

CIS 399: Science of Data Ethics

Spring 2019

Homework 1: Machine Learning Basics and Overfitting

Due February 14, 2019 by 1:30pm

The Goals of the Assignment

From this assignment, we hope you gain experience in:

- Python and Visualizations
- Basic machine learning techniques
- The concept of overfitting
- The concept of fairness metrics (This is just an introduction, we will dive into this in further homeworks)

Getting Started

To start, restart Jupyter (you should refer to the [previous setup guide](#)). Once you have Jupyter running and have opened it up, navigate to the work directory. Open up a terminal instance within Jupyter and type the following:

```
cd work
git clone https://github.com/CIS399-DataEthics/hw1.git
```

This will download the files that you need to complete this homework. Once finished, type `exit`, hit [Enter], and close the Terminal browser window. You should now have a “hw1” folder inside your “work” directory.

In this directory, please open “Homework 1.ipynb” and follow the instructions in that file.

Submission

Prior to submission, make sure you have filled out all the sections and **run each block**. We recommend restarting the kernel and rerunning all the blocks to ensure it works. This can be done by clicking “Kernel -> Restart & Run All”.

When you have completed the above steps, please click “File -> Download as... -> PDF via LaTeX (.pdf)”. This should save a PDF to your Downloads with the contents of the notebook. Please upload this file to Gradescope.