

CS5101 Machine Learning Lab (Aug - Dec 2021)

Assignment 4

(Classification)

Rules:

1. The assignment should be submitted via moodle in .ipynb format.
2. Single notebook should be submitted including all solutions
3. **Plagiarism will be checked and Zero marks will be given if found.**
4. Your code should be neat and well commented
5. Maximum marks for this assignment are 5.
6. Deadline for submission is 11-Sep 23:00 IST

Questions:

1. You are provided with a three class dataset. Learn a Gaussian Naive Bayes classifier (you can use scikit learn) and report the results with observation as mentioned below. Assume data follows gaussian distribution. Data points are two dimensional.
 - Your code should input train and test data from each of the corresponding files and learn a Gaussian Naive Bayes classifier. Class1_train implies train data with label class 1 and similarly for other files.
 - Give following outputs in the python notebook itself with proper headings.
 - Decision region plots learnt by the classifier (plots should have proper labellings and title)
 - Accuracy values for train and test data
 - Confusion matrix for train and test data [2 marks]
 - At last in a text block write your observations about the decision surface obtained.

2. Write Logistic Regression from scratch and do the following.

- Use the following code to generate the dataset:-

```
from sklearn.datasets import make_blobs
X, y = make_blobs(n_samples=100, centers=2, n_features=2, random_state=0)
```

- Give following outputs in the python notebook itself with proper headings.
 - Decision region plots learnt by the classifier (plots should have proper labellings and title)
 - Accuracy values for train and test data
 - Confusion matrix for train and test data [3 marks]
- At last in a text block write your observations about the decision surface obtained.