

Machine learning for data science I

8 June 2026

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. A device has lifetime X modeled by an exponential distribution with unknown rate parameter $\lambda > 0$:

$$p(x|\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

Mean: $\mathbb{E}[X] = \frac{1}{\lambda}$, Var: $\text{Var}(X) = \frac{1}{\lambda^2}$, CDF: $F(x) = 1 - e^{-\lambda x}$

You conduct an experiment with $n = m + c$ devices. For m devices, you observe exact failure times x_1, \dots, x_m . For the remaining c devices, the experiment ends at a known time T , and you only know that they have not failed.

- [5] (a) Briefly describe maximum likelihood estimation (MLE).
- [5] (b) Define and explain the likelihood function $L(\lambda)$.
- [10] (c) Derive the MLE $\hat{\lambda}$ and interpret it in terms of observed data.
- [5] (d) Suppose that both λ and T were unknown. Could you estimate both parameters and how? Explain your choice and comment on its reliability.

2. Consider a binary classification problem where points $x = (x_1, x_2)$ are sampled uniformly from $[-1, 1]^2$ and labeled by the XOR rule: $y = 1$ if $\text{sign}(x_1) \neq \text{sign}(x_2)$ and $y = -1$ otherwise. The resulting decision boundary is not linearly separable, and different hypothesis classes will exhibit different limitations in approximating it.

- [5] (a) Determine the minimum achievable classification error of decision stumps (trees of depth 1, which use a single condition).
- [5] (b) Consider linear classifiers $\hat{y} = \text{sign}(x_1 - x_2 + c)$, which are restricted to decision boundaries with slope 1. Show that such models can improve the expected classification error from the previous question.
- [2] (c) Find the minimum achievable classification error with the model from previous question.
- [5] (d) Describe a low-dimensional feature transformation where XOR becomes linearly separable and construct a classifier $\hat{y} = \text{sign}(w^\top \phi(x) + b)$ achieving zero error in this space.
- [8] (e) Consider a neural network with a single hidden layer containing four neurons

$$h_1 = \text{ReLU}(x_1 + x_2), \quad h_2 = \text{ReLU}(x_1 - x_2), \quad h_3 = \text{ReLU}(-x_1 + x_2), \quad h_4 = \text{ReLU}(-x_1 - x_2),$$

and followed by a single output neuron with a bias term and a $\text{sign}(\cdot)$ activation. Determine the weights of the linear output layer such that the network exactly represents the XOR rule.

3. Given is a posterior distribution $p(\theta)$.

[10] (a) Define Bayesian risk under 0/1 loss and prove that it is minimized by the posterior mode. You may assume a discrete posterior.

[15] (b) Define Bayesian risk under quadratic loss and prove that it is minimized by the posterior mean. You may assume a continuous posterior.

4. Consider the following five clustering methods: k-means, hierarchical clustering with single linkage, hierarchical clustering with average linkage, DBSCAN, gaussian mixture models. All distance metrics are Euclidean. Each cluster is indicated by a different color (black points are not assigned to any cluster). Visualization axes have the same scale.

[10] (a) Briefly describe each of the mentioned clustering methods.

[10] (b) Determine which visualization belongs to which method and explain why.



[5] (c) Sketch the likely clusters for the mentioned clustering methods on the dataset illustrated below and explain your reasoning.



