Machine learning for data science I

19 June 2024

Surname, name (all caps) ____

Student ID: ____

This is a closed book exam.

Write clearly and justify your answers.

Time limit: 105 min.

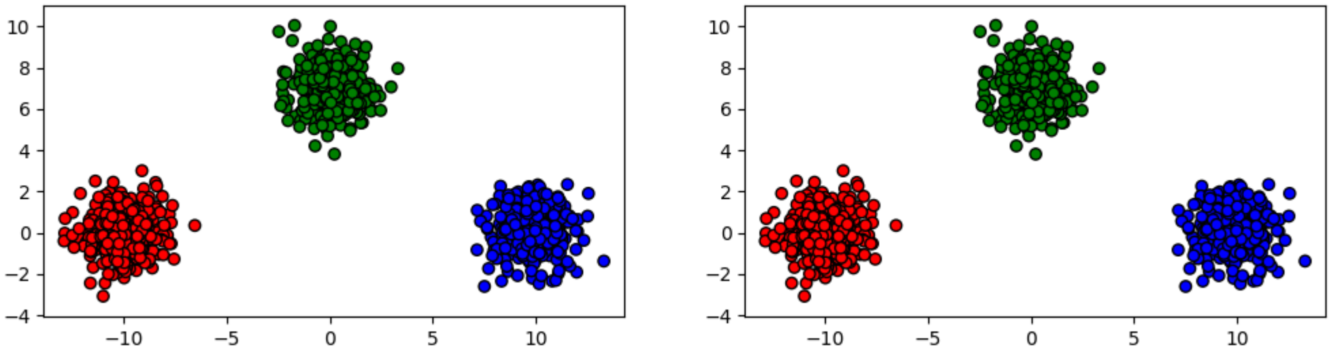
Question:	1	2	3	4	5	Total
Points:	20	20	20	20	20	100
Score:						

1. Two basic approaches to modeling a classification problems with multiple classes are with a multinomial logistic regression model and a one-vs-rest scheme with a logistic regression model. We are interested in predicting the probabilities of different class values. The data set consists of n instances with k features and m possible class values.

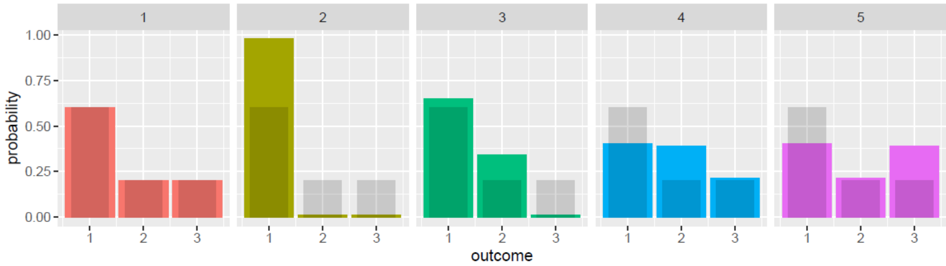
[6] (a) Describe how a multinomial logistic regression model works. How many parameters does it have?

[4] (b) Describe how a one-vs-rest scheme with logistic regression works. How many parameters does it have?

[10] (c) Sketch the regions with similar predicted probability distributions as precisely as possible and explain your reasoning. Draw the regions obtained with a multinomial regression model on the left and a one-vs-rest scheme with a logistic regression model on the right.



2. For each of the following probabilistic loss functions rank the predicted probability distributions according to the expected loss from lowest to highest. There are three possible class values and five probabilistic predictions in different colors. Gray columns represent the true distribution. Define each loss function and explain your reasoning.



The true distribution is $p_0 = [0.6, 0.2, 0.2]$ and the predicted are $p_1 = [0.6, 0.2, 0.2]$, $p_2 = [0.99, 0.01, 0.01]$, $p_3 = [0.64, 0.35, 0.01]$, $p_4 = [0.41, 0.39, 0.2]$, $p_5 = [0.41, 0.2, 0.39]$.

[5] (a) 0-1 loss

[5] (b) log loss

[5] (c) quadratic loss

[5] (d) absolute loss

3. Kernelization of the k-means algorithm with Euclidean distance.

[5] (a) How do we kernelize a machine learning algorithm and why would we do it?

[10] (b) The k-medoids clustering algorithm is similar to the standard k-means algorithm but uses a different definition of a cluster center. It uses cluster's medoid, which is the member that minimizes the sum of distances to all other members. Explain, how can you kernelize the k-medoids algorithm.

[5] (c) How would you kernelize the k-means algorithm (where the cluster center is its mean) - what is different from the k-medoids and how would you overcome this?

4. Answer the following questions related to Laplace approximation of a Beta distribution.

[17] (a) Let the distribution of x be $\text{Beta}(\alpha, \beta)$. Derive the Laplace approximation of the probability density function of x. Write the expressions you obtain explicitly, however, you are not required to simplify them.

[3] (b) Carefully analyze all edge cases where this approximation would run into problems.

The PDF of the Beta distribution is $p(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, x \in [0,1], \alpha > 0, \beta > 0$. Its mean and variance are $\frac{\alpha}{\alpha+\beta}$ and $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$, respectively.

5. Given is a data X (n data instances in rows, m features in columns). Show that the projection vector u with the largest variance $\mathrm{Var}(u^T X^T)$ is an eigenvector of the data's covariance matrix. Assume that the data is already centered. Structure your answer as follows:

[3] (a) define covariance matrix of X

[4] (b) express the variance with covariance matrix

[4] (c) formulate the optimization problem

[3] (d) define Lagrangian function

[3] (e) take the derivate of Lagrangian to show what is the solution of optimization problem

[3] (f) comment on this solution in terms of eigenvectors and eigenvalues