Lecture 6

$y \subseteq \mathbb{R}$

$P = 1$ $\quad w_0 + w_1 x_1$

$\mathcal{H} = \{ \vec{w} \cdot \vec{x} : \vec{w} \in \mathbb{R}^{P+1} \}$ all linear model

$\vec{x} = [1 \; \vec{x}]$

$A : b_0, b_1 = \underset{\vec{w} \in \mathbb{R}^{P+1}}{\arg\min} \underbrace{\left\{ \sum_{i=1}^{n} (y_i - \vec{w} \cdot \vec{x}_i) \right\}}_{SSE}$

Ordinary Least Squares (OLS) Regression

SSE (sum of squared error)

Using calculus:

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$b_1 = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2}$$

$Var[X]$ is estimated by S_x^2

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} (x_i^2 - 2\bar{x} x_i + \bar{x}^2)$$

$$= \frac{1}{n-1} \left(\sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \right)$$

$$= \frac{1}{n-1}\left(\Sigma x_i^2 - n\bar{x}^2\right)$$

For 2 r.v's X, Y

Correlation: $Corr[X,Y] = \dfrac{Cov[X,Y]}{\sqrt{Var[X]\,Var[Y]}}$

Covariance: $Cov[X,Y] = E\left[(X-\mu_x)(Y-\mu_y)\right]$

$Cov[X,Y]$ estimated by S_{xy}

$$S_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}(x_iy_i - x_i\bar{y} - y_i\bar{x} + \bar{x}\bar{y})$$

$$= \frac{1}{n-1}\left(\Sigma x_iy_i - n\bar{x}\bar{y} - \cancel{n\bar{x}\bar{y}} + \cancel{n\bar{x}\bar{y}}\right)$$

$$= \frac{1}{n-1}\left(\Sigma x_iy_i - n\bar{x}\bar{y}\right)$$

$Corr[X,Y]$ is estimated by:

$$r := \frac{S_{xy}}{\sqrt{S_x^2 S_y^2}} = \frac{S_{xy}}{S_x S_y}$$

$$b_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{(n-1)\,S_{xy}}{(n-1)\,S_x^2}$$

$$= \frac{S_{xy}}{S_x^2} = \frac{r\,S_x S_y}{S_x^2}$$

$$= r\,\frac{S_y}{S_x}$$

$$b_0 = \bar{y} - b_1\bar{x} = \bar{y} - r\,\frac{S_y}{S_x}\,\bar{x}$$

$A \Rightarrow g(x) = b_0 + b_1 x$ (how to predict)