HW7

CISC648010 - Fall 2021

Due Date: Nov 30th at 11 PM

1 Decision Tree and Boosting 10 pts each part

This problem **is not** a programming assignment. Consider the following detaset consists of 6 data points :

PATIENT ID	CHEST PAIN?	Healthy Diet?	SMOKES?	EXERCISES?	HEART ATTACK?
1.	yes	yes	no	yes	yes
2.	yes	yes	yes	no	yes
3.	no	no	yes	no	yes
4.	no	yes	no	yes	no
5.	yes	no	yes	yes	yes
6.	no	yes	yes	yes	no

a) Using the above data set, train a decision tree to predict whether a patient is likely to have a heart attack or not. Use the Entropy function as the impurity measure. Grow the tree until the training error becomes zero.
b) Use the above dataset and Adaboost with $T = 2$ to train a classifier. Use the decision stump as the weak learner. Report the value of α_1 and α_2 and decision stump f_1 and f_2.

2 EM algorithm 20 pts

This problem is a programming assignment. Download dataset EM.csv from canvas. Run the following lines of code to load the dataset:

```
import csv
import numpy as np
import matplotlib.pyplot as plt
with open('EM.csv') as csv_file:
    csv_reader = csv.reader(csv_file, delimiter=',', quoting=csv.QUOTE_NONNUMERIC)
    data1 = [ ]
    for row in csv_reader:
        data1.append(row)
data1 = np.array(data1)
```

The dataset includes 1000 feature vectors (the dimension of each feature vector is 2) generated by a Gaussian Mixture Model with $K = 3$. Use the EM algorithm with 50 iterations to estimate the parameters of the Gaussian Mixture Model. You should report $\{\phi_j, \mu_j, \Sigma_j\}_{j=1,2,3}$. Upload your code on canvas to receive the full credit.