Lecture 15 Linear Regression

- Motivation
  - Understand and explain how y varies as a function of x
  - Find a function y = h(x) that best fits our data
  - If the function matching the curve is too complex it may be overfitting
  - Y = XB and a singular point is (xi, xiB)
  - Data generated by a linear function plus noise

$$\vec{y} = h_X(\beta) + \vec{\epsilon}$$

Where **h** is linear in a parameter $\boldsymbol{\beta}$.
Where $\boldsymbol{\epsilon_i}$ are independent $\mathbf{N(0, \sigma^2)}$ distribution.

- Cost Function
  - Given a data set {(x1,y1)…(xn,yn)} and a curve y = h(x) we can evaluate whether it is a good fit to our data through a cost function that compares h(xi) to yi for all I
  - Goal: for a given distance function d, find h where L is the smallest

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

- Assumptions
  - The realtion between x (the independent varible) and y(the dependent variable) is linear in a parameter B
  - εi are independent , identically distributed random variables following a normal distribution
- Goal(s):
  - Learn / estimate B
  - Try to minimize cost function
- Least Squares
- Maximum Likelihood
  - Defining the linear regression problem in terms of probability
  - Define P(Y|h) as the probability of observing Y given that it was sampled from h
    - Goal : find h that maximizes the probability of having observed our data
  - Maximizing L(h) = P(Y|h)
    - Since ε ~ N(0, sigma^2) and Y = XB + ε Y ~ N(XB, sigma^2)
  - An unbiased estimator

$$\begin{aligned} E[\beta_{LS}] &= E[(X^TX)^{-1}X^Ty] \\ &= (X^TX)^{-1}X^TE[y] \\ &= (X^TX)^{-1}X^TE[X\beta + \epsilon] \\ &= (X^TX)^{-1}X^TX\beta + E[\epsilon] \\ &= \beta \end{aligned}$$