CENG 499

Introduction to Machine Learning

Spring 2019-2020

Homework 4: Naive Bayes and Hidden Markov Models

version 1.0

Due date: 12 June 2020, 23:59

1   Introduction

In this assignment, you will have the chance to get hands-on experience with the naive Bayes model and hidden Markov models (HMM). Python is the programming language choice for this homework. Only Python ≥ 3.0 is allowed.

2   Naive Bayes (60 pts)

In this section, you are going to implement a naive Bayes classifier. You will write your code in the function **naive_bayes** in **task1.py**. When grading, we will call the function with different parameters.

- The dataset is a CSV file where each line is an instance. And, the last column is the label of an instance.

- No external library is allowed for this task.

- For the provided dataset, the accuracy value you should achieve is 88,15.

- Since you need to know the number of features and the possible values of a feature to compute the probabilities, you will iterate over the dataset for those values. Because a value of the feature may be present in one set but not present in the other set, you need to iterate over both training and test sets to get the values right.

- The time limit for this task is 1 minute.

3   Hidden Markov Models (40 pts)

In this section, you are going to implement the forward and Viterbi algorithms to solve evaluation and decoding tasks of HMMs. You will write your code in the functions **forward** and **viterbi** in **task2.py**. To test your implementation, you can use **task2_runner.py**.

- As an external library, only NumPy is allowed.

- The provided data consists of NumPy arrays having np.float64 items. In your implementations, do continue to use np.float64 precision.

- The limit is for this task is 1 minute (running task2_runner with all provided data).

- As a reference for HMMs, you can check this **link**.

4   Specifications

- If you exceed the time limit, you will get no points from that task.

- Falsifying results, changing the composition of training and test data are strictly forbidden, and you will receive 0 if this is the case. Your programs will be examined to see if you have actually reached the results and if they are working correctly.

- Using any piece of code that is not your own is strictly forbidden and constitutes as cheating. This includes friends, previous homeworks, or the internet. The violators will be punished according to the department regulations.

- Follow the course page on ODTUClass for any updates and clarifications. Please ask your questions on the discussion forum of ODTUClass instead of e-mailing.

- You have total of 3 late days for **all** your homeworks. For each day you have submitted late, you will lose 10 points. The homeworks you submit late after your total late days have exceeded 3 will not be graded.

5   Submission

Submission will be done via ODTUClass. You will submit a zip file called **hw4.zip** that contains **task1.py** and **task2.py**.