Data Science
Lab Exercise
Prepared By
Dr Muhammad Atif Tahir

1. In this lab, you are going to learn how to classify data points using naïve bayes and support vector machine. The Breast Cancer Wisconsin (wdbc.data) dataset is given

Use 10 Fold Cross Validation to classify Malignant versus Benign. Print mean accuracy and accuracy of individual folds

    i. Using Support Vector Machine (Linear Kernel, C=1)
   ii. Using Support Vector Machine (Polynomial Kernel, degree=2)
  iii. Using Support Vector Machine (RBF Kernel, gamma = 0.5)
  iv. Using Naïve Bayes Classifier

http://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html#sklearn.svm.SVC

```
Accuracy using Fold 1 and Linear SVM 0.894736842105
Accuracy using Fold 2 and Linear SVM 0.929824561404
Accuracy using Fold 3 and Linear SVM 0.947368421053
Accuracy using Fold 4 and Linear SVM 0.964912280702
Accuracy using Fold 5 and Linear SVM 0.964912280702
Accuracy using Fold 6 and Linear SVM 0.964912280702
Accuracy using Fold 7 and Linear SVM 0.982456140351
Accuracy using Fold 8 and Linear SVM 0.929824561404
Accuracy using Fold 9 and Linear SVM 0.947368421053
Accuracy using Fold 10 and Linear SVM 0.964285714286
Mean Accuracy using Linear SVM=0.949060150376
```

```
Accuracy using Fold 1 and Polynomial Kernel: 0.982456140351
Accuracy using Fold 2 and Polynomial Kernel: 0.912280701754
Accuracy using Fold 3 and Polynomial Kernel: 0.947368421053
Accuracy using Fold 4 and Polynomial Kernel: 1.0
Accuracy using Fold 5 and Polynomial Kernel: 0.964912280702
Accuracy using Fold 6 and Polynomial Kernel: 0.947368421053
Accuracy using Fold 7 and Polynomial Kernel: 0.982456140351
Accuracy using Fold 8 and Polynomial Kernel: 0.947368421053
Accuracy using Fold 9 and Polynomial Kernel: 0.947368421053
Accuracy using Fold 10 and Polynomial Kernel: 0.964285714286
Mean Accuracy using Linear SVM=0.959586466165
```

```
Accuracy using Fold 1 and RBF Kernel: 0.19298245614
Accuracy using Fold 2 and RBF Kernel: 0.614035087719
Accuracy using Fold 3 and RBF Kernel: 0.631578947368
Accuracy using Fold 4 and RBF Kernel: 0.508771929825
Accuracy using Fold 5 and RBF Kernel: 0.508771929825
Accuracy using Fold 6 and RBF Kernel: 0.789473684211
Accuracy using Fold 7 and RBF Kernel: 0.719298245614
Accuracy using Fold 8 and RBF Kernel: 0.771929824561
Accuracy using Fold 9 and RBF Kernel: 0.771929824561
Accuracy using Fold 10 and RBF Kernel: 0.767857142857
Mean Accuracy using RBF Kernel =0.627662907268


Accuracy using Fold 1 and Naive Bayes Classifier: 0.894736842105
Accuracy using Fold 2 and Naive Bayes Classifier: 0.859649122807
Accuracy using Fold 3 and Naive Bayes Classifier: 0.912280701754
Accuracy using Fold 4 and Naive Bayes Classifier: 0.929824561404
Accuracy using Fold 5 and Naive Bayes Classifier: 0.964912280702
Accuracy using Fold 6 and Naive Bayes Classifier: 0.964912280702
Accuracy using Fold 7 and Naive Bayes Classifier: 0.982456140351
Accuracy using Fold 8 and Naive Bayes Classifier: 0.964912280702
Accuracy using Fold 9 and Naive Bayes Classifier: 0.964912280702
Accuracy using Fold 10 and Naive Bayes Classifier: 0.964285714286
Mean Accuracy using RBF Kernel =0.940288220551
```