Chun-Biu Li and Taras Bodnar
Department of Mathematics
Stockholm University
E-Mail: taras.bodnar@math.su.se

Statistical Learning

Exam, 2022/01/11

This is an open book exam. You are allowed to use your laptop, your notes and the course book, but not help from other people.

The answers to the tasks should be clearly formulated and structured. All non-trivial steps need to be commented. The solutions should be given in English.

The final grade is determined according to the following table:

Grade	A	B	C	D	E	F
Points	$\geq$45	(45-40]	(40-35]	(35-30]	(30-25]	< 25

Problem 1 [10P]

Consider the linear regression model

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

where

- $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ is the vector of response variables
- $\boldsymbol{X}$ is the $N \times (p+1)$ design matrix
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ is the parameter vector
- $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$ is the vector of errors

Let $\hat{\boldsymbol{\beta}}_{LS}$ be the least-squares (LS) estimator of $\boldsymbol{\beta}$ and let $\hat{\boldsymbol{\varepsilon}}$ be the vector of the residuals.

(a) Let $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$. Prove that $\boldsymbol{H}$ is idempotent. **[2P]**

(b) Prove that $rank(\boldsymbol{H}) = p+1$. **[3P]**

(c) Show that $\boldsymbol{X}^\top\hat{\boldsymbol{\varepsilon}} = \boldsymbol{0}$. **[2P]**

(d) Prove that $\hat{\sigma}^2 = \frac{1}{N-p-1}\hat{\boldsymbol{\varepsilon}}^\top\hat{\boldsymbol{\varepsilon}}$ is an unbiased estimator for σ^2, when $\boldsymbol{X}$ is deterministic. **[3P]**

Problem 2 [10P]

Let G_i have K classes and let $\boldsymbol{Y}: N \times K$ be the indicator response matrix given by

$$\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)^\top \quad \text{with} \quad \boldsymbol{y}_i^\top = (y_{i1} ...., y_{iK}) \quad \text{where} \quad y_{ik} = 1 \text{ if } G_i = k \text{ else } 0.$$

We fit a linear regression model to each of the columns of $\boldsymbol{Y}$ simultaneously by employing the least-squares estimator and collect the estimated coefficients in the matrix

$$\hat{\boldsymbol{B}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y},$$

where the ith column corresponds to the estimated coefficients found by fitting the linear regression model to the i-th column of $\boldsymbol{Y}$ on $\boldsymbol{X}$ which is the design matrix with $p+1$ columns corresponding to the p inputs and a leading column of 1's for the intercept. The classification rule for a new observation with input $\boldsymbol{x}_0$ is the following:

- compute the fitted output $\hat{\mathbf{f}}(\boldsymbol{x}_0)^\top = (1, \boldsymbol{x}_0^\top)\hat{\boldsymbol{B}}$, a K-dimensional vector;
- identify the largest component and classify accordingly:

$$\hat{G}(\boldsymbol{x}_0) = \underset{k \in 1, \ldots, K}{\operatorname{argmax}} \hat{f}_k(\boldsymbol{x}_0)$$

Prove the following equality:

$$\sum_{k=1}^{K} \hat{f}_k(\boldsymbol{x}_0) = 1 \quad \text{for any} \quad \boldsymbol{x}_0,$$

where $\hat{f}_1(\boldsymbol{x}_0), ..., \hat{f}_K(\boldsymbol{x}_0)$ are the elements of $\hat{\mathbf{f}}(\boldsymbol{x}_0)$.**[10P]**

Problem 3 [10P]

Consider the truncated power series representation for cubic splines with K interior knots $\xi_1, \xi_2, ..., \xi_K$ given by

$$\begin{aligned} y_i &= \beta_0 h_0(x_i) + \beta_1 h_1(x_i) + \beta_2 h_2(x_i) + \ldots + \beta_{K+3} h_{K+3}(x_i) + \varepsilon_i \quad \text{with} \\ & \quad h_0(X) = 1,\ h_1(X) = X,\ h_2(X) = X^2,\ h_3(X) = X^3, \\ & \quad h_4(X) = (X - \xi_1)_+^3,\ \ldots\ h_{K+3}(X) = (X - \xi_K)_+^3. \end{aligned}$$

Prove that the natural boundary conditions for natural cubic splines imply the following linear constraints on the coefficients:

(a) $\beta_2 = 0$ and $\beta_3 = 0$, **[3P]**

(b) $\sum_{k=1}^{K} \beta_{3+k} = 0$ and $\sum_{k=1}^{K} \xi_k \beta_{3+k} = 0$. **[7P]**

Problem 4 [10P]

Let

$$\hat{f}(x) = \frac{1}{N}\sum_{i=1}^{N} K_\lambda\left(x - x_i\right)$$

be the kernel estimator of the density $f(x)$, $x \in \mathbb{R}$, based on independent sample $x_1, ..., x_N$ where

- λ is the bandwidth parameter and
- $K_\lambda(u) = \frac{1}{\lambda}K\left(\frac{u}{\lambda}\right)$ with $K(.)$ – a density, symmetric around 0.

Prove the following statements:

(a) The continuity of $K(.)$ implies that $\hat{f}(x)$ is also continuous for $\lambda \neq 0$, **[2P]**

(b) $\hat{f}(x)$ is a density. **[8P]**

Problem 5 [10P]

Consider the regression model

$$y_i = f(\boldsymbol{x}_i) + \varepsilon_i \quad \text{with} \quad f(\boldsymbol{x}_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta},$$

where $\varepsilon_1, ..., \varepsilon_n$ are independent and identically distributed with finite second moment and the design matrix $\boldsymbol{X}$ is assumed to be deterministic. Derive the expression of the expected prediction error under the squared-error loss, defined by

$$\text{Err}(\boldsymbol{x}_0) = \mathbb{E}\left[(Y - \hat{f}(\boldsymbol{x}_0))^2\right],$$

when the regression function $\hat{f}(.)$ is fitted by

(a) the k-nearest-neighbor regression, **[5P]**

(b) the least-squares regression. **[5P]**