# Assignment II

## Machine Learning

## February 2, 2026

**Abstract**

These exercises aim to review the classification problem in supervised learning. Please upload your solutions to Classroom before february 10th.

# Problem 1

Do a little research to answer the following: why is the MSE function, previously used as the cost function when doing regression, no longer popularly used as the cost function in the classification scenario? Keep your answer simple.

# Problem 2

Read chapter 4.5 from [1], focusing on KNN and Logistic Regression models. How do they compare with each other? Write down your conclusions.

# Problem 3

Consider the dataset from [2]. Make a script using sklearn to solve the binary classification problem associated with classifying whether a given horse will make it or not. Use the KNN model from sklearn. Write down all the assumptions and data preprocessing operations you make. Incorporate the F1 score to explain your findings [4].

# Problem 4

Repeat problem 3, but use the Logistic Regression model with sklearn. Make a comparison with your previous results. Also, write down the optimization hyperparameters you choose and explain why you pick them that way.

# Problem 5

Consider the dataset from [3]. Make a script using sklearn to solve the multiclass classification problem associated with classifying beers by style. Use the KNN model. Write down all the assumptions and data preprocessing operations you make. Look up what a confusion matrix is. Then, incorporate it into your analysis [5].

# Problem 6

Look up the terms "One Vs One Classifier" and "One Vs Rest Classifier". Then repeat problem 5 using sklearn for the two aforementioned approaches. Write down all the assumptions and conclusions.

# Problem 7

Consider the general quadratic equation

$$ax^2 + bx + c = 0.$$

Do your best to classify the roots of the equation above in terms of its coefficients. For simplicity, you have to fix the domain for $x$ and propose a suitable domain for the coefficients. Write down all your assumptions and show off your findings using some plots.

# References

[1] G. James et. al.m **An Introduction to Statistical Learning - with Applications in Python**, Springer, 2023.

[2] `https://www.kaggle.com/datasets/yasserh/horsesurvivalprognostication`

[3] `https://www.kaggle.com/datasets/ankurnapa/ankurs-beer-data-set`

[4] `https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html`

[5] `https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html`