Machine learning for data science I

1 September 2025

Surname, name (all caps) ____

Student ID: ____

This is a closed book exam.

Write clearly and justify your answers.

All your answers should fit on the exam pages. You can use auxiliary sheets for thinking but they should not be submitted. Only your clear final answers on the exam will be graded.

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.

[5] (a) How does linear regression model the data - what is the assumed relationship between descriptors and the response (don't forget the errors)?

[5] (b) We determine the parameters of a linear model with the maximum likelihood estimation approach to estimate the mean of the distribution. Write the expression describing the likelihood and the simplified cost function that the optimization problem minimizes?

[5] (c) The likelihood of the data depends on the variance of the distribution. However, the optimization problem does not include the variance at all. Why is that and when did it disappear?

[10] (d) Use a similar MLE approach to also estimate the variance (as a single variable) of the distribution and comment on the obtained result.

2. Compare the kernel SVM model to the nearest neighbours model on a binary classification problem $y_i \in \{-1, 1\}$:

[15] (a) Describe both models and their underlying assumptions.

[10] (b) Define decision functions for both models and compare them. Use $\delta(x, x_i) \in \{0, 1\}$ as an indicator function whether x_i is one of the nearest neighbours of x.

3. For a given quantile $\alpha \in (0,1)$, the Pinball Loss $L_Q(\theta, \hat{\theta}, \alpha)$ is defined as:

$$L_Q(\theta, \hat{\theta}, \alpha) = \begin{cases} \alpha(\theta - \hat{\theta}) & \text{if } \theta > \hat{\theta} \\ (1-\alpha)(\hat{\theta} - \theta) & \text{if } \theta \leq \hat{\theta}. \end{cases}$$

Let $\theta \in (0,1)$ be our parameter and let the posterior density for θ be the unit uniform distribution:

$$p(\theta|y) = U(0,1) = \begin{cases} 1 & \text{for } 0 \leq \theta \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Derive the optimal Bayesian estimator that minimizes the posterior expected loss:

[5] (a) Set up the expression for the posterior expected loss.

[13] (b) Simplify the expression.

[4] (c) Determine the minimum.

[3] (d) Show that it is indeed a minimum.

4. Consider a dataset of hand-written digits that contains 1000 images. Each image consists of 8x8 pixels with gray-scale values and is therefore described by a 64-dimensional feature vector. We would like to perform dimensionality reduction to visualize this data in 2D. For this purpose, we first standardize the features (to zero mean and unit variance). Finally, we apply PCA, MDS, t-SNE or an autoencoder (with ReLU activation).

[12] (a) Briefly describe what distinguishes each of the mentioned dimensionality reduction methods from the others.

[13] (b) The visualizations below show the results of applying each of the methods. Determine which visualization belongs to which method and explain why.