Dimensionality Reduction Algorithms:

Principal Component Analysis (PCA) is a popular technique in machine learning. It relies on the fact that many types of vector-space data are compressible, and that compression can be most efficiently achieved by sampling.

PCA is performed for 'Breast cancer' and 'Customer' dataset. The dataset that I decided to consider apart from Breast cancer is having one categorical and three continuous columns. The dataset is divided training and validation.

I have taken four components to show the results of PCA for the first dataset and two components for the second. Following are the interpretations of the PCA outputs:

Breast Cancer:

PCA= [0.46097265 0.17228283 0.09309775 0.0669506]

By the above-mentioned output, the first, second, third and fourth component is responsible for a variation of 49.09%, 17.2%, 9% & 6% respectively. Therefore, we can assume that the first two are the main components making a total of 63.32% (classification percentage).

Customer:

PCA= [0.76000529 0.23999471]

By the above-mentioned output, the first and the second component is responsible for a variation of 76% & 24% respectively. Therefore, we can assume that the first component is the main component as 76% of classification is taken from the first component.

Breast Cancer:

1)Decision tree

Decision Tree can correctly predict 103 out of 114 instances, resulting in 87.71%% accuracy----------DecisionTreeClassifier : classifier.fit [Done]----------DecisionTreeClassifier : classifier.predict [Done]----------> [DecisionTreeClassifier][[64 8][6 36]]('Accuracy', 0.8771929824561403)

2)Knn:

Knn has an accuracy of 94.73% which means it can predict 111 instances correctly out of 114.

[[70 2]

[4 38]]

Accuracy 0.9473684210526315

3)Gaussian:

Gaussian is having an accuracy of 91% that means out of 114, 104 instances are predicted correctly.

[[69 3]

 [7 35]]

Accuracy 0.9122807017543859

4)SVM:

SVC is having an accuracy of 96% which means that out of 114, 110 instances are predicted correctly.

[[72 0]

 [4 38]]

Accuracy 0.9649122807017544

Customer:

Similarly, I had outputs of decision tree, knn, gaussian and svc for customer dataset.

1)Decision tree:

[[18 12]

 [18 12]]

Accuracy 0.5

Out of 90, 45 instances its predicted correctly.

2)KNN:

[[14 16]

 [17 13]]

Accuracy 0.45

Out of 90, 41 instances are predicted correctly.

3)Gaussian:

[[26 4]

 [20 10]]

Accuracy 0.6

Out of 90, 54 instances are predicted correctly.

4)SVM:

[[30 0]

 [30 0]]

Accuracy 0.5

Out of 90, 45 instances are predicted correctly.