6.390 Introduction to Machine Learning
Recitation Week #2
Issued September 12, 2022

Announcements:

- Welcome to the recitation section of 6.390!
- Don't forget to complete the week 1 lab survey!

Today's Plan:

- Machine Learning Problem Formulation
- Linear Regression

---

1. Suppose that you are given the task to train a model to predict the pollution level in different cities given data points from satellite readings. You are given a dataset comprised of pairs of satellite readings and the pollution level recorded. The satellite readings include many features, e.g., temperature readings, building density, local population, etc. You hypothesize that the pollution level will be equal to a linear combination of the different features measured by the satellites. Your goal is to apply your models to locations where you have the satellite readings, but you do not have the pollution level readings.
   (a) Before training any model, it is important to understand the machine learning pipeline and to go in with a clear plan! Please identify the following:
      - The inputs and outputs of your model
      - The hypothesis class
      - A metric to determine which hypothesis best fits your data

(b) Consider the ordinary least squares objective function,

$$J(\theta, \theta_0) = \frac{1}{n} \sum_{i=1}^{n} (\theta^\top x^{(i)} + \theta_0 - y^{(i)})^2.$$

You use your near-infinite computing power to find the analytical solution to the ordinary least squares objective function:

$$\theta = \left(\tilde{X}^\top \tilde{X}\right)^{-1} \tilde{X}^\top \tilde{Y}.$$

However, you are not able to invert the matrix $\tilde{X}^\top \tilde{X}$.

Brainstorm some potential causes for this predicament. Is there an issue with the satellite readings, the pollution level reads, both, or neither? Is it at all possible to obtain a solution in this regime? What might happen if you employ the random regression algorithm, instead of computing the analytical solution?

(c) You shift gears to a different objective function known as ridge regression,

$$J(\theta, \theta_0) = \frac{1}{n} \sum_{i=1}^{n} (\theta^\top x^{(i)} + \theta_0 - y^{(i)})^2 + \lambda \|\theta\|^2.$$

You arrive at the analytical solution,

$$\theta = \left(\tilde{X}^\top \tilde{X} + n\lambda I\right)^{-1} \tilde{X}^\top \tilde{Y}.$$

What is the role of the hyperparameter, λ, in this objective function? What if different features are at different scales (e.g., parts per million of pollution vs. population density vs. distance of sea level in miles)? Should you apply regularization on the intercept, θ_0? How does ridge regression deal with the case where one feature is a linear function of another? How could you interpret the impact of different features from the values of θ?