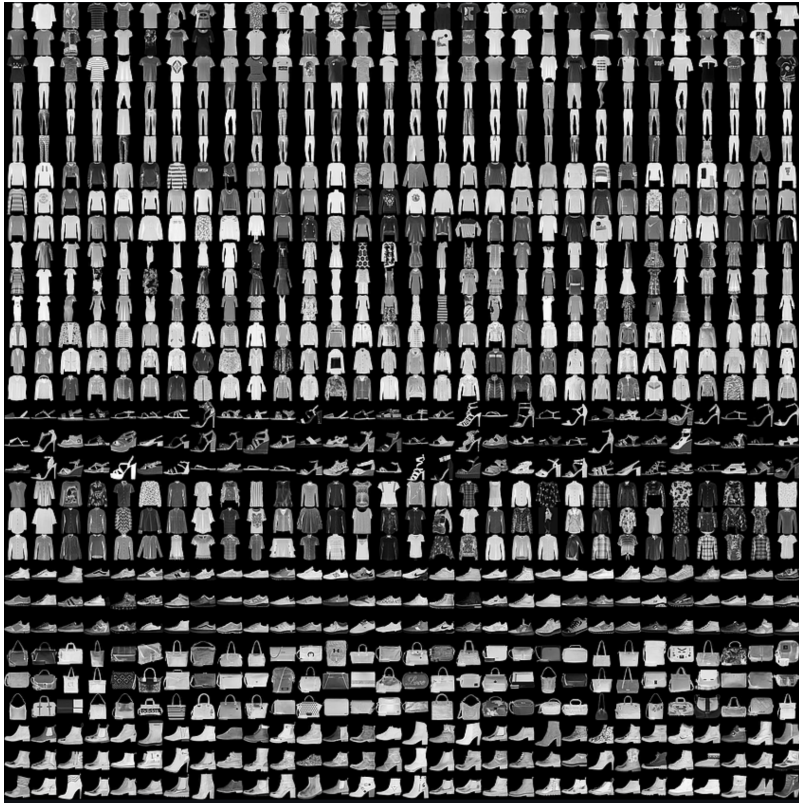


Spring 2024 CS444

Assignment 1: Linear classifiers

Due date: Thursday, February 15, 11:59:59 PM



In this assignment you will implement simple linear classifiers and run them on two different datasets:

1. [Rice dataset](#): a simple categorical binary classification dataset. Please note that the labels in the dataset are 0/1, as opposed to -1/1 as in the lectures, so you may have to change either the labels or the derivations of parameter update rules accordingly.
2. [Fashion-MNIST](#): a multi-class image classification dataset

The goal of this assignment is to help you understand the fundamentals of a few classic methods and become familiar with scientific computing tools in Python. You will also get experience in hyperparameter tuning and using proper train/validation/test data splits.

Download the starting code [here](#).

You will implement the following classifiers (in their respective files):

1. Logistic regression (`logistic.py`)
2. Perceptron (`perceptron.py`)
3. SVM (`svm.py`)
4. Softmax (`softmax.py`)

For the logistic regression classifier, multi-class prediction is difficult, as it requires a one-vs-one or one-vs-rest classifier for every class. Therefore, you only need to use logistic regression on the Rice dataset.

The top-level notebook (CS 444 Assignment-1.ipynb) will guide you through all of the steps. Setup instructions are below. The format of this assignment is inspired by the [Stanford CS231n assignments](#), and we have borrowed some of their data loading and instructions in our assignment IPython notebook.

None of the parts of this assignment require the use of a machine with a GPU. You may complete the assignment using your local machine or you may use [Google Colaboratory](#).

Environment Setup (Local)

If you will be completing the assignment on a local machine then you will need a Python environment set up with the appropriate packages.

We suggest that you use Anaconda to manage Python package dependencies (<https://www.anaconda.com/download>). This guide provides useful information on how to use Conda: <https://conda.io/docs/user-guide/getting-started.html>.

Data Setup (Local)

Once you have downloaded and opened the zip file, navigate to the fashion-mnist directory in assignment1 and execute the get_datasets script provided:

```
$ cd assignment1/fashion-mnist/  
$ sh get_data.sh or $bash get_data.sh
```

The Rice dataset is small enough that we've included it in the zip file.

Data Setup (For Colaboratory)

If you are using Google Colaboratory for this assignment, all of the Python packages you need will already be installed. The only thing you need to do is download the datasets and make them available to your account.

Download the assignment zip file and follow the steps above to download Fashion-MNIST to your local machine. Next, you should make a folder in your Google Drive to hold all of your assignment files and upload the entire assignment folder (including the datasets you downloaded) into this Google drive file.

You will now need to open the assignment 1 IPython notebook file from your Google Drive folder in Colaboratory and run a few setup commands. You can find a detailed tutorial on these steps [here](#) (no need to worry about setting up GPU for now). However, we have condensed all the important commands you need to run into an [IPython notebook](#).

IPython

The assignment is given to you in the CS 444 Assignment-1.ipynb file. As mentioned, if you are using Colaboratory, you can open the IPython notebook directly in Colaboratory. If you are using a local machine, ensure that IPython is installed (<https://ipython.org/install.html>). You may then navigate to the assignment directory in the terminal and start a local IPython server using the jupyter notebook command.

Submission Instructions

Submission of this assignment will involve three steps:

1. If you are working in a pair, only one designated student should make the submission to Canvas and Kaggle. You should indicate your Team Name on Kaggle Leaderboard and team members in the report.
2. You must submit your **output Kaggle CSV files** from each model on the Fashion-MNIST dataset to their corresponding Kaggle competition webpages:
 - [Perceptron](#)
 - [SVM](#)
 - [Softmax](#)

The baseline accuracies you should approximately reach are listed as benchmarks on each respective Kaggle leaderboard.

3. You must upload three files on [Canvas](#):
 1. All of your code (Python files and ipynb file) **in a single ZIP file**. The filename should be **netid_mp1_code.zip**. **Do NOT include datasets** in your zip file.
 2. Your IPython notebook **with output cells** converted to **PDF format**. The filename should be **netid_mp1_output.pdf**.
 3. A brief report **in PDF format** using [this template](#). The filename should be **netid_mp1_report.pdf**.
Don't forget to hit "Submit" after uploading your files, otherwise we will not receive your submission!

Please refer to [course policies](#) on academic honesty, collaboration, late submission, etc.