Data Science
Lab Exercise (Decision Tree)
Prepared By
Dr Muhammad Atif Tahir (Spring 2018)

I. In this lab, you are going to learn how to classify data points using decision tree classifier. Iris data set is given which consists of 3 classes and 150 data points.

(a) Load data set using pandas library

(b) Now, divide your data using hold out approach (80% for training and 20% for testing)

(c) Apply decision tree classifier. See the documentation below. You need to import necessary classes. Print Accuracy, Confusion Matrix and Classification Report

http://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html#sklearn.tree.DecisionTreeClassifier

```
Accuracy=0.9
[[ 7  0  0]
 [ 0 11  1]
 [ 0  2  9]]
             precision    recall  f1-score   support

Iris-setosa       1.00      1.00      1.00         7
Iris-versicolor       0.85      0.92      0.88        12
Iris-virginica       0.90      0.82      0.86        11

avg / total       0.90      0.90      0.90        30
```

(d) Repeat (c) by changing the value of depth (depth=1, 2, 3, 4, 5). Print only accuracy

```
Depth=1, Accuracy= 0.6
Depth=2, Accuracy= 0.866666666667
Depth=3, Accuracy= 0.9
Depth=4, Accuracy= 0.866666666667
Depth=5, Accuracy= 0.866666666667
```

II. Repeat (I) using Occupancy Detection dataset. Ignore Date Attribute. Off course, steps (d) and (g) are not applicable since training / test data is given.
http://archive.ics.uci.edu/ml/datasets/Occupancy+Detection+