CS 165B - Machine Learning, Winter 2023

Assignment #4 | Lead TA: Deepak Nathani
Due Wednesday, March 15th by 10:00am (PDT)

Notes:

- **This assignment is to be done individually.** You may discuss the problems at a general level with others in the class (e.g., about the concepts underlying the question, or what lecture or reading material may be relevant), but the work you turn in must be solely your own.
- Be aware of the late policy in the course syllabus – i.e., *you only have four late days for all the assignments*, so **it is your responsibility to turn in your assignment to GradeScope and CodaLab by the due time**.
- Any updates or corrections will be posted on the Assignments page (of the course web site) and the Piazza, so check there occasionally.
- **All assignments must be clear and legible.** It is recommended to type the solutions on this PDF directly. If you'll be submitting a handwritten assignment, please ensure that it's readable and neat. If your writing is not easily readable, your solution is not easy to follow on the page, or your PDF is not of very high quality, your assignment will not be graded. DO NOT submit picture of your written work. (If you must scan your written work in, use a high-quality scanner. Plan in advance.)
- Be sure to re-read the "**Academic Integrity**" on the course syllabus. You must complete the section below. **If you answered Yes to either of the following two questions, give corresponding full details.**

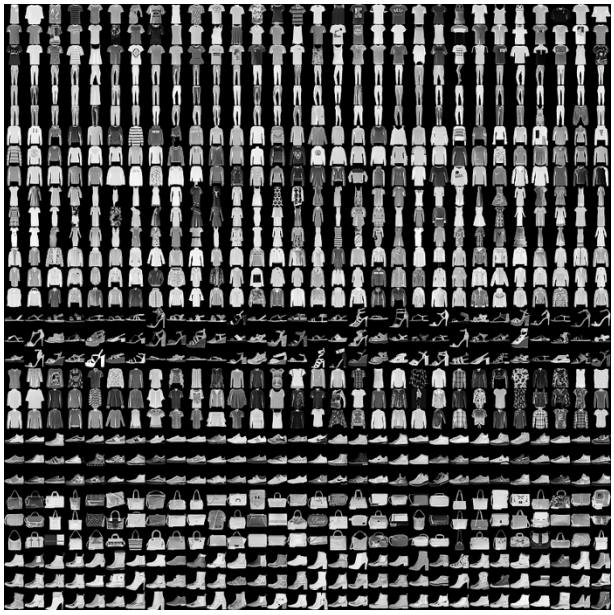
Did you receive any help whatsoever from anyone in solving this assignment? □ Yes □ No

Did you give any help whatsoever to anyone in solving this assignment? □ Yes □ No

---

Programming Assignment [100 points]

The artificial intelligence revolution has arrived, and it should not worry anyone. AI is changing jobs (not replacing them) and augmenting human tasks in consumer goods, manufacturing, customer service and many more verticals. In retail and fashion, the biggest opportunities are in trend forecasting and better supply chain management. In order to utilize AI systems to track the trending of clothing, the computer scientists need to design machine learning algorithms to recognize different clothing objects.



Your task is to implement a machine learning algorithm that predicts the labels, given the image data. There are no constraints on the types of machine learning models.

Please read the following notes before you dive into the assignment.

1. Dataset

The dataset, named as "hw4_train.zip" can be downloaded from the Resource page in Piazza. After decompressing the .zip file, there exist 10 folders named from "0" to "9". Each training and testing image is assigned to one of the following labels, and each folder contains all the training examples for the corresponding label.

Label	Description
0	T-shirt/top
1	Trouser
2	Pullover
3	Dress
4	Coat
5	Sandal
6	Shirt
7	Sneaker
8	Bag
9	Ankle boot

Each (28 pixels*28 pixels) image can be viewed as a 28*28 2-dimensional matrix, where each element in this matrix is an integer within 0~255.

We evaluate your machine learning model using the separate testing dataset. You can download the images of the testing dataset and upload your prediction results to the CodaLab. However, you cannot access the labels of the testing dataset.

2. Requirements

(a) There are no constraints on the types of machine learning models.

(b) There are no constraints on the python libraries you use
However, if you choose to use PyTorch or TensorFlow in this assignment,
we have the version constraints for these two libraries:
For PyTorch, the version you use must >= 1.8.2;
For TensorFlow, the version you use must >= 2.0;

(c) You are not allowed to use other external data except the released dataset. Also, you cannot add more data hand-labeled by yourself when training your machine learning algorithm. **Remember the NO other data rule!**

(d) Please carefully check the "hw4_starter_package.zip" and read the comments in "prediction.py".

3. Submission Instructions

For the assignment #4, you are required to submit **both the prediction file and

source code.**

3.1 Source Code Submission

(a) You are required to submit your original source code in a .zip file to Gradescope.

(b) The format of your source code zip file:
PERMNUM.zip (Replace PERMNUM with you perm number)

- The final prediction file "prediction.txt" you submitted for calculating your hw4 score.
- A "prediction.py" file. Note: Please carefully check the "prediction.py" file in the starter package for the requirements of your submitted "prediction.py". If your file cannot meet the requirements, the score of your hw4 will be 0.
- Any other related files, including your preprocessing code, training source code, parameter file, academic integrity file, etc.

(c) Please DO NOT include the training data in your compressed file.

(d) If you answered Yes to either of the two "Academic Integrity" questions on page 1, please add a .txt file in your submitted .zip file to give corresponding full details.

3.2 Prediction File Submission

(a) You can submit the prediction file to our class CodaLab competition using the following link:
https://codalab.lisn.upsaclay.fr/competitions/11235

(b) The format of your prediction file:
(b.1) The name of your prediction file must be "prediction.txt".
(b.2) The prediction file must have 10000 lines.
(b.3) Each line is an integer prediction label (0 - 9) for the corresponding testing image.
(b.4) The prediction results must follow the same order of the names of testing images (0.png – 9999.png).
(b.5) When submitting, create a zip file from "prediction.txt" and submit the zip file to CodaLab. DO NOT zip the directory containing "prediction.txt", just the txt file itself.

(c) **Important: Register with your UMail account!:** Remember to register a CodaLab Competitions account using your umail account so that the username will be your UCSBNetID. After that, log into CodaLab Competitions and set up your team name (whatever nickname you like). To protect your privacy, only the team names will be shown on the leader-board and your usernames will be anonymous. After your submission finishes running, please choose to submit it to the leader-board. Note that here team name is equivalent to your nickname, and it is still an independent homework assignment.

(d) Note that the start and end time shown on CodaLab competition page is UTC time, and the deadline for submitting your code is **Wednesday, March 15 by 10:00am PDT**. The competition will remain open until four days after the deadline. If you submit your last submission after the deadline, your submission will be counted as a late submission.

4. Grading

(a) In CodaLab, each label will have 1,000 testing samples. The average accuracy is used to evaluate your machine learning models.

(b) After the competition ends, the teaching staff will rank your machine learning models based on the average accuracy, and your final score for assignment #4 will be:

*Score = min(100, 100 * The accuracy of your model / 90*)

Note that this means you will get full score if your accuracy is greater than or equal to 90%.

(c) Bonus: The students who ranks top 3 in assignment #4 will be awarded a book, Deep Learning (Ian Goodfellow, Yoshua Bengio and Aaron Courville. 2016.), which is one of the most popular textbooks in machine learning and artificial intelligence fields.

5. Tips

(a) How to convert images to vectors?

The simplest way is:

```
from PIL import Image
temp = Image.open('XXX.png')
temp.getdata() # return a (28*28 = ) 784-dimensional vector
```

There are lots of other methods to convert images to matrix/vectors, remember Google is your friend.

(b) How to choose parameters for your machine learning models?

One simple and straightforward way is to use the approach of Cross Validation (https://www.openml.org/a/estimation-procedures/1). You are still free to explore other methods.

(c) How to use GPU to train your models?

Please check the official documents of related deep learning libraries to see how to install the libraries and how to run your model with GPU support. There are also freely available GPU computing resources for university students from Amazon Web Services and Google Could, and you can apply it using your UCSB email. **The teaching staff will not help you to set up your GPU environment due to the variable and complex system environment.**