

# COSC 410 S24 Lab 2: Classification and Evaluation metrics

This Lab is due by 11PM EST on the day of the lab (Feb 5/6 depending on your lab). You should work in groups of two or three for this lab.

## Introduction

The purpose of this lab, and the associated HW1, is to strengthen your understanding of evaluation metrics more broadly the standard Machine Learning pipeline. By completing the lab and the HW, you will demonstrate that you understand:

- how to compute precision, recall and accuracy for multiple classes
- how to fit a logistic regression classifier using sklearn
- when to use train, validation and test splits
- how to implement k-fold cross validation
- how to generate hypotheses about feature relevance and interpret model failure
- the basics of model selection and hyperparameter tuning

## Structure

This lab has two parts:

1. Work through an ipynb notebook to train and evaluate a binary classifier on a toy dataset.
2. Look through the requirements for HW1 and brainstorm how to approach:
  - evaluating a multi-class classifier
  - implementing k-fold cross validation

## Provided Files

- Lab2.ipynb and dat.csv for the lab
- HW1.ipynb and iris.csv for HW1

## Part 1

Work through Lab2.ipynb with your partner. Then submit Lab2.ipynb with the cell outputs saved to Gradescope/Moodle for your lab section.

## Part 2

**Make sure to leave yourself at least 15-20 minutes to work on this part**

Homework 1 requires you to generalize the approach you adopted in Lab2.ipynb to the Iris dataset. This dataset differs from the toy dataset you worked on in the lab in two ways:

1. It has three classes instead of two.
2. It is much smaller (has only 150 rows instead of 1000).

Discuss the following questions with your partner. You can take notes of your discussion, but do not actually start writing any code for Homework 1.

- Will the functions you wrote to compute accuracy, precision and recall work for three classes? Why or why not? If they will not work, what modifications will you need to make such that they can be used for three (or more generally  $n$ ) classes?
- Homework 1 asks you to implement k-fold cross validation. Why is this more important for the Iris dataset than it was for the toy dataset?
- Implementing k-fold cross validation involves the following three steps. How might you approach implementing each of these steps?
  1. Dividing the training data into k-folds. Think specifically about how you can add a column to the dataframe which indicates what the folds are.

2. Compute evaluation metrics for a given fold.
3. Average evaluation metrics across all folds.

In your response to 2 and 3, think carefully about the distinction between a model type (e.g., Logistic Regression with specific hyperparameter settings) and model instance (the specific model you will use to train and evaluate on a given fold).