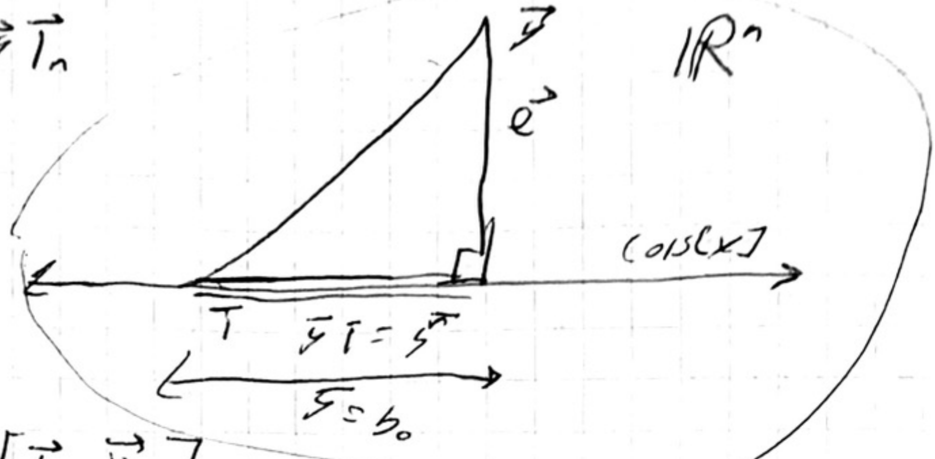
Lecture 10

Let's examine the null model $p=0$ so that $X=[\vec{1}_n] \Rightarrow \vec{b}=b_0=\bar{y}$

$$H = X\underbrace{(X^TX)^{-1}}_{\frac{1}{n}}X^T = \frac{1}{n}\vec{1}\vec{1}^T = \frac{1}{n}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix}$$

$$\hat{y} = H\vec{y} = \begin{bmatrix} \bar{y} \\ \bar{y} \\ \vdots \\ \bar{y} \end{bmatrix} = \bar{y}\vec{1}_n$$

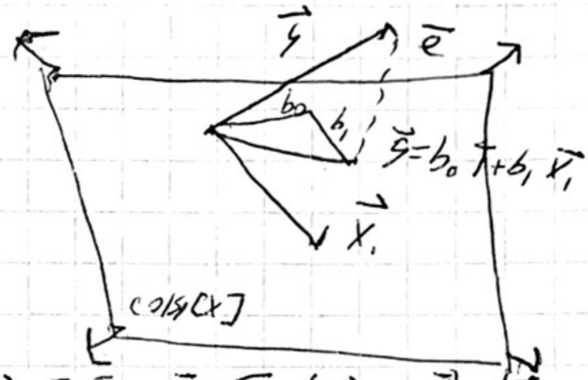


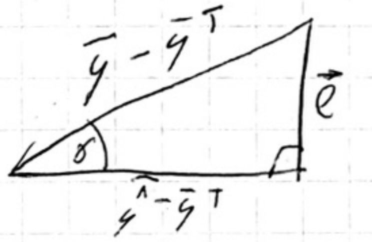
Consider $p=1$ so that $X = [\vec{1} \ \ \vec{x}_1]$

Py theorem tells us

$$\|\hat{y}\|^2 + \|\vec{e}\|^2 = \|\vec{y}\|^2$$

$$\cos^2(\theta) = \frac{\|\hat{y}\|^2}{\|\vec{y}\|^2}$$





$$\vec{e} := \vec{y} - \hat{y} = \vec{y} - \bar{y}\vec{1} + \bar{y}\vec{1} - \hat{y} = (\vec{y} - \bar{y}\vec{1}) - (\hat{y} - \bar{y}\vec{1})$$

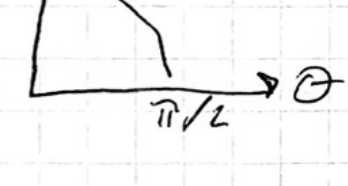
$$\text{Proj}_{\text{colsp}[X]}(\vec{y} - \bar{y}\vec{1}) = H(\vec{y} - \bar{y}\vec{1}) = H\vec{y} - \bar{y}H\vec{1}$$
$$= \hat{y} - \bar{y}\vec{1}$$

$$\Downarrow$$

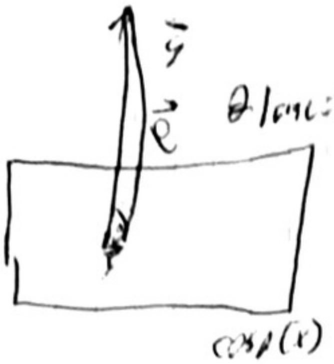
$$\|\vec{y} - \bar{y}\vec{1}\|^2 = \|\hat{y} - \bar{y}\vec{1}\|^2 + \|\vec{e}\|^2, \quad R^2 = \frac{SST - SSE}{SST} = \frac{SSR}{SST} = \cos^2(\theta) \in [0,1]$$

$$\underbrace{\sum (y_i - \bar{y})^2}_{SST} = \underbrace{\sum (\hat{y}_i - \bar{y})^2}_{SSR} + \underbrace{\sum e_i^2}_{SSE}$$

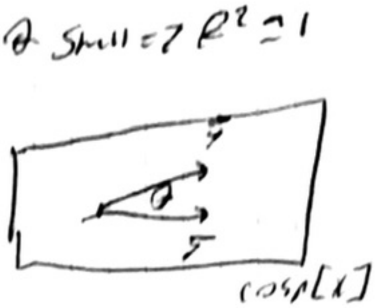
$$SST = SSR + SSE$$



Sum of Squares total   Sum of Squares regression   Sum of Squares error



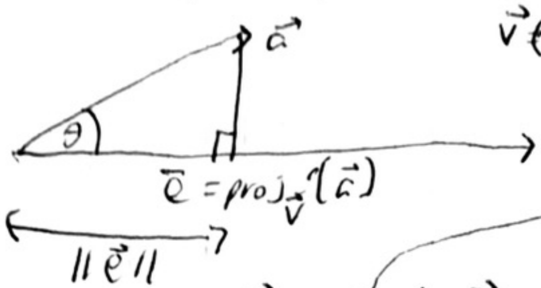
θ large $\Rightarrow R^2 \approx 0$



θ small $\Rightarrow R^2 \approx 1$

Back to linear algebra

By law of cosines



$\vec{v} \in R^n$

$$\cos(\theta) = \frac{\vec{a}\cdot\vec{v}}{\|\vec{a}\|\,\|\vec{v}\|} = \frac{\|\vec{e}\|}{\|\vec{a}\|}$$

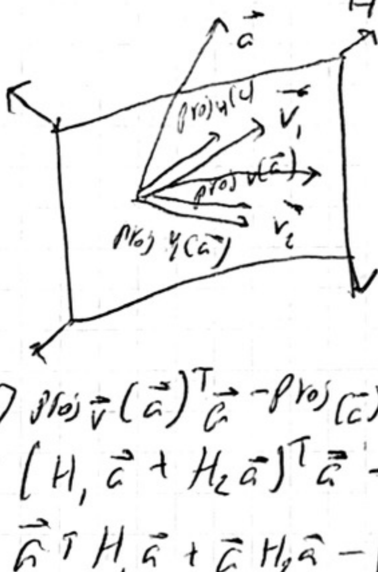
$$\|\vec{e}\| = \frac{\vec{a}\cdot\vec{v}}{\|\vec{v}\|} \quad \text{def cosine}$$

$$\Rightarrow \vec{e} = \|\vec{e}\|\cdot\frac{\vec{v}}{\|\vec{v}\|} = \frac{\vec{a}\cdot\vec{v}}{\|\vec{v}\|^2}\vec{v} = \frac{\vec{a}^T\vec{v}\vec{v}}{\|\vec{v}\|^2} = \frac{\vec{v}\vec{v}^T}{\|\vec{v}\|^2}\vec{a} = H\vec{a}$$

$$H = \frac{1}{\|\vec{v}\|^2}\vec{v}\vec{v}^T = \left[\frac{v_1}{\|\vec{v}\|^2}\vec{v} \;\middle|\; \frac{v_2}{\|\vec{v}\|^2}\vec{v} \;\middle|\; \cdots \;\middle|\; \frac{v_n}{\|\vec{v}\|^2}\vec{v}\right]; \quad \text{rank}[H] = 1$$

$$HH = \left(\frac{1}{\|\vec{v}\|^2}\vec{v}\vec{v}^T\right)\left(\frac{1}{\|\vec{v}\|^2}\vec{v}\vec{v}^T\right) = \left(\frac{1}{\|\vec{v}\|^4}\,\vec{v}\underbrace{\vec{v}^T\vec{v}}_{\|\vec{v}\|^2}\vec{v}^T\right) = \frac{1}{\|\vec{v}\|^2}\vec{v}\vec{v}^T = H$$

$$V = [\vec{v}_1 \mid \vec{v}_2] \quad \underbrace{proj_V(\vec{a})}_{H\vec{a}} \overset{?}{=} \overbrace{proj_{\vec{v}_1}(\vec{a})}^{H_1\vec{a}} + \overbrace{proj_{\vec{v}_2}(\vec{a})}^{H_2\vec{a}} = (H_1 + H_2)\vec{a}$$



Will always project onto colsp[V] but H may not be the correct length (It can over/under count). The correct length gives you the right angle

$$proj_V(\vec{a})^T(\vec{a} - proj_V(\vec{a})) = 0$$

angle between $\vec{v}, \vec{v}$

$$\|\vec{u}+\vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 + 2\|\vec{u}\|\,\|\vec{v}\|\cos(\theta)$$

$$\Rightarrow proj_{\vec{V}}(\vec{a})^T\vec{a} - proj_V(\vec{a})^T proj_V(\vec{a})$$

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

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

angle between $\vec{v}_1, \vec{v}_2$

$\in [0,1]$

$(H_1\vec{a})^T(H_1\vec{a})$ | $(H_2\vec{a})^T(H_2\vec{a})$

$\vec{a}^T H H_1\vec{a}$ | $\vec{a}^T H_1 H_2\vec{a}$

$\vec{a}^T H_1\vec{a}$ | $\vec{a}^T H_2\vec{a}$

The only way to make this expression zero is if $\cos(\theta)=0$ i.e. θ = a right angle. Thus, the full projection is a sum of the component projections if the components are orthogonal

Let $V=[\vec{v}_1|\vec{v}_2|\dots|\vec{v}_d] \in \mathbb{R}^{n\times d}$, $\forall i,j \; \vec{v}_i\cdot\vec{v}_j=0$ $\quad V = [\vec{v}_1 \cdots \vec{v}_d]$

$\Rightarrow \text{proj}_{\text{colsp}[V]}(\vec{a}) = \text{proj}_{\vec{v}_1}(\vec{a}) + \dots + \text{proj}_{\vec{v}_d}(\vec{a})$

$$= \frac{\vec{v}_1\vec{v}_1^T}{\|\vec{v}_1\|^2}\vec{a} + \dots + \frac{\vec{v}_d\vec{v}_d^T}{\|\vec{v}_d\|^2}\vec{a}$$

$$= \left(\frac{\vec{v}_1\vec{v}_1^T}{\|\vec{v}_1\|^2} + \dots + \frac{\vec{v}_d\vec{v}_d^T}{\|\vec{v}_d\|^2}\right)\vec{a}$$

$$= \left(\vec{v}_1\vec{v}_1^T + \dots + \vec{v}_d\vec{v}_d^T\right)\vec{a} = \underbrace{\left(\sum_{i=1}^{d} v_i v_i^T\right)}_{H}\vec{a}$$

If $\|\vec{v}_1\| = \|\vec{v}_2\| = \dots = \|\vec{v}_d\| = 1$ i.e. all units together

$Q = [\vec{v}_1|\dots|\vec{v}_d]$ which is an orthonormal matrix

$$\underbrace{Q^TQ}_{d\times d} = \begin{bmatrix} \leftarrow \vec{v}_1^T \rightarrow \\ \vdots \\ \leftarrow v_d^T \rightarrow \end{bmatrix}\begin{bmatrix} \uparrow & & \uparrow \\ \vec{v}_1 & \cdots & \vec{v}_d \\ \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \ddots & 1 & & \ddots & 0 \\ & \ddots & 1 & \vdots & 0 \\ & & & \ddots & \\ & & & & 1 \end{bmatrix} = I_d$$

$$\underbrace{QQ^T}_{n\times n} = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_d \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}\begin{bmatrix} \leftarrow v_1^T \rightarrow \\ \leftarrow v_2^T \rightarrow \\ \vdots \\ \leftarrow v_d^T \rightarrow \end{bmatrix} = \vec{v}_1\vec{v}_1^T + v_2v_2^T + \dots + \vec{v}_d v_d^T = H$$

$$= \begin{bmatrix} A_1 & A_2 & \cdots & A_d \end{bmatrix}\begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_d \end{bmatrix} = A_1B_1 + A_2B_2 + \dots + A_dB_d$$

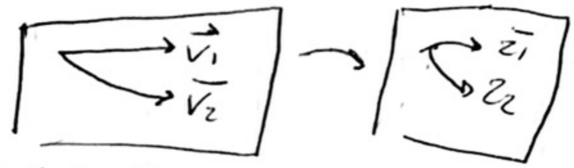
$\Rightarrow QQ^T = V(V^TV)^{-1}V^T = H$

where the columns of Q are the orthonormalized columns of $V=[v_1|\dots|v_d]$.
Further $\text{colsp}[Q] = \text{colsp}[V]$ since the column vectors in Q represents a change of basis of the column vectors of V

$$\text{proj}_{\text{colsp}[Q]}(\vec{a}) = Q\underbrace{(Q^TQ)^{-1}}_{I}Q^T = QQ^T$$

How can we convert matrix V to matrix Q? There is a computational algorithm called "Gram-Schmidt" and during the computation, you can collect a matrix that is the change of basis

$$\underset{n\times d}{V} = \underset{n\times d}{Q}\,\underset{d\times d}{R} \Rightarrow V R^{-1} = Q$$



$\text{colsp}[V] = \text{colsp}[Q]$

This is also called Q-R decomposition of a matrix. R will be upper triangular and full rank (and invertible)