ECE 657A Assignment 2

Submitted By :
Amandeep Kaur (21044104)
Bhupesh Dod (21046099)

Question 2: Naïve Bayes Classifier

Solution: The libraries are imported; the datasets are loaded. Combined the two-wine dataset into a single dataset. Now, after applying standardization on data and split into test and training sets, Dimensionality reduction, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) is performed on Dataset for Feature Extraction. The new versions (PCA and LDA) of dataset and the original raw dataset are used to compare the performance of different classification models.

The Multinomial and Complement Naïve Bayes library used for decision tree is sklearn.naive_bayes.

The Multinomial and Complement Naïve Bayes can't take negative input values so min-max scaler pre-processing is performed on all 6 datasets before applying algorithm.

Naïve Bayes on ABALONE DATASET

It can be seen from the table that accuracy with Complement Naïve Bayes is higher than Multinomial Naïve Bayes on ABALONE dataset and the highest accuracy is 23.47% with LDA components = 5

	Abalone_Raw_MNB	Abalone_PCA_MNB	Abalone_LDA_MNB	Abalone_Raw_CNB	Abalone_PCA_CNB	Abalone_LDA_CNB
With 5 Fold	16.19	16.00	16.00	17.56	20.02	23.53
Without 5 Fold	18.18	17.99	17.99	19.43	19.33	25.07

While the accuracy with KNN (k=73) is 28.32 % in ABALONE RAW dataset, much higher than Naïve Bayes in all cases

Model	**Model/Parameter**	**abalone-raw**	**abalone-pca**	**abalone-lda**
			Principal Components = 3	LDA Components = 5
kNN	K = 73	28.32%	27.17%	27.84%
Multinomial Naïve Bayes		16.19%	16.00%	16.00%
Complement Naïve Bayes		17.56%	20.02%	23.53%

Naïve Bayes on WINE DATASET

It can be seen from the table that accuracy with Multinomial Naïve Bayes is higher than Complement Naïve Bayes on WINE dataset (the trend is opposite as seen in ABALONE case) and the highest accuracy is 47.89% with Raw Wine.

	Wine_Raw_MNB	Wine_PCA_MNB	Wine_LDA_MNB	Wine_Raw_CNB	Wine_PCA_CNB	Wine_LDA_CNB
With 5 Fold	47.89	44.46	44.46	44.40	43.08	42.59
Without 5 Fold	47.26	41.23	41.23	42.83	44.74	44.68

While the accuracy with KNN (k=46) is 68.15% in WINE RAW dataset, much higher than Naïve Bayes in all cases.

Model	**Model/Parameter**	**wine-raw**	**wine-pca**	**wine-lda**
			Principal Components = 10	LDA Components = 5
kNN	K = 46	68.15%	67.38%	68.12%
Multinomial Naïve Bayes		47.89%	44.46%	44.46%
Complement Naïve Bayes		44.40%	43.08%	42.59%

It can be concluded that performance of KNN classifier is better than Naïve Bayes Classifier on both, ABALONE and WINE datasets.