# Lecture 11 (MLE)

## 1 Analytical Solution

$(X^T X)^{-1}$ is pseudo-inverse since $(X^T X)^{-1} \cdot X = I$.

## 2 Normalisation (Standardisation) of Data

Change $x_i$ to $x_i'$ to have 0 mean and unit variance.

## 3 Probabilistic Interpretation (MLE - Maximum Liklihood Estimate)

1. Predict the distribution $(x_i, y_i)$ came from
2. Idea is to add noise to prediction $y_i = \theta^T x_i + N(0, \sigma^2)$
3. Compute $\prod_{i=1}^{m} P(y_i|x_i;\theta)$ which is called the liklihood estimate
4. To find the optimal $\theta$, *argmax* is taken for the estimate

$$P(y_i|x_i;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \theta^T x_i)^2}{2\sigma^2}\right)$$

$$\implies \underset{\theta}{argmax} \left(\log(\prod_{i=1}^{m} P(y_i|x_i;\theta))\right) = \underset{\theta}{argmax} \left(\sum_{i=1}^{m} \left(\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \frac{-1}{2\sigma^2} \left(y_i - \theta^T x_i\right)^2\right)\right)$$

$$\implies \underset{\theta}{argmax} LL(\theta) = \underset{\theta}{argmin} \left(y_i - \theta^T x_i\right)^2$$

In general, the algorithm computes $\underset{\theta}{argmax} \log(\prod_{i=1}^{m} P(y_i|x_i;\theta))$ and $\log(\prod_{i=1}^{m} P(y_i|x_i;\theta)) = LL(\theta)$ (log liklihood)