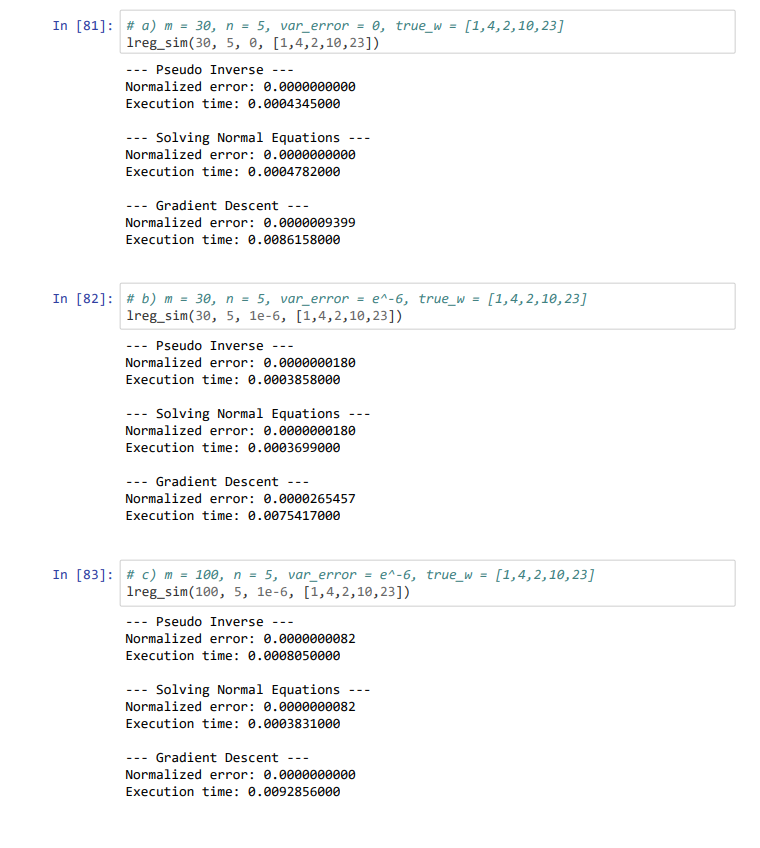
Homework 1

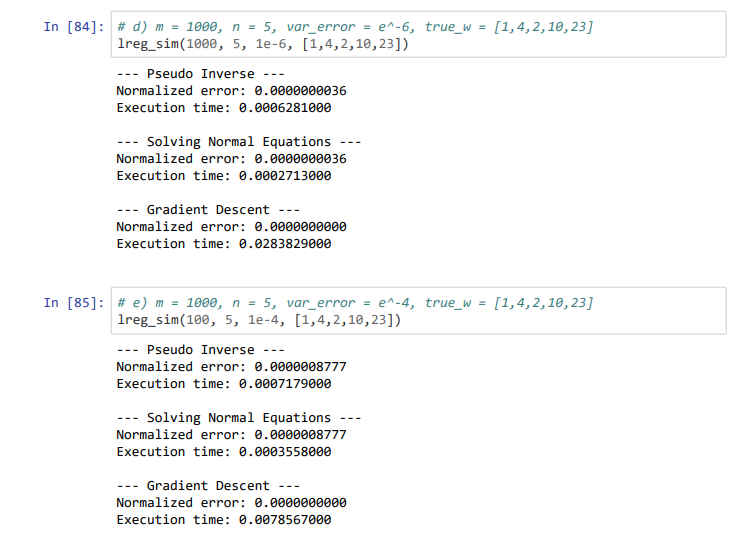
EE425X - Machine Learning: A signal processing perspective

An Nguyen

# Simulated Data

The independent variables are generated from a standard normal distribution. All entries of the data are independent identically distributed. The dependent variable (y) is derived from the independent variables plus a random error term. The error term follows a normal distribution with mean 0 and a given variance.





The normalized errors and execution times are reported above. Here are a few observations from our results:

* The normalized error is 0 when the error terms are all 0 (we get a perfect fit of the regression line).
* The normalized error is negatively correlated with the sample size (more data points give us better estimates of the parameters vector)
* Pseudo Inverse method is faster than Solving Normal Equations method for smaller sample size. Both methods are much faster than the Gradient Descent method in our experiment.

# Real Data

With the following parameters, we apply the Gradient Descent method to estimate θ

* max\_iter: 10000
* learning\_rate: 10-11

The choices of max\_iter and learning\_rate are arbitrary (trial and error) so that learning rate is not too large (learning rate > 10-10 provides a NaN result as its product with the gradient exceeds the limit of python). Meanwhile, larger number of iterations only produces minimally better result.

* The mean square error: 1344.51
* Estimated parameters vector: [2.6e-3, 3.6e-1, 6.6e-3, 1.95, 6.0e-4, 4.9e-2]
* Solution of the least square error: [-1.3e-03, -4.2e-01, 36, 0.1, -1.5e02, 1.3e+02]

The last term of the parameters vector is the intercept of the regression function.

# Extra Credit

With standardized features the mean square error is reduced to 23.03

# 4 Extra Credit