BIN2023R01 – INTRODUCTION TO DATAMINING & MACHINE LEARNING FOR BIOINFORMATICS

Lab Exercise 6- Classification with logistic regression

Aim: To build a logistic regression model that can classify and predict the given dataset.

Procedure:

1. Utilize the necessary packages to construct the model.
2. Conduct feature selection on the provided datasets.
3. Perform quality analysis and preprocess the data.
4. Construct the logistic regression model.
5. Assess the model's performance.
6. Make predictions for user-defined data.

Questions:

1. What is logistic regression, and how does it differ from linear regression?
2. What types of problems can logistic regression be used for?
3. What are the assumptions made by logistic regression?
4. What are the limitations of simple linear regression compared to multiple linear regression?
5. What is the significance of the parameters (coefficients) learned by logistic regression?
6. Interpret the model's performance metrics including accuracy, error rate, true positive rate, true negative rate, false positive rate, and false negative rate.
7. Which features were selected and excluded during the model construction process?
8. Analyze the outcomes of the user-defined dataset. How can we determine the model's predictive accuracy for unseen data lacking a target variable?

Soft copy deadline: 4th March 11:59PM
Hard copy deadline: 5th March 3:15PM