multilinear regression

row → house

features ↓

$$X = \begin{pmatrix} & & \\ & & \\ & & \end{pmatrix}$$

N samples

k features

Y N targets

$$y = m_1 x_1 + \cdots + m_k x_k + b$$ ← model

what is b? what are the m_i?

↑ bias ↑ weights

"Best" fit is measured by minimizing

$$\text{Loss} = \sum_{i=1}^{N} \left(y_i - (m_1 x_{i1} + m_2 x_{i2} + \cdots + m_k x_{ik} + b) \right)^2$$

$$X^* = \begin{pmatrix} & 1 \\ X & \vdots \\ & 1 \end{pmatrix} \quad N \times (k+1)$$

$$M = \begin{pmatrix} m_1 \\ \vdots \\ m_k \\ b \end{pmatrix} \quad k+1$$

$$Y = X^* M$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} x_{i1} & x_{i2} & \cdots & x_{ik} & 1 \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_k \\ b \end{pmatrix} =$$

$$Y - X^* M = \begin{pmatrix} \vdots \\ y_i - \sum_{j=1}^{k+1} x_{ij} m_j \\ \vdots \end{pmatrix} \qquad m_{k+1} = b$$

$$L = \| Y - X^* M \|^2$$
$$= \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{k+1} x_{ij} m_j \right)^2$$

Side remark

$$f(x_1, \ldots, x_{k+1})$$

$$\nabla f = \begin{bmatrix} \partial f / \partial x_1 \\ \vdots \\ \partial f / \partial x_{k+1} \end{bmatrix} \qquad \nabla f = 0$$

$$\left(\text{all } \frac{\partial f}{\partial x_i} = 0 \right)$$

Want $\nabla L = 0$

Compute

$$\frac{\partial L}{\partial m_t} \qquad t = 1, \ldots, k+1$$

$$\frac{\partial L}{\partial m_t} = 2 \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{k+1} x_{ij} m_j \right) x_{it}$$
$$= 2 \sum_{i=1}^{N} x_{it} \left(y_i - \sum_{1}^{k+1} x_{ij} m_j \right)$$

side remark

$A: N \times k \qquad B: k \times s$

$C = A \cdot B \qquad C: N \times s$

$$C_{ij} = \sum_{t=1}^{k} a_{it} b_{ts}$$

$$X^* = (x_{ij}) \qquad N \times (k+1)$$

$$(X^*)^T = (x_{ji}) \qquad (k+1) \times N$$

$$A = \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix} \qquad A^t = \begin{pmatrix} a & c & e \\ b & d & f \end{pmatrix}$$

3×2 $\qquad$ 2×3

$$\frac{\partial L}{\partial m_t} = 2 \sum_{i=1}^{N} x^T_{ti} \left[y_i - \sum_{j=1}^{k+1} x_{ij} m_j \right]$$

$$\nabla L = \begin{bmatrix} \partial L / \partial m_1 \\ \partial L / \partial m_2 \\ \vdots \\ \partial L / \partial m_{k+1} \end{bmatrix} = 2 \underbrace{X^{*T}}_{(k+1) \times N} \Big(\underbrace{\overset{N \times 1}{Y} - \overset{N \times (k+1) \cdot (k+1) \times 1}{X^* M}}_{N \times 1} \Big) = 0$$

Solution to Least Sqs prob is

$$(X^*)^T Y - (X^*)^T X^* M = 0$$

$$\underbrace{(X^*)^T X^*}_{D} M = (X^*)^T Y$$

$$D = (X^*)^T X \qquad (k+1) \times (k+1) \text{ matrix}$$

$$DM = \underbrace{(X^*)^T Y}_{k+1\text{-vector}}$$

$$\begin{pmatrix} 3 & 1 \\ -2 & 5 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

assume D invertible

$$M_{LS} = D^{-1}\left[(X^*)^T Y\right]$$

Take "best" values $\hat{M}_{LS}$ for M

predicted $\hat{Y} = X^* M_{LS}$

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

$$(X^*)^T\left[Y - XM\right] = 0$$

condition for best M.

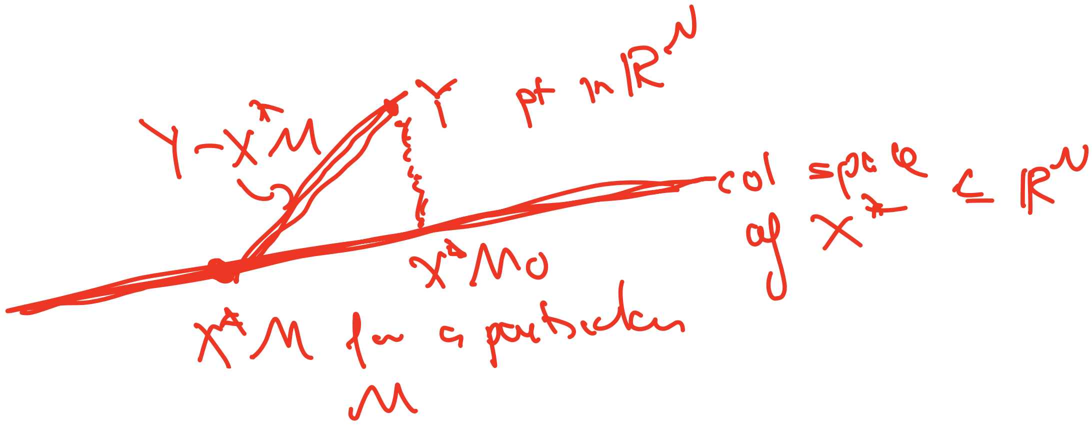
Think about

$Y \Leftrightarrow$ a point in $\mathbb{R}^N$

Columns of X^* are points in $\mathbb{R}^N$

as M varies over $\mathbb{R}^{k+1}$

X^*M varies over a $k+1$ dim'l subspace of $\mathbb{R}^N$ column space of X^*.



$\|Y - X^* M\|^2 \Leftrightarrow$ distance^2 from Y to $X^* M$

minimizing means find CLOSEST PT in COL SPACE of X^* TO target point Y

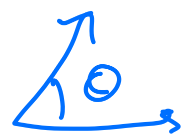
seems like this happens when line from Y to closest pt $X^* M_0$ is $\perp$ to col space.

Dot product
$w, v \in \mathbb{R}^N$

$w = (w_1, \ldots, w_N)$
$v = (v_1, \ldots, v_N)$

$w \cdot v = \sum w_i v_i$

$|w \cdot v| = \|w\| \|v\| \cos\theta$



$w \cdot v = 0 \Longleftrightarrow w \perp v$ $(\theta = \pi/2, -\pi/2)$

w, v orthogonal $\Longleftrightarrow w \cdot v = 0$.

Y to be orthogonal to col space of X^*

Y ortho to each col of X^*.

$$X^* M = \sum_{i=1}^{k+1} m_i X_{\cdot,i} \quad \leftarrow i^{th} \text{ col}$$

$$Y \cdot X_{\cdot,i} = 0 \longrightarrow Y \cdot (X^* M)$$

since $Y \cdot \sum m_i X_{\cdot,i}$

$$= \sum m_i (Y \cdot X_{\cdot,i})$$

$$= 0$$

Want $(Y - X^* M)$ orth to all cols of X^*

in fact

$$(X^*)^T (Y - X^* M) = 0$$

$$\Updownarrow$$

c

cols of X^* are $\perp$ to $Y - X^* M$

at optimal point.

$$M_{LS} = D^{-1} [(X^*)^T Y]$$

$$D = (X^*)^T (X^*)$$