Linear Regression

measure (x_i, y_i) $\quad i = 1, 2 \ldots n$ $\quad$ # of data points

model form $\quad y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

objective $\quad \min \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i) \right)^2$

$$(y_i - \hat{y}_i)^2$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad , \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

$$SXY = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \quad , \quad SXX = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

"uncorrected"

$$\hat{\beta}_1 = \frac{SXY}{SXX} \quad , \quad \hat{\beta}_0 = \bar{y} - \beta_1 \bar{x}$$

data = (x_i, y_i) $\quad$ model "like" $\quad \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

residual, $r_i = y_i - \hat{y}_i$

residual sum of square, $RSS = \sum_{i=1}^{n} r_i^2$

(use this space to make any additional comments for any of the parts)

Multi-parameter regression

initially Linear form.

measure $(x_{1i}, x_{2i}, \dots x_{ki}, y_i)$ $i = 1, 2, \dots n \rightarrow$ data points

model form.
$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \dots + \beta_k x_{ki} + \varepsilon_i$$

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \qquad \begin{array}{c} \\ n \downarrow \end{array} \; X = \begin{array}{c} k \longrightarrow \\ \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1n} \\ \vdots & x_{21} & x_{22} & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & \dots & x_{nk} \end{bmatrix} \end{array} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

$$E = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$Y = XB + E$$

$\hookrightarrow$ assume errors are independent

& random distribution of mean = 0

trying to minimise 'E'

$\rightarrow$ solve $\dfrac{\partial E'E}{\partial B} = 0$

$$\hat{B} = (X^T X)^{-1} X^T Y$$

$$\hat{Y} = X\hat{B}$$

$$r_i = Y - \hat{Y}$$

$$RSS = Y^T Y - \hat{B}^T X^T X \hat{B}$$

Non-linear (least square) fitting/optimisation

$\hat{y}_i = h(x_i)$ "model" $y_i = h(x_i, \theta)$

$$\min_{\theta} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

$A \rightarrow B$ t C_A C_B T constant

"1st order" $\frac{dC_A}{dt} = -k C_A \rightarrow C_A = C_{A0} e^{-kt}$

$y = p(1) * \exp(p(2) * t)$

$p(1) = C_A(1)$

$\ln\left(\frac{C_A}{C_{A0}}\right) = -kt \Rightarrow k = \frac{-\ln\left(\frac{C_A}{C_{A0}}\right)}{t} = p(2)$

ref.