleiding Machine Learning

Exercises, Bishop Ch. 2

Exercise 1

(Bishop 2.1)

Exercise 2

(Bishop 2.5)

Exercise 3

Bishop 2.6 (show only the mean $\mathbb{E}[\mu] = a/(a+b)$.

Exercise 4

Suppose we have two coins, A and B, and we do not know whether these coins are fair.

1. Let μ be the probability the coin comes up H(eads). Give an expression for the likelihood of a data set $\mathcal{D}$ of N observations of independent tosses of the coin.

Suppose we have observed the following results of two series of coin tosses:

coin:	data $\mathcal{D}$:
A	H,T,T,H,T,T,T
B	H

2. What is the maximum likelihood estimate for μ_A, the probability that a toss with coin A results in H(eads)? And for μ_B? Based on these maximum likelihood estimates, what is the probability that the next toss of coin A will result in H(eads)? And the next toss with coin B? Do these results make sense?

3. Let us now take a Bayesian approach. Find an expression for $p(\mu|\mathcal{D})$ using Bayes' rule and show that a prior proportional to powers of μ and $(1-\mu)$ will lead to a posterior that is also proportional to powers of μ and $(1-\mu)$. Are you free to choose whatever prior you like?

Such a prior exists and is called the Beta distribution with hyperparameters a and b:

$$\mathrm{Beta}(\mu|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1} \qquad 0 \leq \mu \leq 1 \tag{1}$$

in which $\Gamma(x)$ is the gamma function with property $\Gamma(x+1) = x\Gamma(x)$.

4. Give combinations (a, b) for a prior that expresses: a) total ignorance, b) high confidence in a reasonably fair coin. For each prior and each coin, calculate the posterior probability density of μ given the observed coin tosses $\mathcal{D}$ and plot the results (for example by using the `betapdf` command in MatLab). Do these results make more sense than the ML estimates?

Exercise 5

(Bishop 2.14) Niet verplicht

Exercise 6

Bishop 2.29

Exercise 7

Bishop 2.38

Exercise 8

We consider two distributions, with one defined conditional on the other, as

$$\begin{aligned} p(u) &= \mathcal{N}(u|\mu_0, \sigma^2) & (2) \\ p(v|u) &= \mathcal{N}(v|c \cdot u, s^2) & (3) \end{aligned}$$

where μ_0, σ^2, c and s^2 are constant model parameters.

1. The conditional distribution $p(u|v)$ is also a Gaussian. Which equations from Bishop are relevant for computing this function?
2. Write down an expression for the distribution $p(u|v)$ and show that the mean $\mu_{u|v}$ and variance $\sigma^2_{u|v}$ of this distribution are given by

$$\mu_{u|v} = \frac{\frac{\mu_0}{\sigma^2} + \frac{cv}{s^2}}{\frac{1}{\sigma^2} + \frac{c^2}{s^2}} \tag{4}$$

$$\frac{1}{\sigma^2_{u|v}} = \frac{1}{\sigma^2} + \frac{c^2}{s^2} \tag{5}$$

3. Compute $p(v)$.
4. Compute $p(u, v)$. Hint: using the right equations, the calculation does not get very messy.