UNIVERSITY OF TWENTE.

EXAM: Statistical Learning

November 5th, 2021, 8.45-11.45 h.

Instructor Hanyuan (Hans) Hang

During the exam, you're allowed to have a one-page cheat sheet, handwritten A4 on both sides, and a calculator. Answer the questions on separate sheets of paper. Do not forget to number the sheets and to write your name on each of them.

Part A: Basic concepts

(a) [1 point] Describe the k-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. [3 points] In the linear regression model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon,$$

the least squares estimator exists under the assumption that $\boldsymbol{X}^\top \boldsymbol{X}$ is invertible. If $\boldsymbol{X}^\top \boldsymbol{X}$ is not invertible, show that for any $\boldsymbol{\beta}$ there exists a $\boldsymbol{\beta}' \neq \boldsymbol{\beta}$ such that

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon = \boldsymbol{X}\boldsymbol{\beta}' + \varepsilon.$$

What does this mean concerning estimation of the vector $\boldsymbol{\beta}$?

2. [3 points] Consider the elastic-net optimization problem:

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda\big[\alpha\|\boldsymbol{\beta}\|_2^2 + (1-\alpha)\|\boldsymbol{\beta}\|_1\big].$$

Show how one can turn this into a lasso problem, using an augmented version of $\boldsymbol{X}$ and $\boldsymbol{y}$.

3. Recall that the probability density function (p.d.f.) of an exponential distributed random variable with parameter $\theta > 0$ (denoted in the following by $\mathrm{Exp}(\theta)$) is given by

$$\theta^{-1}e^{-x/\theta} \cdot \mathbf{1}\{x \geq 0\}.$$

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, $X_1, \ldots, X_N \sim \mathrm{Exp}(\theta)$ modeling the life spans of the N mobile phones. The observed data is

$$S_i = \mathbf{1}\{X_i \geq s\}, \qquad i = 1, \ldots, N.$$

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.

(b) [1 point] Show that the p.d.f. of $X_i|\{X_i \geq s\}$ is

$$x \mapsto \theta^{-1}e^{(s-x)/\theta} \cdot \mathbf{1}\{x \geq s\}.$$

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

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

(d) [1 point] Show that for any real number a, there holds

$$\int_a^\infty ue^{-u}\,du = (a+1)e^{-a}.$$

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

$$\mathrm{E}_{\theta'}\big[X_i|\{X_i \geq s\}\big] = s + \theta'.$$

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

$$\mathrm{E}_{\theta'}\big[X_i | \{X_i < s\}\big] = \theta' - \frac{se^{-s/\theta'}}{1 - e^{-s/\theta'}}.$$

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

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

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