Statistical Machine Learning 2016

Exercises, week 6

6 October 2016

Exercise 1

Find the eigenvalues and a set of mutually orthogonal eigenvectors of the symmetric matrix:

$$\begin{bmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{bmatrix}$$

Exercise 2

Consider a discrete variable x that can take K values, $x \in \{1, \ldots, K\}$. If we denote the probability of $x = k$ by the parameter θ_k, then the distribution of x is given by

$$P(x = k|\boldsymbol{\theta}) = \theta_k \tag{1}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^T$ and the parameters are constrained to satisfy

$$\theta_k \geq 0 \qquad \text{and} \qquad \sum_{k=1}^{K} \theta_k = 1 \tag{2}$$

1. Explain why the parameters should satisfy these constraints.

Now consider a dataset χ of N independent observations, $\chi = \{x_1, \ldots, x_N\}$.

2. Show that the log-likelihood $\ln P(\chi|\boldsymbol{\theta})$ is of the form

$$\ln P(\chi|\boldsymbol{\theta}) = \sum_{k=1}^{K} m_k \ln \theta_k \tag{3}$$

   What are the m_k's (in terms of the x_i's, k's etc.)?

3. Show that the maximum likelihood solution $\boldsymbol{\theta}^*$ is given by

$$\theta_k^* = \frac{m_k}{N} \tag{4}$$

   Hint: Use a Lagrange multiplier for the constraint $\sum_{k=1}^{K} \theta_k - 1 = 0$.

Exercise 3

The beta distribution is

$$\mathrm{Beta}(\mu|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1} \qquad 0 \leq \mu \leq 1 \tag{5}$$

in which $\Gamma(x)$ is the gamma function (a well defined mathematical function, see book, www exercise 1.17). The gamma function is a generalization of the factorial function $(n-1)!$ as it satisfies

$$\Gamma(x+1) = x\Gamma(x) \tag{6}$$

We are looking for an expression for the expectation value in terms of a and b

$$\langle\mu\rangle = \int_0^1 \mu \; \mathrm{Beta}(\mu|a,b) d\mu \tag{7}$$

Since the beta distribution is normalised, we can start from the relation

$$\int_0^1 \mu^{a-1}(1-\mu)^{b-1} d\mu = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{8}$$

1. Show that the expectation value is given by

$$\langle\mu\rangle = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \tag{9}$$

   Hint: you do not actually have to compute any integrals.

2. Use this result and the property $\Gamma(x+1) = x\Gamma(x)$ to show that

$$\langle\mu\rangle = \frac{a}{a+b} \tag{10}$$

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:

$$\mathrm{Beta}(\mu|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1} \qquad 0 \leq \mu \leq 1 \tag{11}$$

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

Consider a mixture of K Gaussian densities of the form

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{12}$$

Show that if the mixing coefficients satisfy

$$\pi_k \geq 0 \quad \text{and} \quad \sum_k \pi_k = 1$$

then the mixture of Gaussians in (12) is positive and normalized. (You may assume that the components of the mixture are normalized)