Machine learning for data science I

29 August 2024

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. We are dealing with a regression problem of modeling some blood parameter. We want to do this based on a single binary indicator variable, which indicates whether a patient received some drug or not. The data set is provided in the table below.

[8] (a) Determine the coefficients of a linear regression model. Explain your solution.

[7] (b) Determine the coefficients of a linear regression model if we don't model the intercept. Explain your solution.

[5] (c) Comment on the use of linear regression for modeling this problem.

x	y
1	0.5
0	8.0
1	1.0
1	2.5
0	5.0
0	4.0
1	4.0
0	3.0

2. Answer the following question about log loss.

[3] (a) Given a predicted probability distribution $p = [p_1, \ldots, p_n]$ and the outcome $y = [y_1, \ldots, y_n]$ (zeros with a single 1 indicating the outcome) write the expression that describes the log loss. What is the range of possible values?

[2] (b) Write the expression that describes the risk if the true distribution is $r = [r_1, \ldots, r_n]$.

[3] (c) What is a proper scoring rule?

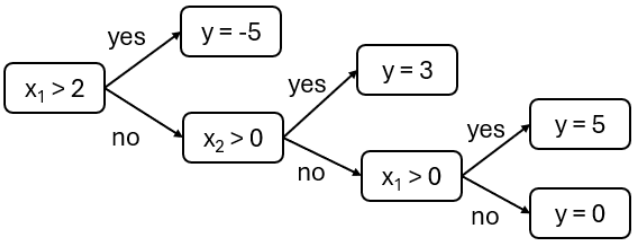
[12] (d) Prove that log loss is a proper scoring rule.

3. Answer the following questions about partial dependence plots.

[4] (a) Explain what is a partial dependence plot (PDP)?

[12] (b) We have a problem with three features $x_1, x_2, x_3 \in [-5, 5]$ and a continuous target $y \in \mathbb{R}$. All feature values are equally dense. Draw the PDP for x_1 for the following models:

- linear regression $y = 4x_1 + x_2 + 1 + \epsilon, \epsilon \sim N(0, 2)$
- linear regression $y = 4x_2 + x_3 + 1 + \epsilon, \epsilon \sim N(1, 3)$
- regression tree



[4] (c) Given a PDP for a feature, how would you derive a single number that would indicate the feature's global importance?

[20] 4. List four approaches to Bayesian computation, i.e. computing/approximating the posterior $p(\theta|y)$. Briefly describe each approach and list some of its advantages and disadvantages that distinguish it from the other approaches.

5. Consider the following statements. Answer if they are true or false, and explain (justify) your answer in one sentence.

[2] (a) PCA reduces the data to two components.

[2] (b) If we are given some categorical label for each of the data instances, we would be able to compare various dimensionality reduction techniques using silhouette index.

[2] (c) Instead of an attribute-based data, an input to t-SNE could also be a distance matrix.

[2] (d) Instead of an attribute-based data, an input to PCA could also be a distance matrix.

[2] (e) We are given distances between the European cities, and we would like to recreate a map showing their mutual positions. We should use t-SNE for this task instead of PCA or MDS.

[2] (f) We would like to denoise the data. Of the three dimensionality reduction techniques that include PCA, MDS, and t-SNE, we will choose PCA for this task.

[2] (g) We would like to display the neighboring structures in our data set within two-dimensional visualization. Of the three dimensionality reduction techniques that include PCA, MDS, and t-SNE, we will choose t-SNE for this task.

[3] (h) Starting with attribute-valued data, the advantage of the MDS to t-SNE when used for reducing the data space to two dimensions is that we can interpret the two resulting MDS coordinates, whereas we can not do this for t-SNE.

[3] (i) Dimensionality reduction by t-SNE provides a better clustering structure than that of the MDS.