M342W

$$SSE = \underset{1\times n}{\vec{y}^T}\ \underset{n\times 1}{\vec{y}} - 2\underset{1\times(p+1)}{\vec{w}^T}\ \underset{(p+1)\times n}{X^T}\ \underset{n\times 1}{\vec{y}} + \underset{1\times(p+1)}{\vec{w}^T}\ \underset{(p+1)\times n}{X^T}\ \underset{n\times(p+1)}{X}\ \underset{(p+1)\times 1}{\vec{w}}$$

$$\begin{bmatrix} \frac{\partial SSE}{\partial w_0} \\ \frac{\partial SSE}{\partial w_1} \\ \vdots \\ \frac{\partial SSE}{\partial w_p} \end{bmatrix} \overset{set}{=} \vec{O}_{p+1} =: \frac{\partial SSE}{\partial \vec{w}}$$ and solve for $b_0, b_1, \ldots, b_p$

Linear: • let $\vec{x} \in \mathbb{R}^n$. let $a \in \mathbb{R}$ be a constant wrt $\vec{x}$.

$\Rightarrow$ ⓪ $\frac{\partial a}{\partial \vec{x}} = \vec{O}_n$.

• let $\vec{a} \in \mathbb{R}^n$ constant wrt $\vec{x}$. ① $\frac{\partial}{\partial \vec{x}}[\vec{a}^T\vec{x}] = \begin{bmatrix} \partial/\partial x_1[a_1x_1 + a_2x_2 + \ldots + a_nx_n] \\ \vdots \end{bmatrix}$, $\vec{a}^T\vec{x} = \vec{x}^T\vec{a}$

$$= \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \vec{a} \neq \vec{a}^T$$

• let $a, b \in \mathbb{R}$ be constants wrt $\vec{x}$.

② $\frac{\partial}{\partial \vec{x}}[af(\vec{x}) + bg(\vec{x})] = \begin{bmatrix} \partial/\partial x_1[af(\vec{x}) + bg(\vec{x})] \\ \vdots \end{bmatrix} = \begin{bmatrix} a\,\partial/\partial x_1[f(\vec{x})] + b\,\partial/\partial x_1[g(\vec{x})] \\ \vdots \end{bmatrix}$

$= a\frac{\partial}{\partial \vec{x}}[f(\vec{x})] + b\frac{\partial}{\partial \vec{x}}[g(\vec{x})]$

• let $A \in \mathbb{R}^{n\times n}$, symmetric, constant wrt $\vec{x}$

$\frac{\partial}{\partial \vec{x}}[\vec{x}^T A \vec{x}]$ This scalar expression is "quadratic form" and it is a common expression and very well-studied.

$$A\vec{x} = \begin{bmatrix} \leftarrow \vec{a}_1 \rightarrow \\ \leftarrow \vec{a}_2 \rightarrow \\ \vdots \\ \leftarrow \vec{a}_n \rightarrow \end{bmatrix}\begin{bmatrix} \uparrow \\ \vec{x} \\ \downarrow \end{bmatrix} = \begin{bmatrix} \vec{a}_1\cdot\vec{x} \\ \vec{a}_2\cdot\vec{x} \\ \vdots \\ \vec{a}_n\cdot\vec{x} \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_n + \ldots + a_{2n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n \end{bmatrix}$$

$$\vec{x}^T(A\vec{x}) = [x_1\ x_2\ \cdots\ x_n]\begin{bmatrix} \vec{a}_1\cdot\vec{x} \\ \vec{a}_2\cdot\vec{x} \\ \vdots \\ \vec{a}_n\cdot\vec{x} \end{bmatrix} = x_1\vec{a}_1\cdot\vec{x} + x_2\vec{a}_2\cdot\vec{x} + \ldots + x_n\vec{a}_n\cdot\vec{x}$$

$= x_1(a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n) + x_2(a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n) + \ldots$

$\ldots x_n(a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n)$ next page

$$\frac{\partial}{\partial x_1}[\ \ ] = 2a_{11}x_1 + \overbrace{a_{12}x_2 + \ldots + a_{1n}x_n + a_{21}x_2}^{\substack{(\text{Symmetric})\\ \text{equal}}} + \ldots + \underbrace{a_{n1}x_n}_{\text{equal}}$$

$$= 2a_{11}x_1 + 2a_{12}x_2 + \ldots + 2a_{1n}x_n$$

$$= 2\vec{a}_1 \cdot \vec{x}$$

$$\frac{\partial}{\partial x_2}[\ \ ] = 2\vec{a}_2 \cdot \vec{x}$$

$$\vdots$$

(3) $$\frac{\partial}{\partial \vec{x}}[\vec{x}^T A \vec{x}] = \begin{bmatrix} 2\vec{a}_1 \cdot \vec{x} \\ 2\vec{a}_2 \cdot \vec{x} \\ \vdots \\ 2\vec{a}_n \cdot \vec{x} \end{bmatrix} = 2A\vec{x}$$

$$\frac{\partial}{\partial \vec{w}}[\overbrace{\vec{y}^T\vec{y} - 2\vec{w}^T X^T \vec{y} + \vec{w}^T X^T X \vec{w}}^{SSE}] \overset{\text{Rule (2)}}{=} \frac{\partial}{\partial \vec{w}}[\cancel{\vec{y}^T\vec{y}}]_{\text{0 by rule (0)}} - 2\frac{\partial}{\partial \vec{w}}[\vec{w}^T(X^T\vec{y})] + \frac{\partial}{\partial \vec{w}}[\vec{w}^T(X^TX)\vec{w}]$$

$(X^TX)^T = X^T(X^T)^T = X^TX \Rightarrow$ Symmetric

$$\overset{\text{Rule (1)}}{=} -2X^T\vec{y} + \frac{\partial}{\partial \vec{w}}[\vec{w}^T(X^TX)\vec{w}] \overset{\text{Rule (3)}}{=} -\cancel{2}X^T\vec{y} + \cancel{2}X^TX\vec{w} \overset{\text{set}}{=} \vec{0}_{p+1}$$ and solve for $\vec{b}$

$$\Rightarrow \underbrace{(X^TX)^{-1}X^TX}_{I_{p+1}}\vec{w} = (X^TX)^{-1}X^T\vec{y} \Rightarrow \boxed{\vec{b} = (X^TX)^{-1}X^T\vec{y}} \Rightarrow \underbrace{\hat{y}_* = g(\vec{x}_*) = \vec{x}_* \vec{b}}_{\text{predictions}}$$

$\vec{b}$ is the best $\vec{w}$

In order to compute the OLS coefficients ($\vec{b}$), you need X^TX to be invertible. (X^TX is $(p+1)\times(p+1)$ square matrix). This means, $\text{rank}[X^TX] = p+1$ i.e. "full rank" i.e. all columns linearly independent.

Thm: $\text{rank}[A^TA] = \text{rank}[A]$

This means $\text{rank}[X] = p$, i.e. the columns of X are linearly independent.

$$X = \begin{array}{c} \overbrace{\hspace{6em}}^{\text{feature measurements on all } n \text{ subjects}} \\ \begin{bmatrix} 1 & \uparrow & \uparrow & & \uparrow \\ 1 & \vec{x}_{\cdot 1} & \vec{x}_{\cdot 2} & \cdots & \vec{x}_{\cdot p} \\ \vdots & \downarrow & \downarrow & & \downarrow \\ 1 & & & & \end{bmatrix} \end{array}$$

If X is full rank that means .... there is no exact data duplication e.g. x_1: height in inches, x_2: height in cm.

What if you do have a feature that is linearly dependent with the other features in X? You just drop it. Then X will be full rank and you're good to estimate the OLS coefficients.

$$\vec{y} = \underbrace{\hat{\vec{y}}}_{= X\vec{b}} + \vec{e} \Rightarrow \vec{e} = \vec{y} - \hat{\vec{y}}, \quad SSE = \sum_{i=1}^{n} e_i^2 = \vec{e}^T \vec{e}$$

$$MSE = \frac{1}{n-(p+1)} SSE, \quad RMSE = \sqrt{MSE}, \quad R^2 = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}$$

$$= \frac{S_y^2 - S_e^2}{S_y^2} \quad \text{(same)}$$

~~p+1 is somet~~

You sometimes say the model has $p+1$ "degrees of freedom" (i.e. the number of parameters, $w_0, w_1, \ldots, w_p$, is $p+1$) and $p+1 = \dim[\text{colsp}[X]]$.