N. Bertschinger
M. Kaschube
V. Ramesh

Machine Learning I
Exercise Sheet 2
Due on Wed, Nov 14, 12:15

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

Exercise 1. *Show that if two random variables X and Y are independent, then their covariance $\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$ is zero. You may assume that the random variables are discrete.*

2 points

Exercise 2. *Suppose you have observed N samples $x_1, \ldots, x_N$ drawn from a Gaussian distribution. Compute the maximum likelihood estimators for the mean and variance of the data, i.e.*

$$\max_{\mu,\sigma^2} \log \prod_{n=1}^{N} p(x_n; \mu, \sigma^2)$$

where $p(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$.

2 points

Exercise 3. *In this exercise we want to sample random vectors from a two-dimensional normal distribution $\mathcal{N}(\mu, \mathbf{\Sigma})$, with $\mu = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ and $\mathbf{\Sigma} = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$. Sampling from a distribution means the probability of drawing a given vector is proportional to the pdf of this distribution. Since the two components are uncorrelated, we can easily do this by drawing two random numbers using a built-in random number generator for univariate normal distributions. Then we transform the samples to match the desired mean and variances.*

a) Sample 1000 values from the two-dimensional normal distribution. 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').

b) Estimate the mean and variance along each dimension from the data using their Maximum Likelihood estimates (derived in Exercise 2 and provided in the lecture slides). How close are the estimated parameters to the real ones? Use 2, 20, 200 data points for your estimate. Compute also the covariance for these sets and compare with Exercise 1.

4 points

N. Bertschinger
M. Kaschube
V. Ramesh

Machine Learning I
Exercise Sheet 2
Due on Wed, Nov 14, 12:15

Exercise 4. *Show that an arbitrary square matrix with elements* w_{ij} *can be written in the form* $w_{ij} = w^S_{ij} + w^A_{ij}$ *where* w^S_{ij} *and* w^A_{ij} *are symmetric and anti-symmetric matrices, respectively, satisfying* $w^S_{ij} = w^S_{ji}$ *and* $w^A_{ij} = -w^A_{ji}$ *for all* i *and* j*.*

1 points

Exercise 5. *Show that a real, symmetric matrix of size* $D \times D$ *has* $D(D+1)/2$ *independent parameters.* **1 points**