$$S = \{(x_i, y_i)\} \qquad X = \begin{pmatrix} x_0 & 1 \\ x_1 & 1 \\ \vdots & \vdots \\ x_{n-1} & 1 \end{pmatrix} \qquad y = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{n-1} \end{pmatrix}$$

$$f(x) = p_0 x + p_1 \qquad \beta = \begin{pmatrix} p_0 \\ p_1 \end{pmatrix}$$

1) $SE\{f\}$ (eq. $SE\{\beta\}$) $= \| X\beta - y \|^2$

$$\nabla_\beta SE = X^T (X\beta - y)$$

$$\nabla_\beta SE = 0$$

$$X^T X \beta = X^T y$$

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

2)

$$\hat{y} = X\hat{\beta}$$

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

$$r = \hat{y} - y$$

$$\beta^* = \underset{\beta}{\operatorname{argmin}} \| X\beta - r \|^2$$

$$X^T(X\beta^* - r) = 0$$

$$X^T X \beta^* = X^T(\hat{y} - y)$$

$$\beta^* = (X^T X)^{-1} X^T \left(X(X^T X)^{-1} X^T y - y \right)$$

$$= \cancel{(X^T X)^{-1}}\, \cancel{X^T} X (X^T X)^{-1} X^T y - (X^T X)^{-1} X^T y$$

$$= \hat{\beta} - \hat{\beta}$$

$$= 0$$

Alternatively, we can prove 2) using a different strategy:

$$\hat{y} = X\hat{\beta}$$

$$\hat{y} = \underbrace{X(X^T X)^{-1} X^T}\, y$$

$$\hat{y} = Hy$$

(the "hat matrix", puts the hat on y)

$$H = X(X^T X)^{-1} X^T$$

now we prove that H is a <u>projection</u>:

$$HH = H$$

(ie: iterating the procedure does nothing)

$$HH = X(X^T X)^{-1} \cancel{X^T X (X^T X)^{-1}} X^T = X(X^T X)^{-1} X^T = H$$

We then use linearity, and show that the same model can't extract any signal from the residual:

$$\hat{r} = Hr$$

$$Hr = H(\hat{y} - y)$$

$$Hr = HH\hat{y} - H\hat{y}$$

$$= H\hat{y} - H\hat{y}$$

$$= 0$$