Test
Statistical Learning
October 30, 2019, 8.45-11.45 h.
Instructor Johannes Schmidt-Hieber

Part A: Basic concepts

(a) [1 point] Describe the nearest neighbor classifier.

(b) [1 point] What is the statement of the Gauss-Markov theorem?

(c) [1 point] How does foreward-stagewise selection work?

(d) [2 points] What is the masking phenomenon in classification?

Part B: Theory

1. [1 point] Given $\alpha = 0.05$, consider a multiple hypothesis testing problem with M individual hypothesis tests. Suppose that for any of the individual tests the p-value is smaller than α. To bound the false discovery rate (FDR) by α, which of the null hypotheses will the Benjamini-Hochberg procedure reject?

2. [2 points] Exclude the intercept from the linear model and consider the LASSO estimator

$$\widehat{\beta}^{\mathrm{LASSO}} \in \|y - X\beta\|_2^2 + \lambda\|\beta\|_1$$

with penalty $\lambda > 2\max_{i=1,\ldots,p} |X_i^\top y|$, where X_i denotes the i-th column of the design matrix X. Show that the LASSO functional has the unique solution

$$\widehat{\beta}^{\mathrm{LASSO}} = 0.$$

3. [3 points] In the linear regression model $Y = X\beta + \varepsilon$, the least squares estimator exists under the assumption that $X^\top X$ is invertible. If $X^\top X$ is not invertible, show that for any β there exists a $\beta' \neq \beta$ such that

$$Y = X\beta + \varepsilon = X\beta' + \varepsilon.$$

What does this mean concerning estimation of the vector β?

4. Recall that the p.d.f. of an exponential distributed random variable with parameter $\theta > 0$ (denoted in the following by $\text{Exp}(\theta)$) is given by $\theta^{-1}\exp(-x/\theta)\mathbf{1}(x > 0)$. It is well known that the life span distribution of many items (mobile phones, cars, light bulbs, ...) follows an exponential distribution. Suppose we observe for n mobile phones whether they still work after a fixed time $s > 0$ and we are interested in recovering the parameter θ. Thus, in this case, the full (unobserved) dataset consists of n i.i.d. exponential random variables, that is, $X_1, \ldots, X_n \sim \text{Exp}(\theta)$ modeling the life spans of the n mobile phones. The observed data are $S_i = \mathbf{1}(X_i \geq s)$, $i = 1, \ldots, n$. This means $S_i = 1$ if the i-th mobile phone still works after time s and $S_i = 0$ otherwise.

   (a) [1 point] Derive the log-likelihood of the full data model, where we observe $X_1, \ldots, X_n$.

   (b) [1 point] Show that the p.d.f. of $X_i|(X_i \geq s)$ is $x \mapsto \theta^{-1}\exp((t-x)/\theta)\mathbf{1}(x \geq t)$.

   (c) [1 point] Show that the p.d.f. of $X_i|(X_i < s)$ is

$$x \mapsto \frac{e^{-x/\theta}}{\theta(1 - e^{-s/\theta})}\mathbf{1}(0 \leq x < s).$$

   (d) [1 point] Show that for any real number a, $\int_a^\infty ue^{-u}du = (a+1)e^{-a}$.

   (e) [1 point] Prove that for any $\theta' > 0$,

$$E_{\theta'}[X_i|(X_i \geq s)] = s + \theta'.$$

   (f) [1 point] Prove that for any $\theta' > 0$,

$$E_{\theta'}[X_i|(X_i < s)] = \theta' - \frac{se^{-s/\theta'}}{1 - e^{-s/\theta'}}.$$

   (g) [2 points] Show that the E-step in the EM-algorithm is given by

$$-n\log\theta - \frac{\theta^{(t)}n}{\theta} - \frac{s}{\theta}\sum_{i=1}^n S_i + \frac{se^{-s/\theta^{(t)}}}{\theta(1 - e^{-s/\theta^{(t)}})}\Big(n - \sum_{i=1}^n S_i\Big).$$

   (h) [1 point] Derive the M-step of the EM-algorithm.