

CS670/470 Individual Assignment 1

UMass-Boston

9/12/2017

1 Educational Goal

Preliminary understanding of classification workflow in machine learning.

2 Detail

Project goal: Get a head start on how to use existing machine learning libraries to do classification.

Due Date: 4:00 pm, September 21, 2017

Programming language: Python.

Package: scikit-learn.

3 Scikit-Learn

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. If you are installing python for this first time, it comes packaged along with many other libraries, features, and an IDE (Spyder) in the Anaconda scientific distribution of Python, available free at <https://www.anaconda.com/download/>. It can be installed to an existing Python environment using pip, with dependencies on numpy and scipy. Detailed instructions and documentation are available on the sklearn site, <http://scikit-learn.org/stable/install.html>.

Iris Data Set:

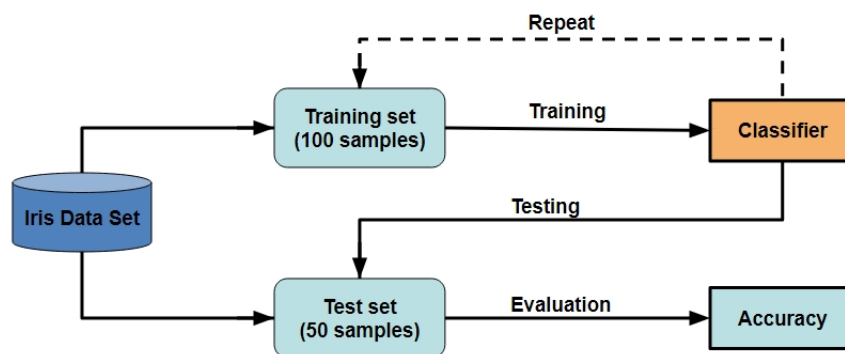
4 Class label



Table 1: Iris Data Set.

Data Set Characteristics:	Multivariate	Number of Instances:	150
Associated Tasks:	Classification	Number of Attributes:	4
Types of irises:	Setosa, Versicolour, and Virginica.		
Features of irises:	Sepal Length, Sepal Width, Petal Length and Petal Width.		
Wikipedia:	https://en.wikipedia.org/wiki/Iris_flower_data_set		

5 Approach



Step 1. Load iris data and construct the training set, testing 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 KNN classifier and train it on the training set: Create your KNN classifier and use the training set you got 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).

Step 4. (Optional): Use any plot you want to show your classification results. (bonus points possible!)

6 Classifier

K-Nearest Neighbor: <http://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>

7 Submission Requirements

1. Brief description of what is classification in machine learning.
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.