

CS 4320: Machine Learning

Assignment Explore Data

One of the major tasks in machine learning is exploring and cleaning the data, *before* learning a model to represent the data.

In this assignment you will operate on a data set intended for supervised learning. You will explore the features of the data.

Use your personal data set available on Canvas in the `data-exploration` folder.

Required Steps

Aquire Data

Download *your* personal data file from Canvas.

Visualize the Data

Create a report containing the following displays. Note this exercise is intended to give you an opportunity to use numpy, pandas, and matplotlib. Process and plot the data with these tools. Use another document authoring system to combine the plots into a PDF document. For each plot, write a short (2-3 sentences) description of the plot, and your observations.

Put your data, and all code used to process the data in the `data-exploration` directory of your git repository.

Plot histograms for the values of each feature and the label.

Ploy scatter plots for the label (y-axis) vs each feature (x-axis).

Submission

- Submit the report (as PDF) to Canvas.
- Submit the code via source code repository.