SVM Implementation

1. Import libraries
2. Import dataset
3. Explore the data to gain insights about the data
   (View dimension, preview the dataset,
   view the column names of the data frame,
   remove leading spaces from column names,
   rename column names,
   check the distribution of the target_class column,
   View the percentage distribution of the target_class column,
   View the summary of the dataset,
   Explore missing values in variables,
   view summary statistics in numerical variables,
   draw boxplots to visualize outliers,
   Check the distribution of variables)
5. Split data into separate training and test set
6. Feature Scaling
7. Run SVM with default hyperparameters
   Run SVM with rbf kernel and C=1.0 and gamma=auto
   Run SVM with rbf kernel and C=100.0
   Run SVM with rbf kernel and C=1000.0
8. Run SVM with linear kernel
   Run SVM with linear kernel and C=1.0
   Run SVM with linear kernel and C=100.0
   Run SVM with linear kernel and C=1000.0
9. Run SVM with polynomial kernel
   Run SVM with polynomial kernel and C=1.0
   Run SVM with polynomial kernel and C=100.0
   Run SVM with polynomial kernel and C=1000.0
10. Run SVM with sigmoid kernel

Run SVM with sigmoid kernel and C=1.0

Run SVM with sigmoid kernel and C=100.0

Run SVM with sigmoid kernel and C=1000.0

11. Compare the train-set and test-set accuracy
12. Check for overfitting and underfitting
13. visualize confusion matrix with seaborn heatmap
14. Generate classification report
15. Evaluate classification accuracy
16. Evaluate Classification error
17. Plot ROC Curve
18. Compute ROC AUC
19. Stratified k-fold cross-validation with shuffle split with linear kernel
21. Stratified k-fold cross-validation with shuffle split with rbf kernel
22. Stratified k-fold cross-validation with shuffle split with polynomial kernel
23. Stratified k-fold cross-validation with shuffle split with sigmoid kernel

Naive Bayes implementation

1. Import libraries
2. Import dataset(adult.csv)
3. Exploratory data analysis
    - View dimensions of the dataset
    - Rename column names
    - View summary of the dataset
    - Explore categorical variables and view them
        - check missing values in categorical variables
        - view frequency counts of values in categorical variables
        - view frequency distribution of categorical variables
        - Explore workclass variable -
            - check labels in the workclass variable

check the frequency distribution of values in the workclass variable

replace '?' values in the workclass variable with `NaN`

Explore occupation variable

check the frequency distribution of values in the occupation variable

replace '?' values in occupation variable with `NaN`

Explore native_country variable

check labels in the native_country variable

check the frequency distribution of values in the native_country variable

replace '?' values in the native_country variable with `NaN`

Check missing values in categorical variables again

Explore Numerical Variables

view the numerical variables

check missing values in numerical variables

4. Declare feature vector and target variable

5. Split data into separate training and test set

6. print the percentage of missing values in the categorical variables in the training set

7. print categorical variables with missing data

8. impute missing categorical variables with the most frequent value

9. Encode categorical variables

10. Feature Scaling

11. Model training

12. Predict the results

13. Check the accuracy score

14. Compare the train-set and test-set accuracy

15. Visualize confusion matrix with seaborn heatmap

16. Generate Classification Report

17. Evaluate Classification accuracy

18. Evaluate Classification error

19. Calculate class probabilities and store the probabilities in data frame with appropriate heading

20. Plot histogram of predicted probabilities
21. Plot ROC Curve
22. Compute ROC AUC
23. Calculate cross-validated ROC AUC
24. Applying 10-Fold Cross Validation
25. Compute the Average cross-validation score