

Lab 8 (due: Mar 28)

MACHINE LEARNING - COSC 4360

Department of Computer Science and Electrical Engineering
Spring 2025

Exercises

Create a **New Project** for every exercise. Take a screenshot of the source code along with its output and place the **source code** and the **screenshot** in a **zipped folder** named **LastNameFirstName_Lab8**

Exercise 1

In a highway, the **minimum** speed limit is 30 mph and the **maximum** is 75 mph. The following data set named *speedLimits.csv* describes whether drivers got a Ticket (T) or No Ticket (NT) given the speed of their vehicle. Create a **scatter** plot with the data points from the .csv file; use *green* color for NT and *red* color for T, please see Fig. 1 below. Find the optimal **Kernel** (except for *precomputed*) for the **SVM** that separates the two classes.

Note: Use *test_size=0.1* and *random_state=0*

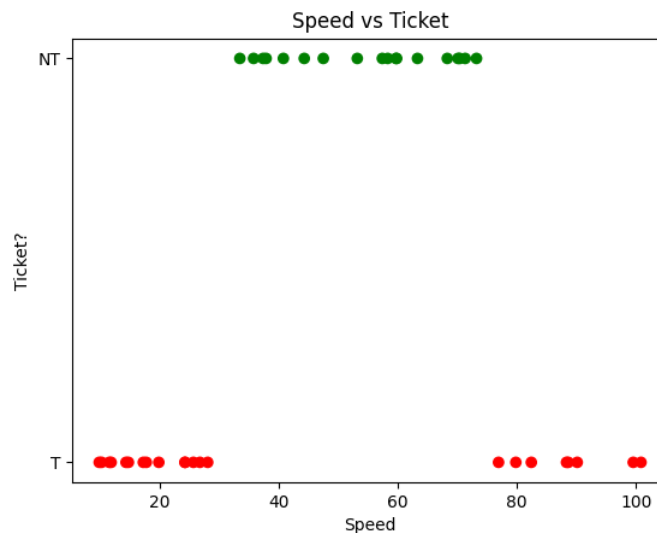


Figure 1: Scatter plot for Speed vs Ticket or No Ticket

Exercise 2

Given the following dataset: *breast-cancer-wisconsin-data.csv* and *breast-cancer-wisconsin-names*, remove all rows that contain a ? (this should reduce the number of rows from 699 to 683). Ignore the first column (ID number) and assign the remaining columns to X (features) and Y (label). Apply **Standardization**, then **PCA** with PC=2, and perform classification using the **SVM** supervised learning algorithm. Create a **scatter** plot with the two different classes and a **hyperplane** line (use 'linear' kernel). Your plot should be similar to Fig. 2 below. Print the **accuracy** of the model as well as the **Confusion Matrix**.

Note: Use the default *test_size=0.25* and *random_state=42*

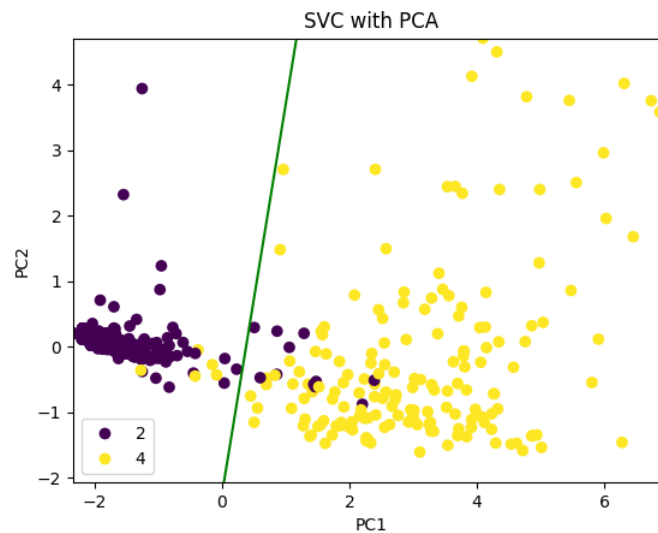


Figure 2: Classification using SVM with PCA

Note: Submit through **Canvas**