

Lab Sheet 1

This is an introduction into standard libraries used in ML and how to load an existing data set. Before you start, I recommend installing Jupyter Lab and using it for this and all future labs.

Task 1: Load Libraries

Load the following libraries: * `numpy` for linear algebra * `pandas` for data processing, reading csv, etc. * `pyplot` for plotting your data * `seaborn` for more plotting

Task 2: Read in the data

Read in the `loans.csv` dataset and save it in a dataframe. Print out the first five rows to the console, just to see what kinda dataset you're dealing with here.

Task 3: Inspect the data

For a quick inspection of your data, try the following: 1. Check for NaN values and length of dataset (amount of observations). 2. Get some basic stats for all numeric variables, etc. count, mean, max, min, sd, etc. 3. Use `seaborn` for a `countplot` (bar chart) of one of your categorical variables (e.g. `Property`). 4. Use `pyplot` for a `ggplot` (scatter plot) of two of your numerical variables (e.g. `Age` and `Amount`) and see if there is any kind of correlation.

Task 4: Fit a model

Use `LinearRegression` from the `sklearn.linear_model` module (or any other model of your choice) to fit.