Taras Bodnar and Johannes Heiny
Department of Mathematics
Stockholm University
E-Mail: taras.bodnar@math.su.se

Statistical Learning

Exam, 2024/01/11

This is an open-book exam. You are allowed to use the course book and your ten A4 (two-sided) pages with hand-written notes, 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 of rank $p+1$,
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ 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) Prove that $\mathbb{E}(\hat{\boldsymbol{\beta}}_{LS}) = \boldsymbol{\beta}$. **[2P]**

(b) Show that $\mathbb{C}ov(\hat{\boldsymbol{\beta}}_{LS}) = \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$, when $\boldsymbol{X}$ is deterministic and $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I}_N)$. **[3P]**

(c) Prove that $\hat{\boldsymbol{\beta}}_{LS} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1})$, when $\boldsymbol{X}$ is deterministic and $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I}_N)$. **[2P]**

(d) Show that $\hat{\boldsymbol{\beta}}_{LS}$ and $\hat{\boldsymbol{\varepsilon}}$ are independent, when $\boldsymbol{X}$ is deterministic and $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I}_N)$. **[3P]**

Problem 2 [10P]

Consider a logistic regression with $K > 2$ classes, that is $G \in \{1, ..., K\}$. Using that

$$\begin{aligned}
\log \frac{\mathbb{P}(G = 1|\boldsymbol{X} = \boldsymbol{x})}{\mathbb{P}(G = K|\boldsymbol{X} = \boldsymbol{x})} &= \boldsymbol{x}^\top \boldsymbol{\beta}_1 \\
\log \frac{\mathbb{P}(G = 2|\boldsymbol{X} = \boldsymbol{x})}{\mathbb{P}(G = K|\boldsymbol{X} = \boldsymbol{x})} &= \boldsymbol{x}^\top \boldsymbol{\beta}_2 \\
&\vdots \\
\log \frac{\mathbb{P}(G = K-1|\boldsymbol{X} = \boldsymbol{x})}{\mathbb{P}(G = K|\boldsymbol{X} = \boldsymbol{x})} &= \boldsymbol{x}^\top \boldsymbol{\beta}_{K-1}
\end{aligned}$$

Compute the probability of each class $\mathbb{P}(G = k|\boldsymbol{X} = \boldsymbol{x})$ for $k = 1, ..., K$ as a function of $\boldsymbol{\beta}_1$,..., $\boldsymbol{\beta}_{K-1}$ where $\boldsymbol{x}$ is an observed vector of predictors.

Problem 3 [10P]

Consider a regression with a basis expansion

$$y_i = \beta_0 h_0(x_i) + \beta_1 h_1(x_i) + \beta_2 h_2(x_i) + \beta_3 h_3(x_i) + \varepsilon_i, \quad i = 1, ...N.$$

Show that a regression with linear splines for $K = 2$ knots can be presented as a regression with a basis expansion, where the basis functions are the following (t_+ denotes the positive part of t):

$$h_0(X) = 1,\ h_1(X) = X,\ h_2(X) = (X - \xi_1)_+,\ h_3(X) = (X - \xi_2)_+.$$

Problem 4 [10P]

Let

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

be the kernel estimator of the density $g(x)$, $x \in \mathbb{R}$, based on an independent and identically distributed 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 $K(.)$ being a density, symmetric around 0.

Let $K(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$ and let the true density $g(x)$ be the density of the normal distribution with mean μ and variance σ^2. Compute $\mathbb{E}[\hat{g}(x)]$ for $\lambda > 0$ and a fixed point $x \in \mathbb{R}$.

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

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

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

(a) the least-squares regression, **[5P]**

(b) the ridge regression. **[5P]**