MACHINE LEARNING AND PATTERN RECOGNITION

**Lab Week 7**

Instructions

**Step 1**: Import all libraries

* Numpy
* Pandas
* Matplotlib, Pyplot
* Seaborn
* SVC
* Train\_test\_split
* Confusion\_matrix, classification\_report, accuracy\_score, balanced\_accuracy,\_score

**Step 2**: Load all the data from the three given csv files for ODI, T20 and test match types.

**Step 3**: Concatenate runs from all the match type and store in **Runs**. Similarly do for strike rate and store in **SR** and Grounds in **ground**.

**Step 4**: get the match type data from all csv and store in **match\_type.**

***match\_type = np.concatenate((np.repeat("ODI", len(odi['Runs'])), np.repeat("T20", len(t20['Runs'])), np.repeat("TEST", len(test['Runs']))))***

**Step 5**: Create a data frame **data\_cricketer** and store Runs, SR and **match\_type** in it.

**Step 6**: From **data\_cricketer**, strore **Runs** and **SR** in **X** and **match\_type** in **Y.**

**Step 7**: Split the data X and Y in **x\_train, x\_test, y\_train , y\_test.** Use **test size=0.3** and **random state=0**

**Step 8:** Create SVC model with kernel Poly and its degree 6 (Adjust according to your model performance). Then fit the model and make prediction.

**Step 9**: Compute confusion matrix, accuracy and balanced accuracy (<https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html> )

* ***confusion\_matrix()***
* ***accuracy\_score()***
* ***balanced\_accuracy()***

***Step 10: Plot confusion matrix***

* Recommend to us ***sns.heatmap()***

**Step 11:** Also print classification report that includes precision, recall, f1 score and accuracy. See the output image. (Your output may slightly differ)

* Use ***classification\_report()***

A screenshot of a graph

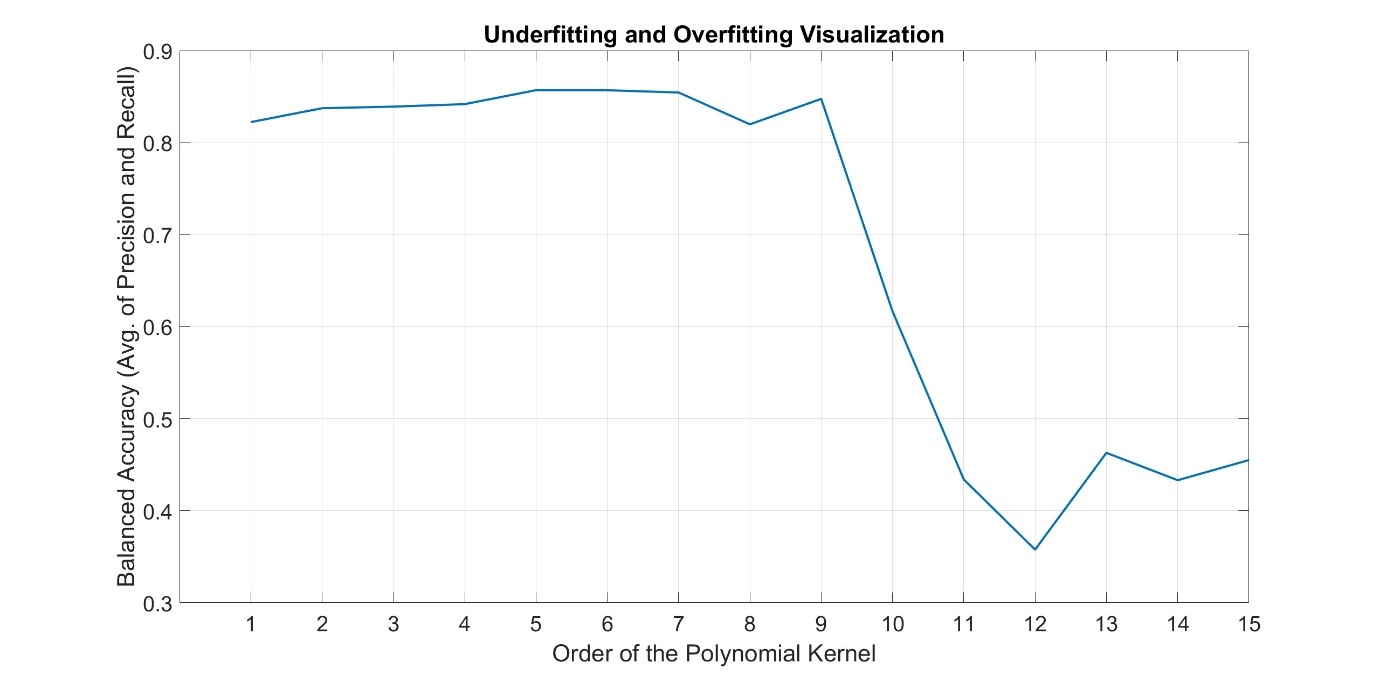
Description automatically generated

A graph of different colored circles

Description automatically generated**Step 12**: Create a scatter plot for Runs scored on x-axis and strike rate on y-axis with match type as a grouping. See reference output plot given below (Any one of two images is fine).

A graph showing numbers and numbers

Description automatically generated with medium confidence

Note: See why we are using polynomial degree of 6th order. If you can visualize it, plot this graph.

Deliverables

1. Scatter plot for strike rate vs runs
2. Confusion matrix as shown below in output reference image
3. Other plots shown above

Questions

1. What are Support Vectors in SVM? How do they help in making SVM less prone to outliers?
2. What is the difference between a hard margin and a soft margin in SVM?
3. How does SVM deal with non-linear classification boundaries?
4. What is the role of gamma in SVM? How does the value of gamma impact underfitting/overfitting?
5. How are SVMs used for multi-class classification problems?