Shahid Beheshti University

Machine Learning Fundamentals

2nd Assignment – 2 May 2024

Hello everyone, I trust you are all in good health and spirits. This is the second assignment for our Machine Learning course. The deadline for this assignment is **9 May 2, 2024.** I encourage all students to adhere to this timeline for submissions. Should you have any questions concerning the exercises, please feel free to reach out.

Part 1

a) Write the **exact formulation** of Bias-Variance tradeoff.

b) Under what conditions might logistic regression outperform linear discriminant analysis, and vice versa?

c) Discuss the impact of the curse of dimensionality on the K-Nearest Neighbor algorithm.

d) Explain the concept of nested cross-validation and its benefits.

e) If we use a combination of the penalty term in Lasso and Ridge regression, what will happen in classification. Describe that. Also compare norm-one and norm-two and write about the intuition. If we replace Norm-p or norm-infinity instead of norm-1 or norm-2, what is the result?

f) How can the bootstrap method be used to estimate the standard error of a sample mean? (extra point)

g) Discuss how cross-validation can be used to assess the bias-variance trade-off of a model. (extra point)

Part 2

In this part you will work with the [Titanic-spaceship](#) dataset and perform a binary classification with your own knowledge.

- You need to conduct exploratory data analysis to gain **insights** from your data. (Which may occur during the feature engineering phase.)

- You should perform data cleaning and feature engineering on the dataset, and **describe each step** you take and **why** in the process in your Jupyter notebook or a separate document.

- It is necessary to train and evaluate an arbitrary model **before** and **after** feature engineering to examine the **effect** of your modifications on model's performance.

- You should use cross-validation to evaluate the performance of your model and perform an **ablation** study on your work.

Plagiarism will not be tolerated. Homework submissions will be cross-checked against other students' submissions. Additionally, **the use of AI to <u>fully</u> generate answers or code** for assignments is strictly forbidden.