

# CS670/470 Individual Assignment 2

---

UMass-Boston

3/21/2019

## 1 Educational Goal

Preliminary understanding of classification workflow in machine learning; using the state-of-the-arts machine learning libraries to implement the Naive Bayes method.

## 2 Details

### Project Objectives:

1. Get a head start on how to use existing machine learning libraries to do classification using the Gaussian Naive Bayes.
2. Familiar with an online team-based collaborative platform.

**Due Date:** 2:00 pm, 3/28/2019

**Programming language:** Python.

**Package and framework:** scikit-learn, colab.

## 3 Background of Scikit-Learn & Colaboratory & Gaussian Naive Bayes

**Scikit-Learn, or Sklearn**, offers tools for machine learning and data analysis in Python including implementations of many of the algorithms that will be discussed in this course. More details at <https://scikit-learn.org/stable/>

**Colaboratory, or colab**, is a free Jupyter notebook environment that requires no setup and runs entirely in the cloud. With Colaboratory you can write and execute code, save and share your analyses, and access powerful computing resources, all for free from your browser. More details at <https://colab.research.google.com/notebooks/welcome.ipynb>

**Classifier – Gaussian Naive Bayes:** [https://en.wikipedia.org/wiki/Naive\\_Bayes\\_classifier#Gaussian\\_naive\\_Bayes](https://en.wikipedia.org/wiki/Naive_Bayes_classifier#Gaussian_naive_Bayes)

## 4 Data

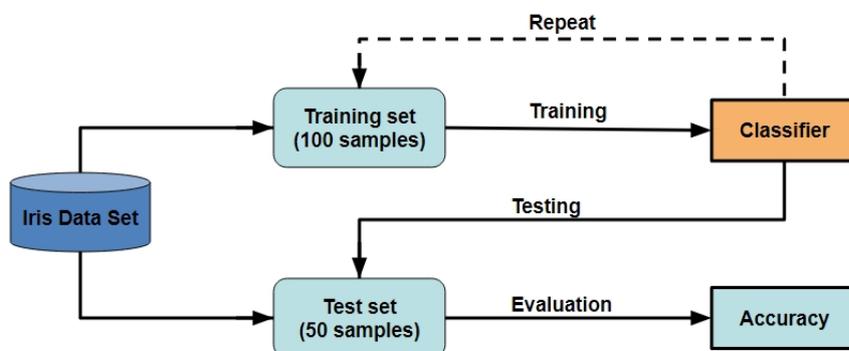
Table 1: Iris Data Set.

Data Set Characteristics: Multivariate	Number of Instances: 150
Associated Tasks: Classification	Number of Features: 4
Class labels of the Iris Data set:	Setosa, Versicolour, and Virginica.
Features of the Iris Data set (4 features):	Sepal Length, Sepal Width, Petal Length and Petal Width.
Wikipedia:	<a href="https://en.wikipedia.org/wiki/Iris_flower_data_set">https://en.wikipedia.org/wiki/Iris_flower_data_set</a>

### 4.1 Class label



## 5 Approach



**Step 1. Load iris data and construct the training set, test set:** Use "`sklearn.datasets.load_iris`" to load the iris data set. Then split it into 2 parts (100 samples for training and 50 for testing) for your experiments.

**Step 2. Create your Gaussian Naive Bayes classifier and train it on the training set:** Create your Gaussian Naive Bayes classifier using the sk-learn libraries and use the training set you constructed in Step 1 to train your classifier.

**Step 3. Test your classifier on the test set and get your evaluation results:** After training, use the test set to evaluate your classifier's performance (accuracy and F-1 score).

**Step 4. ROC curve:** Depict the ROC curve to show your classification results.

## **6 Submission Requirements**

1. Share your colab notebook page with [teaching.yong@gmail.com](mailto:teaching.yong@gmail.com).
2. Write an experiment report to discuss your experimental results.
3. Only soft copy is required. Submit the soft copy of the report through your UMassOnline account.