$\underset{1\times n}{\vec{y}}$ $\underset{n\times 1}{\vec{y}}$ $\underset{1\times(p+1)}{\vec{w}^T}$ $\underset{(p+1)\times n}{X^T}$ $\underset{n\times 1}{\vec{y}}$ $\underset{1\times(p+1)}{\vec{w}^T}$ $\underset{(p+1)\times n}{X^T}$ $\underset{n\times(p+1)}{X}$ $\underset{(p+1)\times 1}{\vec{w}}$ 2/24/21

$$SSE = \vec{y}^T\vec{y} - 2\vec{w}^T X^T \vec{y} + \vec{w}^T X^T X \vec{w}$$

$$\frac{\partial SSE}{\partial \vec{w}} := \begin{bmatrix} \frac{\partial SSE}{\partial w_0} \\ \frac{\partial SSE}{\partial w_1} \\ \frac{\partial SSE}{\partial w_p} \end{bmatrix} \overset{Set}{=} \vec{0}_{p+1} \text{ and solve for } b_0, b_1, \ldots, b_p$$

Let $\vec{x} \in \mathbb{R}^n$, let $a \in \mathbb{R}$ be a constant with $\vec{x} \Rightarrow \frac{\partial}{\partial \vec{x}}[a] = \vec{0}_n$ (0)

Let $\vec{a} \in \mathbb{R}^n$ constant with $\vec{x}$

(1) $\frac{\partial}{\partial \vec{x}}[\vec{a}^T\vec{x}] = \begin{bmatrix} \partial/\partial x_1 [a_1x + a_2x_2 + \ldots + a_nx_n] \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \vec{a} \neq \vec{a}^T$

$\vec{x}^T\vec{a}$

let $a, b \in \mathbb{R}$ constants with $\vec{x}$

(2) $\frac{\partial}{\partial \vec{x}}[af(\vec{x}) + bg(\vec{x})] = \begin{bmatrix} \partial/\partial x_1 [af(\vec{x}) + bg(\vec{x})] \end{bmatrix}$

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

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

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

$$\frac{\partial}{\partial \vec{x}}\left[\vec{x}^T A \vec{x}\right], \quad A\vec{x} = \begin{bmatrix} \leftarrow \vec{a}_1 \rightarrow \\ \leftarrow \vec{a}_2 \rightarrow \\ \vdots \\ \leftarrow \vec{a}_n \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow \\ \vec{x} \\ \downarrow \end{bmatrix} = \begin{bmatrix} \vec{a}_1 \cdot \vec{x} \\ \vec{a}_2 \cdot \vec{x} \\ \vdots \\ \vec{a}_n \cdot \vec{x} \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \dots a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots a_{2n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 \dots a_{nn}x_n \end{bmatrix}$$

This scalar expression, $\vec{x}$-tr $A\vec{x}$ is called a "quadratic form" and its a common expression and very well-studied

$$\vec{x}^T(A\vec{x}) = \begin{bmatrix} x_1 x_2 \dots x_n \end{bmatrix} \begin{bmatrix} \vec{a}_1 \cdot \vec{x} \\ \vec{a}_2 \cdot \vec{x} \\ \vdots \\ \vec{a}_n \cdot \vec{x} \end{bmatrix} = x_1 a_1 \cdot \vec{x} + x_2 \vec{a}_2 \cdot \vec{x}_2 + \dots + x_n \vec{a}_n \cdot \vec{x}$$

$$= x_1(a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n) + x_2(a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n) + \dots + x_n(a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n)$$

$$\frac{\partial}{\partial x_1}[\,] = 2a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + a_{21}x_2 + \dots + a_{n1}x_n = 2a_{11}x_1 + 2a_{12}x_2 + \dots + 2a_{1n}x_n$$

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

$$\frac{\partial}{\partial x_2}[\,] = a_{12}x_1 + a_{21}x_1 + 2a_{22}x_2 + \dots + a_{2n}x_n + \dots + a_{n2}x_n = 2a_{12}x_1 + 2a_{22}x_2 + \dots + 2a_{2n}x_n$$

$$= 2\,\vec{a}_2 \cdot \vec{x}$$

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

b is the optimal / "best" "w"

SSE

$$\frac{\partial}{\partial \vec{w}}\left[\vec{y}^T\vec{y} - 2\vec{w}^T X^T \vec{y} + \vec{w}^T X^T X \vec{w}\right] \overset{\text{rule \#2}}{=} \frac{\partial}{\partial \vec{w}}\left[\vec{y}^T\vec{y}\right] - 2\frac{\partial}{\partial \vec{w}}\left[\vec{w}^T(X^T\vec{y})\right] + \frac{\partial}{\partial \vec{w}}\left[\vec{w}^T(X^TX)\vec{w}\right]$$

$\frac{\partial}{\partial \vec{w}}\left[\vec{y}^T\vec{y}\right] = 0$ by rule #0

rule #1 ↓

$$= -2X^T\vec{y} + \frac{\partial}{\partial \vec{w}}\left[\vec{w}^T(X^TX)\vec{w}\right] \overset{\text{rule \#3}}{=} -2X^T\vec{y} + 2X^TX\vec{w} \overset{\text{set}}{=} \vec{0}_{p+1}$$

$(X^TX)^T = X^T(X^T)^T = X^TX \Rightarrow$ Symm.

and solve for $\vec{b}$

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

$I_{p+1}\ A\vec{x} = \vec{b}$

predictions

In order to compute the OLS coefficients (vector b), you need X^TX, a $(p+1)\times(p+1)$ square matrix, to be invertible. Equivalently, rank $[X^TX] = p+1$ ie "full rank" ie all columns are linearly independent. Since there a thm: rank $[X^TX]$ = rank $[X]$, this mean rank $(X) = p+1$, ie the columns of X are linearly indep.

$X = \left[\ \vec{x}_{\cdot 1}\ \ \vec{x}_2\ \ \ldots\ \ \vec{x}_{\cdot p}\ \right]$

features measurements on all n subjects

If X is full rank that means..... there is no exact data duplications eg X_1: height measured in inches and X_2: height measured in centimeters. What if you do have a features 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.

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

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

$MSE = n-(p+1)$

$= \frac{s_y^2 - s_e^2}{s_y^2}$ (same)

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