Bishop's University

CS 596 – Research Topics On Computer Science

Assignment 1: Machine Learning Basics

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 and Pytorch. You will also get experience in hyper parameter tuning and using proper train/test/validation data splits. For that purpose, you will implement some linear classifiers and you will validate them on the well-known CIFAR-10 image classification dataset. The classifiers you will implement are the following:

- K-nearest neighbor,
- Support Vector Machine,
- Linear regression,
- Logistic regression.

Recommendations

Because this assignment represents initiation to Python programming for the most of students, you can rely, reuse and integrate in the assignment one or more the following implementations:

- Tutorial To Implement k-Nearest Neighbors in Python From Scratch
- K-NN classification - PyTorch API
- Linear SVM with PyTorch
- PyTorch: Linear and Logistic Regression Models
- Logistic Regression on MNIST with PyTorch
- Linear Regression using PyTorch

The starting code is provided. The top-level notebook (`CS596 Assignment-1.ipynb`) will guide you through all 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.

If your local machine does not support GPU programming, you are encouraged to use Google Colaboratory to do this assignment; since it provides free access to a Tesla K80 (for running short jobs).

Here are two interesting links about Google Colaboratory with the instructions to install Pytorch:

https://towardsdatascience.com/getting-started-with-google-colab-f2fff97f594c

https://colab.research.google.com/notebooks/welcome.ipynb

By this environment, you will be able to do GPU programming and to run your programs through the cloud computing provided by Google.

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. This guide provides useful information on how to use Conda.

Data Setup (Local)

Once you have downloaded the zip file, navigate to the cifar10 directory in MP1 and execute the get_dataset script provided:

```
cd MP1/cifar/
./get_datasets.sh
```

Data Setup (For Colaboratory)

If you are using Google Colaboratory for this assignment you will need do some additional setup steps.

Download the assignment zip file and follow the steps above to download CIFAR-10 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 cifar10 dataset 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).

IPython

The assignment is given to you in the `CS596 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. You may then navigate the assignment directory in terminal and start a local ipython server using the `jupyter notebook` command.

Submission

You will have to submit all the programmed solutions, the CSV files associated to the four linear classifiers that contain the prediction results. Please, provide a pdf report following the format of the provided word template that highlights the team members, the contribution of each one of them, and the obtained accuracy for each classifier. The function that computes this accuracy is already provided in the starting code.

Good luck ☺