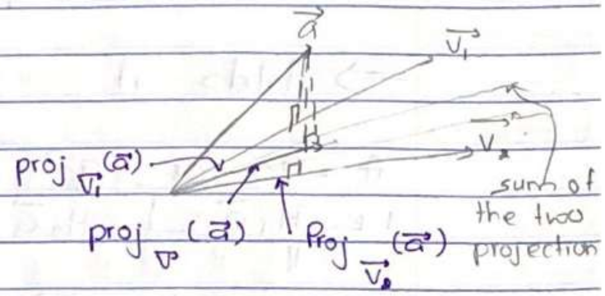
Lecture - 11 03/03/2020

$V = [\vec{v_1} \mid \vec{v_2}]$

$\text{colsp}[V] := \text{span}[\{\vec{v_1}, \vec{v_2}\}]$

$\text{Proj}_V(\vec{a}) \overset{?}{=} \text{proj}_{\vec{v_1}}(\vec{a}) + \text{proj}_{\vec{v_2}}(\vec{a})$

$H\vec{a} = \underset{\substack{\parallel \\ C_1\vec{v_1}}}{H_1\vec{a}} + \underset{\substack{\parallel \\ C_2\vec{v_2}}}{H_2\vec{a}}$



$\Longrightarrow H = H_1 + H_2$

Is $H_1 + H_2$ an orthogonal proj. ?

Two Properties.

I $\text{Proj}_W(\vec{a}) = C_1\vec{v_1} + C_2\vec{v_2} \in \text{colsp}[V]$

II $\text{Proj}_V(\vec{a}) \perp \vec{a} - \text{proj}_V(\vec{a})$

$\Longrightarrow \text{proj}_V(\vec{a})^T(\vec{a} - \text{proj}_V(\vec{a})) = 0.$

$\text{proj}_V(\vec{a})^T\vec{a} - \text{proj}_V(\vec{a})^T \cdot \text{proj}_V(\vec{a}) = 0.$

$\Longrightarrow \text{proj}_V(\vec{a})^T\vec{a} - \|\text{proj}_V(\vec{a})\|^2 = 0$

$\Longrightarrow ((H_1 + H_2)\vec{a})^T\vec{a} - \|(H_1 + H_2)\vec{a}\|^2 = 0.$

$\Longrightarrow (H_1\vec{a} + H_2\vec{a})^T\vec{a} - \|H_1\vec{a} + H_2\vec{a}\|^2 = 0.$

$\Longrightarrow (\vec{a}^T H_1^T + \vec{a}^T H_2^T)\vec{a} - \|H_1\vec{a} + H_2\vec{a}\|^2 = 0$

$\Longrightarrow (\vec{a}^T H_1^T\vec{a} + \vec{a}^T H_2^T\vec{a}) - \|H_1\vec{a} + H_2\vec{a}\|^2 = 0$

$$\Longrightarrow \underbrace{\vec{a}^T H_1^T}_{(H_1\vec{a})^T} H_1 \vec{a} + \underbrace{\vec{a}^T H_2^T}_{(H_2\vec{a})^T} H_2 \vec{a} - \| H_1\vec{a} + H_2\vec{a} \|^2 = 0$$

$$\Longrightarrow \| H_1\vec{a} \|^2 + \| H_2\vec{a} \|^2 - \| H_1\vec{a} + H_2\vec{a} \|^2 = 0$$

$$\Longrightarrow \| H_1\vec{a} \|^2 + \| H_2\vec{a} \|^2 - \| H_1\vec{a} \|^2 - \| H_2\vec{a} \|^2 + 2 \| H_1\vec{a} \| \| H_2\vec{a} \| \cos(\theta(H_1\vec{a}, H_2\vec{a})) = 0$$

$\Longrightarrow$ Holds if

$\theta = 90^\circ$ or -90°

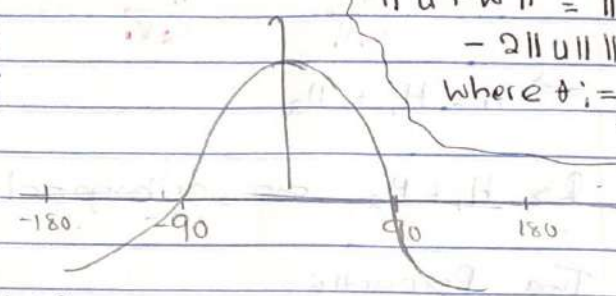
i.e. $H_1\vec{a} \perp H_2\vec{a}$

$\quad\;\; \|$

$\Longrightarrow c_1\vec{v}_1 \perp c_2\vec{v}_2$

$\Longrightarrow \vec{v}_1 \perp \vec{v}_2$

$\| \vec{u} + \vec{w} \|^2 = \|u\|^2 + \|w\|^2 - 2\|u\| \|w\| \cos(\theta)$ where $\theta :=$ ∡ between $\vec{v}, \vec{w}$



Let $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_d$ $\in \mathbb{R}^n$ $d \leq n$ be mutually orthogonal

$\vec{v}_{j_1} \perp \vec{v}_{j_2} \quad \forall j_1 \neq j_2 \qquad V = [\vec{v}_1 | \ldots | \vec{v}_d]$

$$\begin{aligned}
\operatorname{proj}_V(\vec{a}) &= \operatorname{proj}_{\vec{v}_1}(\vec{a}) + \ldots + \operatorname{proj}_{\vec{v}_d}(\vec{a}) \\
&= H_1\vec{a} + \ldots + H_d\vec{a} \\
&= \frac{\vec{v}_1\vec{v}_1^T}{\|\vec{v}_1\|^2}\vec{a} + \ldots + \frac{\vec{v}_d\vec{v}_d^T}{\|\vec{v}_d\|^2}\vec{a} \\
&= \underbrace{\left(\frac{\vec{v}_1\vec{v}_1^T}{\|\vec{v}_1\|^2} + \ldots + \frac{\vec{v}_d\vec{v}_d^T}{\|\vec{v}_d\|^2} \right)}_{H} \vec{a}
\end{aligned}$$

let $\|\vec{v}_1\| = \ldots = \|\vec{v}_d\| = 1$

$$Q = V = [\vec{v}_1 | \dots | \vec{v}_d]$$

[If orthonormal matrix then $V = Q$]

unit length and orthogonal

orthogonal = orthonormal

$$\text{proj}_V(\vec{a}) = \underbrace{(\vec{v}_1\vec{v}_1^T + \dots + \vec{v}_d\vec{v}_d^T)}_{H}\vec{a}$$

$$H = (v_{11}\vec{v}_1 | v_{12}\vec{v}_1 | \dots | v_{1n}\vec{v}_1) + (v_{21}\vec{v}_2 | v_{22}\vec{v}_2 | \dots | v_{2n}\vec{v}_2) + \dots + (v_{d1}\vec{v}_d | v_{d2}\vec{v}_d | \dots | v_{dn}\vec{v}_d)$$

$$\Longrightarrow [\vec{v}_1 | \vec{v}_2 | \dots | \vec{v}_d] \begin{bmatrix} \vec{v}_1^T \\ \vec{v}_2^T \\ \vdots \\ \vec{v}_d^T \end{bmatrix} = \underset{n\times d}{Q}\,\underset{d\times n}{Q^T} = \underset{n\times n}{H} \overset{\text{general}}{=} V(V^TV)^{-1}V^T$$

$$\sum_{j=1}^{d} \vec{v}_j\vec{v}_j^T$$

$$\boxed{\begin{array}{l} \text{If } \operatorname{colsp}[\underset{n\times d}{V}] = \operatorname{colsp}[\underset{n\times d}{Q}] \\ \Longrightarrow V(V^TV)^{-1}V^T = QQ^T \end{array}}$$

Proof :=

$$Q(Q^TQ)^{-1}Q^T = QQ^T$$

$$\begin{bmatrix} \leftarrow q_1^T \rightarrow \\ \leftarrow q_2^T \rightarrow \\ \vdots \\ \leftarrow q_d^T \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ q_1 & q_2 & \dots & q_d \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} = I_d.$$

$X \in \mathbb{R}^{n\times(p+1)}$ change of basis.

$$X \longrightarrow Q$$

$$\underset{n\times(p+1)}{X} = \underset{n\times(p+1)}{Q}\,\underset{(p+1)(p+1)}{R} \Longrightarrow_{\text{full rank}} Q = XR^{-1}.$$

R: upper triangular

Q − R decomposition.

$$X = QR = Q\begin{bmatrix} c & d & e \\ 0 & f & g \\ 0 & 0 & h \end{bmatrix}$$ upper triangular matrix $\|X_1\|^2$

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

$$\Longrightarrow (X^T X)\vec{b} = X^T \vec{y}$$

$$\Longrightarrow ((QR)^T (QR))\vec{b} = (QR)^T \vec{y}$$

$$\Longrightarrow R^T \underbrace{Q^T Q}_{I} R\vec{b} = R^T \underbrace{Q^T \vec{y}}_{\vec{z}}$$

$$\Longrightarrow R^T R \vec{b} = R^T \vec{z}$$

$$\Longrightarrow (R^T)^{-1} R^T R \vec{b} = (R^T)^{-1} R^T \vec{z}$$

$$\Longrightarrow R\vec{b} = \vec{z}$$

$$\begin{bmatrix} c & d & e \\ 0 & f & g \\ 0 & 0 & h \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

$$hb_2 = z_3 \Longrightarrow b_2 = \frac{z_3}{h}$$

$$fb_1 - gb_2 = z_2$$

$$\Longrightarrow fb_1 = z_2 - g\, z_1/n$$

$$\Longrightarrow b_1 = \frac{z_2 - g\, z_1/n}{f}$$

$$SST = SSR + SSE \qquad SST = SSR + SSE$$

$\sum (y_i - \bar{y})^2$ - a fixed function of y.

Does not change.

$SSR \uparrow$ $SSE \downarrow$

$R^2 \uparrow$, RMSE $\downarrow$

$$SSR = \sum_{i=1}^{n} (\hat{y} - \bar{y})^2 = \sum_{i=1}^{n} \hat{y}_i^2 - 2\bar{y} \sum_{i=1}^{n} \hat{y}_i + \sum_{i=1}^{n} \bar{y}^2.$$

$$= \|\vec{\hat{y}}\|^2 - 2n\bar{y}^2 + n\bar{y}^2 = \|\hat{y}\|^2 - n\bar{y}^2.$$

$$\sum \hat{y}_i = \vec{\hat{y}}^T \vec{1}_n = (H\vec{y})^T \vec{1}_n = \vec{y}^T H \vec{1}_n = \vec{y}^T \vec{1}_n = \sum y_i = n\bar{y}$$

$$\|\vec{a} + \vec{b}\|^2 = \|a\|^2 + \|b\|^2. \quad \vec{a} \perp \vec{b}$$

$$= \|H\vec{y}\|^2 - n\bar{y}^2 = \left\| \left(\sum_{j=0}^{p} \text{proj}_{q_j} (\vec{y}) \right) \right\|^2 - n\bar{y}^2.$$

$$X = \begin{bmatrix} \vdots & \uparrow & & \uparrow \\ \vdots & x_1 & \cdots & x_p \\ 1 & \downarrow & & \downarrow \end{bmatrix} \xrightarrow{G.S.} Q = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ q_0 & q_1 & \cdots & q_p \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

$$= \underbrace{\|\text{proj}_{\vec{q_0}} (\vec{y})\|^2}_{>0} + \underbrace{\|\text{proj}_{\vec{q_1}} (\vec{y})\|^2}_{>0} + \cdots + \underbrace{\|\text{proj}_{\vec{q_p}} (\vec{y})\|^2}_{>0} - n\bar{y}^2$$

new column

$\uparrow$ Rank

$$X = \left[\begin{array}{cccc|c} \vdots & \uparrow & & \uparrow & \uparrow \\ \vdots & x_1 & \cdots & x_p & \vec{x_*} \\ 1 & \downarrow & & \downarrow & \downarrow \end{array}\right] \xrightarrow{G.S.} Q = \left[\begin{array}{cccc|c} \uparrow & \uparrow & & \uparrow & \\ q_0 & q_1 & \cdots & q_p & \vec{q_*} \\ \downarrow & \downarrow & & \downarrow & \end{array}\right]$$

s.t $\vec{x_*}$ linearly independent of X.

$$= \underbrace{\|\text{proj}_{\vec{q_0}} (\vec{y})\|^2}_{>0} + \underbrace{\|\text{proj}_{\vec{q_1}} (\vec{y})\|^2}_{>0} + \cdots + \underbrace{\|\text{proj}_{\vec{q_p}} (\vec{y})\|^2}_{>0} +$$

due to the new column $\rightarrow$ $\underbrace{\|\text{proj}_{\vec{q_*}} (\vec{y})\|^2}_{>0} - n\bar{y}^2$

$SST = SSR + SSE$

$\uparrow$ increase $\downarrow$ decrease

$SST = SSR_* + SSE_*$

$\Longrightarrow SSR_* > SSR$ & $SSE_* < SSE$

$H = QQ^T$

$H_\perp = I - H = Q_\perp Q_\perp^T$

Let's add all $n-(p+1)$ columns to get.

$$X = \underbrace{\begin{bmatrix} \quad \end{bmatrix}}_{p+1} \underbrace{\begin{bmatrix} \quad\quad \end{bmatrix}}_{n-(p+1)} \in \mathbb{R}^{n \times n} \text{ full rank.}$$

$$\hat{y} = H\vec{y} = X(X^TX)^{-1}X^T\vec{y} = \overbrace{XX^{-1}}^{I}\overbrace{(X^T)^{-1}X^T}^{I}\vec{y}$$

$$= \underset{H}{\underset{\shortparallel}{I}}\,\vec{y} = \vec{y}$$

Is I orthogonal proj. matrix?

$I^T = I$

$II = I$ ✓

Bad for generalization error / future performance

$$\Longrightarrow SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum 0^2 = 0$$

Predicted means performance on $\mathbb{D}$ which is in sample "overfitting"

$\Longrightarrow R^2 = 1, \quad RMSE = 0.$

Core concept in data science plague of data science in the real world

Why is this bad?

New columns are most related to $\vec{z}_j$'s,

the

$$Y = g(x) + \underbrace{\overbrace{\left(h^*(\vec{x}) - g(\vec{x})\right)}^{\text{estimation error}} + \underbrace{\overbrace{\left(h^*(\vec{x}) - f(\vec{x})\right)}^{\text{misspecification error}} + \underbrace{\overbrace{\left(t(\vec{z}) - f(\vec{x})\right)}^{\text{Ignorance error}}}_{\delta}}_{\mathcal{E}}}_{e}$$

Assume overfit by adding random columns?

$\implies X_1, \ldots, X_p$ are the data that truly matters.

$\implies f(\vec{x})$ is fixed & $h^*(\vec{x})$ is fixed.

$$h^*(\vec{x}) = \vec{x}\,\vec{\beta} \implies \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p.$$

$$\{ f(\vec{x}) = \sin(y, x_2) + x_3^{\,1} + \ldots + \ldots \}$$

$\implies$ overfitting error is estimator error.

$g(x) = \vec{x}\,\vec{b}$

As you overfit $\vec{b}$ & $\vec{\beta}$ diverge.