N. Bertschinger
M. Kaschube
V. Ramesh

Machine Learning I
Exercise Sheet 3
Due on Mon, November 21, 12:15

If you turn in your solutions via email, please send to:
kamyshanska@fias.uni-frankfurt.de
and include [ML1-1617] in the subject line.

Exercise 1. *Consider a multivariate normal distribution in variable* $\mathbf{x}$ *with mean* μ *and covariance* $\mathbf{\Sigma}$*. Show that if we make the linear transformation* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ *then the transformed variable* y *is distributed as:*

$$p(y) = \mathcal{N}(\mathbf{A}\mu + \mathbf{b}, \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T) \ .$$

3 points

Exercise 2. *Show that we can convert a normal distribution with mean* μ *and covariance* $\mathbf{\Sigma}$ *to a new distribution with mean* $\mathbf{0}$ *and covariance* $\mathbf{I}$ *using the linear transformation* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ *where*

$$\mathbf{A} = \mathbf{\Sigma}^{-1/2}$$

$$\mathbf{b} = -\mathbf{\Sigma}^{-1/2}\mu \, .$$

This is known as the whitening transform. Note that M *is the square root of a matrix* Q*, i.e.* $M = Q^{1/2}$*, if the matrix product* $MM = Q$*.*

2 points

Exercise 3. *Now, we want to exploit the knowledge acquired in Exercises 1 and 2 to conceive a method for sampling random vectors from an arbitrary multivariate normal distribution* $\mathcal{N}(\mu, \mathbf{\Sigma})$*. Sampling from a distribution means, the probability of drawing a given vector is proportional to the pdf of this distribution. First, we sample a vector of length* N *(number of dimensions) from a normal distribution* $\mathcal{N}(\mathbf{0}, \mathbf{I})$*. Since all components are uncorrelated, we can easily do this by drawing* N *random numbers using a built-in random number generator for univariate normal distributions. Then we transform the samples with appropriate matrix* $\mathbf{B}$*, such that the new distribution has the desired covariance. Finally, we shift the samples to the desired mean value* ν*.*

a) How are $\mathbf{B}$*,* $\mathbf{\Sigma}$ *and* ν *related?*

b) Sample 1,000 values from the two-dimensional normal distribution with mean vector $\mu = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ *and the covariance matrix* $\mathbf{\Sigma} = \begin{pmatrix} 4 & -0.5 \\ -0.5 & 2 \end{pmatrix}$*.*

Plot a two-dimensional histogram and explain why this plot shows approximately the desired distribution. Use $M \times M$ quadratic bins of equal size. The value of a given bin is defined as the number of samples that fall into this bin, divided by the total number of samples. To plot the histogram, represent bin values by grey values (or color) and plot these grey values into a two-dimensional coordinate system (like an 'image').

c) *Estimate the mean and covariance matrix from the data using their Maximum Likelihood estimates. How close are the estimated parameters to the real ones? Use 2, 20, 200 data points for your estimate.*

d) *Plot the likelihood as a function of two parameters of your choice, keeping all other parameters constant.*

5 points