Department of Computer Science & Engineering
Final Year B. Tech. (CSE) – I : 2021-22
4CS462 : PE2 - Data Mining Lab

Assignment No. 6
By DM21G03
(2018BTECS00082 : Hritik Belani , 2019BTECS00209 : Sailee Akim)

Date: 08/10/2021

- **Title**
  Data Analysis Tool

- **Objective**
  Designing and Implementing Classifiers

- **Specification**
  Python: 3+
  Dataset: Iris, Breast Cancer

- **Introduction & Theory**

  Regression Classifier:
  Logistic Regression is a supervised learning algorithm that is used when the target variable is categorical. Hypothetical function h(x) of linear regression predicts unbounded values. But in the case of Logistic Regression, where the target variable is categorical, we have to strict the range of predicted values. Consider a classification problem, where we need to classify whether an email is a spam or not. So, the hypothetical function of linear regression could not be used here to predict as it predicts unbound values, but we have to predict either 0 or 1.

  Naïve Bayesian:
  Naive Bayes is among one of the very simple and powerful algorithms for classification based on **Bayes Theorem** with an assumption of independence among the predictors. The Naive Bayes classifier assumes that the presence of a feature in a class is not related to any other feature. Naive Bayes is a classification algorithm for binary and multi-class classification problems.

  K-nn:
  K Nearest Neighbors Classification is one of the classification techniques based on instance-based learning. Models based on instance-based learning to generalize beyond the training examples. To do so, they store the training
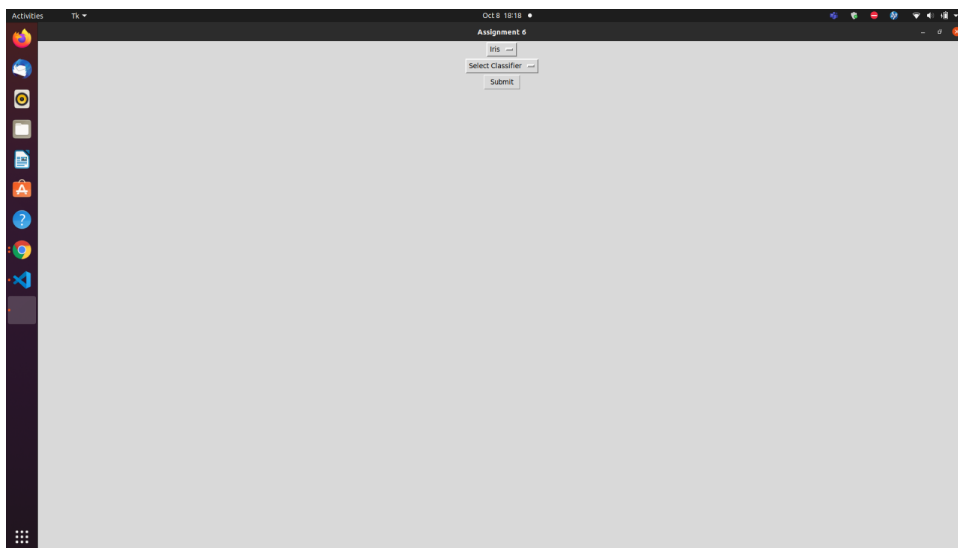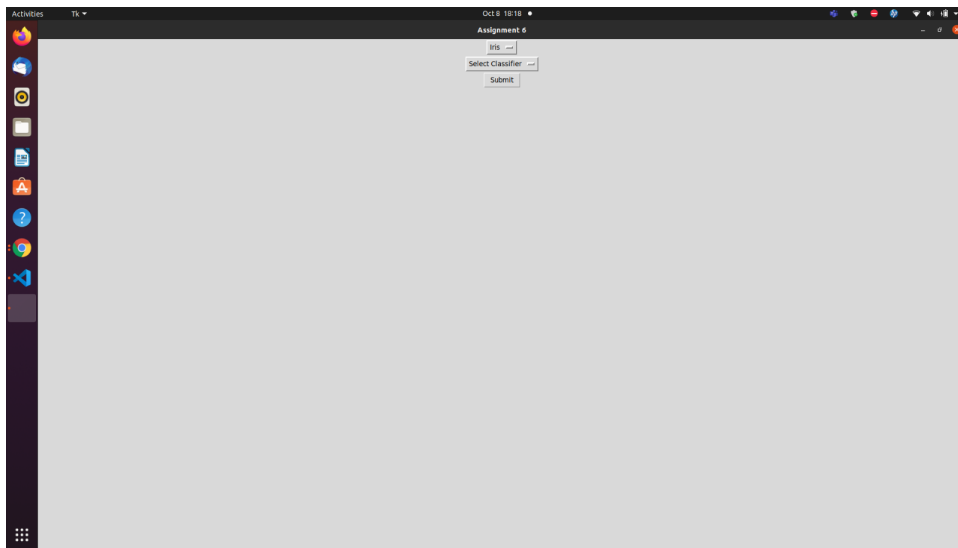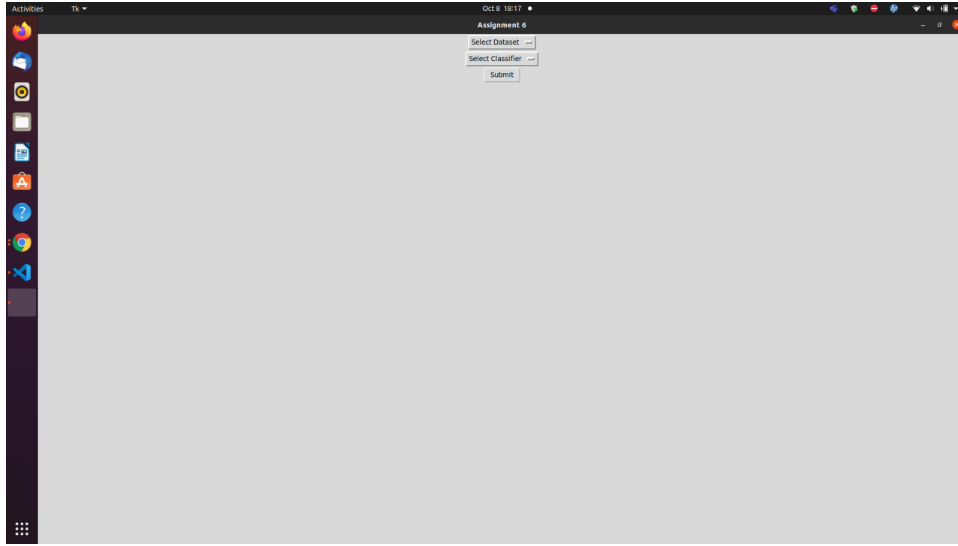
examples first. When it encounters a new instance (or test example), then they instantly build a relationship between stored training examples and this new instant to assign a target function value for this new instance. Instance-based methods are sometimes called lazy learning methods because they postponed learning until the new instance is encountered for prediction.
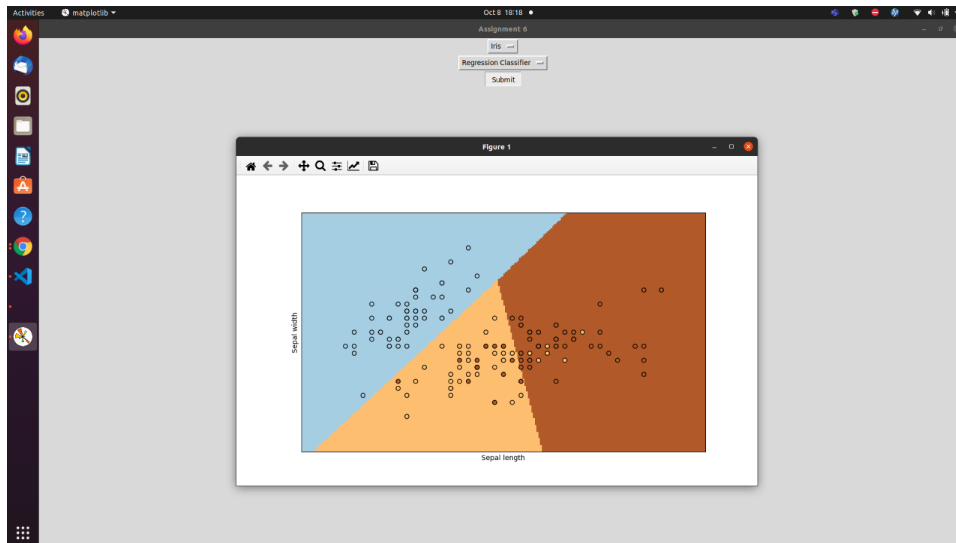
Ann:

ANN Classification is an example of **Supervised Learning**. Known class labels help indicate whether the system is performing correctly or not. This information can be used to indicate a desired response, validate the accuracy of the system, or be used to help the system learn to behave correctly. The known class labels can be thought of as *supervising* the learning process; the term is not meant to imply that you have some sort of interventionist role.

- **Procedure**

  1. Design and implement the following classifiers:
       a) Regression classifier.
       b) Naïve Bayesian Classifier.
       c) k-NN classifier (Take k = 1,3,5,7)
       d) Three-layer Artificial Neural Network (ANN) classifier (use back propagation). Plot error graph (iteration vs error).
  2. Tabulate the results in confusion matrix and evaluate the performance of above classifier using following metrics:
       a) Recognition rate
       b) Misclassification rate
       c) Sensitivity d) Specificity
       e) Precision & Recall
  3. Use the following data sets from UCI machine learning repository :
       a) IRIS
       b) Breast Cancer

Assignment 6

Select Dataset ⌄

Select Classifier ⌄

Submit

Assignment 6

Iris ⌄

Select Classifier ⌄

Submit

Assignment 6

Iris ⌄

Select Classifier ⌄

Submit

Figure 1 (Regression Classifier)

Iris ▾

Regression Classifier ▾

Submit

Assignment 6

Iris ▾

Naive Bayesian Classifier ▾

Submit

Accuracy:0.9555555555555556

Confusion Matrix:[[12 0 0]
 [0 16 1]
 [0 1 15]]

	precision	recall	f1-score	support
0	1.00	1.00	1.00	12
1	0.94	0.94	0.94	17
2	0.94	0.94	0.94	16
accuracy			0.96	45
macro avg	0.96	0.96	0.96	45
weighted avg	0.96	0.96	0.96	45

Assignment 6

Iris ▾

k-nn Classifier ▾

Submit

Accuracy:0.9666666666666667

Confusion Matrix:[[16 0 0]
 [0 4 1]
 [0 0 9]]

	precision	recall	f1-score	support
0	1.00	1.00	1.00	16
1	1.00	0.80	0.89	5
2	0.90	1.00	0.95	9
accuracy			0.97	30
macro avg	0.97	0.93	0.95	30
weighted avg	0.97	0.97	0.97	30

- **Conclusion**

  Thus, studied and implemented different classifiers such as Regression classifier, Naïve Bayesian classifier, k-nn classifier and artificial neural network classifier in Python from scratch.

- **Reference**
  - **https://www.geeksforgeeks.org/implementation-of-k-nearest-neighbors-from-scratch-using-python/**
  - **https://www.geeksforgeeks.org/ml-naive-bayes-scratch-implementation-using-python/**
  - **https://www.geeksforgeeks.org/implementation-of-logistic-regression-from-scratch-using-python/**