Machine learning for data science I

23 June 2025

Surname, name (all caps) ______________________________

Student ID: ______________________________

This is a closed book exam.

Write clearly and justify your answers.

Time limit: 90 min.

Question:	1	2	3	4	Total
Points:	25	25	25	25	100
Score:					

1. Answer the following questions about linear regression.

[10] (a) Consider a simple linear regression model with a single parameter β and single feature x_i, i.e. $y_i = \beta x_i + \epsilon_i$. Derive a closed form solution for this simple case.

[5] (b) Linear regression models typically assume an additive normally distributed noise in the dependent variable. Consider what would happen in the previous case if the source of noise was in the measurement of the independent variable x_i (instead of the dependent variable y_i, which would have no noise). Suppose we would estimate β naively in the same manner as before. Explain how would that influence the obtained result.

[10] (c) Derive the relationship between the naively estimated β' and the true β - what is the expected value for a naive estimation of β'?

Hint: You can assume a large number of samples and therefore make the following approximation: $\sum_{i=1}^{n} A_i \approx n\,\mathbb{E}[A]$.

2. Naive Bayes is a probabilistic classification method based on Bayes' theorem, with the naive assumption that all features are conditionally independent given the class. The classifier predicts the class that maximizes the product of the prior and the feature likelihoods: $\hat{y} = \arg\max_y [p(y) \prod_{i=1}^{n} p(x_i \mid y)]$.

   Gaussian Naive Bayes method estimates $p(x_i \mid y)$ and $p(y)$ for continuous features by fitting a normal distribution with a mean and variance specific to that combination of class value y and feature i.

   You can assume that we are dealing with a binary classification problem.

[5] (a) How would you estimate the probabilities $p(x_i = t \mid y = c)$ and $p(y = c)$ if the features were categorical?

[5] (b) Derive the expression for log-odds of a Logistic Regression classifier.

[15] (c) Show that the Gaussian Naive Bayes classifier is in fact equivalent to Logistic Regression under certain conditions. What are these conditions?
Hint: Derive the expression for log-odds of a GNB and compare it with LR.

3. Answer the following questions about Laplace Approximation (LA) of posterior probability distribution $p(x)$ (let $p(x)$ be a multivariate density).

[5] (a) What is the purpose of LA - why do we need it?

[5] (b) Describe the algorithm of LA.

[5] (c) LA can fit the mean well, but severely overestimate the variance. Draw an example univariate density where this would happen and the corresponding LA to that density. Let the drawing be approximately to scale (densities have to integrate to the same area).

[5] (d) Draw an example where LA fits the mean well, but severely underestimates the variance.

[5] (e) Draw an example where LA doesn't approximate the mean well.

4. Understanding and Interpreting t-SNE.

[8] (a) t-SNE models low-dimensional similarity using a Student t-distribution with one degree of freedom:

$$q_{ij} = \frac{(1 + \|\mathbf{z}_i - \mathbf{z}_j\|^2)^{-1}}{\sum_{k \neq l}(1 + \|\mathbf{z}_k - \mathbf{z}_l\|^2)^{-1}}$$

Why is a Student t-distribution used instead of a Gaussian in the low-dimensional space? Address both mathematical and visualization-related reasons.

[9] (b) Compare t-SNE and MDS with respect to:

- What they aim to preserve.
- Treatment of local vs. global structure.
- Interpretation of distances in the resulting plots.

[8] (c) You run t-SNE twice on the same dataset with different random seeds. The cluster structures are the same, but their positions differ. Is this a problem? Why or why not? What aspects of the output remain meaningful?