ML ASSIGNMENT_2 REPORT

Q-1

As it was mentioned in the questionnaire that the model has to design without using any libraries.
The same has been implemented and with the help of python.

Model was trained by a self defined function and its overall accuracy was found to be 97%.

Class Wise accuracy:
setosa-100%
versicolor-92%
virginica-100%

Confusion matrix:

actual
Setosa- [[19 0 0]
versicolor-[0 12 1]
Verginica- [0 0 13]]

Q-2

As per the question requirements.Class Wise plot were generated for both train and test set.

I have dropped a few variables to have visibility in plot.

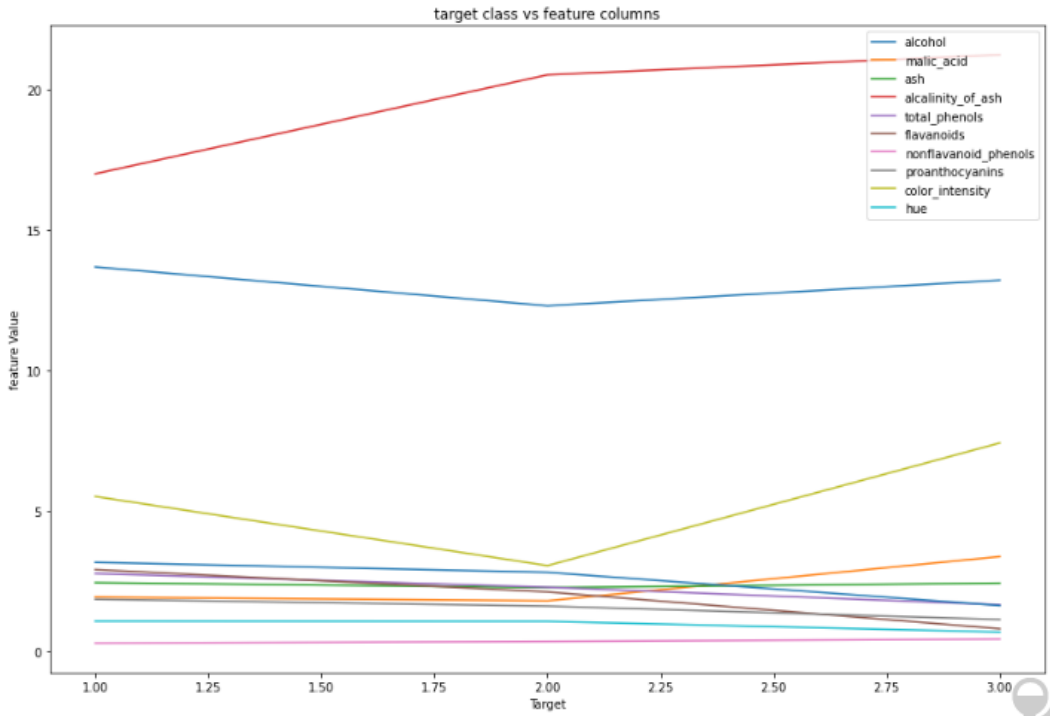
Observation from both the plot:

1-except alcalinity_ash,alcohol,color_intensity all other feature values are almost the same in all the target classes.

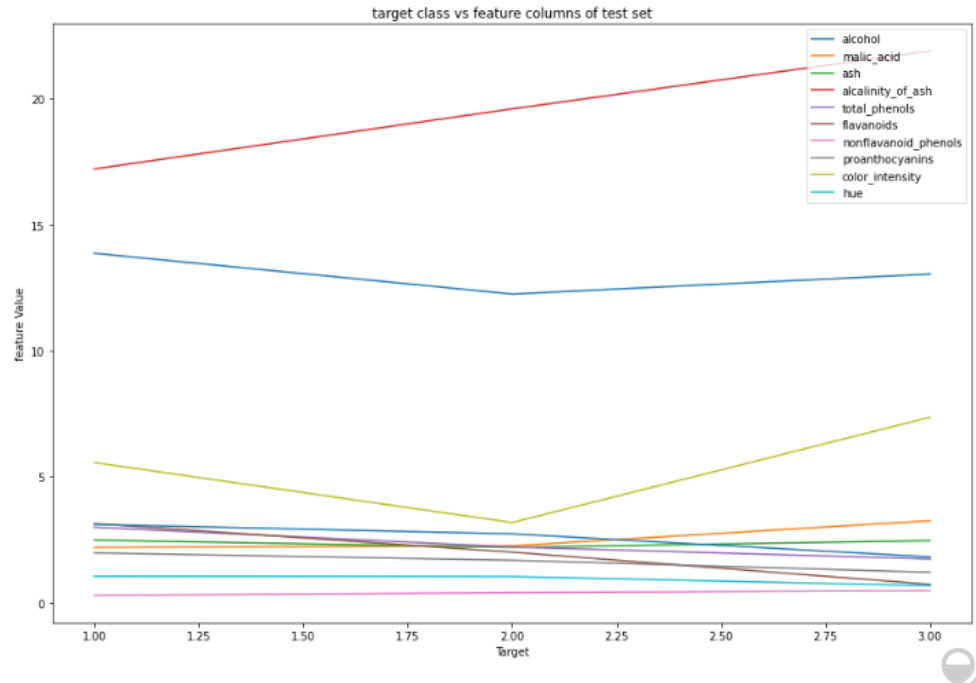
2-alcohol and color_intensity decrease in class 2.

3-alcalinity_ash remains the same in both class 2 and 3 which is more than class 1.

Train set line plot:



Test set line plot:



Model was trained according to given dataset and following output were generated:

Accuracy:

Gaussian Naive Bayes model accuracy(in %): 94.4

Confusion matrix:

actual

Setosa -[18, 1, 0],
versicolor- [0, 19, 2],
verginica- [0, 0, 14]]

Class Wise priors:

Target
setosa 0.317073
versicolor 0.406504
Virginica 0.276423

Class Wise mean and variance:

mean and standard deviation for each feature grouped by Target

```
In [18]: X_train_mean=X_train.groupby(y_train).apply(np.mean)
         X_train_std=X_train.groupby(y_train).apply(np.std)
```

```
In [19]: X_train_mean
```

Out[19]:

Target	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flavanoids	nonflavanoid_phenols	proanthocyanins	color_intensity	h
1	13.673590	1.928974	2.441795	16.989744	104.282051	2.76641	2.904872	0.289231	1.851026	5.508718	1.066
2	12.292000	1.796200	2.265400	20.508000	94.740000	2.27860	2.112000	0.349200	1.607200	3.047000	1.064
3	13.200882	3.367353	2.425000	21.220588	99.617647	1.65500	0.802647	0.433529	1.131765	7.410882	0.685

```
In [20]: X_train_std
```

Out[20]:

Target	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flavanoids	nonflavanoid_phenols	proanthocyanins	color_intensity	hu
1	0.451411	0.610484	0.220650	2.216580	9.575473	0.287058	0.384413	0.061243	0.438610	1.297527	0.1092
2	0.504218	0.857713	0.315773	3.592650	16.882903	0.546553	0.721557	0.119597	0.511316	0.820430	0.1806
3	0.569052	1.043209	0.166455	2.159763	11.762389	0.365596	0.303366	0.121241	0.388250	2.186514	0.1087

B-Part:

Model was trained first in the ratio of 40:40:20 and than in 80:10:10

It was found that in the case of a 40:40:20 trained model, accuracy was the same as it was without ration split. I.e **Accuracy** is 94.4% and it has the same confusion matrix as it was with part-A.

Confusion matrix:

```
                  actual
Setosa       -[18,  1,  0],
versicolor-   [ 0, 19,  2],
verginica-    [ 0,  0, 14]]
```

But when we use 80:10:10 ration , accuracy was slightly decreased and same case with confusion matrix aslo.

Accuracy in this case is 92.59

Confusion matrix:\

```
                  actual
Setosa     [[18,  1,  0],
versicolor  [ 1, 18,  2],
Verginica   [ 0,  0, 14]]
```

Thank you