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 2

Exercise 1: Maximum-Likelihood Estimation (7.5+7.5+7.5+7.5 P)

We consider the problem of estimating using the maximum-likelihood approach the parameters $\lambda, \eta > 0$ of the probability distribution:

$$p(x, y) = \lambda\eta e^{-\lambda x - \eta y}$$

supported on $\mathbb{R}^2_+$. We consider a dataset $\mathcal{D} = ((x_1, y_1), \ldots, (x_N, y_N))$ composed of N independent draws from this distribution.

(a) *Show* that x and y are independent.

(b) *Derive* a maximum likelihood estimator of the parameter λ based on $\mathcal{D}$.

(c) *Derive* a maximum likelihood estimator of the parameter λ based on $\mathcal{D}$ under the constraint $\eta = 1/\lambda$.

(d) *Derive* a maximum likelihood estimator of the parameter λ based on $\mathcal{D}$ under the constraint $\eta = 1 - \lambda$.

Exercise 2: Multiple Linear Regression (15+5+5+5 P)

Consider the multiple linear regression problem $y = \mathbf{x}^\top \boldsymbol{\beta} + \epsilon$, where $\mathbf{x} \in \mathbb{R}^d$ are the predictor variables, $y \in \mathbb{R}$ is the response variable, and $\boldsymbol{\beta} \in \mathbb{R}^d$ are the linear regression coefficients. We have again a dataset $\mathcal{D} = ((\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N))$ of N independent draws of pairs (x_i, y_i). We summarize data and noise into the vectors $\mathbf{y} = (y_1, \ldots, y_N)^\top \in \mathbb{R}^N$ and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_N)^\top \in \mathbb{R}^N$, and the matrix $X = (\mathbf{x}_1, \ldots, \mathbf{x}_N)^\top \in \mathbb{R}^{N \times d}$.
In the lecture, we have derived the maximum-likelihood solution for $\boldsymbol{\beta}$ under the assumption of zero-mean Gaussian distributed noise (denoted by $\epsilon \sim \mathcal{N}(0, \sigma^2)$):

$$\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \mathbf{y} \; .$$

(a) *Show* that $\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2 (X^\top X)^{-1})$; i.e., $\hat{\boldsymbol{\beta}}$ is Gaussian distributed with mean $\boldsymbol{\beta}$ and covariance matrix $\sigma^2 (X^\top X)^{-1}$.

(b) *Discuss* the benefit of knowing the full distribution of $\hat{\boldsymbol{\beta}}$ rather than only the estimate itself. What additional statements about $\boldsymbol{\beta}$ can be made (hint: variable selection)? Assume that σ^2 is known and does not need to be estimated.

(c) Assume we have measured a new datapoint, $\mathbf{x}_*$. We use our regression model to predict the response for $\mathbf{x}_*$: $\hat{y}_* = {\mathbf{x}_*}^\top \hat{\boldsymbol{\beta}}$. *Derive* the distribution of $\hat{y}_*$.

(d) *Discuss* the benefit of also knowing that distribution in an application of your choice.

Exercise 3: Programming (40 P)

Download the programming files on ISIS and follow the instructions.