

# Machine learning for data science I

31 August 2023

Surname, name (all caps) \_\_\_\_\_

Student ID: \_\_\_\_\_

This is a closed book exam.

Write clearly and justify your answers.

Time limit: 105 min.

Question:	1	2	3	4	5	Total
Points:	20	20	20	20	20	100
Score:						

1. Explain your answers to the following questions related to the k-nearest neighbors method:

- [4] (a) What are the time and space complexities of *building* a kNN model of a data set with  $n$  instances and  $f$  features?
- [4] (b) What is the time complexity of making a *prediction* for a single instance using the Euclidean distance?
- [4] (c) Propose two approaches (one exact and one approximate) for making predictions with a sublinear time complexity in terms of the number of instances  $n$ .
- [4] (d) Can you design a better approach for the special case where  $k = n$ ? Comment on its time and space complexity of building the model and making predictions with it.
- [4] (e) Describe a data set where the kNN method with  $k = 1$  would perform poorly in terms of the accuracy of its predictions.

2. Consider a regression problem where two features are perfectly correlated (suppose that  $x_2 = ax_1$ ,  $a \geq 1$ ). We will model the process with a regularized linear regression and will use an *extremely small* regularization parameter  $\lambda$  (without regularizing the intercept). Assume that the optimal parameters  $\beta_i$  of a non-regularized model are all positive. Explain your answers to the following questions about the parameters  $\beta'_i$  of a regularized model.

- [4] (a) What is the effect of using an extremely small regularization parameter  $\lambda$  in linear regression (ignoring potential precision errors in the computation) compared to not using regularization at all?
- [8] (b) How will L1 regularization (Lasso) affect the model weights in the correlated case with an extremely small regularization weight mentioned above; which  $\beta'_1, \beta'_2$  are chosen?
- [8] (c) What about L2 regularization (Ridge); which  $\beta'_1, \beta'_2$  are chosen?

3. Consider a binary classification data set with  $n = 1000$  instances. Somehow, we found another data instance, but we don't know its class value. We decided to include it into the data set as two instances - one with a positive class value and the other with a negative class value. Explain what effects can such addition have on different classification models (compared to the same model built on the original data set).

[4] (a) k-nearest neighbors

[4] (b) logistic regression

[4] (c) support vector machine (with hard and soft margins)

[4] (d) decision tree

[4] (e) random forest

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

- [4] (a) What is the purpose of LA - why do we need it?
- [4] (b) Describe the algorithm of LA.
- [4] (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).
- [4] (d) Draw an example where LA fits the mean well, but severely underestimates the variance.
- [4] (e) Draw an example where LA doesn't approximate the mean well.

5. Explain your answers to the following questions related to the k-means clustering method:

- [4] (a) Describe the purpose of clustering methods - what do we want to achieve with them?
- [6] (b) Briefly describe the two most popular clustering methods: k-means and hierarchical clustering. Explain one advantage of each of these two methods.
- [5] (c) K-means method is solving an optimization problem. Does it find the optimal solution? Explain why it does find the optimum or provide an illustration of a counter-example.
- [5] (d) How would you adapt the k-means method to handle a data set with weighted instances (without creating duplicates of data points)? How does your method compare to increasing the data set with copies of data points proportional to their weights?