Pattern Recognition and Machine Learning

Lab - 11 Assignment
SVM (Non Graded)

Guidelines for submission

1. Perform all tasks in a single colab file.
2. Create a report regarding the steps followed while performing the given tasks. The report should not include excessive unscaled preprocessing plots.
3. Try to modularize the code for readability wherever possible
4. Plagiarism will not be tolerated

Question 01: [100 marks]

The dataset is available here. You need to build a prediction model to identify fake notes based on several features extracted from bank-notes' specimen pictures.

Q.1 Pre-Process the dataset by handling missing values and normalizing the data. Split in the ratio 70:20:10 for train-test-validation. [10 Marks]

Q2. Train an SVM classifier using Sklearn library. Analyze the performance (classification accuracy) for different values of 'C'. (Choose at least 5 different values of C) [25 marks]

It is recommended to plot the decision boundaries for the various values of C to get a better understanding of how C affects the learning process.

Q3. Use various types of kernels(RBF, Linear, Quadratic etc) and train the SVM model using the Sklearn library. Plot the decision boundary for different svm models trained. [25 marks]

Note: While solving questions involving plotting the decision boundary, it is recommended to first perform dimension reduction of the dataset into two/three dimensions using LDA. And then plot the decision boundaries on the dataset, to gain a better visual understanding.

Instead of this, you can also choose to plot the decision boundary using the 2 features from the dataset, who have the highest correlation with the target.

Q4) Use the online tool: Link. It is an interactive SVM visualizer. Initially, one needs to manually add data points to the graph and then choose various values of hyperparameters (C, gamma, kernel, etc.). After that, the tool outputs the decision boundary and also highlights the support vectors.

For this question, create 2 types of datasets, according to your preferences, and analyze how each hyperparameter affects the SVM model.

To do so, try changing the kernel, C, gamma, etc. and add screenshots of the outputs in the report. Try to perform hyperparameter tuning to generate the best possible decision boundary for each dataset created.

Also, mention your observations in the report along with potential reasons for your observations. [20+20 marks]

Resources
1. SVM (sklearn)

2. Interactive Demo - SVM