Residual Sum of Squares

• Its the sum of rest of $e_i = y_i - \hat{y}_i$

$$RSS = e_1^2 + e_2^2 + \dots + e_n^2 = \sum_{i=1}^n e_i^2$$

• The *RSS* expression is derived from the <u>Maximum Likelihood Estimation</u> by assuming that the error terms takes on a <u>normal</u> <u>distribution</u> bell curve

$$\hat{eta}^{ ext{MLE}} = rg \max_{eta} \sum_{i=1}^n \log[p(X_i|eta)]$$

RSS Derivation

1. Assumption of Linear Regression

$$Y_i = eta_0 + eta_1 X_i + arepsilon_i$$

- ullet $arepsilon_i \sim \mathcal{N}(0,\sigma^2)$
- i.i.d.(independent, identically distributed)

1. Likelihood Function

$$L(eta,\sigma^2) = \prod_{i=1}^n rac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/2\sigma^2}$$

Where:

- $ullet x_i = Y_i$ The observed Data
- $\mu = \hat{Y}_i = \hat{eta}_0 + \hat{eta}_1 X_i$ The expected value
- The normal distribution is on Y

$$L(eta,\sigma^2) = \prod_{i=1}^n rac{1}{\sqrt{2\pi\sigma^2}} e^{-(Y_i-eta_0-eta_1X_i)^2/2\sigma^2}$$

$$L(eta,\sigma^2) = \prod_{i=1}^n rac{1}{\sqrt{2\pi\sigma^2}} e^{-\sum (y_i-\hat{y}_i)^2/2\sigma^2}$$

•
$$(y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Log Likelihood

$$egin{split} \log(L(eta,\sigma^2)) &= \sum_{i=1}^n rac{1}{\sqrt{2\pi\sigma^2}} - \sum (y_i - \hat{y}_i)^2/2\sigma^2 \ &= -rac{n}{2} {\log(2\pi\sigma^2)} - \sum_{i=1}^n (y_i - \hat{y}_i)^2/2\sigma^2 \end{split}$$

Maximazing with respect to β_0

$$egin{aligned} rac{dL}{d\hat{eta}_0} &= rac{1}{\sigma^2} \sum (y_i - \hat{eta}_0 - \hat{eta}_1 X_i) \ rac{1}{\sigma^2} \sum_{i=1}^n (y_i - \hat{eta}_1 X_i) - \hat{eta}_0 \end{aligned}$$

Soliving For 0

$$rac{1}{\sigma^2}\sum_{i=1}^n(y_i-\hat{eta}_1X_i)-n\hat{eta}_0=0$$

$$egin{aligned} n\hat{eta}_0 &= \sum y_i - \hat{eta}_1 X_i = nar{y} - \hat{eta}_1 nar{x} \ \hat{eta}_0 &= ar{y} - \hat{eta}_1ar{x} \end{aligned}$$

Maximazing with respect to β_1

$$egin{aligned} rac{dL}{d\hat{eta}_1} &= -rac{1}{2\sigma^2} \sum (-2X_i(y_i - \hat{eta}_0 - \hat{eta}_1 X_i)) \ rac{dL}{d\hat{eta}_1} &= rac{1}{\sigma^2} \sum (X_i(y_i - ar{y} + \hat{eta}_1 ar{x} - \hat{eta}_1 X_i)) \end{aligned}$$

Soliving For 0

$$egin{aligned} rac{1}{\sigma^2} \sum (X_i(y_i - ar{y} + \hat{eta}_1ar{x} - \hat{eta}_1X_i)) &= 0 \ \sum X_i(y_i - \hat{y}) - \hat{eta}_1 \sum X_i(x_i - ar{x}) &= 0 \ \hat{eta}_1 &= rac{\sum y_i - \hat{y}}{\sum x_i - ar{x}} \ \hat{eta}_1 &= rac{\sum (x_i - ar{x})(y_i - ar{y})}{\sum (x_i - ar{x})^2} \ \hat{eta}_1 &= rac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)} \end{aligned}$$

- $ullet f^n \; rac{d}{df^n} = n rac{d}{df} f^{n-1}$
- When maximizing a **Likelihood function** and solving for 0 we ignore the constant terms $\frac{1}{\sigma^2}$

Maximazing with respect to σ^2

$$rac{dL}{d\hat{y}_i} = -rac{n}{2\sigma^2} + rac{1}{2(\sigma^2)^2}\sum (y_i - \hat{y}_i)^2$$

Soliving For 0

$$egin{split} -rac{n}{2\sigma^2} + rac{1}{2(\sigma^2)^2} \sum (y_i - \hat{y}_i)^2 &= 0 \ rac{n}{2\sigma^2} &= rac{1}{2(\sigma^2)^2} \sum (y_i - \hat{y}_i)^2 \ rac{2(\sigma^2)^2}{2\sigma^2} &= rac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \ \sigma^2 &= rac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} &= rac{ ext{RSS}}{n} &= ext{MSE} \end{split}$$