Homework 1b: Linear Regression part 2.
EE425X - Machine Learning: A Signal Processing Perspective

Homework 1 focused on learning the parameter θ for linear regression. In this homework we will first understand how to use the learnt parameter to predict the output for a given query input. We will also understand bias-variance tradeoff and how to decide the model dimension when limited training data is available. This HW will rely heavily on the code from the previous homework.

Generate Data Code: Generate $m + m_{test}$ data points satisfying

$$y = \theta^T \boldsymbol{x} + e$$

with θ being ONE fixed n length vector for all of them. Use $n = 100$, $\theta = [100, -99, 98, -97...1]'$, $\sigma_e^2 = 0.01||\theta||_2^2$, $e \sim \mathcal{N}(0, \sigma_e^2)$, $\boldsymbol{x} \sim \mathcal{N}(0, I)$, and assume mutual independence of the different inputs and noise values (e).

1. Use code from Homework 1 (using any one approach is okay) to learn θ. Vary m and show a plot of both estimation error in θ,

   $$||\theta - \hat{\theta}||_2^2/||\theta||^2$$

   and a second plot of the "Monte Carlo estimate" of the prediction error on the test data (test data MSE).

   $$\text{Normalized-Test-MSE} := \mathbb{E}[(y_{test} - \hat{y})^2]/\mathbb{E}[y_{test}^2], \text{ with } \hat{y} := \hat{\theta}^T \boldsymbol{x}_{test}$$

   Monte Carlo estimate means: compute $(y_{test} - \hat{y})^2$ for m_{test} different input-output pairs and then average the result.

   (a) Vary m: use $m = 80$, $m = 100$, $m = 120$, $m = 400$. If your code is unable to return an estimate of θ, you can report the errors to be ∞ (and for the plot just use a large value say 100000 to replace ∞.

   (b) Repeat this experiment with $\sigma_e^2 = 0.1||\theta||_2^2$.
   Thus this part will produce four plots.

2. In this second part, suppose you have only $m = 80$ training data points satisfying $y = \theta^T \boldsymbol{x} + e$, with $n = 80$. What you will have concluded from part 1 is that you cannot learn θ correctly in this case because m is even smaller than n.

   Let us assume you do not have the option to increase m. What can you do? All you can do is reduce n to a value $n_{small} \leq m$. Experiment with different values of n_{small} to come up with the best one. Do this experiment for two values of σ_e^2: $\sigma_e^2 = 0.01||\theta||_2^2$ and $\sigma_e^2 = 0.1||\theta||_2^2$.

   How to decide which entries of $\boldsymbol{x}$ to throw away? For now, just throw away the last $n - n_{small} + 1$ entries. So for $n_{small} = 1$, let $\boldsymbol{x}_{small}$ be just the first entry, and so on. So for $n_{small} = 30$ for example, $\boldsymbol{x}_{small}$ will be the first 30 entries of $\boldsymbol{x}$. There are many other better ways which we will learn about later in the course.

   Start with $n_{small} = 1$ and keep increasing its value and each time compute Normalized-Test-MSE by learning a value of θ first (using $m = 80$ of course). Obtain a plot. Use the plot and what you learn in class to decide what value of n_{small} is best.

3. Interpret your results based on the Bias-Variance tradeoff discussion. See Section 11 of Summary-Notes and what will be taught in the next few classes.