FCIM.FIA16.1 Spring 2023
Lab 3: Linear Regression

Handed out: Monday, January 30, 2023
Due: Friday, March 3, 2023

Oh these prices..

The company is sending you to live at the "Golden Rule" space station. And while the variable gravity doesn't seem to bother you, the prices for housing there do. You would like to know how much an apartment could cost but all the consulting companies ask for incredible amounts of money, so the solution is thus obvious – you'll need to find out yourself.

Being a qualified AI specialist you decide to over-engineer it a bit and create a program that could predict the price of an apartment given a set of characteristics of the apartment complex it is part of. You get your hands on some unformatted historical data about apartment complexes on the space station and their prices. The old man who supplied the data said that the fields that you might be interested are `complexAge` in column 3, `totalRooms` in column 4, `totalBedrooms` in column 5, `complexInhabitants` in column 6, `apartmentsNr` in column 7 and `medianCompexValue` in column 9.

Because running personal projects on production machines is a big no-no at the company, you've dug up an old server that can run Google Colab Notebooks and has `Scikit-learn`, `Pandas`, `Numpy` and `Matplotlib` pre-installed as the only libraries (beggars can't be choosers). You'll need to create a model that would perform linear regression on the available data. As a result, the model should be able to predict the `medianCompexValue` of an apartment complex, given a set of it's characteristics. Perform some simple statistical tests to determine the accuracy of your model.

Reporting

At the end of this lab, you will need to submit a *report* describing what you have implemented. The *report* must be uploaded on Else, in the according assignment activity. You should use the provided template. Suggested development environments are Google Colab or JupyterLab.

Grading policy

Task 1 Import your data. Analyze it via common statistical approaches. Cleanse the data if necessary. **(2p.)**

Task 2 Train your model by applying linear regression. **(1p.)**

Task 3 Show the prediction power of your model by attempting to predict the price of a new house. **(1p.)**

Task 4   Re-train your model. Use Ridge, Lasso or Elastic Net regularization. **(2p.)**

Task 5   Score and compare the scores of the models you have implemented. Interpret the result. **(2p.)**

Task 6   During your work, provide visualizations that would help better understand your data, your results etc. **(1p.)**

Report & Presentation   Clear explanations, report formatting, code quality etc. **(1p.)**

Plagiarism will not be tolerated.
Good Luck!