Martina Favero
Department of Mathematics
Stockholm University
E-Mail: martina.favero@math.su.se

Statistical Learning

Exam, 2026/01/17

This is an open-book exam. You are allowed to use the course book (or a printed version) and your one A4 (two-sided) page with hand-written notes.

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.

This written exam consists of 5 problems, each worth 20 points. The final grade is determined according to the following table:

Grade	A	B	C	D	E	F
Points	≥90	[80,90)	[70,80)	[60,70)	[50,60)	< 50

Please number all sheets of paper that you hand in, so that their order is easy to recover (just in case).

Lycka till!

Problem 1 [20P]

Consider the regression model

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

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

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

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

(a) **[10P]** k-nearest neighbor regression,

(b) **[10P]** ridge regression.

Problem 2 [20P]

Show that a regression with a linear spline for $K = 2$ knots can be presented equivalently as a regression with a basis expansion

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

where the basis functions are the following ($t+$ denotes the positive part of t):

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

Hint: find explicit relationships between the coefficients $\beta_{00}, \beta_{10}, \beta_{01}, \beta_{11}, \beta_{02}, \beta_{12}$ of the regression with a linear spline and the coefficients $\beta_0, \beta_1, \beta_2, \beta_3$ of the regression with a basis expansion.

Problem 3 [20P]

Consider the kernel estimator

$$\hat{g}(x, y) = \frac{1}{N} \sum_{i=1}^{N} K_{\lambda_x}(x - x_i) K_{\lambda_y}(y - y_i), \qquad x, y \in \mathbb{R},$$

of the density g, based on an i.i.d. sample $(x_1, y_1) \ldots, (x_N, y_N)$ from a distribution with density g. Here $\lambda_x, \lambda_y > 0$ are the bandwidth parameters and $K_\lambda(u) = \frac{1}{\lambda} K(\frac{u}{\lambda})$ is the Kernel, with K being a continuous density function which is symmetric around zero.

Consider the regression function

$$f(x) = \mathbb{E}[Y \mid X = x] = \int y \; g(y \mid x) \; dy$$

and its estimator

$$\hat{f}(x) = \int y \; \hat{g}(y \mid x) \; dy \quad \text{where} \quad \hat{g}(y \mid x) = \frac{\hat{g}(x, y)}{\int \hat{g}(x, y) dy}.$$

(a) **[2P]** Show that $\int \hat{g}(x, y) dy = \hat{g}(x)$, where $\hat{g}(x) = \frac{1}{N} \sum_{i=1}^{N} K_{\lambda_x}(x - x_i)$ is the kernel estimator of $g(x)$.

(b) **[4P]** Show that $\int y K_{\lambda_y}(y - y_i) dy = y_i$.

(c) **[8P]** Show that $\hat{f}$ is equal to the Nadaraya-Watson kernel-weighted average

$$\frac{\sum_{i=1}^{N} K_{\lambda_x}(x - x_i) y_i}{\sum_{i=1}^{N} K_{\lambda_x}(x - x_i)}.$$

(d) **[6P]** Explain how the weights of the Nadaraya-Watson kernel-weighted average compare to the weights of the $k-$nearest neighbor estimator.

Problem 4 [20P]

The Bayes rule

$$f(x) = \text{argmin}_c \mathbb{E}_{Y|X}[L(Y, c) \mid X = x]$$

minimises the expected prediction error $EPE(f) = \mathbb{E}[L(Y, f(X))]$, with respect to a loss function L. In this problem we compute the Bayes rule for three different loss functions.

Assume $X \in \mathbb{R}^p$ is a random input vector and $Y \in \mathbb{R}$ is a random output variable.

(a) **[7P]** Derive the Bayes rule with respect to the squared loss function

$$L(Y, f(X)) = (Y - f(X))^2.$$

(b) **[7P]** Derive the Bayes rule with respect to the loss function

$$L(Y, f(X)) = \mid Y - f(X) \mid .$$

Assume now that, instead of Y, we have a categorical output variable G.

(c) **[6P]** Derive the Bayes rule with respect to the 0-1 loss function

$$L(G, \hat{G}(X)) = \begin{cases} 0 & \text{if} \quad G = \hat{G}(X) \\ 1 & \text{if} \quad G \neq \hat{G}(X) \end{cases}$$

Problem 5 [20P]

(a) **[10P]** Briefly explain the main idea of quadratic discriminant analysis. Show that the degree-of-freedom of quadratic discriminant analysis equals to

$$(K-1)\left[\frac{p(p+3)}{2} + 1\right],$$

where K is the number of classes and p is the dimension of the predictor variables.

(b) **[10P]** For the LASSO regression problem, one has to solve

$$\hat{\beta}^{\text{LASSO}} = \arg\min_{\beta} \left(\sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right)$$

Find the relations between the parameters β_0, β_j's and β_0^c, β_j^c's such that the above optimization problem can be stated equivalently as

$$\hat{\beta}^c = \arg\min_{\beta^c} \left(\sum_{i=1}^{N} \left(y_i - \beta_0^c - \sum_{j=1}^{p} (x_{ij} - \overline{x}_j)\beta_j^c \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j^c| \right).$$