Mark Podolskij
mark.podolskij@uni.lu

HOMEWORK 4

Exercise 1. Assume that the vector (X_1, X_2, X_3, X_4) has a 4-dimensional normal distribution with mean 0 and arbitrary covariance matrix. Show that

$$\mathrm{cov}(X_1X_2, X_3X_4) = \mathrm{cov}(X_1, X_3)\mathrm{cov}(X_2, X_4) + \mathrm{cov}(X_1, X_4)\mathrm{cov}(X_2, X_3).$$

Exercise 2. Let X be a k-dimensional vector with $\mathbb{E}[X] = 0 \in \mathbb{R}^k$ and $\mathbb{E}[XX^T] = \Sigma \in \mathbb{R}^{k\times k}$. We already know from the lecture that the first principal component $\beta_1 \in \mathbb{R}^k$ is the eigenvector of Σ that corresponds to the largest eigenvalue λ_1 of Σ. Show that the vector $\beta_2 \in \mathbb{R}^k$ such that $\|\beta_2\|^2 = 1$, $\langle \beta_1, \beta_2\rangle = 0$, which maximises $\mathrm{var}(\beta_2^T X)$, is the eigenvector of Σ that corresponds to the second largest eigenvalue λ_2 of Σ.

Exercise 3. Let $X_1, \ldots, X_n$ be a sequence of i.i.d. random variables with common distribution $\mathcal{N}_k(\mu, \Sigma)$. Assume that the mean vector $\mu \in \mathbb{R}^k$ is known. Show that the maximum likelihood estimator of $\Sigma \in \mathbb{R}^{k\times k}$ is given by

$$\widehat{\Sigma}_{\mathrm{ML}} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)(X_i - \mu)^T.$$

Exercise 4. (Numerical exercise to solve with R or Python).

Use the same prostate dataset as in the last sheet, which can be found at the link : "https ://hastie.su.domains/ElemStatLearn/". We want to perform the regularization method known as Adaptive Lasso, which seeks to minimize

$$\|Y - Xb\|_2^2/n + \lambda\sum_{j=1}^{k}\hat{\omega}_j|b_j|,$$

where λ is the tuning parameter (chosen through 10-fold cross validation), b_j are the k estimated coefficients and $\hat{\omega}_j$ are Adaptive Weights. We recall that

$$\hat{\omega}_j = \frac{1}{(|b_j^{in}|)^\gamma},$$

where b_j^{in} is an initial estimate of the coefficients and γ is a positive constant for adjustment of the Adaptive Weights.

1. Build the Ridge Regression model for the variable prostate antigen (lpsa). Use it to obtain the initial estimates of the coefficients b_j^{in}.
2. Create the Adaptive Weights $\hat{\omega}_j$ for $\gamma = 0, 5, 1$ and 2. Choose the best λ through 10-fold cross validation.
3. Use the glmnet function to execute the adaptive Lasso. Plot the area under ROC (receiver operating characteristic) curve, also called AUC, and report the values of minimum λ (obtained for miminum AUC).