

IML REPORT LAB -7

B21ME047

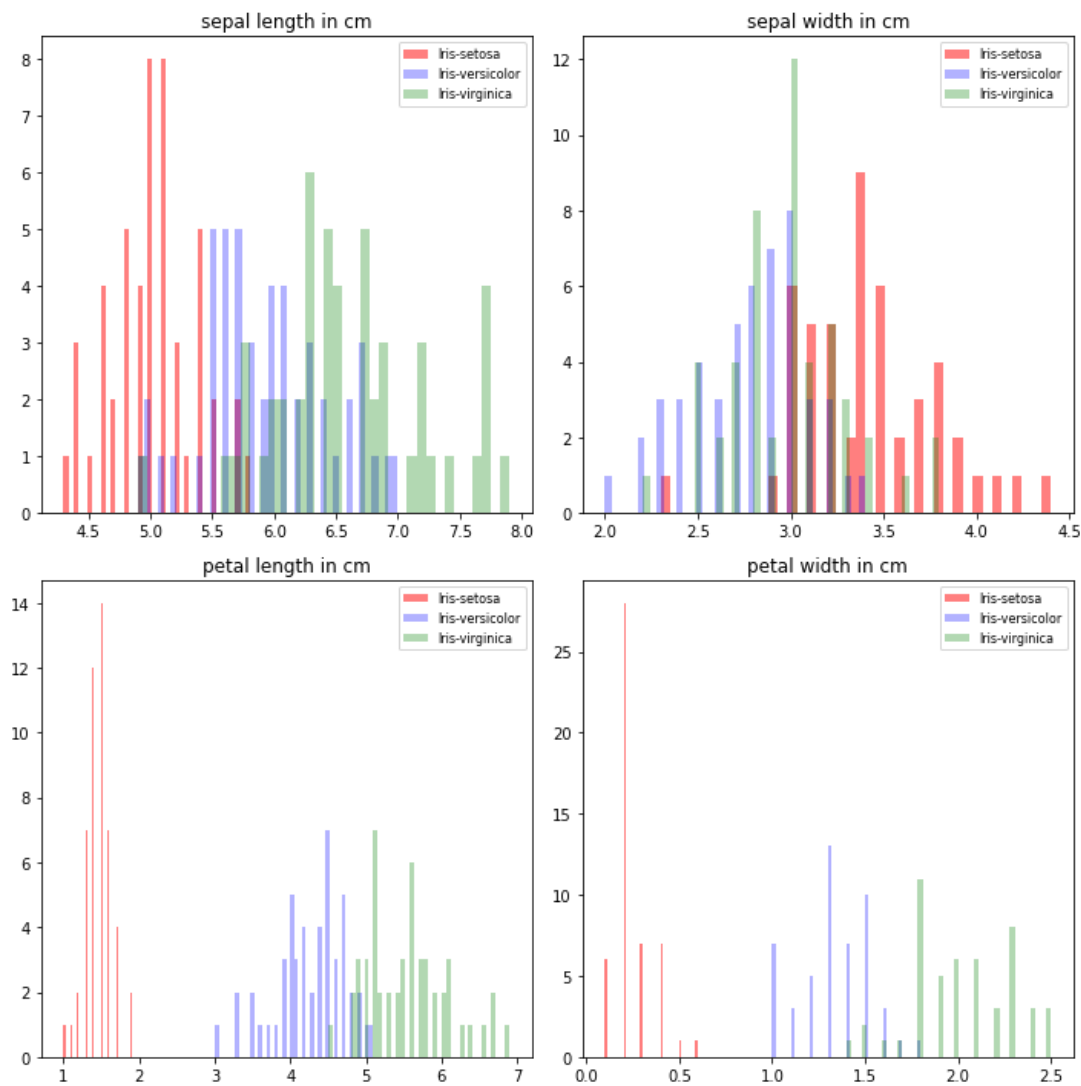
PRANJAL VERMA

Q.1

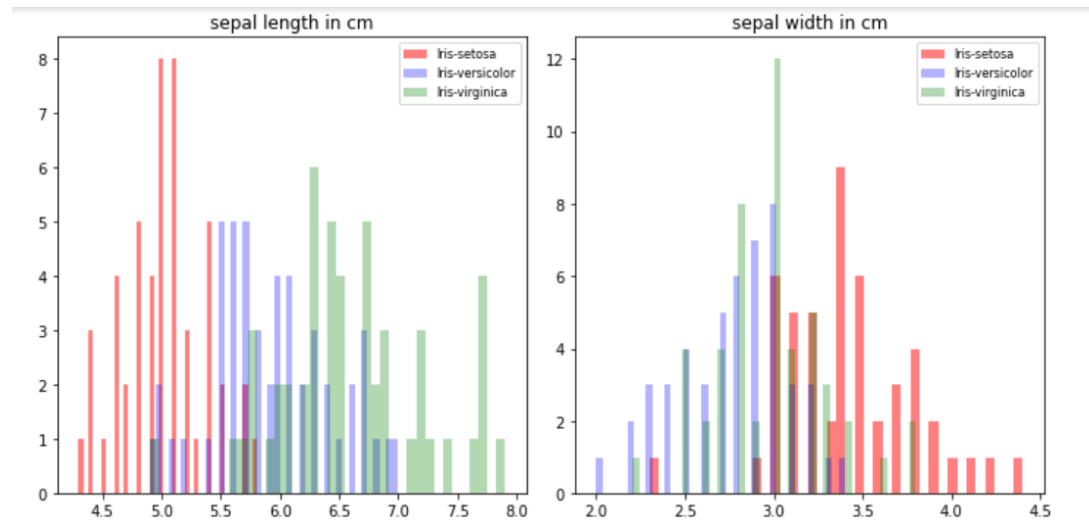
I first checked for all the null values in the data set and found none.

I then split the iris database into data and target consisting of species column and then split the data train and test set in the ratio of 80:20.

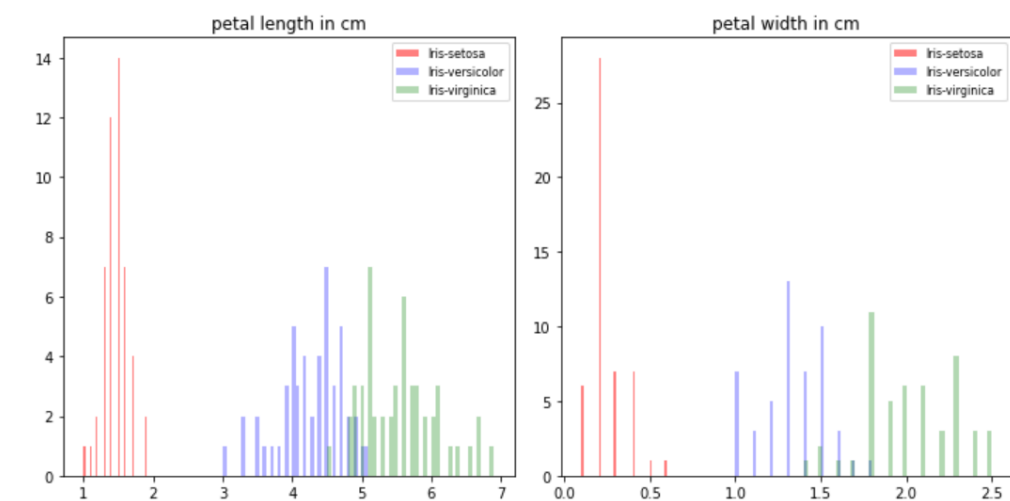
I looked for relationship between various features and the target column and plotted a graph.



We see that for columns "'sepal length in cm', 'sepal width in cm' target variable is less likely to be determined due to overlapping.



We see that for columns 'petal length in cm' , 'petal width in cm' target variable can be determined accurately due to no overlapping so these can be used as dimensions in PCA with high accuracy.



[using-principal-component-analysis-pca-for-machine-learning](#)

Q.2

I used KNN classifier to predict accuracy on the test data and it came out to be 90%.

Q.3

I applied PCA algorithm on the given features with total components $n = 3$ and transformed the test and trained data.

<https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>

library used :

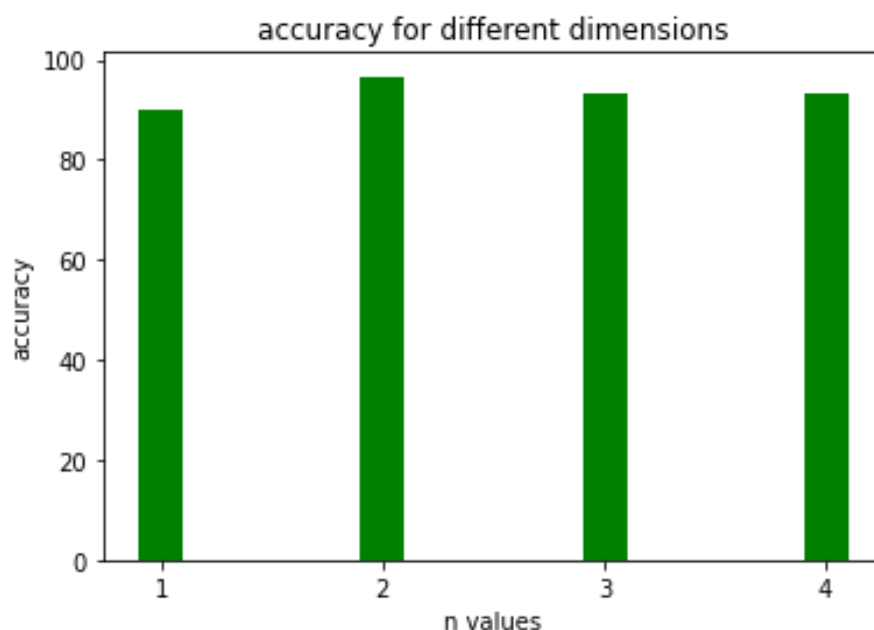
```
from sklearn import decomposition
from sklearn import datasets
```

Q.4

I again used the kNN classifier but this time on the transformed dataset from the PCA and its accuracy came out to be 86.66%.

Q.5

I examined the accuracy for different values of projection dimension and the made a histogram to make a comparison.



We can see that accuracy is maximum for $n = 2$ with 96% accuracy.