HW 3: Coding

Due Date: 03/28/2021

Lecturer: Dr. Abd-AlRahman AlMomani *Data-Driven Science & ML*

In the attached file (HW03_Data.mat) you find a dataset that has A, y and x, such that:

$$y = Ax + \text{noise} \tag{1.1}$$

In the following problems, you need to use ONLY A and y, to estimate the parameters, and compare the estimated parameter to the true solution x given to you when needed.

Problem 1:

Use the Iterative Hard-Threshold (IHT) to estimate the solution $\hat{x}$.

First, show the performance of (IHT) under different noise levels. Assume the cardinality constrain that

$$\|\hat{x}\|_0 = 5$$

. then, find the solution after adding a white gaussian noise with $SNR = 1, 2, ..., 20$.

For each SNR value, find the solution, and compute the relative error

$$error = \frac{\|x - \hat{x}\|}{\|x\|}$$

Then, plot the SNR value on the x-axis and the error on the y-axis. (Check if you need log-scale the y-axis).

Write your comments and observations.

Specifically, explain why the curve is not smooth? How to get better statistical confidence about the results?

Problem 2:
Check the built-in matlab function:

model = sequentialfs(fun,X,y)

which is a Sequential feature selection function using custom criterion. Read the documentation of this function, and suggest an objective function "fun" to select the best features (same date as problem 1) with K-folds cross-validation.

Write your comments and conclusions.

For Python users, check the function: sklearn.feature_selection.SequentialFeatureSelector

Going the extra mile by extending the application of this function to several examples, designing a good objective function, and validating/comparing the results can substitute your Midterm exam grade. Consult me to discuss the details if you wish to take this opportunity.