Problem Set 2

Soumyajit Pal

August 2024

1   Hyperparameter Tuning

Consider the built-in Diabetes dataset from scikit-learn. Carry out the following tasks:

1. Load the dataset from sklearn.

2. Prepare a data frame for the given dataset.

3. Report the attribute names in the dataset, along with the total number of samples in the dataset.

4. Is it a dataset suited for classification or regression? Report the target variable.

5. Choose any two algorithms, depending on the dataset and report its performance on the dataset. Mention the performance metric you have chosen for evaluation.

6. Choose any two hyperparameters of the algorithms selected in Step 5.

7. Perform hyperparameter tuning using GridSearchCV.

8. Perform hyperparameter tuning using RandomizedSearchCV.

9. What inferences can you draw from Steps 5, 7 and 8?

10. Repeat all the above steps for the Breast Cancer Wisconsin dataset.