ECE 657A Assignment 2
Submitted By :
Amandeep Kaur (21044104)
## Bhupesh Dod (21046099)

Question 1 : Representation Learning.

Solution : The libraries are imported, the datasets are loaded. Combined the two wine datas into a single dataset. Performed the Exploratory Data Analysis and searched for missing values and no missing values were found. Now, apply standardization on data and split into test and training sets. The best value for knn from assignment 1, k =73, weights = "distance". This value of k will be used for Abalone data and for Wine data, the best value is k=46. These values will be used in order to perform knn on different datasets. For Abalone Raw the accuracy for k =73 is 28.32%. i.e. **for KNN Abalone Raw accuracy 28.32%, at k=73.**

### 1.1 ABALONE DATA

1.1a Abalone PCA

The plot shows two PCA components of Abalone after performing PCA Dimensionality Reduction and using the hue as the Rings variable to distinguish between different datapoints.
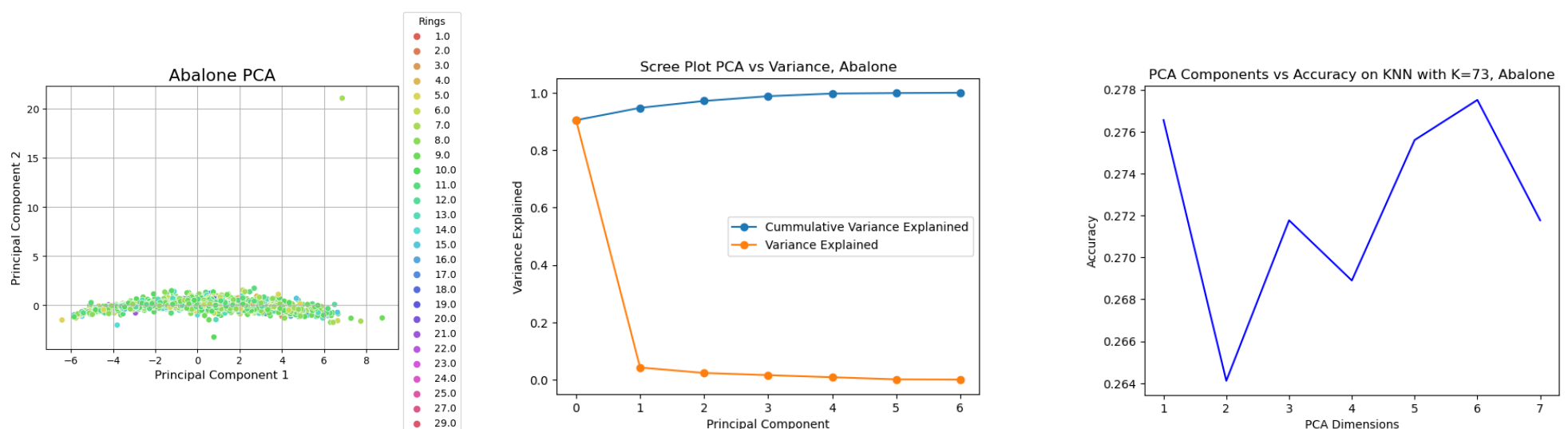
The Scree Plot shows the variance explained and cumulative variance explained represented by the PCA eigenvectors. 95% of the variance is explained by 3 PCA components. Now, KNN is applied on Abalone PCA and the results are as follows. The line graph displays the Accuracy vs Number of PCA Dimensions. The Accuracy of KNN increases as the PCA components are increased however it decreased when the components were increased to 2 from 1. Overall, the best setting of KNN gives the best accuracy of 27.75% at 6 PCA components. It is just slightly up from 27.17% which is at 3 PCA components explaining the 95% of variance.The maximum variance is explained by 6 PCA Components and the accuracy is 27.75%. So, for **KNN Abalone PCA accuracy 27.17%, at k=73.**

1.1b Abalone LDA

The plot shows two LDA components of Abalone after performing LDA Dimensionality Reduction and using the hue as the Rings variable to distinguish between different datapoints.
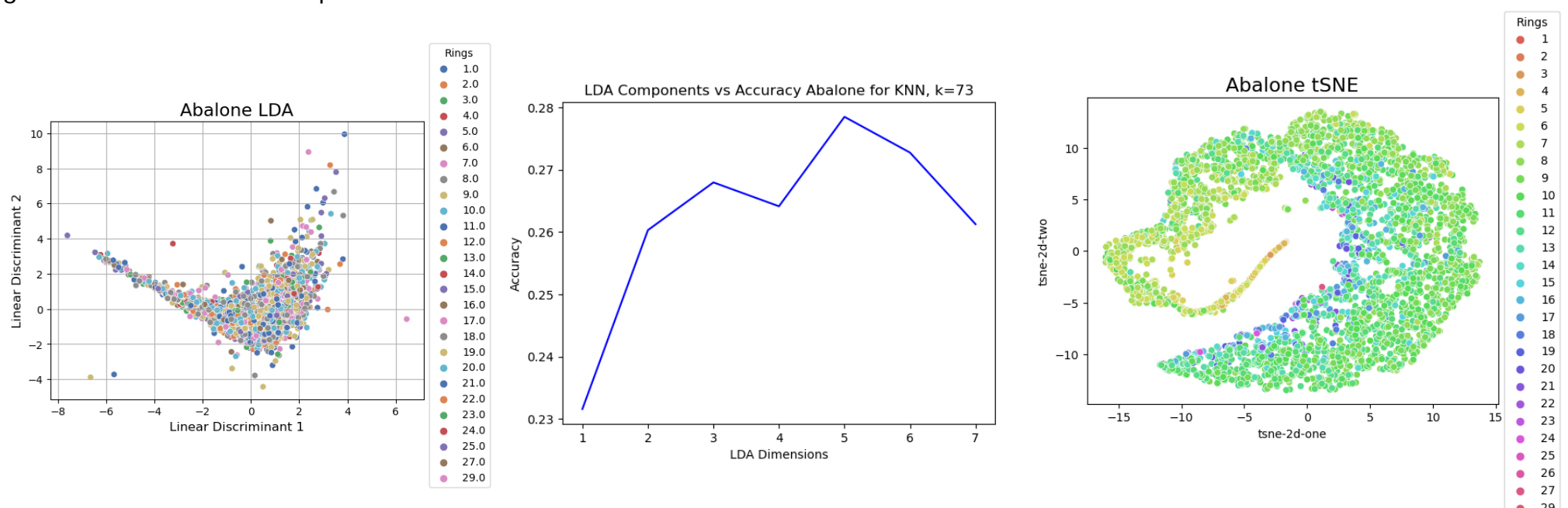
As the LDA Dimensions increase the accuracy increases except for Components = 4. However the best Accuracy for KNN with k =73 is at LDA = 5 components. At LDA = 5, accuracy = 27.84% which is just slightly better than PCA = 2 components. So, for **KNN Abalone LDA accuracy 27.84%, at k =73.**

1.1c t-SNE Abalone

t-SNE is unsupervised dimensionality reduction method. The above graph shows the abalone data after projecting to a lower dimension using t-SNE. As t-SNE maintains the manifold distance, the data in low dimension of t-SNE varies highly from that in PCA and LDA.
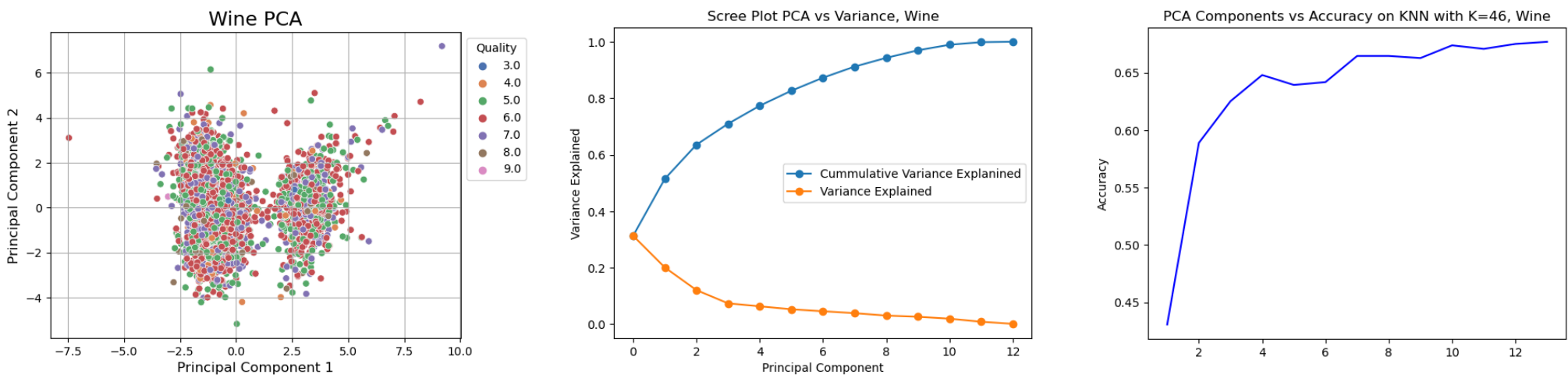
1.2 WINE DATA

1.2a Wine Raw

After checking the best parameters for knn, k = 46 with weights = "distance" gives the best accuracy of 68.15%. **KNN Wine Raw accuracy 68.15%, at k = 46.**

1.2b Wine PCA

The plot shows two PCA components of Wine after performing PCA Dimensionality Reduction and using the hue as the Quality variable to distinguish between different datapoints.

For 95% explained variance, PCA gives 10 components. The Scree Plot shows the variance explained and cumulative variance explained represented by the PCA eigenvectors. 95% of the variance is explained by 10 PCA components. Now, KNN is applied on Wine PCA and the results are in the graph. The graph displays the Accuracy vs Number of PCA Dimensions. The Accuracy of KNN increases as the PCA components are increased. Overall, the best setting of KNN gives the best accuracy of 67.69% at 13 PCA components. However, at the 10 components PCA explaining 95% variance, accuracy is 67.38% which is just slightly less than 13 PCA. The accuracy of Wine PCA has decreased from Wine Raw. **KNN Wine PCA accuracy = 67.38%, at k = 46.**

1.2c Wine LDA

The plot shows two LDA components of Abalone after performing LDA Dimensionality Reduction and using the hue as the Rings variable to distinguish between different datapoints.
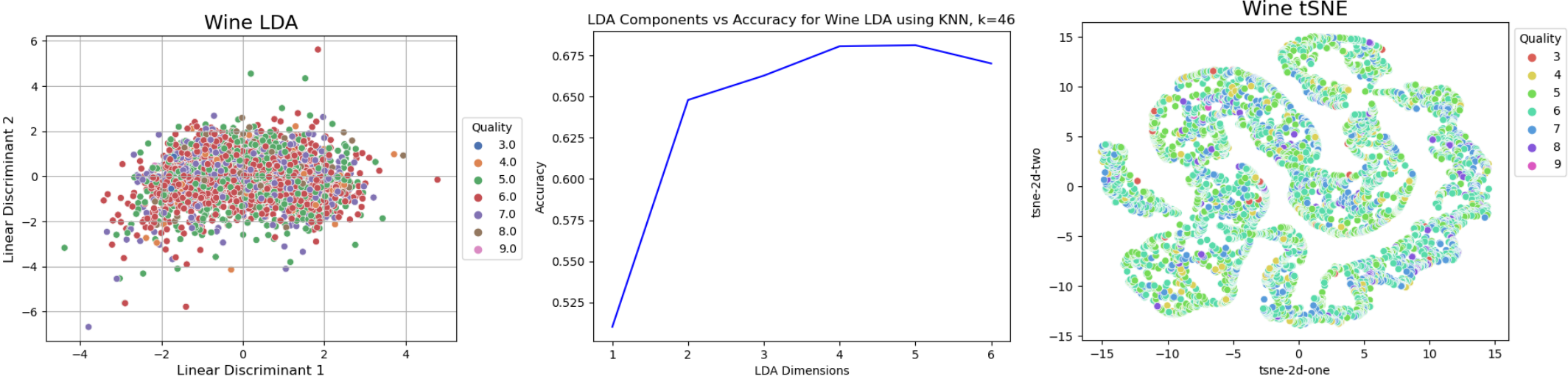
The maximum accuracy is 68.12% at 5 LDA. Which is almost the same as the Wine Raw. **KNN Wine LDA accuracy = 68.12%, at k=46.**

1.2d t-SNE Wine

t-SNE is unsupervised dimensionality reduction method. The above graph shows the wine data after projecting to a lower dimension using t-SNE. As t-SNE maintains the manifold distance, the data in low dimension of t-SNE varies highly from that in PCA and LDA.