

Graduate Certificate in Artificial Intelligence with Machine Learning
AIGC 5002 - Machine Learning and Deep Learning
Fall 2023

Lab 3:
Logistic Regression
September 27, 2023

Submission guidelines:

- For this lab, you will need to submit 1 PDF file by the end of lab time.
- After you complete all the exercises, convert your Jupyter Notebook (.ipynb) to PDF. Name the PDF as follows: `firstname_lastname_LAB3.pdf`
- Go to the course Blackboard → Labs folder → Lab 3 and submit the pdf.

Part 1: Logistic Regression

1. Observe the in-class demo.
2. Begin by researching online repositories that host datasets. Some examples include Kaggle, UCI Machine Learning Repository, and Google Dataset Search.
3. Look for datasets tagged with “Binary Classification”, or browse through datasets in the respective repositories to find ones that suit binary classification tasks.
4. Select a dataset that is labeled for binary classification. This means that the target variable should have only two classes.
5. Make sure the dataset has a sufficient number of instances and features to make the exercise meaningful. (Minimum 200 rows)
6. Display the first few and last few rows of the dataset to get a sense of what the data looks like.
7. Determine the number of instances and features in the dataset.
8. Examine the features (columns) of the dataset, and determine their data types (e.g., numerical, categorical).
9. Research “Class Imbalance” and understand its significance. Research mitigation methods.
10. Split your dataset and explain your approach.
11. Design and fit a Logistic regression model
12. Compute the evaluation metrics and discuss the ones most critical in your problem.
13. Compute the confusion matrix and analyze it.

Enjoy!
