CST8506 - Lab 2
Support Vector Machines

Due Date: Check Brightspace for due dates.

Introduction

The goal of this lab is to classify Seed dataset using SVM (using standardized dataset, transformed dataset by applying PCA and LDA). The dataset is available at https://archive.ics.uci.edu/dataset/236/seeds. Download and create a csv file just as you did in Lab 1. Wherever required, you must use the last three digits of your student number as the random_state.

Steps (all these steps should be done in Python):

1. Load the Seeds Dataset file using Pandas csv read functionality.
2. Print number of instances, number of attributes and first 5 rows of the dataset.
3. Split the dataset into train and test sets.
4. Standardize data in the dataset.
5. Fit SVM models using all kernel types on the training set, make predictions on the test set, and determine the best-performing kernel based on step 6.
6. Print the confusion matrix (CM) and the accuracy for all the kernel used. Here After use the best kernel for further analysis.
7. Apply PCA to find the principal components of the standardized dataset.
8. Find and print the accuracy and CM after applying SVM on the transformed dataset.
9. Using the first 2 components, plot two 2D graphs – first one is the train set (color-coded by the class) and the second one with the support vectors too (color coded by the class).
10. Apply LDA to reduce the dimensionality of the standardized dataset.
11. Find and print the accuracy and CM after applying SVM on the transformed dataset.
12. For the new dataset, plot two 2D graphs – first one is the train set (color-coded by the class) and the second one with the support vectors too (color coded by the class).
13. By manual calculations, find the first principal component and the new values for the following dataset:

Length	Width
5	3
8	5
3	7
1	6
2	9

This step must by done on paper and should be handwritten. This should be neatly written and should be readable. After finishing the task, take photos or scan it and attach it to the answer document. I don't care about the way that you include your calculations, but I should be able to see it in the answer document.

In order to get grades,

1. You should be ready with your Python code and results.
2. Submit your jupyter / colab notebook file and the answer document to Brightspace.