UNIVERSITE DE GENEVE Section de Mathématiques
Faculté des sciences

Optimization with Applications I
TP 3

Download the dataset **House Prices Advanced**. This includes many parameters of houses as well as their actual sale price. It is split into a train and test dataset. You will be asked to fit a model to the train dataset and evaluate it on the test dataset. Some of the dataset values are not numerical but rather categorical. we will ignore these for our purposes here - so please filter them out and work only with the parameters that are numerical and fit into the context of a linear regression.

We call y the vector of house prices and X the regression matrix of relevant parameters.

Exercise 1 (Least Squares)**.** Find the least squares optimizer $\arg\min_\alpha ||y - X\alpha||_2^2$. Which factors have the most influence on home pricing ? How good does the estimator perform on the test data ?

Exercise 2 (Ridge regression)**.** Write a method that finds the optimizer

$$\hat{\alpha}_\lambda = \arg\min_\alpha ||y - X\alpha||_2^2 + \lambda||\alpha||_2^2$$

for any $\lambda > 0$.

Plot the mean square error on the test set as a function of the λ you trained your estimator on. Is this what you expected ?

Exercise 3 (Steepest Descent)**.** Implement a steepest descent algorithm that approximates $\hat{\alpha}_\lambda$ by following the gradient. Use $||\nabla f(\alpha)||_2^2 \leq \epsilon$ as a stopping condition. Compare the results to those of Exercise 1 and again plot the mean square error plot on the test set as a function of λ.

Important : Every function and exercise must be tested. Plug in some values for which you know the correct answers and compare the output of your function.