Johannes Heiny
Department of Mathematics
Stockholm University
E-Mail: johannes.heiny@math.su.se

Statistical Learning

Exam, 2025/01/11

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, 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.

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	$\geq$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]

In this problem, we consider kernel density estimation. Let

$$\hat{f}(x) = \frac{1}{N}\sum_{i=1}^{N} K_\lambda(x - x_i)\,, \qquad x \in \mathbb{R},$$

be the kernel estimator of the density f, based on an i.i.d. sample $x_1, \ldots, x_N$ from a (real-valued) distribution with density f. Here $\lambda > 0$ is the bandwidth parameter.

(a) **[2P]** Assume that K is a density which is symmetric around zero. Show that $u \mapsto \frac{1}{\lambda}K\left(\frac{u}{\lambda}\right)$ defines a density.

(b) **[18P]** Let $K_\lambda(u) = \frac{1}{\lambda}K\left(\frac{u}{\lambda}\right)$. Compute $\mathbb{E}[\hat{f}(x)]$ in the following two cases.

   (1) $K(u) = \frac{1}{\sqrt{2\pi}}\exp(-u^2/2)$ and f being the density of the normal distribution with mean μ and variance σ^2.

   (2) $K(u) = f(u) = \frac{1}{2}\mathbf{1}_{[-1,1]}(u)$.

Problem 2 [20P]

(a) **[10P]** 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 ridge regression problem, one has to solve

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

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

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

Problem 3 [20P]

Gauss–Markov theorem: Consider the linear regression setting $y = X\beta + \varepsilon$ of Chapter 3. The noise ε satisfies $\mathbb{E}[\varepsilon_i] = 0$ and $\operatorname{Var}(\varepsilon_i) = \sigma^2 < \infty$ for all i. In addition, we assume $\operatorname{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$.

Show that if $\hat{\mathbf{V}}$ is the covariance matrix of the least squares estimator $\hat{\beta} = (X^\top X)^{-1}X^\top y$ of β and $\widetilde{\mathbf{V}}$ is the covariance matrix of any other linear unbiased estimator, then $\widetilde{\mathbf{V}} - \hat{\mathbf{V}}$ is a positive semidefinite matrix.

Problem 4 [20P]

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

$$f(X) = \sum_{j=0}^{K+3} \beta_j h_j(X) \,,$$

where

$$h_j(X) = \begin{cases} 1\,, & \text{if } j = 0, \\ X^j\,, & \text{if } j = 1, \dots, 3, \\ (X - \xi_{j-3})_+^3\,, & \text{if } j = 4, \dots, K+3. \end{cases}$$

Prove that the natural boundary conditions for natural cubic splines (Section 5.2.1) imply the following linear constraints on the coefficients:

(a) $\beta_2 = 0$ and $\beta_3 = 0$,

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

Problem 5 [20P]

Let $z = (z_1, \dots, z_N)^\top \in \mathbb{R}^N_{\geq 0}$ and $\mu > 0$. We consider the following optimization problem:

$$\min_x \|x - z\|_2^2 \quad \text{subject to} \quad x \in \mathbb{R}^N \text{ and } \|x\|_2^2 \leq \mu \,.$$

(a) **[1P]** Show that if $\|z\|_2^2 \leq \mu$, the solution is trivial.

(b) **[3P]** Let $x^* = (x_1^*, \dots, x_N^*)^\top$ be a minimizer of the above problem. Show that $x_i^* \leq z_i$ for all i.

(c) **[6P]** Prove that the minimum is attained at a unique point.

(d) **[10P]** Find the unique solution to the optimization problem.
Hint: In the lecture, we showed that the optimization problem

$$\arg\min_\beta \|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2$$

is equivalent to

$$\arg\min_\beta \|y - X\beta\|_2^2 \quad \text{subject to} \quad \|\beta\|_2^2 \leq t,$$

for $t = \|\hat{\beta}_\lambda^{\text{ridge}}\|_2^2$. Note that, to receive full points, this hint may only be used for part (d).