Abir Islam

CS506

Professor Galletti

506 Apr 4, 2024 : Linear Regression

- Motivation
    - Working with continuous variables rather than discrete classes
    - We want to model the behavior of one data feature as a function of another
    - We want to model the general trend/pattern rather than try to fit a function to all of the data points
- Assumptions
    - For a given value of a variable X, we can assume that the variable we want to capture, Y, follows a normal distribution with a mean at X□ and a particular standard deviation
    - For Linear Regression purposes, we assume the data follows a linear function with a small degree of randomness as derived from a normal distribution
    - This normal distribution is consistent for all values of X□, not necessarily X, meaning we can capture more complex, non-linear in X relationships
    - IMPORTANT: our function should be linear in the parameter □, not necessarily X
- Randomness can arise from simply not understanding the full complexity of data, where manipulation of certain dimensions of the data could possibly reveal a more definite pattern

- Cost Function
    - Minimum Distance
        - We can implement a distance function to evaluate the fit of our regression, for example:

$$\boxed{L(h) = \sum_i d(h(x_i), y_i)}$$

        - Find the sum of the distance between all points and the linear function we described
        - Essentially, we want to take the function we were given, take the derivative and set it equal to 0, and then find the value of □ that produces the minimum value
        - This is known as the least squares method, ultimately giving us:

$$\boxed{\beta_{LS} = (X^T X)^{-1} X^T y}$$

        - Where □$_{LS}$ is the least squares approximation of our original □ parameter
    - Maximum Likelihood
        - Rather than minimizing distance of points to our model, we can maximize the likelihood of observing the dataset given the model
        - P(Y | h)