Estimation for Linear Models:

Intuition: best model has small RSS

Choose $\hat{\beta}$ to minimize RSS:

$$\min_{\beta} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \iff \min_{\beta} \sum_{i=1}^{n} \{y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})\}^2$$

To find a critical point, differentiate w.r.t. each β_j & set $= 0$:

$$\frac{\partial}{\partial \beta_0} RSS = \sum_{i=1}^{n} 2\{y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})\}(-1) = 0$$

$$\frac{\partial}{\partial \beta_1} RSS = \sum_{i=1}^{n} 2\{y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})\}(-x_{i1}) = 0$$

$$\vdots$$

$$\frac{\partial}{\partial \beta_p} RSS = \sum_{i=1}^{n} 2\{y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})\}(-x_{ip}) = 0$$

$$\Rightarrow \quad \sum_{i=1}^{n} (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})(1) = \sum_{i=1}^{n} y_i (1)$$

$$\sum_{i=1}^{n} (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})(x_{i1}) = \sum_{i=1}^{n} y_i \cdot x_{i1}$$

$$\vdots$$

$$\sum_{i=1}^{n} (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})(x_{ip}) = \sum_{i=1}^{n} y_i x_{ip}$$

Note: for a vector $\underset{\sim}{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$,

$$X'\underset{\sim}{a} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

$$\Rightarrow \quad X'X\beta = X'y$$

$$X'X\beta = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \cdots + \beta_p x_{1p} \\ \beta_0 + \beta_1 x_{21} + \cdots + \beta_p x_{2p} \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + \cdots + \beta_p x_{np} \end{bmatrix}$$

$$X'y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\Rightarrow (X'X)^{-1}X'X\beta = (X'X)^{-1}X'y$$

$$\Rightarrow \beta = (X'X)^{-1}X'y$$

- Formally, this is a critical point and may not minimize RSS. We could verify it gives a minimum by showing the Hessian is positive definite.
- So our estimate of β is $\hat{\beta} = (X'X)^{-1}X'y$

~~So our fitted values are~~ $\cancel{X\hat{\beta} = X(X'X)^{-1}X'y}$

Ex. 1: Suppose we fit a model with no explanatory variables, only an intercept:

$$y_i = \beta_0 + \varepsilon_i$$
$$\varepsilon_i \sim \text{Normal}(0, \sigma^2)$$

- Write down the design matrix X
- Find $\hat{\beta}$

Ex. 2: Suppose we fit a one-way ANOVA model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$x_i = \begin{cases} 0 & \text{if obs. } i \text{ in ~~baseline group~~ first treatment group} \\ 1 & \text{if obs. } i \text{ in second treatment group} \end{cases}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

~~Write down the design matrix~~

Suppose we have $n=3$ observations with obs. 1 in the first treatment group and obs 2,3 in the second treatment group.

- Write down the design matrix X.
- Find $\hat{\beta}$.