Introduction to Data Analysis and Machine Learning with Python

Homework 7

Sat Apr 27 2024
Due: May 6 2024

Download the datafile "data_HW7.txt" from the course website (under Session 8 Materials). Read the data file in Python. The datafile has three features ("x1","x2" and "x3") for objects of two classes (0 or 1). For each class, there are 200 objects in the data file.
Set up an MLPClassifier:

```
mlp = MLPClassifier(solver='adam', hidden_layer_sizes=(5,5), max_iter=5000)
```

(It is important to use exactly these parameters)
Repeat the following sequence of steps a number of times (e.g, 100 trials):

- Use `train_test_split()` to split the data into 50% training data and 50% test data
- Use `mlp.fit()` to train the model
- Predict the target values for both the train and test data
- Calculate the accuracy (using `metrics.accuracy_score()`) of the prediction for both the train and test data

You will find that the accuracy varies from trial to trial!

1. What is the average accuracy for the test sample?
    a. About 53%
    b. About 63%
    c. About 73%
    d. About 83%
    e. About 93%

2. What is the average accuracy for the training sample?
    a. About 56%
    b. About 66%
    c. About 76%
    d. About 86%
    e. About 96%

3. Investigate the origin of the fluctuations of the accuracy result. Is it
    a. Due to the train_test_split() step
    b. Due to mlp.fit() step
    c. Due to both

4. Change the hidden_layer_sizes from (5,5) to (20,20) and max_iter from 5000 to 50000 and repeat the 100 trials. What do you observe for the accuracies?
    a. Both stay the same
    b. Training set accuracy goes up, test set accuracy goes up
    c. Training set accuracy goes down, test set accuracy goes up
    d. Training set accuracy goes up, test set accuracy goes down

e. Training set accuracy goes up, test set accuracy stays the same

5. Try to install the PyTorch framework (https://pytorch.org) on your computer (ideally, such that you can run it in a jupyter notebook). See https://pytorch.org/get-started/locally/ . Verify the installation by running:

```
import torch
print(torch.__version__)
```