Lecture Notes for Theory of Linear Models

Vector Space and Projection

Longhai Li
Department of Mathematics and Statistics
University of Saskatchewan

Vector and Projection

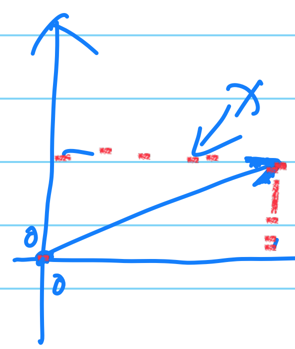
- Vector and Geometry
- Inner Product and Perpendicular
- Projection to a Single Vector
- Pythagorean theory
- Shortest distance property of projection

Vector

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$ column vector

a point in $\mathbb{R}^n$

$\overrightarrow{0x}$

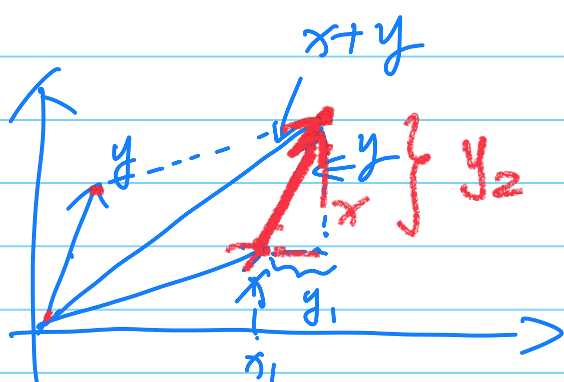


$(x_1, x_2)^T = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$n = 2$

$\mathbb{R}^n$

Addition

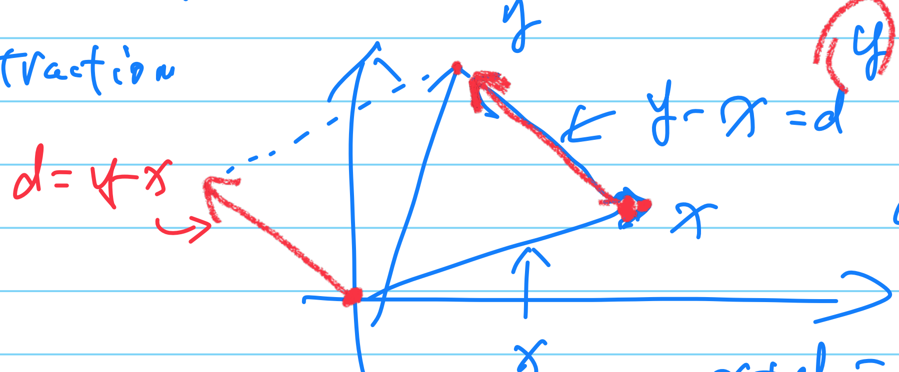


$x + y$

$$= \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

Subtraction

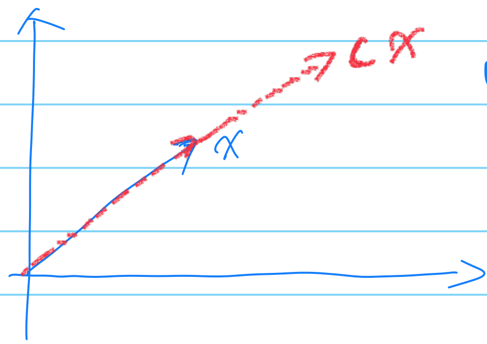
$y - x$



$d = y - x$

$x + d = y$

Multiplication by a scalar

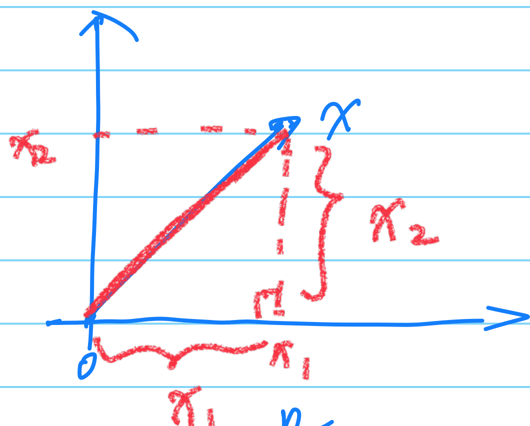


written with matrix multiplication

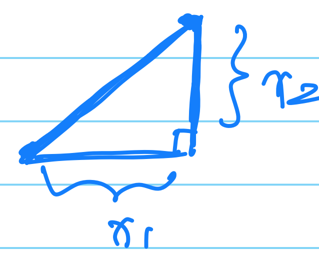
$$\underset{n\times 1}{x}\ \underset{1\times 1}{[c]}$$

$$cx = x[c] \text{ not } [c]x$$

Length of Vector (Euclidean Distance)



$x = (x_1, \ldots, x_n)'$

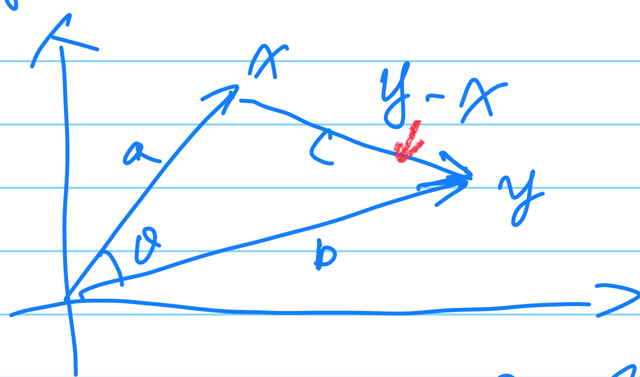


$$\|x\|^2 = \sum_{i=1}^{n} x_i^2$$

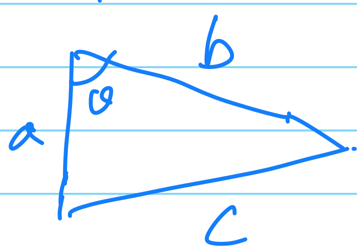
$$\|x\| = \sqrt{\|x\|^2} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

Euclidean distance

Angle (Inner Product)

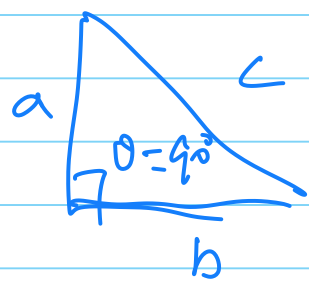


$$\theta = 90^\circ \ \left(\frac{\pi}{2}\right)$$



$$c^2 = a^2 + b^2 \quad \text{P.T.}$$

$$c^2 = a^2 + b^2 - 2ab\cos\theta$$



plugging $a = \|x\|$, $b = \|y\|$, $c = \|x - y\|$:

$$\|y - x\|^2 = \|x\|^2 + \|y\|^2 - 2\|x\| \cdot \|y\| \cos\theta$$

$$\|y-x\|^2 = \sum_{i=1}^{n} (x_i-y_i)^2$$

$$= \sum_{i=1}^{n} \left(x_i^2+y_i^2-2x_iy_i\right)$$

$$= \|x\|^2+\|y\|^2-2\cdot x'y$$

$$y'x = x'y = \sum_{i=1}^{n} x_iy_i$$

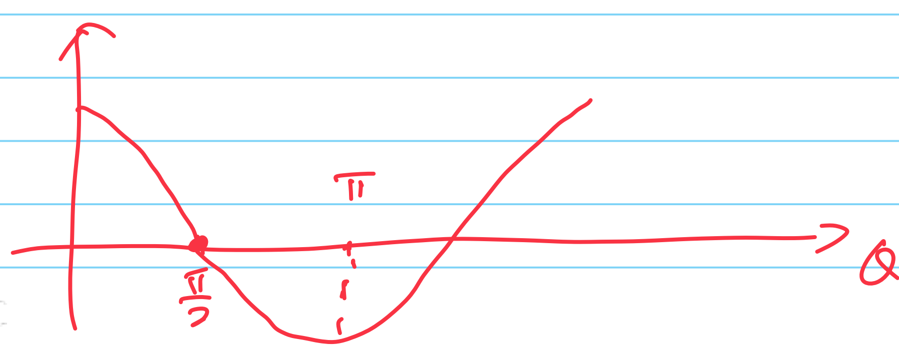
$$= x\cdot y = \langle x, y\rangle = \langle y, x\rangle$$

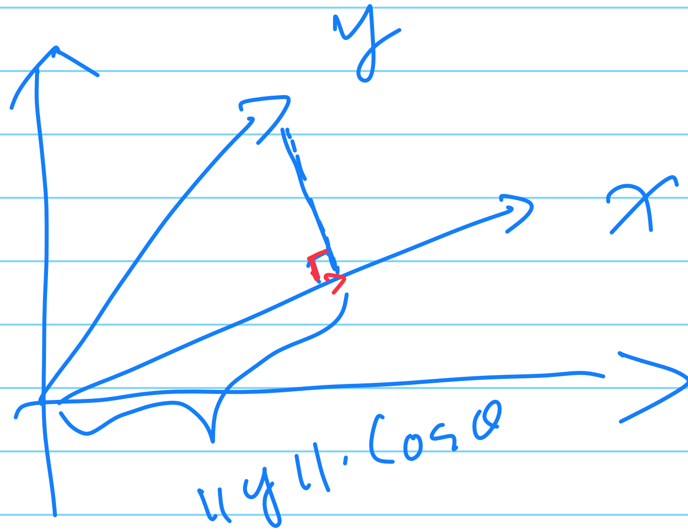
is called inner product

$$\cancel{\|x\|^2}+\cancel{\|y\|^2}\cancel{-}\cancel{2}\cdot x'y = \cancel{\|x\|^2}+\cancel{\|y\|^2}$$
$$\cancel{-}\cancel{2}\,\|x\|\cdot\|y\|\cdot\cos\theta$$

$$\underline{x'y} = \|x\|\cdot\|y\|\cdot\cos\theta$$

$$= \|y\|\cdot\cos\theta\cdot\|x\|$$





$$x'y = \boxed{\|y\| \cdot \cos\theta} \cdot \|x\|$$

$$\|y\| \cdot \cos\theta = \frac{x'y}{\|x\|} = \left\langle \frac{x}{\|x\|}, y \right\rangle$$

co-ordinate

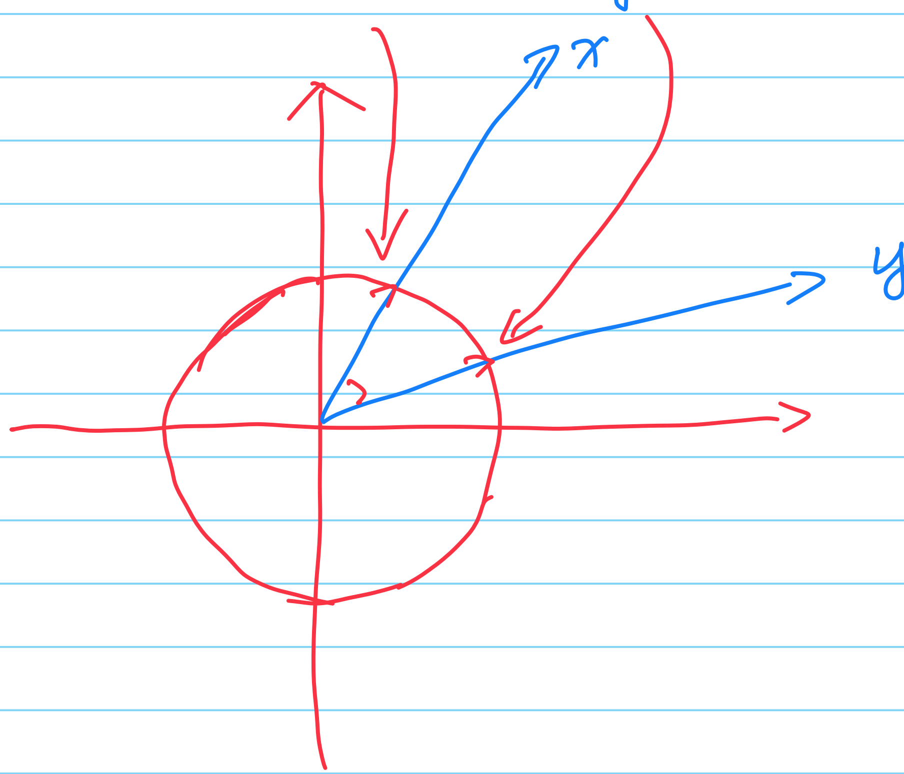
is the length of the projection
(+/-)
of y onto x.

$$x' y = \|x\| \cdot \|y\| \cdot \cos\theta$$

$$\cos\theta = \frac{x' y}{\|x\| \cdot \|y\|}$$

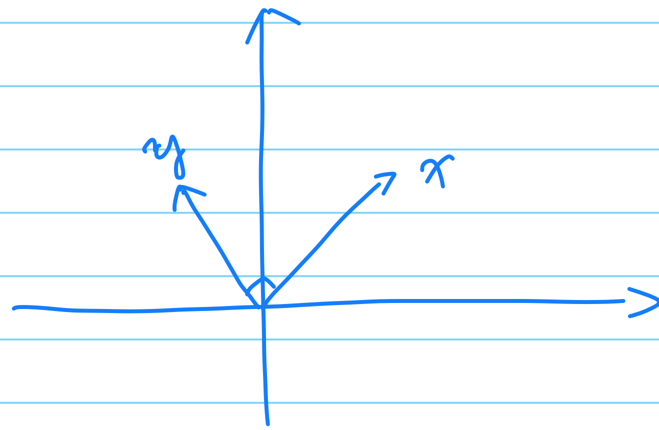
$$= \left(\frac{x}{\|x\|}\right)' \cdot \left(\frac{y}{\|y\|}\right)$$

$$= \left\langle \frac{x}{\|x\|}, \frac{y}{\|y\|} \right\rangle$$

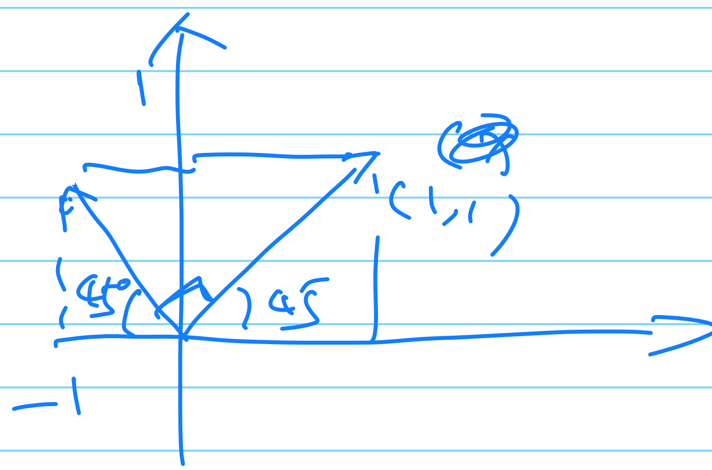


Perpendicular

$x \perp y \iff x' \cdot y = 0$



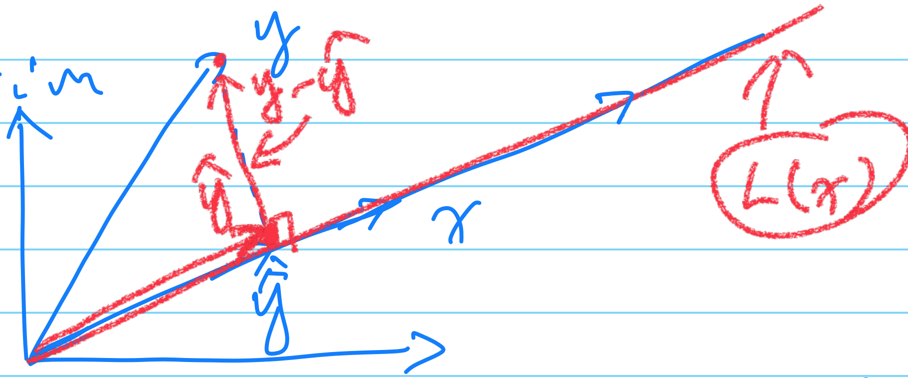
Example



$x = (1, 1)', \quad y = (-1, 1)$

$x'y = 1 \times (-1) + 1 \times 1 = 0 \Rightarrow x \perp y$

Projection



Def:

$\hat{y}$ is the projection of y onto $L(x)$ if

$\hat{y}$ is a vector in $L(x) = \{cx \mid c \in \mathbb{R}\}$

i.e, $\hat{y} = c \cdot x$ for $c \in \mathbb{R}$

such that $\underline{y - \hat{y}} \perp x$

Let's find an expression of $\hat{y}$

$$x' \cdot (y - \hat{y}) = 0$$

$$x' y - x' \cdot (c x) = 0$$

$$x' y = c \cdot x' x = c \cdot \|x\|^2$$

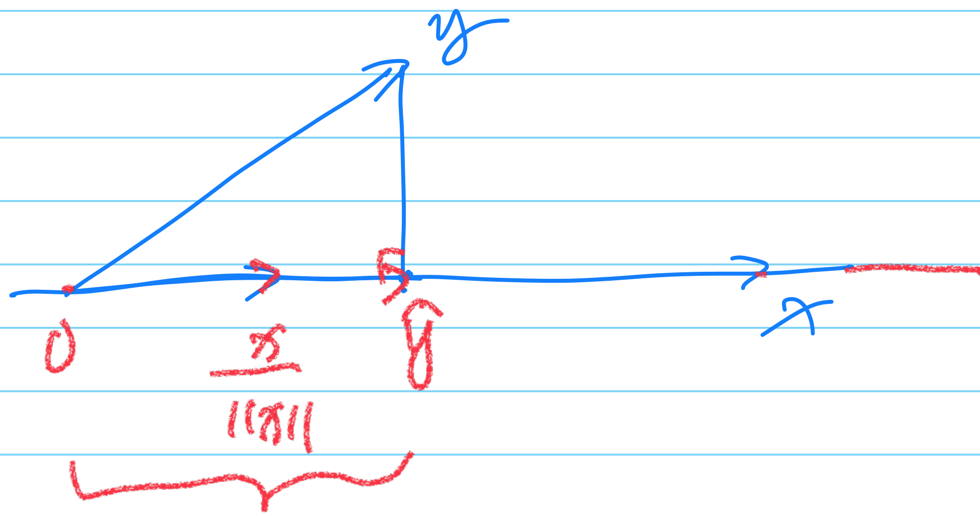
$$c = \frac{x' \cdot y}{\|x\|^2}$$

$$\hat{y} = \frac{x' \cdot y}{\|x\|^2} \cdot x$$

$$= \left(\frac{x}{\|x\|}\right)' y \cdot \left(\frac{x}{\|x\|}\right)$$

scale

direction



$\langle y, \frac{x}{\|x\|} \rangle$

Notation:

$$\hat{y} = \text{proj}(y \mid x) = P(y \mid x)$$

$$= \frac{x'y}{\|x\|} \cdot \frac{x}{\|x\|}$$

$$= \frac{x'y}{\|x\|^2} \cdot x$$

$$= x \frac{x'y}{\|x\|^2}$$

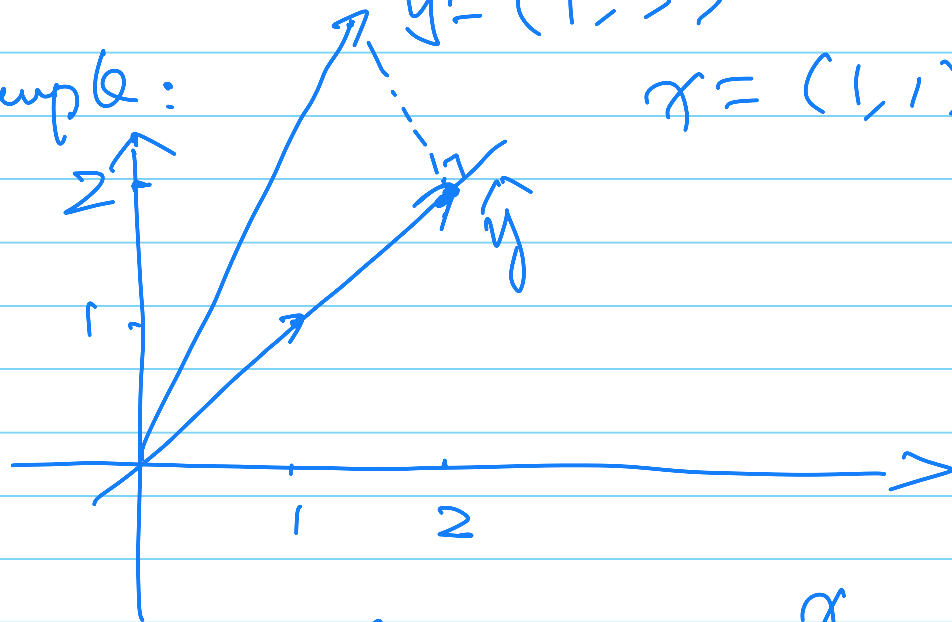
$$= \underset{p\times p}{\frac{xx'}{\|x\|^2}} \underset{p\times 1}{y} = P_x \cdot y$$

$x \in \mathbb{R}^p$, $\underbrace{\underset{p\times 1}{x} \cdot \underset{1\times p}{x'}}_{p\times p}$

Example:

$y = (1, 3)$

$x = (1, 1)' = j_2$



$$\hat{y} = \left\langle \frac{x}{\|x\|}, y \right\rangle \cdot \frac{x}{\|x\|}$$

$$= \left\langle \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \end{pmatrix} \right\rangle \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

$$= \frac{4}{\sqrt{2}} \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

$$= \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

$$\hat{y} = \frac{x \cdot x'}{\|x\|^2} \, y$$

$$P_x = \begin{pmatrix}1\\1\end{pmatrix} \cdot (1, 1) / 2$$

$$= \frac{1}{2}\begin{pmatrix}1 & 1\\1 & 1\end{pmatrix}$$

$$\hat{y} = \frac{1}{2}\begin{pmatrix}1 & 1\\1 & 1\end{pmatrix} \cdot \begin{pmatrix}1\\3\end{pmatrix}$$

$$= \frac{1}{2}\begin{pmatrix}4\\4\end{pmatrix} = \begin{pmatrix}2\\2\end{pmatrix}$$

Example:

$y = (y_1, \ldots, y_n)'$

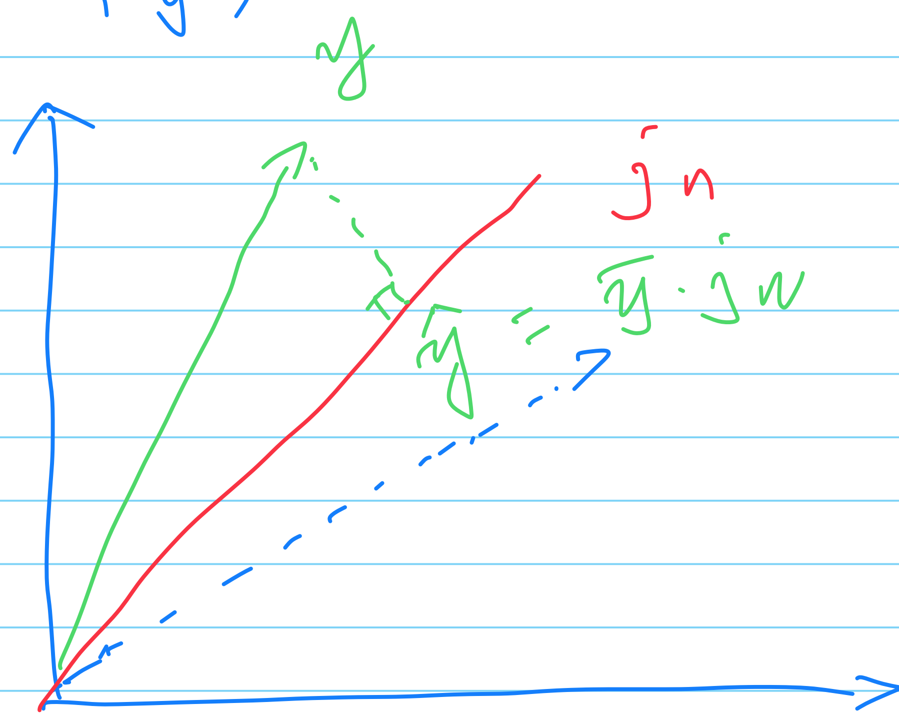
$j_n = (1, 1, \ldots, 1)'$

$\text{proj}(y \mid j_n)$

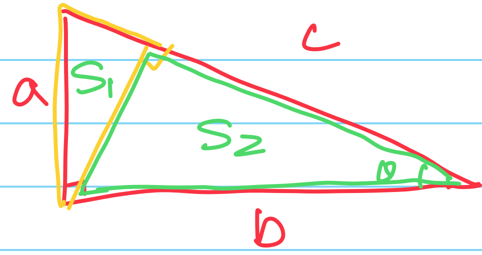
$$= \frac{j_n j_n'}{\|j_n\|^2} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$= \frac{1}{n} \cdot \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & & 1 \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

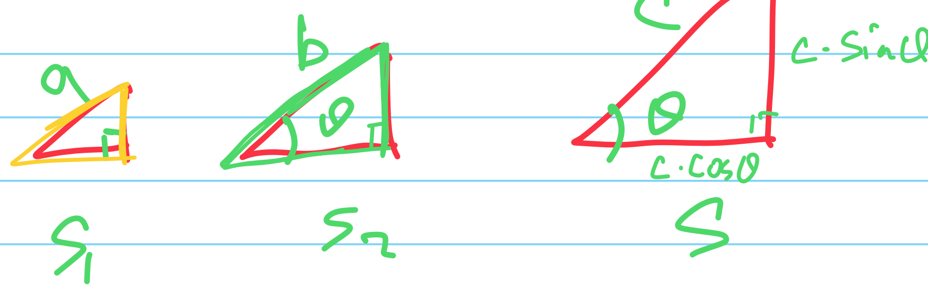
$$= \begin{pmatrix} \bar{y} \\ \vdots \\ \bar{y} \end{pmatrix} = \bar{y} \cdot j_n$$



Pythagorean Theorem in Geometry



$$S = S_1 + S_2$$



$S_1 = a^2 \cdot k$, where $k = \frac{1}{2}\cos\theta \cdot \sin\theta$

$S_2 = b^2 \cdot k$

$S = c^2 \cdot k$

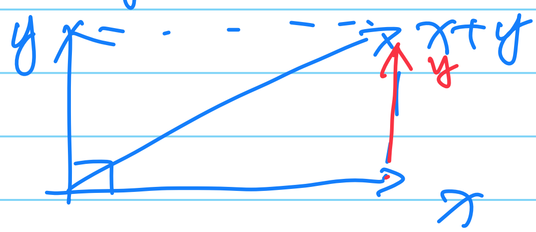
$c^2 \cdot k = a^2 \cdot k + b^2 \cdot k$

$c^2 = a^2 + b^2$

Pythagoreen Theorem (P.T.)

If $x \perp y \iff x'y = 0$

then $\|x+y\|^2 = \|x\|^2 + \|y\|^2$



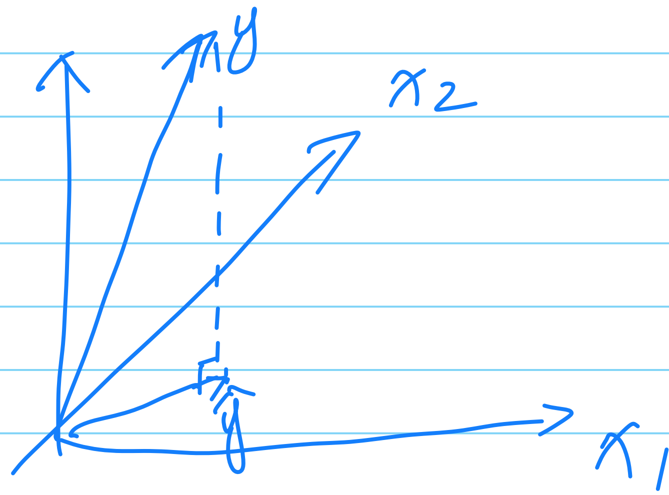
Pf:

$$\|x+y\|^2 = (x+y)'(x+y)$$

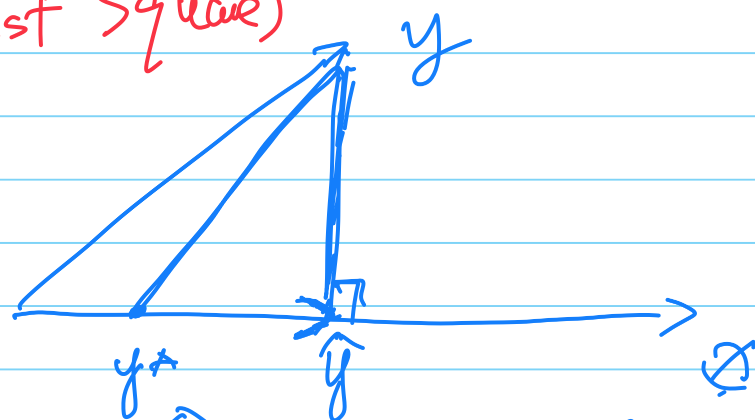
$$= x'x + x'y + y'x + y'y$$

$$= \|x\|^2 + \|y\|^2 + 2 \cdot \cancelto{0}{y'x}$$

$$= \|x\|^2 + \|y\|^2$$



shortest distance prop. of projection
(Least Square)



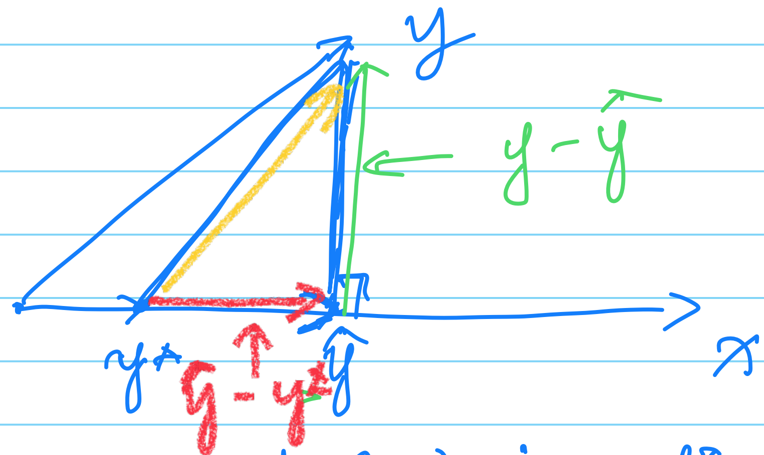
$p(y|x) = \hat{y}$ is defined as follows:

$\hat{y} = cx$ s.t. $\underline{\hat{y} - y} \perp x$

$\hat{y}$ is the vector in $L(x)$ that is closest to y.

For any $y^* \in L(x)$, $\|y - \hat{y}\| \le \underline{\|y - y^*\|}$

pf:



Suppose $y^* \in L(x)$, i.e. $y^* = b x$ for some $b \in \mathbb{R}$

$y-\hat{y} \perp x \Rightarrow y-\hat{y} \perp cx$, for any $c \in \mathbb{R}$

$y-\hat{y} \perp y^*$, $(y-\hat{y})' y^* = 0$

$y-\hat{y} \perp \hat{y}$, $(y-\hat{y})' \hat{y} = 0$

$y-\hat{y} \perp \hat{y}-y^*$, $(y-\hat{y})'(\hat{y}-y^*) = 0$

$$y - y^* = \underline{y-\hat{y}} + \underline{\hat{y}-y^*}$$

By P.T.,

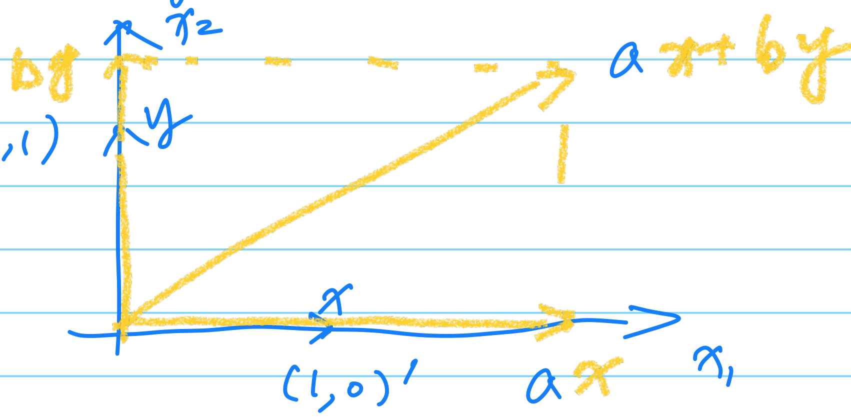
$$\|y-y^*\|^2 = \|y-\hat{y}\|^2 + \|\hat{y}-y^*\|^2 \geq \|y-\hat{y}\|^2$$

Basics of Vector Space

- **Vector Space**
- **Vector Space Spanned by Vectors**
- **Rank/Dimension of Vector Space**

Vector space

Example



$$x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad y = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$L(x, y) = \mathbb{R}^2$$

V, a subset of $\mathbb{R}^n$, is a vector space if

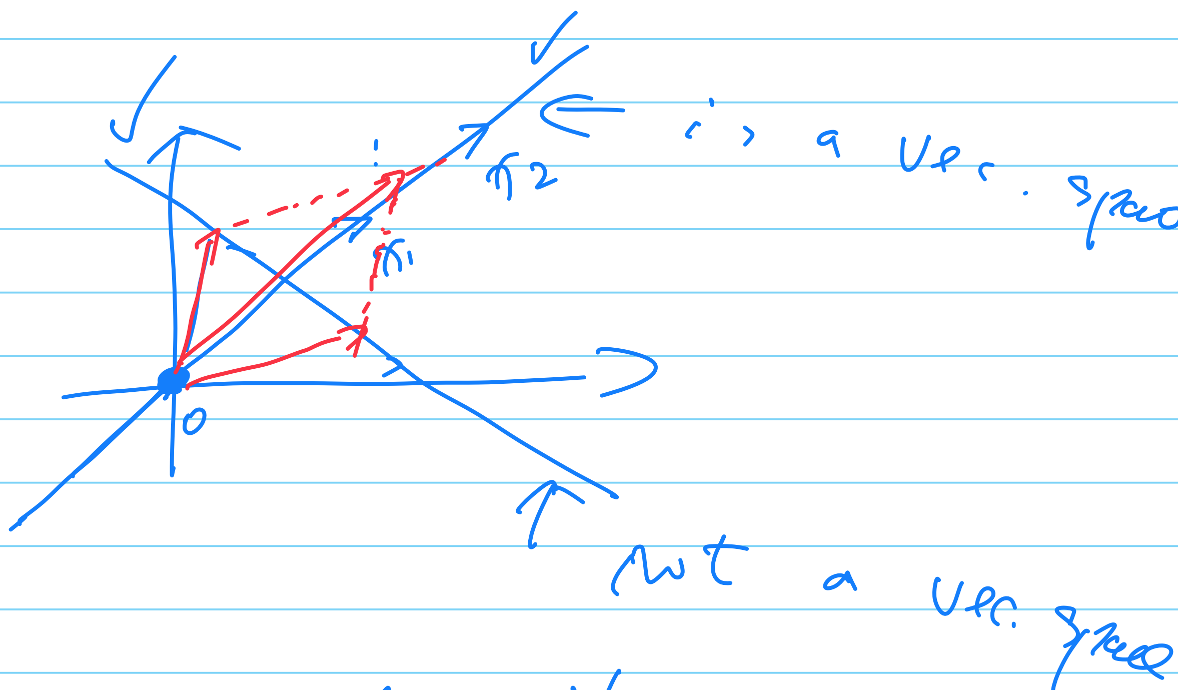
(1) $x_i, x_j \in V \Rightarrow x_i + x_j \in V$

(2) $x \in V \Rightarrow c \cdot x \in V$

(including $c = 0$)

closed under addition & scaling

Example



If $x_1, \dots, x_k \in V$

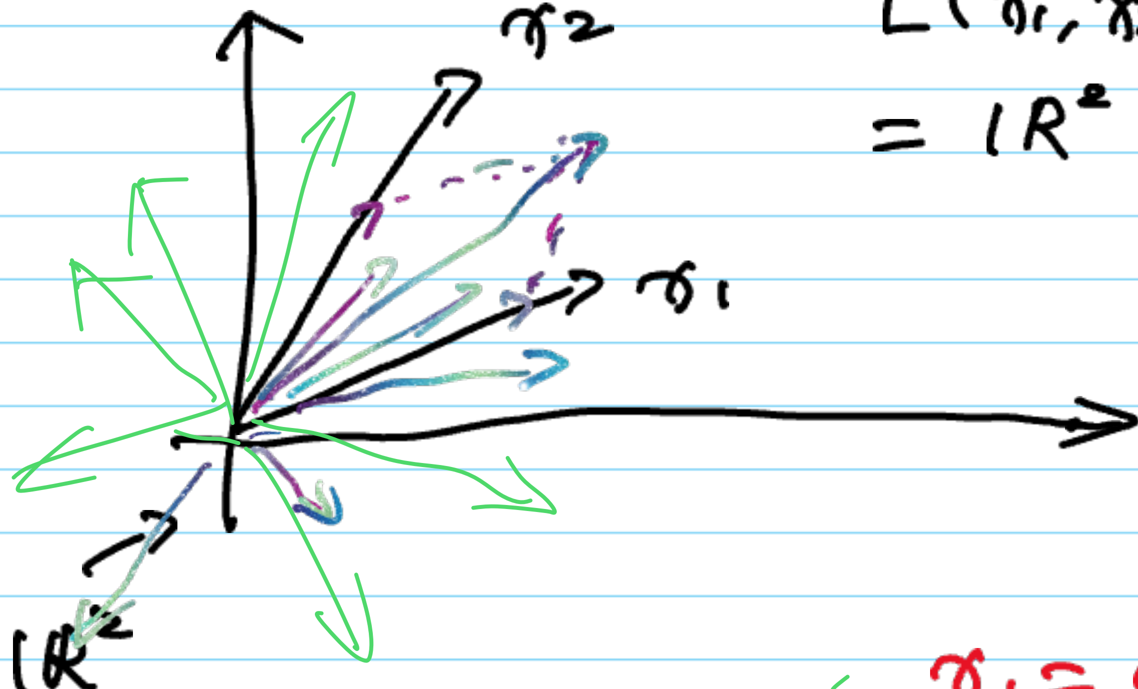
then $c_1 x_1 + c_2 x_2 + \dots + c_k x_k \in V$

closed under linear combination

spanned vector space

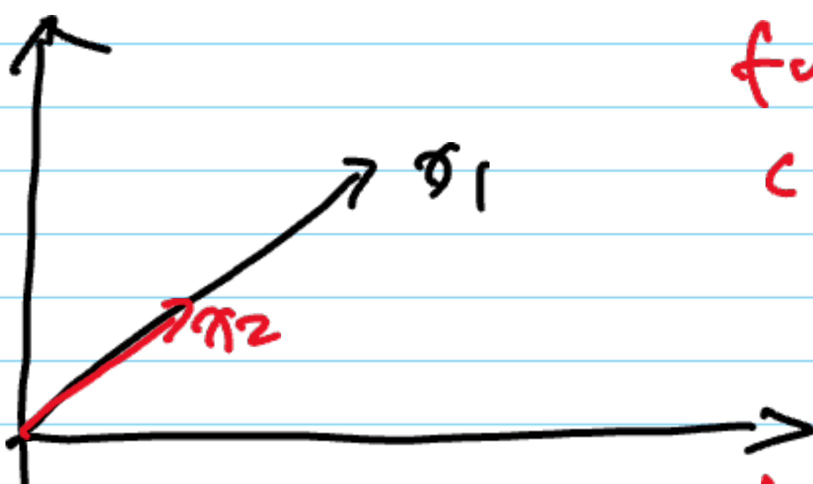
$$L(x_1, \cdots, x_p)$$

$$= \{ x \mid x = c_1 x_1 + \cdots + c_p x_p, \ c_i \in \mathbb{R} \}$$

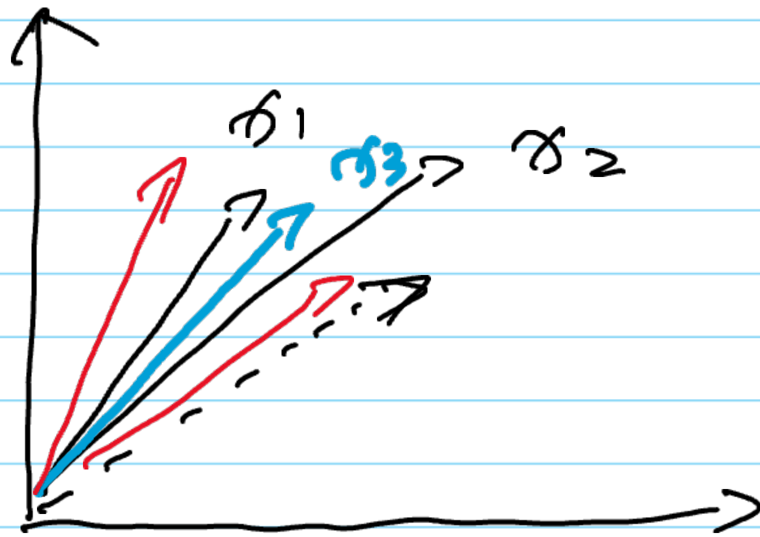


$$L(x_1, x_2) = \mathbb{R}^2$$

$x_1 = c \cdot x_2$ for some $c \in \mathbb{R}$



$$L(x_1, x_2) = L(x_1) = L(x_2)$$



(1) $x_3 = c_1 x_1 + c_2 x_2$

$L(x_1, x_2, x_3) = L(x_1, x_2)$

(2) $x_3 \notin L(x_1, x_2)$

$L(x_1, x_2, x_3) = \mathbb{R}^3$

Column space & row space

$X = (x_1, x_2, \dots, x_p)$

$\text{column}(X) = c(X) = L(x_1, \dots, x_p)$

$$X = \begin{pmatrix} r_1' \\ r_2' \\ \vdots \\ r_n' \end{pmatrix}$$

$\text{row}(X) = r(X) = L(r_1, r_2, \dots, r_n)$

Linear independence (LIN)

$x_1, \dots, x_p$ are LIN if

$$\sum_{i=1}^{p} c_i x_i = 0 \Rightarrow c_i = 0$$

$x_1, \cdots, x_p$ are NOT LIN If

$\exists\, i,\quad x_i \in L(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_p)$

s.t. $\exists\, b_1, b_2, \cdots, b_{i-1}, b_{i+1}, \cdots, b_p$ s.t.

$$x_i = b_1 x_1 + b_2 x_2 + \cdots + b_{i-1} x_{i-1} + b_{i+1} x_{i+1} + \cdots + b_p x_p$$

$x_1, \cdots, x_p$ X: $n \times p$ matrix

$X = (x_1, \cdots, x_p)$, how many linearly indep. (LIN) vectors ?

$\text{rank}(X) =$

(1) # of LIN vect. in $x_1, \cdots, x_p$

(2) $\text{Dim}(L(x_1, \cdots, x_p))$

Properties of $\mathrm{rank}(X)$

X: $n \times p$ matrix

(1) $\mathrm{rank}(X) = \mathrm{rank}(X')$

Another equivalence of (1):

$$\mathrm{Dim}\left(c(X)\right) = \mathrm{Dim}\left(r(X)\right)$$

(2) $\mathrm{rank}(X) \leq \min(n, p)$

Proof that column rank is equal to row rank:

Let A be an $m \times n$ matrix. Let the column rank of A be r, and let $\mathbf{c}_1, \ldots, \mathbf{c}_r$ be any basis for the column space of A. Place these as the columns of an $m \times r$ matrix C. Every column of A can be expressed as a linear combination of the r columns in C. This means that there is an $r \times n$ matrix R such that $A = CR$. R is the matrix whose ith column is formed from the coefficients giving the ith column of A as a linear combination of the r columns of C. In other words, R is the matrix which contains the multiples for the bases of the column space of A (which is C), which are then used to form A as a whole. Now, each row of A is given by a linear combination of the r rows of R. Therefore, the rows of R form a spanning set of the row space of A and, by the Steinitz exchange lemma, the row rank of A cannot exceed r. This proves that the row rank of A is less than or equal to the column rank of A. This result can be applied to any matrix, so apply the result to the transpose of A. Since the row rank of the transpose of A is the column rank of A and the column rank of the transpose of A is the row rank of A, this establishes the reverse inequality and we obtain the equality of the row rank and the column rank of A.

Source: https://en.wikipedia.org/wiki/Rank_(linear_algebra)

$$\underset{m\times n}{A} = \underset{m \times r}{(c_1, \ldots, c_r)} \cdot \underset{r \times n}{R}$$

$$= C \cdot \begin{bmatrix} b_1' \\ \vdots \\ b_r' \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{j=1}^{r} c_{1j} b_j' \\ \vdots \\ \sum_{j=1}^{r} c_{mj} b_j' \end{bmatrix} = \begin{bmatrix} a_1' \\ \vdots \\ a_n' \end{bmatrix}$$

where $a_i' = \sum_{j=1}^{r} c_{ij} b_j'$

We can easily see that $a_i \in \mathrm{row}(R)$

Example:

$$X = \begin{array}{cccc} x_1 & x_2 & x_3 & \\ \left(\begin{matrix} 1 & 4 & 6 \\ 2 & 8 & 12 \end{matrix}\right) & & & \begin{matrix} \leftarrow r_1' \\ \leftarrow r_2' \end{matrix} \end{array}$$

$x_2 = 4x_1 \qquad x_3 = 6x_1$

$\text{rank}(X) = 1$

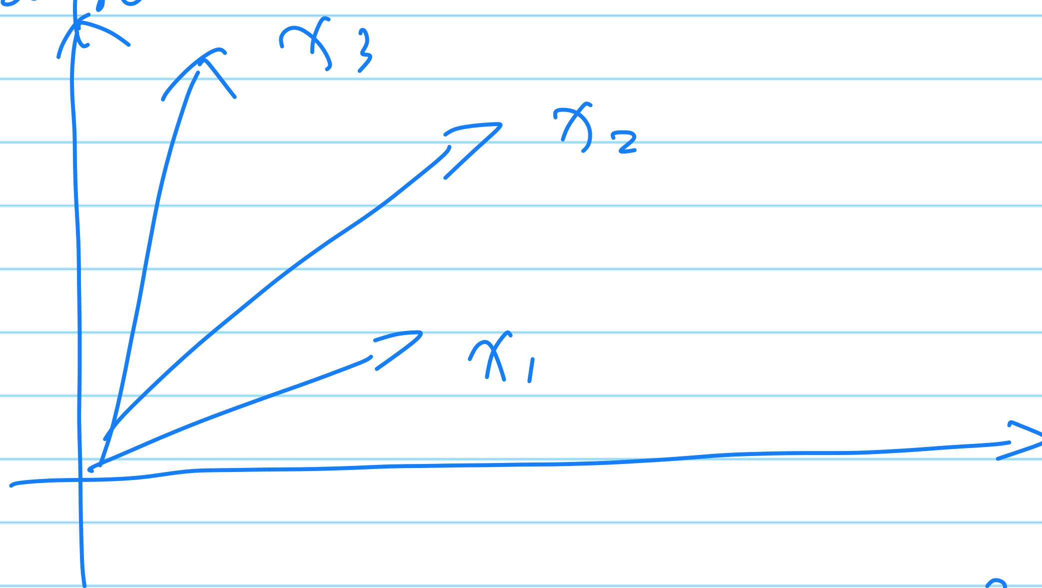
$r_2 = 2 \cdot r_1$

To illustrate the proof, we can write X as follows:

$$X = \begin{pmatrix} 1 \\ 2 \end{pmatrix} (1, 4, 6)$$

$$= \begin{bmatrix} 1 \cdot (1, 4, 6) \\ 2 \cdot (1, 4, 6) \end{bmatrix}$$

Example

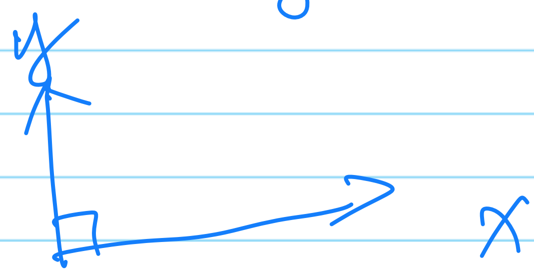


$x_1, x_2, \ldots, x_{100} \in \mathbb{R}^2$

$\mathrm{Dim}(\mathrm{col}(\underbrace{[x_1, x_2, \ldots, x_{100}]}_{2 \times 100}))$

$= \mathrm{Dim}\left(\mathrm{col} \underbrace{\begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_{100}' \end{bmatrix}}_{100 \times 2}\right) \leq 2$

$x \perp y \iff x'y = 0$ or $\langle x, y \rangle = 0$

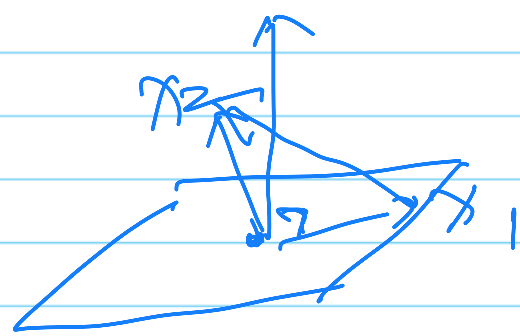


Orthog. to a subspace (Def.)

$y \perp V \iff \forall x \in V, \; x'y = 0$ or $x \perp y$

$x_1, \ldots, x_n \in V$

$\mathbb{R}^3$



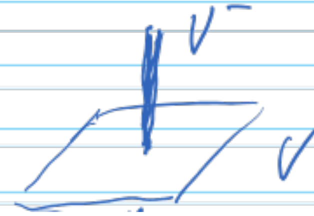
$y \perp x_i$

$y' x_i = 0$

$\Rightarrow y' \left(\sum_{i=1}^{n} c_i x_i \right) = 0$

Orthog. Complement (Def)

$V^\perp = \{ x \in \mathbb{R}^n \mid x \perp V \}$



Kernel & Image space

$$X = (x_1, \ldots, x_p), \quad x_i \in \mathbb{R}^n$$

$$= \begin{pmatrix} r_1' \\ \vdots \\ r_n' \end{pmatrix}, \quad r_i \in \mathbb{R}^p$$

$$\mathrm{im}(X) = L(x_1, \ldots, x_p)$$

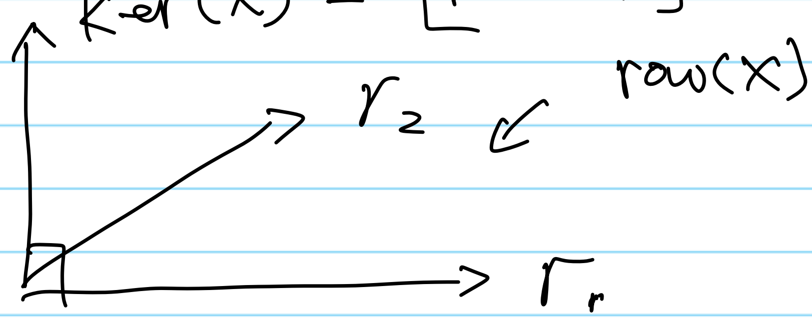
$$= \{ X\beta \mid \beta \in \mathbb{R}^p \} \subseteq \mathbb{R}^n$$

$$\mathrm{Ker}(X) = \{ \beta \in \mathbb{R}^p \mid X\beta = 0 \} \subseteq \mathbb{R}^p$$

$$= \left\{ \beta \in \mathbb{R}^p \;\middle|\; \begin{pmatrix} r_1' \\ \vdots \\ r_n' \end{pmatrix} \beta = 0 \right\}$$

$$= \{ \beta \in \mathbb{R}^p \mid r_1'\beta = 0, \ldots, r_n'\beta = 0 \}$$

$$\mathrm{Ker}(X) = [\mathrm{row}(X)]^{\perp}$$



(3) Nullity Theorem

$$\text{Nullity}(X) = \text{Dim}(\text{Ker}(X))$$

$$\text{Nullity}(X) + \text{rank}(X) = p$$

$$\mathbb{R}^p = \text{Ker}(X) \oplus \text{Ker}(X)^\perp$$

$$= [\text{row}(X)]^\perp \oplus \text{row}(X)$$

$$p = \text{Nullity}(X) + \text{rank}(X)$$

Understanding Nullity Theorem with SVD

Suppose

$$X = \overbrace{\begin{pmatrix} \underset{r\times r}{\Lambda} & 0 \\ 0 & \underbrace{0}_{p-r} \end{pmatrix}}^{p} \begin{array}{l} \left.\vphantom{\Lambda}\right\} n \\ \left.\vphantom{0}\right\} n-r \end{array}$$

$r = \operatorname{rank}(X)$

Note: SVD, $X = U \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} V$

$$\begin{array}{cc} & \begin{array}{cc} \overbrace{\qquad\quad}^{r} & \overbrace{\qquad\qquad\qquad}^{p-r} \end{array} \\ r\left\{\right. & \begin{pmatrix} \Lambda & 0 & \cdots & 0 \\ & 0 & \cdots & 0 \\ 0 \quad 0 & 0 & \cdots & 0 \\ 0 \quad 0 & 0 & \cdots & 0 \end{pmatrix} \end{array} \underset{p\times 1}{\beta} = 0$$

The solution is all β of this form:

$$\left\{ \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \beta_{r+1} \\ \vdots \\ \beta_p \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}0\\0\\0\end{matrix}}\right\} r \\ \left.\vphantom{\begin{matrix}0\\0\\0\end{matrix}}\right\} p-r \end{matrix} \;\middle|\; \beta_i \in \mathbb{R} \right\}$$

$$p - r = \text{Nullity}(X)$$

A useful method for comparing rank:

$$\text{rank}(A) \leq \text{rank}(B)$$

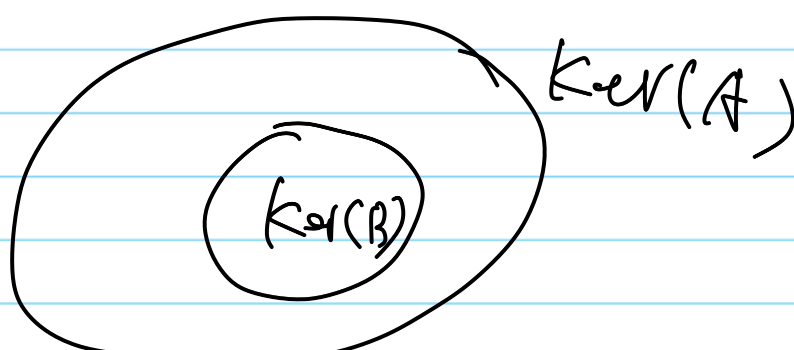
$$\Leftrightarrow \text{Nullity}(A) \geq \text{Nullity}(B)$$

$$\Leftrightarrow \text{Ker}(A) \supseteq \text{Ker}(B)$$

$$\Leftrightarrow \text{“} B\beta = 0 \Rightarrow A\beta = 0 \text{”}$$

$$\text{Ker}(B) = \{\beta \mid B\beta = 0\}$$

$$\text{Ker}(A) = \{\beta \mid A\beta = 0\}$$



Dim of col(X) → Dim of row(X)

→ Dim of $[\text{row}(X)]^{\perp}$

(4) $\text{rank}(XZ) \le \min(\text{rank}(X), \text{rank}(Z))$

Pf: $\underset{p\times m}{Z} = (z_1, \cdots, z_m)$, $\underset{n\times p}{X} = (x_1, \cdots, x_p)$

$$XZ = (Xz_1, \cdots, Xz_m)$$

$$Xz_j = \sum_{i=1}^{p} x_i z_j^{(i)} \in C(X)$$

$$z_j = \begin{pmatrix} z_j^{(1)} \\ \vdots \\ z_j^{(n)} \end{pmatrix}$$

$$\text{rank}(XZ) \le \text{rank}(X)$$

Similarly $\text{rank}(\underset{\tilde{X}}{Z'}\underset{\tilde{Z}}{X'}) \le \text{rank}(Z') = \text{rank}(Z)$

$\text{rank}(XZ) = \text{rank}(Z'X')$

Another proof:

$$Z\beta = 0 \Rightarrow XZ\beta = 0$$

so $\ker(Z) \subseteq \ker(XZ)$

$\Rightarrow \text{nullity}(Z) \le \text{nullity}(XZ)$

$\Rightarrow \text{rank}(Z) \geq \text{rank}(XZ)$

(5) $A: n \times n$, $|A| = 0 \Leftrightarrow rank(A) < n$

$|A| \neq 0 \Leftrightarrow rank(A) = n$

(A^{-1} exists, non singular)

A is invertible : $Ax = y$ has the unique solution $x = A^{-1}y$

$Ker(A) = \{\beta \mid A\beta = 0\} = NULL = \left\{\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}\right\}$

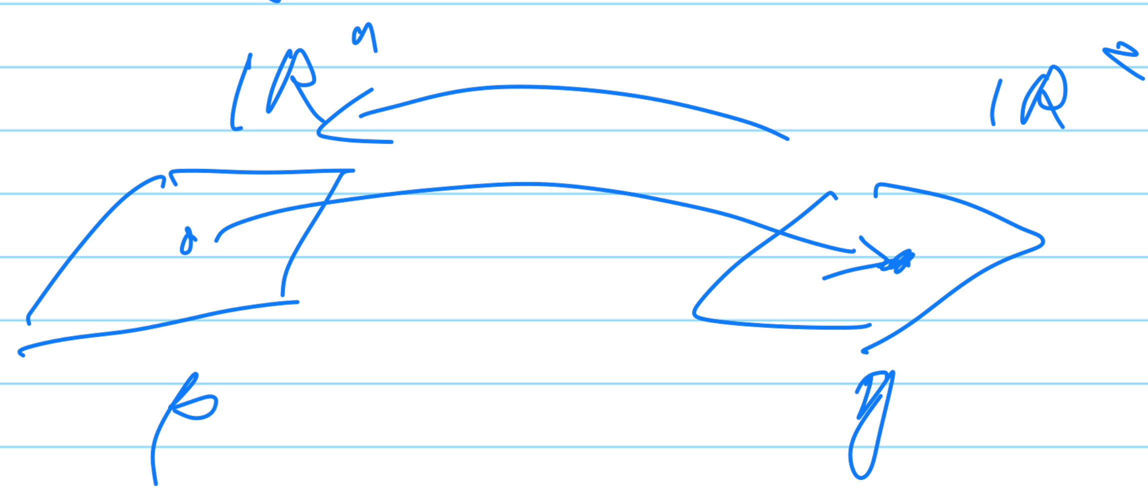
"$A\beta = 0 \Rightarrow \beta = 0$"

"$A\beta_1 = A\beta_2 \Rightarrow \beta_1 = \beta_2$"

$\beta_1 \neq \beta_2 \Rightarrow A\beta_1 \neq A\beta_2$

"$\forall y \in \mathbb{R}^n$, $\exists \beta \in \mathbb{R}^n$, s.t. $A\beta = y$

$\beta = A^{-1}y$"



(6) $\text{rank}(AX) = \text{rank}(X)$, if $|A| \neq 0$

Pf:

$\text{rank}(AX) \leq \text{rank}(X)$

using nullity theorem,

$\text{rank}(X) \leq \text{rank}(AX)$

$\Leftrightarrow \text{nullity}(X) \geq \text{nullity}(AX)$

$\Leftrightarrow AX\beta = 0 \Rightarrow X\beta = 0$

The last statement is true b.c. A^{-1} exists

This implies that

$\text{row}(AX) = \text{row}(X)$

$$A = \begin{pmatrix} a_1' \\ \vdots \\ a_n' \end{pmatrix}$$

$$AX = \begin{pmatrix} a_1'X \\ \vdots \\ a_n'X \end{pmatrix}$$

$X'a_i \in \text{row}(X)$

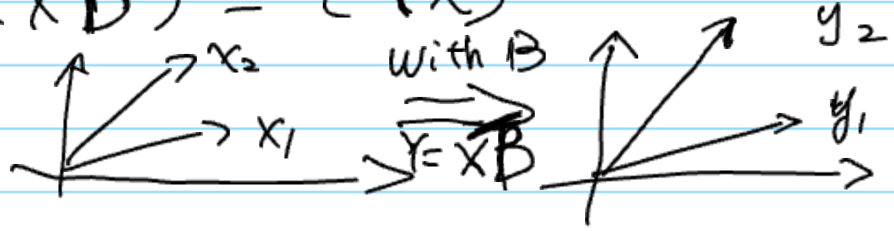
Equivalent statement of (6)

B : $p \times p$ matrix, B^{-1} exists (invertible)

(6.1) $\text{rank}(XB) = \text{rank}(X)$ b.c.

$\text{rank}(XB) = \text{rank}(B'X') = \text{rank}(X') = \text{rank}(X)$

(6.2) $C(XB) = C(X)$

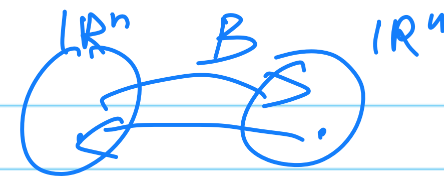


where B : $p \times p$ matrix and B^{-1} exists

$L(x_1, x_2) = L(y_1, y_2)$

if $(x_1, x_2) \overset{Y = XB}{\Longleftrightarrow} (y_1, y_2)$ is $\underline{\text{1-1 \& onto}}$
invertible

A direct proof:



$B\gamma = \beta$

$\forall\ y \in C(X)$,

$\exists\ \beta \in \mathbb{R}^p$ s.t. $y = X\beta$

Since B is invertible, $\exists\ \gamma$ s.t.

$\beta = B\gamma$.

Therefore, $y = XB\gamma = (XB)\gamma$

$\in C(XB)$

Therefore, $C(X) \subseteq C(XB)$

$B = (b_1, \dots, b_p) : p \times p$, $b_j \in \mathbb{R}^p$

$$\underset{n\times p}{X}\ \underset{p\times p}{B} = X(b_1, \dots, b_p) = (Xb_1, Xb_2, \dots, Xb_p)$$

$Xb_j \in C(X)$, Therefore,

$C(XB) \subseteq C(X)$

putting together, $C(XB) = C(X)$

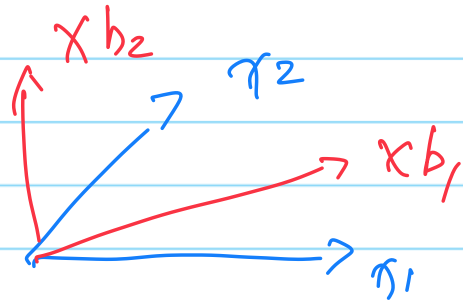
Examples:

(1)

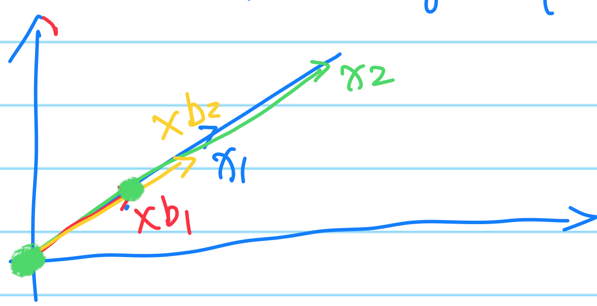
$x_1, \quad x_2 \in \mathbb{R}^2$

x_1 & x_2 LIN

$X = [x_1, x_2]$



(2) $x_2 = c \cdot x_1$, linearly dependent



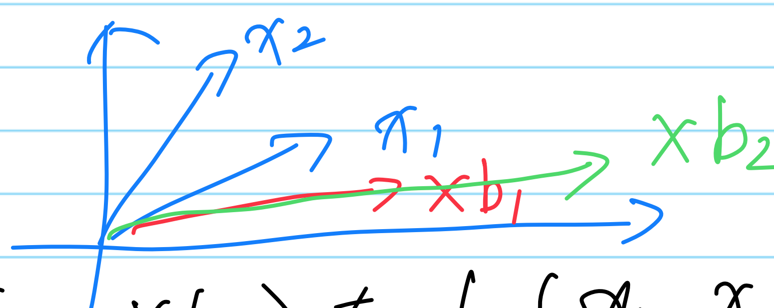
$b_j = \begin{pmatrix} b_{1j} \\ b_{2j} \end{pmatrix}$

$[b_1, b_2]$ is invertible

$Xb_j = x_1 \cdot b_{1j} + x_2 \cdot b_{2j}$

$X = (x_1, x_2)$, $L(Xb_1, Xb_2) = L(x_1, x_2)$

(3) $2b_1 = b_2$, $B = (b_1, b_2)$



$L(Xb_1, Xb_2) \neq L(x_1, x_2)$

(7) $\mathrm{rank}(\underset{n\times p}{X}\ \underset{p\times n}{X'}) = \mathrm{rank}(\underset{p\times n}{X'}\ \underset{n\times p}{X}) = \underline{\mathrm{rank}(X) = \mathrm{rank}(X')}$

Furthermore, $C(XX') = C(X)$

Pf: $\mathrm{rank}(X'X) \le \mathrm{rank}(X)$

$\mathrm{rank}(X'X) \ge \mathrm{rank}(X)$?

$\Leftrightarrow \mathrm{nullity}(X'X) \le \mathrm{nullity}(X)$?

$\Leftrightarrow X'X\beta = 0 \Rightarrow X\beta = 0$?

$\Leftrightarrow$ "If $X'X\beta = 0 \Rightarrow \beta'X'X\beta = 0 \Rightarrow \|X\beta\|^2 = 0$
$\Rightarrow X\beta = 0$"

Since $\mathrm{rank}(X'X) = \mathrm{rank}(X)$, we have
$\mathrm{rank}(XX') = \mathrm{rank}(Y'Y) = \mathrm{rank}(Y) = \mathrm{rank}(X)$

$\uparrow$ Let $Y = X'$

$C(XX') \subseteq C(X)$

$\mathrm{rank}(XX') = \mathrm{rank}(X)$

$\mathrm{Dim}(C(XX')) = \mathrm{Dim}(C(X))$

$C(\underline{XX'}) = C(X)$

Questions:

X : $n \times p$ matrix

$\text{rank}(X) = p$, i.e. full column rank.

(1) $\underset{p \times n}{X'} \underset{n \times p}{X}$ is invertible ?

$$= \begin{pmatrix} x_1' \\ \vdots \\ x_p' \end{pmatrix} (x_1, \dots, x_p) \quad : \; p \times p$$

(2) $\text{rank}\Big(\underset{n \times p}{X} \cdot \underset{p \times p}{(X'X)^{-1}} \underset{p \times n}{X'} \Big) = p$?

$\Downarrow$

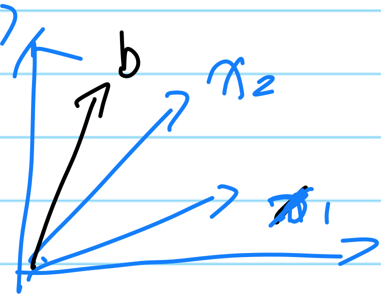
$B \cdot B'$ $\quad$ B invertible

$(XB) \cdot (XB)'$

(3) $C\big(X \cdot (X'X)^{-1} X' \big) = C(X)$?

(8) $\text{rank}\{[X, b]\} \geq \text{rank}(X)$

$X = (x_1, x_2)$



$\mathbb{R}^2$

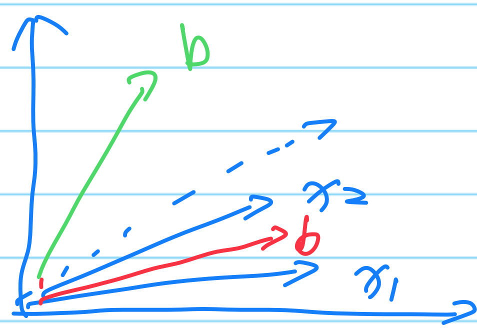
(9) $\text{rank}([X, b]) = \text{rank}(X)$

$\iff b \in C(X)$

$\iff \exists\ \beta,\ s.t.\ \underline{X\beta = b}$

$\iff X, b$ are consistent

$\iff X\beta = b$ has a solution.



$\mathbb{R}^3$

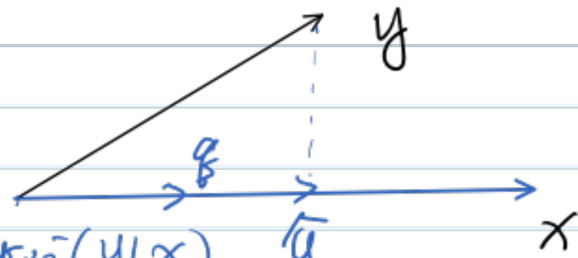
Example

$$[X, b] = \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & b \end{array} \\ \begin{pmatrix} 1 & 4 & 1 \\ 2 & -1 & 2 \\ 1 & 2 & 1 \end{pmatrix} \end{array}$$

$X\beta = b$

Projection onto Vector Space via Orthonormal Basis

projection to $L(x)$



$\hat{y} = proj(y|x)$: $\hat{y}$

$\hat{y} = c\,x$ for some $c \in \mathbb{R}$, $\hat{y} \in L(x)$

$y - \hat{y} \perp L(x)$

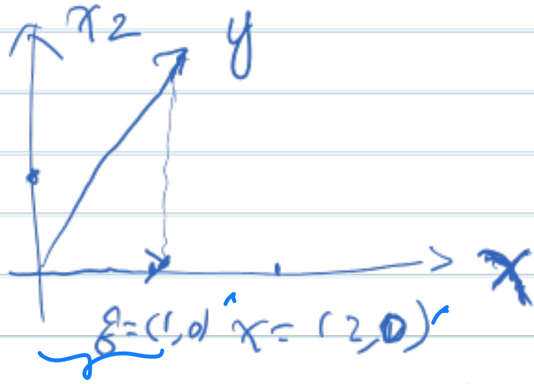
$$\hat{y} = \frac{x'y}{\|x\|^2} \cdot x = \frac{x x'}{\|x\|^2} \cdot y$$ (how to lin. Transform y)

$$= \left\langle \frac{x}{\|x\|}, y \right\rangle \cdot \frac{x}{\|x\|}$$

$$= \underline{\langle q, y \rangle} \cdot q, \text{ where } q = \frac{x}{\|x\|}, \ \|q\| = 1$$

↑ base of $L(x)$

Example

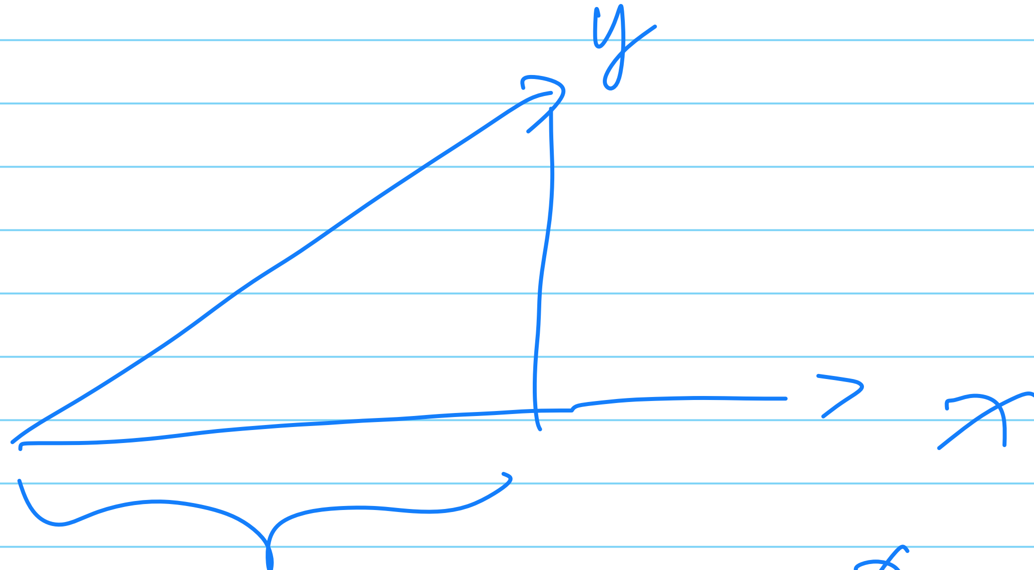


$\langle q, y \rangle = y_1$

$$\hat{y} = y_1 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} y_1 \\ 0 \end{pmatrix}$$

"projection is just dropping dimensions"

$\langle q, y \rangle$



$$\langle q, y\rangle = \langle \frac{x}{\|x\|}, y\rangle$$

$$\hat{y} = \langle q, y\rangle \cdot q$$

where $\|q\| = 1$

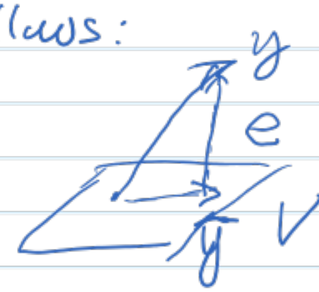
Definition

Proj. to a subspace $V \subseteq \mathbb{R}^n$

$\text{proj}(y|V) = \hat{y}$ is as follows:

1) $\hat{y} \in V$

2) $y - \hat{y} \perp V$



$V = L(x_1, \cdots, x_p)$

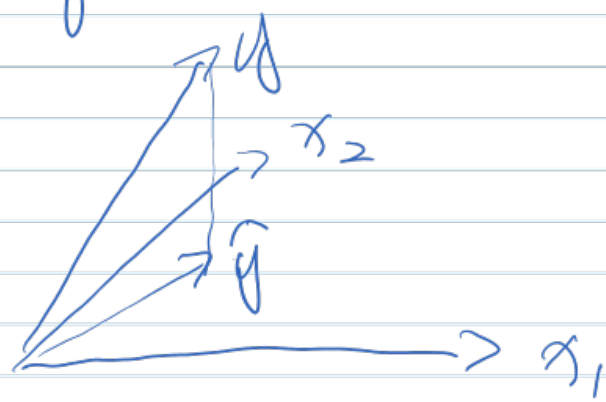
what's $\text{proj}(y|V)$?

Theorem: $V = L(x_1, \cdots, x_p)$

$\text{proj}(y|V) = \hat{y}$

$\Leftrightarrow y - \hat{y} \perp x_i$ for all $i = 1, \cdots, p$

$\hat{y} \in L(x_1, \cdots, x_p)$



pf: $\forall x \in L(x_1, \cdots, x_p) = V$

$x = \sum_{i=1}^{p} c_i x_i$, for some $c_i \in \mathbb{R}$

$\Rightarrow$) suppose $\hat{y} = \text{proj}(y|V)$ as defined above, $y - \hat{y} \perp V$

$x_i \in V$, so $y - \hat{y} \perp x_i$

$(\Leftarrow)$ $y - \hat{y} \perp x_i \Rightarrow y - \hat{y} \perp \sum_{i=1}^{p} c_i x_i \Rightarrow y - \hat{y} \perp V$

$(y - \hat{y})' x_i = 0 \Rightarrow (y - \hat{y})' \sum c_i x_i$

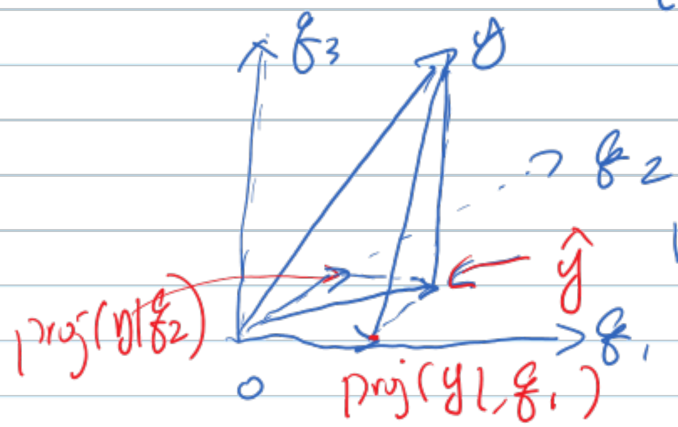
$= \sum c_i (y - \hat{y})' x_i$

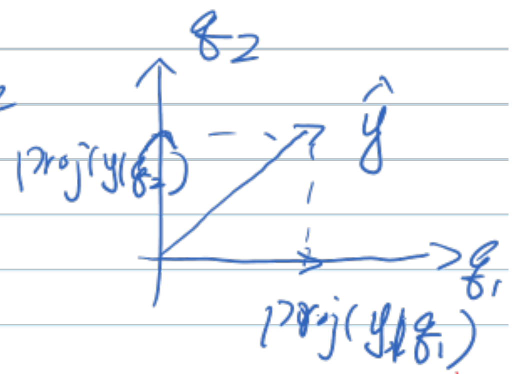
Theorem:

suppose $q_1, q_2, \dots, q_k$ is an **orthonormal** basis for $V = L(x_1, \dots, x_p)$

$[\, k \le p,\ k = \mathrm{rank}([x_1, \dots, x_p]) \,]$.

Then $\mathrm{proj}(y \mid V) = \sum_{i=1}^{k} \mathrm{proj}(y \mid q_i)$

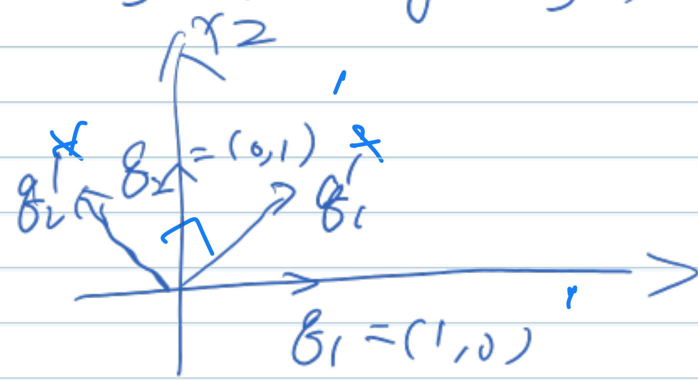




What's orthogonormal basis?

$L(q_1, q_2, \dots, q_k) = L(x_1, \dots, x_p)$

$q_i \perp q_j$ for any $i \ne j$, $\|q_i\| = 1$



Vector form for $\hat{y}$:

$$\hat{y} = \text{proj}(y \mid V) = \sum_{i=1}^{k} \text{proj}(y \mid q_i)$$

$$= \sum_{i=1}^{k} \langle y, q_i \rangle \cdot q_i \quad (\|q_i\| = 1)$$

$$= \sum_{i=1}^{k} \frac{\langle y, q_i \rangle}{\|q_i\|^2} \cdot q_i, \text{ if } \|q_i\| \neq 1$$

Pf: suppose $\|q_i\| = 1$ for $i = 1, \dots, k$

$\hat{y} \in V$. we will show

$$y - \hat{y} \perp V$$

$$\iff y - \hat{y} \perp q_j, \text{ for } j = 1, \dots, k$$

$$\langle y - \hat{y}, q_j \rangle$$

$$= \langle y, q_j \rangle - \langle \sum_{i=1}^{k} \langle y, q_i \rangle q_i, q_j \rangle$$

$$= \langle y, q_j \rangle - \sum_{i=1}^{k} \langle y, q_i \rangle \cdot \langle q_i, q_j \rangle$$

$$= \langle y, q_j \rangle - \langle y, q_j \rangle \langle q_j, q_j \rangle$$

$$= \langle y, q_j \rangle - \langle y, q_j \rangle = 0$$

$$\boxed{\langle q_i, q_j \rangle = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}}$$

proj(y|V) in matrix form. $I_k = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}_{k\times k}$

suppose $\underbrace{\|q_i\| = 1, \; q_i \in \mathbb{R}^n}_{k}$

$$\text{proj}(y|V) = \sum_{i=1}^{k} \left(\underset{n\times 1}{q_i} \cdot \left[\underset{1\times n}{q_i'} \; \underset{n\times 1}{y} \right] \right)$$

$$= (q_1, \cdots, q_k) \begin{pmatrix} q_1' \\ \vdots \\ q_k' \end{pmatrix} \underset{n\times 1}{y}$$

$$= \underset{n\times k}{Q} \cdot \underset{k\times n}{Q'} \underset{n\times 1}{y}$$

$$= Q^* \underset{n\times n}{\begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}} (Q^*)' y$$

$$= \left(\sum_{i=1}^{k} q_i q_i' \right) \cdot y$$

where $Q = \underset{n\times k}{(q_1, \cdots, q_k)}$: $n\times k$, partial basis of $\mathbb{R}^n$

$Q^* = \underset{n\times n}{(q_1, \cdots, q_k, q_{k+1}, \cdots q_n)}$: $n\times n$

Note: $Q'Q = I_k$, $\quad Q^*(Q^*)' = (Q^*)'Q^* = I_n$

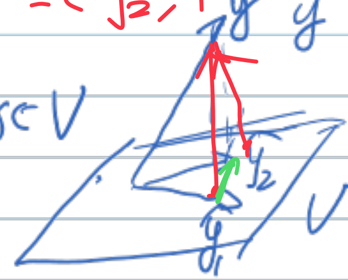
Uniqueness of projection

Theorem: $\hat{y}_1$, $\hat{y}_2$ are two projections of y onto V. Then $\hat{y}_1 = \hat{y}_2$.

Pf:

$$\overbrace{\langle y - \hat{y}_1, x \rangle}^{\langle y, x\rangle - \langle \hat{y}_1, x\rangle} = \underbrace{\langle y - \hat{y}_2, x \rangle}_{\langle y, x\rangle - \langle \hat{y}_2, x\rangle} = 0$$

$\forall x \in V$



$\Rightarrow \langle \hat{y}_1, x \rangle = \langle \hat{y}_2, x \rangle \quad \forall x \in V$

$\Rightarrow \langle \hat{y}_1 - \hat{y}_2, x \rangle = 0, \ \forall x \in V$

$\Rightarrow \langle \hat{y}_1 - \hat{y}_2, \hat{y}_1 - \hat{y}_2 \rangle = 0$

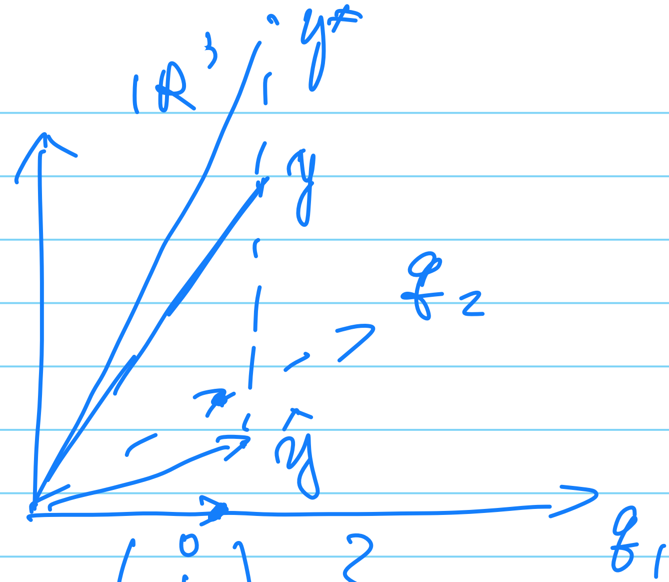
[$\langle x+y, z \rangle = \langle x, z\rangle + \langle y, z \rangle$]

$\Rightarrow \| \hat{y}_1 - \hat{y}_2 \|^2 = 0$

$\Rightarrow \hat{y}_1 - \hat{y}_2 = 0$

Example:

$y \in \mathbb{R}^3$



$$V = L\left\{ \underset{q_1}{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}}, \underset{q_2}{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}} \right\} \qquad y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

$$\text{proj}(y \mid V) = \text{proj}(y \mid q_1) + \text{proj}(y \mid q_2)$$

$$= y_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} y_1 \\ y_2 \\ 0 \end{pmatrix}$$

Example $y_{ij} = u_i + \varepsilon_{ij}$

$\boxed{y_{11}\ y_{12}}$ $\quad$ $\boxed{y_{21}, y_{22}}$ $\quad$ $\boxed{y_{31}, y_{32}}$

G1 $\qquad$ G2 $\qquad$ G3

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_6 \end{bmatrix}$$

$$y = [x_1 \quad x_2 \quad x_3] \cdot u + \varepsilon$$

$\in L(x_1, x_2, x_3)$

$$\text{proj}(y \mid L(x_1, x_2, \cdots, x_3))$$

$[x_1, x_2, x_3]\hat{u}$

$\parallel$

$\hat{y}$

$$= \sum_{i=1}^{3} \text{proj}(y \mid x_i)$$

b.c. x_1, x_2, x_3 are orthogonal.

i.e. $x_i' x_j = 0, i \neq j$

$$\text{proj}(y \mid x_i) = \frac{\langle y, x_i \rangle}{\|x_i\|^2} \cdot x_i$$

$$\langle y, x_1 \rangle = y_{11} + y_{12}$$

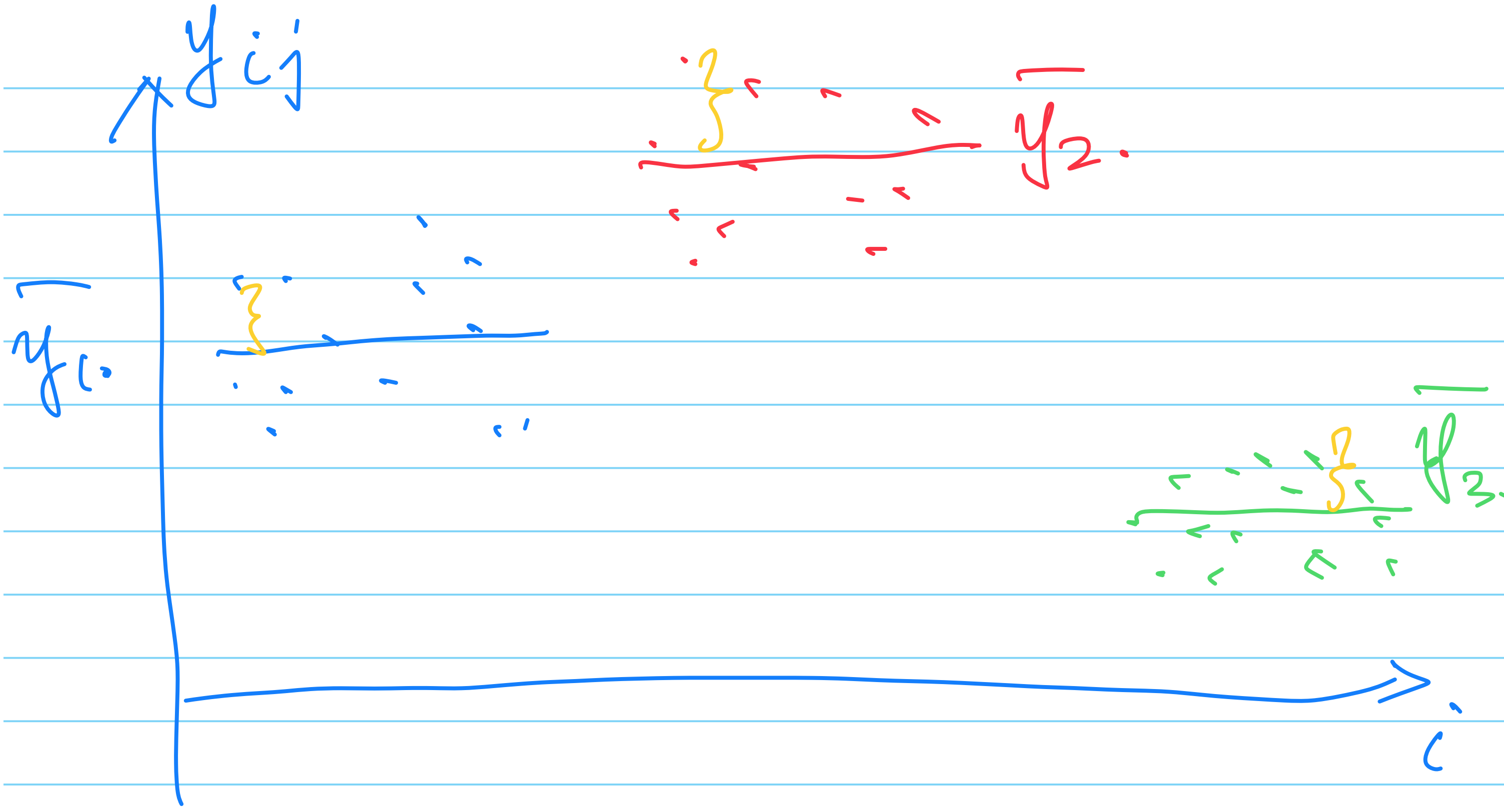
$$\|x_1\|^2 = 1 + 1 = 2$$

$$\frac{\langle y, x_1 \rangle}{\|x_1\|^2} = \frac{y_{11} + y_{12}}{2} = \bar{y}_{1\cdot}$$

$$\operatorname{proj}(y \mid x_1) = \begin{bmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{1\cdot} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \bar{y}_{1\cdot} \cdot x_1$$

$$\operatorname{proj}(y \mid L(x_1, x_2, x_3)) =$$

$$\begin{bmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{1\cdot} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \bar{y}_{2\cdot} \\ \bar{y}_{2\cdot} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \bar{y}_{3\cdot} \\ \bar{y}_{3\cdot} \end{bmatrix} = \begin{bmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \\ \bar{y}_{2\cdot} \\ \bar{y}_{2\cdot} \\ \bar{y}_{3\cdot} \end{bmatrix}$$



$$\hat{y} = \bar{y}_{1\cdot}\, x_1 + \bar{y}_{2\cdot}\, x_2 + \bar{y}_{3\cdot}\, x_3$$

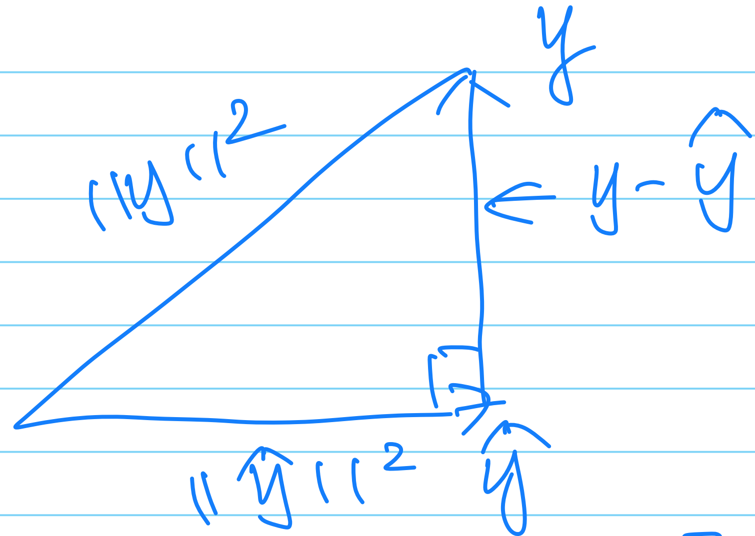
$$= (\underbrace{\bar{y}_{1\cdot}, \cdots, \bar{y}_{1\cdot}}_{n_1}, \underbrace{\bar{y}_{2\cdot}, \cdots, \bar{y}_{2\cdot}}_{n_2}, \underbrace{\bar{y}_{3\cdot}, \cdots, \bar{y}_{3\cdot}}_{n_3})'$$

$$y - \hat{y} = \begin{bmatrix} y_{11} - \bar{y}_{1.} \\ y_{12} - \bar{y}_{1.} \\ \vdots \\ y_{31} - \bar{y}_{3.} \\ y_{32} - \bar{y}_{3.} \end{bmatrix} \begin{matrix} \} \to SS_1 \\ \} \to SS_2 \\ \} \to SS_3 \end{matrix}$$

$$\|y - \hat{y}\|^2 = SS_1 + SS_2 + SS_3$$

where $SS_i = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2$

sum square with groups.



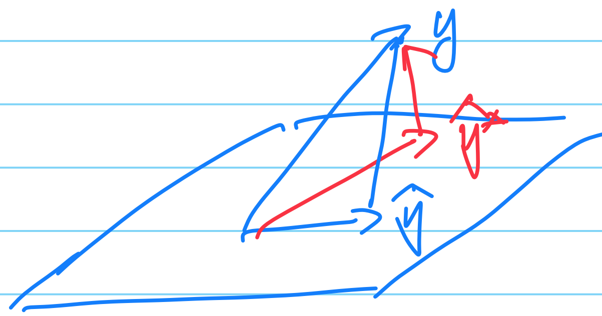
$$\|y-\hat{y}\|^2 = \|y\|^2 - \|\hat{y}\|^2$$

$$= \sum_i \sum_j y_{ij}^2 - \sum_i n_i \bar{y}_{i\cdot}^2$$

$$= \sum_i \sum_j y_{ij}^2 - \sum_i \frac{y_{i\cdot}^2}{n_i}$$

where $y_{i\cdot} = n_i \cdot \bar{y}_{i\cdot}$

projection is the least-squared prediction



Theorem:

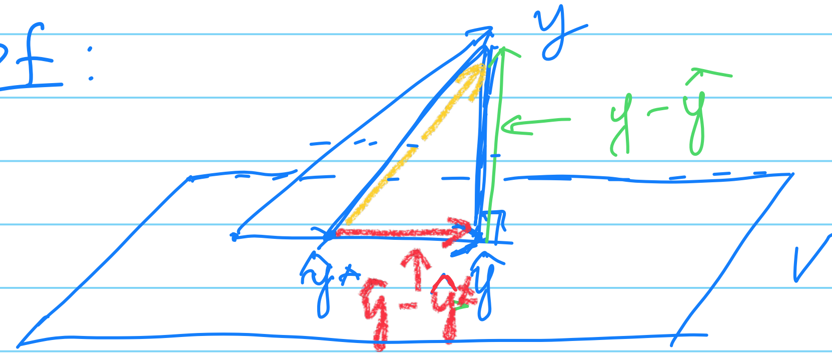
$proj(y|x) = \hat{y}$ is defined as follows:

$\hat{y} \in V$ s.t. $\hat{y} - y \perp V$

$\hat{y}$ is the vector in V that is closest to y. That is,

for any $\hat{y}^* \in V$, $\|y - \hat{y}\|^2 \leq \|y - \hat{y}^*\|^2$

Pf :



1) $\hat{y}-\hat{y}^* \in V$ (since $\hat{y}$ & $\hat{y}^* \in V$)

2) $y-\hat{y} \perp V$ (definition of $\hat{y}$)

$\Rightarrow y-\hat{y} \perp \hat{y}-\hat{y}^*$

$$y-\hat{y}^* = \underline{y-\hat{y}} + \underline{\hat{y}-\hat{y}^*}$$

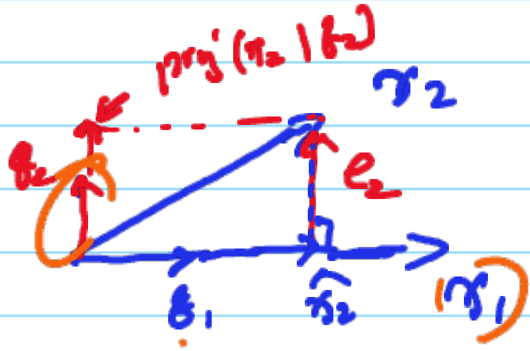
By Pythagorean Theorem,

$$\|y-\hat{y}^*\|^2 = \|y-\hat{y}\|^2 + \|\hat{y}-\hat{y}^*\|^2 \geq \|y-\hat{y}\|^2$$

Gram-Schmidt Orth.

(QR factorization)

$\mathbb{R}^2$



$$q_1 = x_1 / \|x_1\|$$

$$\hat{x}_2 = \langle x_2, q_1 \rangle \cdot q_1$$

$$e_2 = x_2 - \hat{x}_2 \perp q_1$$

$$q_2 = \frac{e_2}{\|e_2\|}$$

$$L(q_1, q_2) = L(x_1, x_2) \qquad x_2 = \langle x_2, q_2 \rangle \cdot q_2 + \langle x_2, q_1 \rangle \, q_1 .$$

$$x_1 = \underbrace{\langle x_1, q_1 \rangle}_{= \|x_1\|} q_1 + 0 \cdot q_2$$

$$(x_1, x_2) = (q_1, q_2) \begin{pmatrix} \langle x_1, q_1 \rangle & \langle x_2, q_1 \rangle \\ 0 & \langle x_2, q_2 \rangle \end{pmatrix}$$

$\perp$

$$X = Q \cdot R$$

orthog. upper-triangle.

QR factorization

$$(x_1, \cdots, x_p) = (q_1, \cdots, q_k) \begin{array}{c} \begin{bmatrix} \langle x_1, q_1 \rangle & \langle x_2, q_1 \rangle & \cdots & \langle x_p, q_1 \rangle \\ 0 & \langle x_2, q_2 \rangle & \cdots & \langle x_p, q_2 \rangle \\ 0 & 0 & \cdots & \langle x_p, q_3 \rangle \\ \vdots & \vdots & & \\ 0 & 0 & \cdots & \langle x_p, q_k \rangle \end{bmatrix} \\ \begin{matrix} \ \ r_1 \qquad & r_2 \qquad & \qquad\qquad r_p \end{matrix} \end{array}$$

$$X = Q \qquad R$$

basis

r_j is the coef. of x_j in $L(q_1, \cdots, q_k)$

$n \times k$ $\qquad$ $k \times p$

$\{q_1, \cdots, q_k\}$ is an orth. basis for $L(x_1, \cdots, x_p)$

$$\begin{cases} e_j = x_j - \mathrm{proj}(x_j \mid q_1, \cdots, q_{j-1}) \\ q_j = \dfrac{e_j}{\|e_j\|} \end{cases}$$

$$q_1 = \frac{x_1}{\|x_1\|}.$$

Projection matrix of projection onto $c(X)$

- Normal equation
- Projection matrix

Normal equation

Let $X = (x_1, \cdots, x_p)$: $n \times p$ matrix

We want to project y to $C(X)$

That is, we want to find $\beta \in \mathbb{R}^p$ s.t.

$$y - X\beta \perp C(X)$$

$$\Leftrightarrow \; y - X\beta \perp x_i, \text{ for } i = 1, \cdots, p$$

$$\Leftrightarrow \; x_i'(y - X\beta) = 0, \text{ for each } i$$

$$\Leftrightarrow \; X'(y - X\beta) = 0$$

$$\Leftrightarrow \; \underset{p\times n}{X'} \underset{n\times 1}{y} = X'X \underset{n\times 1}{\beta} \quad \leftarrow \text{normal equation}$$

when $(X'X)^{-1}$ exists, that is

$x_1, \cdots, x_p$ are LIN.

$$\hat{\beta} = (X'X)^{-1}X'y \quad \leftarrow \text{LS est.}$$

Then, another expression for $\text{proj}(y|C(X))$

$$\text{proj}(y|C(X)) = X\cdot\hat{\beta} = X\cdot(X'X)^{-1}X'y$$

$P = X\cdot(X'X)^{-1}X'$ is the proj. matrix onto $C(X) = C(P)$ (?)

Connection with $P = QQ'$:

when $\text{rank}(X) = p$, with QR factorization, we can write

$$\underset{n\times p}{X} = \underset{n\times p}{Q}\cdot\underset{p\times p}{R}$$

[also R is invertable

$Q'Q = I_p$]

i.e. columns of Q are orthogonal

$$P = X(X'X)^{-1}X'$$

$$= Q\cdot R(R'Q'QR)^{-1}R'Q'$$

$$= Q\,R(R'R)^{-1}R'Q'$$

$$= QQ'$$

I_p

why $C(P) = C(X)$?

$\underset{n \times p}{X} = \underset{n \times p}{Q} \underset{p \times p}{R}$, $\text{rank}(R) = p$

$C(X) = C(Q)$

$P = Q \cdot Q'$

$C(P) = C(Q)$

so $C(P) = C(X)$

Projection Matrix

- Projection matrix in general
- Symmetric and Idempotent Matrix

Def :

A square matrix P: $n \times n$ is a projection matrix onto $C(P)$ if

$\forall\, y \in \mathbb{R}^n$, $y - Py \perp C(P)$

Note that $Py \in C(P)$.

Examples:

1) $y = (y_1, y_2, y_3)'$

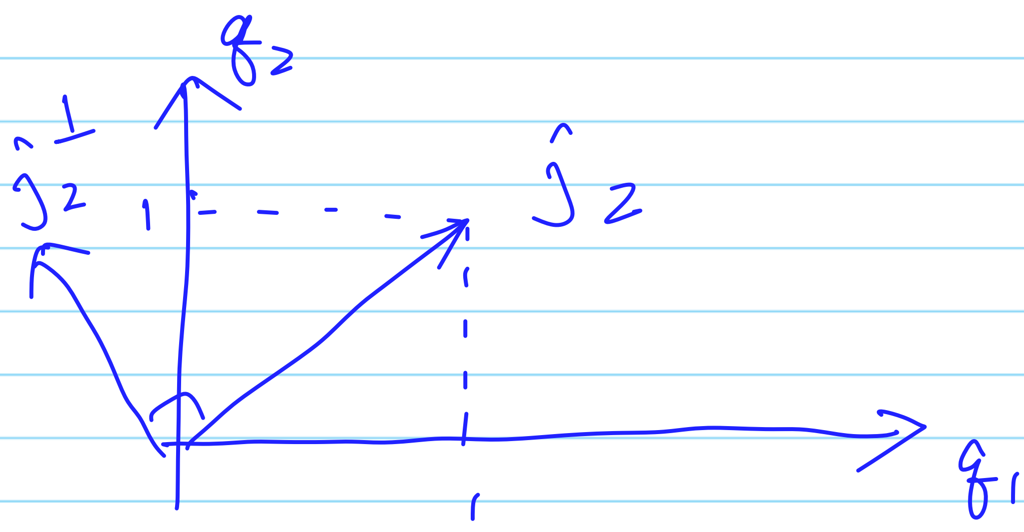
$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad Py = \begin{pmatrix} y_1 \\ 0 \\ y_3 \end{pmatrix}$$

2) $$P_{j_n} = \frac{1}{n} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

$$= \frac{1}{n} \hat{j}_n \hat{j}_n', \quad \hat{j}_n = (1, \cdots, 1)'$$

$$P = I_n - P_{\hat{j}_n} \quad ?$$

$\mathbb{R}^2$



Theorem: P is a projection matrix onto $V = C(P)$, iff

(1) P is symmetric

(2) $P^2 = P$ (idempotent)

$(P = P^k$, for $k = 2, \dots)$ $y - Py = (I - P)y$

Pf: "$\Rightarrow$"

$(y - Py) \perp Pz \quad \forall\, y, z \in \mathbb{R}^n$

$\Rightarrow y'(I - P')Pz = 0, \quad \forall\, y, z \in \mathbb{R}^n$

$\Leftrightarrow (I - P')P = 0 \Leftrightarrow P = P'P$

$P'P$ is symmetric, so, P is symmetric.

$P = P'P \Leftrightarrow P = P^2$

"$\Leftarrow$" $\forall\, y, z \in \mathbb{R}^p$

a vector in $C(P)$

$\langle y - Py, Pz \rangle = y'(I - P')Pz$

$= y'(P - P'P)z,$

$= y'(P - P^2)z$, b.c. $P' = P$

$= 0$, b.c. $P = P^2$

Theorem: P is a proj matrix onto $C(P)$.

iff "$\forall y \in C(P)$, $Py = y$

$\forall z \in C(P)^\perp$, $Pz = 0$"

pf of "$\Rightarrow$"

suppose $y \in C(P)$, $\exists z \in \mathbb{R}^n$, s.t. $y = Pz$

$$Py = P \cdot Pz = Pz = y$$

suppose $w \perp C(P) \Rightarrow w \perp Pw$

$\Rightarrow w'Pw = 0 \Rightarrow w'P'Pw = 0 \quad (P = P'P)$

$\Rightarrow \|Pw\| = 0 \Rightarrow Pw = 0$

proof of "$\Leftarrow$":

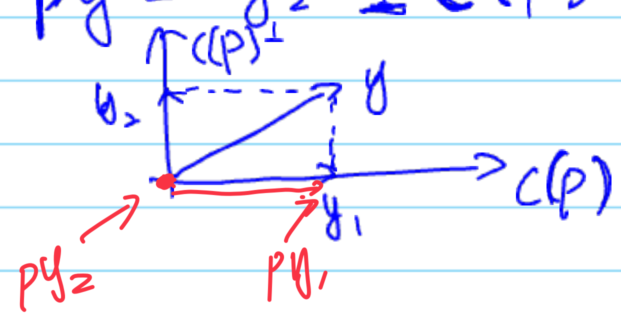
$\forall y \in \mathbb{R}^n$,

$y = y_1 + y_2 \qquad y_1 \in C(P),\ y_2 \perp C(P)$

e.g. $y_1 = \text{proj}(y|P)$, $y_2 = y - y_1$

$$Py = Py_1 + Py_2 = y_1 + 0 = y_1$$

$$y - Py = y_2 \perp C(P)$$

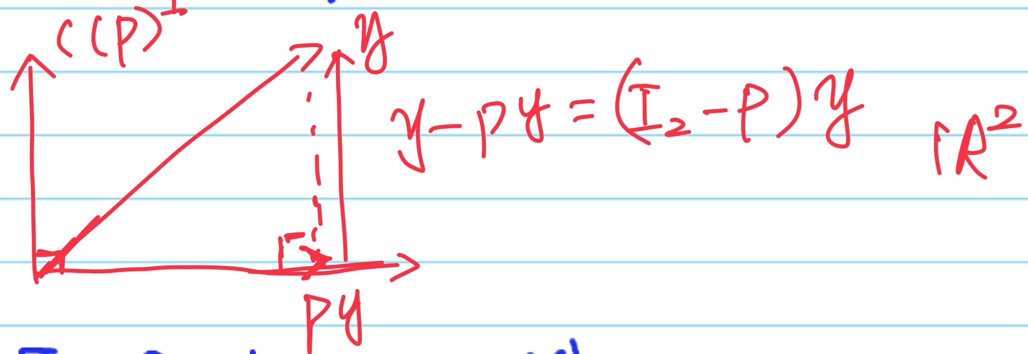


$\mathbb{R}^2$

Projection onto Complement subspace

Thm: Let $\underset{n\times n}{P}$ be a proj matrix onto $C(P) \in \mathbb{R}^n$

Then $I_n - P$ is a proj matrix onto $C(I_n-P) = C(P)^\perp$



Pf: (1) $I_n - P$ is symmetric

(2) $(I_n - P)^2 = I_n - P - P + P^2 = I_n - P$

(3) $C(I_n - P) = C(P)^\perp$:

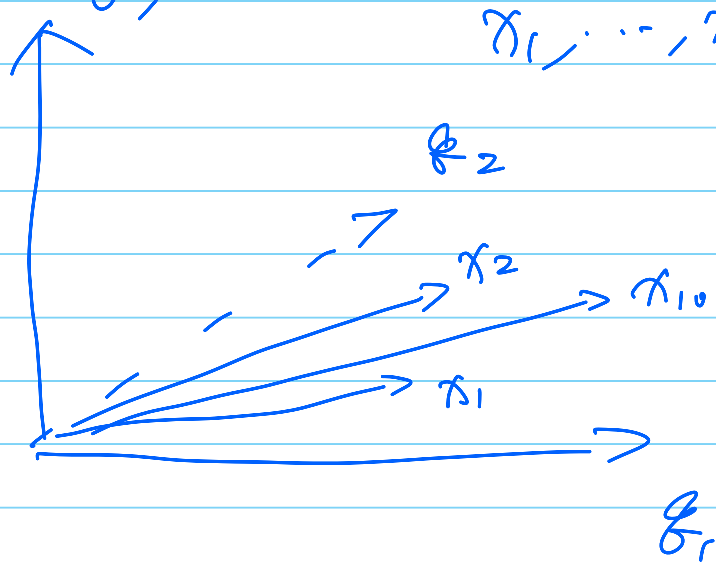
"$\Rightarrow$" $\forall z \in C(I_n - P)$, $\exists x$, s.t. $z = (I_n - P)x$

$z = x - Px \perp C(P)$

"$\Leftarrow$" $\forall y \perp C(P)$, $Py = 0$, $\Rightarrow y - Py = y$

Since $y = y - Py = (I-P)y$, $y \in C(I-P)$

Example



$x_1, \dots, x_{100} \in L(q_1, q_2)$ $\mathbb{R}^3$

$$L(x_1, \dots, x_{100})^{\perp} = C(I_3 - P_X)$$

P_X = projection matrix onto $C(X)$
$= QQ'$

where Q is an orthonormal basis of $C(X)$

If $x_1, \dots, x_{100} \in L(q_1, q_2)$

then $C(I_3 - P_X) = L(q_3)$

Examples:

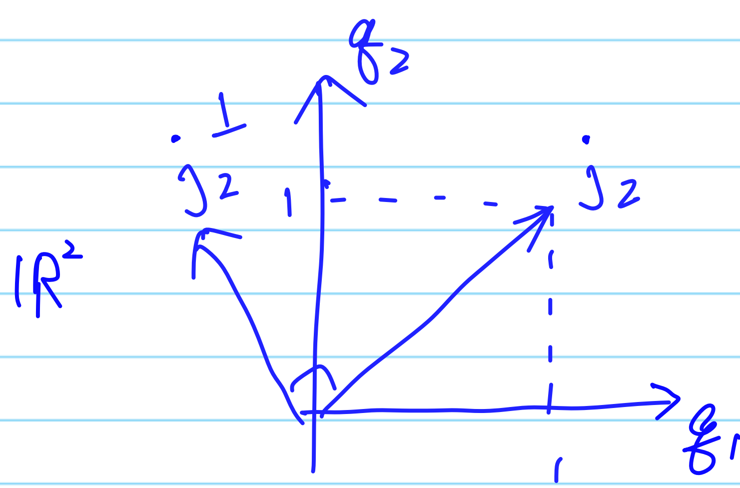
$$\hat{j}_n = (1, 1, \cdots, 1)'$$

$$P_{\hat{j}_n} = \frac{1}{n} \hat{j}_n \hat{j}_n',$$

$$= \frac{1}{n} \begin{pmatrix} 1 & 1, & \cdots, & 1 \\ 1 & 1, & \cdots, & 1 \\ 1 & 1, & \cdots, & 1 \end{pmatrix}$$

$$P = I_n - P_{\hat{j}_n}$$

$$= P_{\hat{j}_n^{\perp}}$$



$$C(I_n - P_{\hat{j}_n})$$

$$= C(P_{\hat{j}_n})^{\perp}$$

$$= \hat{j}_n^{\perp}$$

Projection onto nested subspaces

- Projection onto orthogonal complement space
- Projection onto nested subspaces

Nested Stat. Model

$$\begin{array}{c} y \\ \begin{bmatrix} \vdots \\ \vdots \\ \vdots \end{bmatrix} \end{array} = \begin{array}{cc} X_1 & X_2 \\ \multicolumn{2}{c}{\left[\begin{array}{c|c} \cdots & \cdots \\ \cdots & \cdots \\ \cdots & \cdots \end{array}\right]} \end{array} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \varepsilon$$

$H_0: \; y \sim X_1, \qquad SSE_0$

$H_1: \; y \sim X_1 + X_2, \qquad SSE_1$

$$c(X_1) \subseteq c([X_1, X_2])$$

Projections onto nested spaces

Thm: If P_0 is a proj matrix onto $C(P_0)$

P_1 is a - - - - - - - - $C(P_1)$

$C(P_0) \subseteq C(P_1)$ [$y = X_0\beta + \varepsilon$, $y = X_1\beta + \varepsilon$, $C(X_0) \subseteq C(X_1)$]

Then $P_1 P_0 = P_0 P_1 = P_0$

Pf: $\forall y \in \mathbb{R}^n$, $P_0 y \in C(P_0) \subseteq C(P_1)$

$\Rightarrow P_1(P_0 y) = P_0(y)$

$\Rightarrow P_1 P_0 = P_0$ $\quad$ P_0 is symmetric

then $P_0 = P_1 P_0 = (P_1 P_0)' = P_0' P_1' = P_0 P_1$

Thm: If P_0 is a proj matrix onto $C(P_0)$

P_1 is a - - - - - - - - $C(P_1)$

$C(P_0) \subseteq C(P_1)$

then $\underline{P_1 - P_0}$ is a proj mat. onto

$C(P_1 - P_0) = \underline{[C(P_0)]^{\perp} \cap C(P_1)}$

Pf 1: $[\, C(P_1 - P_0) \perp C(P_0) \,]$

(1) $(P_1 - P_0)' = P_1' - P_0' = P_1 - P_0$ symmetric

(2) $(P_1 - P_0)^2 = P_1^2 - P_0 P_1 - P_1 P_0 + P_0^2$

$= P_1 - 2P_0 + P_0 = P_1 - P_0$

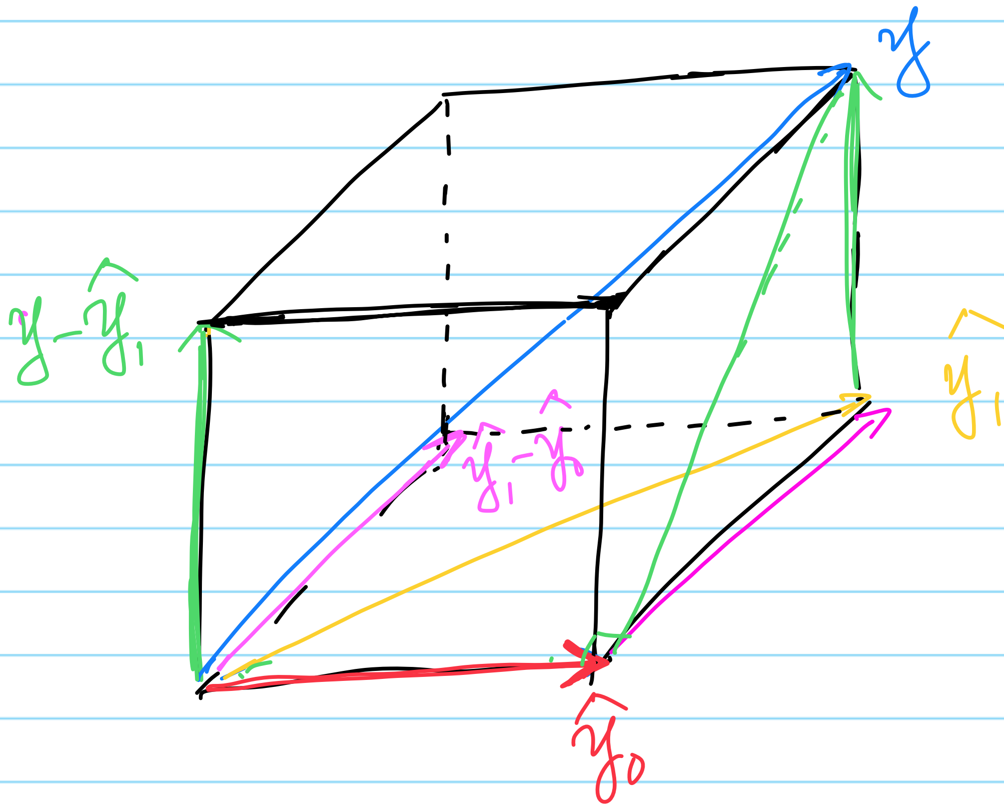
(3) $C(P_1 - P_0) = C(P_0)^{\perp} \cap C(P_1)$?

$\Leftrightarrow C(P_1 - P_0) \perp C(P_0)$?

$\forall\, y, z \in \mathbb{R}^n$, $\langle (P_1 - P_0)y,\, P_0 z \rangle = y'(P_1 - P_0) \cdot P_0 z$

$= y'(P_1 P_0 - P_0^2) z = y'(P_0 - P_0) z = 0$

$C(P_1 - P_0) \subseteq C(P_1)$ is obvious : . . .



$$y = \hat{y}_0 + (\hat{y}_1 - \hat{y}_0) + (y - \hat{y}_1)$$
$$= P_0 y + (P_1 y - P_0 y) + (I - P_1) \cdot y$$

Another pf of $\hat{y}_1 - \hat{y}_0 \perp \hat{y}_0$:

$$\hat{y}_0 = \text{proj}(\hat{y}_1 \mid C(P_0)) = P_0(P_1 y)$$

Therefore, $\hat{y}_1 - \hat{y}_0 \perp \hat{y}_0$

Remark:

suppose $P_1 = [x_1, \cdots, x_p] : n \times p$

$c(P_0) \subseteq c(P_1)$

$c(P_0)^\perp c(P_1)$

$= c(P_1 - P_0) = c(P_1 - P_0 P_1)$

$= c(P_1 - \text{proj}(P_1 | P_0))$, where

$P_1 - \text{proj}(P_1 | P_0)$

$= [x_1 - \text{proj}(x_1 | P_0), \cdots, x_p - \text{proj}(x_p | P_0)]$

$= [x_1 - P_0 x_1, \cdots, x_p - P_0 x_p]$

$= [x_1, \cdots, x_p] - P_0 \cdot [x_1, \cdots, x_p]$

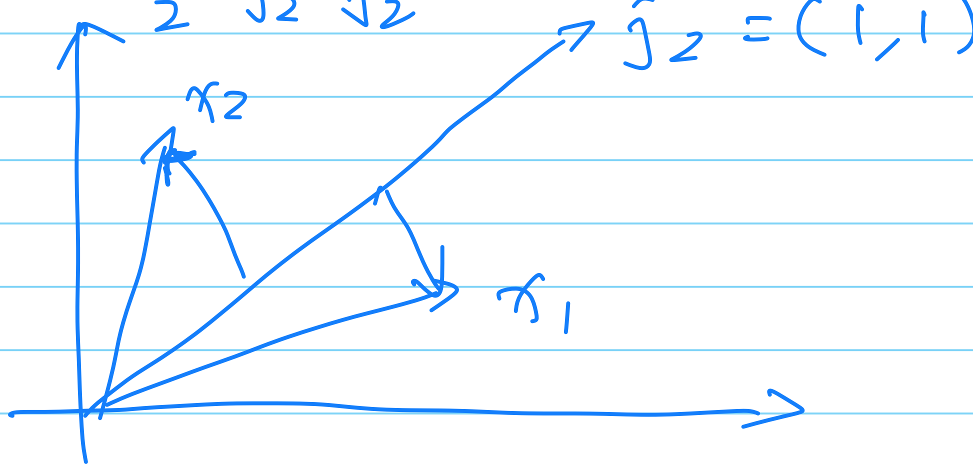
$= P_1 - P_0 P_1 = P_1 - P_0$

In words, the subspace generated by $\{x_1 - P_0 x_1, \cdots, x_p - P_0 x_p\}$ is the same as $c(P_0)^\perp c(P_1)$

Example:

$P_1 = [x_1, x_2]$, $x_i \in \mathbb{R}^2$

$P_0 = \frac{1}{2} \hat{j}_2 \cdot \hat{j}_2'$

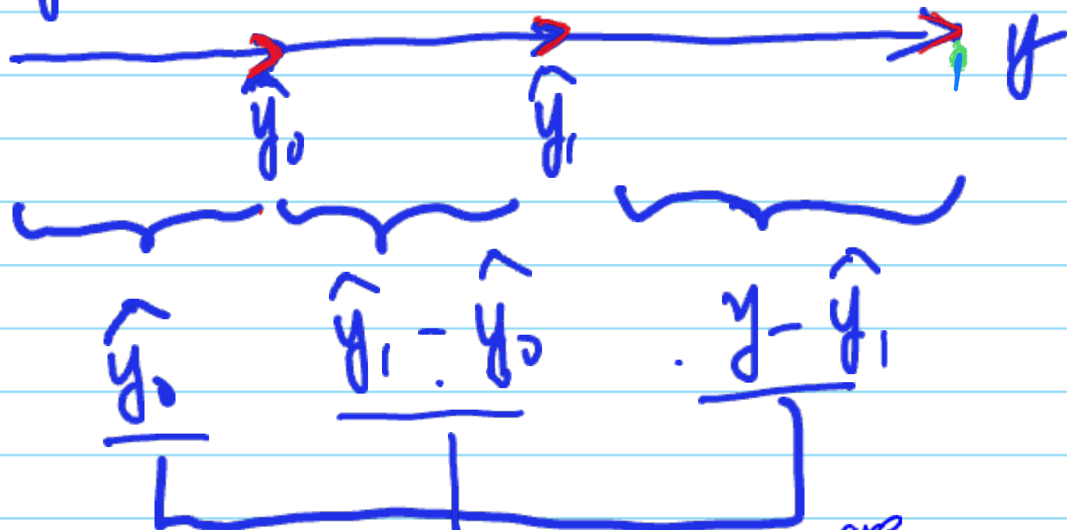


$C(P_0)^{\perp} C(P_1)$

$= C([x_1 - \text{proj}(x_1 | \hat{j}_2),\ x_2 - \text{proj}(x_2 | \hat{j}_2)])$

An illustrative figure

$\hat{y}_0 = P_0 y$, $\hat{y}_1 = P_1 y$, $C(P_0) \subseteq C(P_1)$



these three pieces are orthogonal

$$y = \hat{y}_0 + \hat{y}_1 - \hat{y}_0 + y - \hat{y}_1$$

$$\|y\|^2 = \|\hat{y}_0\|^2 + \|\hat{y}_1 - \hat{y}_0\|^2 + \|y - \hat{y}_1\|^2$$

$$\|\hat{y}_1 - \hat{y}_0\|^2 = \|\hat{y}_1\|^2 - \|\hat{y}_0\|^2$$

$$\|y - \hat{y}_1\|^2 = \|y\|^2 - \|\hat{y}_1\|^2$$

Similar to $(b-a)^2 = b^2 - a^2$

Example: (one-way ANOVA)

An example of data

$$
\underbrace{\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{bmatrix}}_{y}
\quad
\underbrace{\begin{bmatrix} g_1=1 \\ g_2=1 \\ g_3=2 \\ g_4=2 \\ g_5=3 \\ g_6=3 \end{bmatrix}}_{g \text{ (group index)}}
\quad
\begin{array}{c}
\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \\
\;\; x_1 \quad\; x_2 \quad\; x_3
\end{array}
\quad
\underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}}_{j_n}
$$

$j_n \in L(x_1, x_2, x_3)$.

$x_i = \mathbb{1}(g = i)$, indicator of group i

H_0: $y_{ij} = \mu + \varepsilon_{ij}$ $\quad [\, y = \underbrace{j_n}_{X} \cdot \underbrace{[\mu]}_{\beta} + \varepsilon \,]$

H_1: $y_{ij} = \mu_i + \varepsilon_i$

In matrix,

H_0:

$$y = j_n \cdot \mu + \varepsilon, \quad j_n = (1, 1, \ldots, 1)'$$

H_1:

$$y = [x_1, x_2, x_3] \cdot \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} + \varepsilon$$

projections:

Under H_0: $\quad \text{proj}(y \mid j_n) \equiv P_0 y$

under H_1: $\quad \text{proj}(y \mid L(x_1, x_2, x_3))$
$\equiv P_1 y$

$L(j_n) \subseteq L(x_1, x_2, x_3)$

since $j_n = x_1 + x_2 + x_3$

That is, H_0 is a reduced model of H_1.

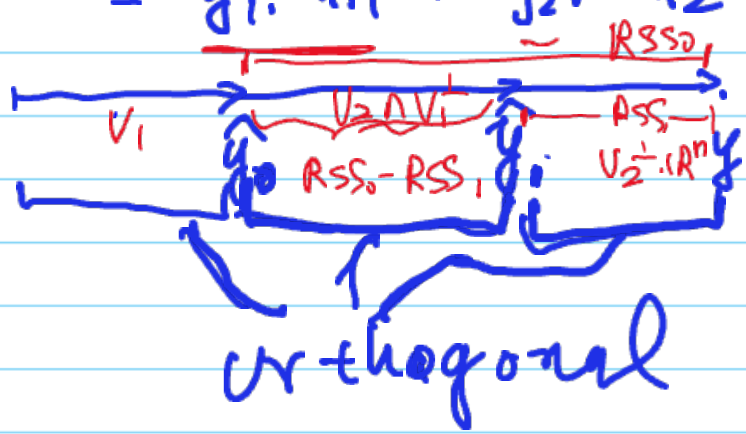
$$P_0 = \frac{1}{n} j_n j_n'$$

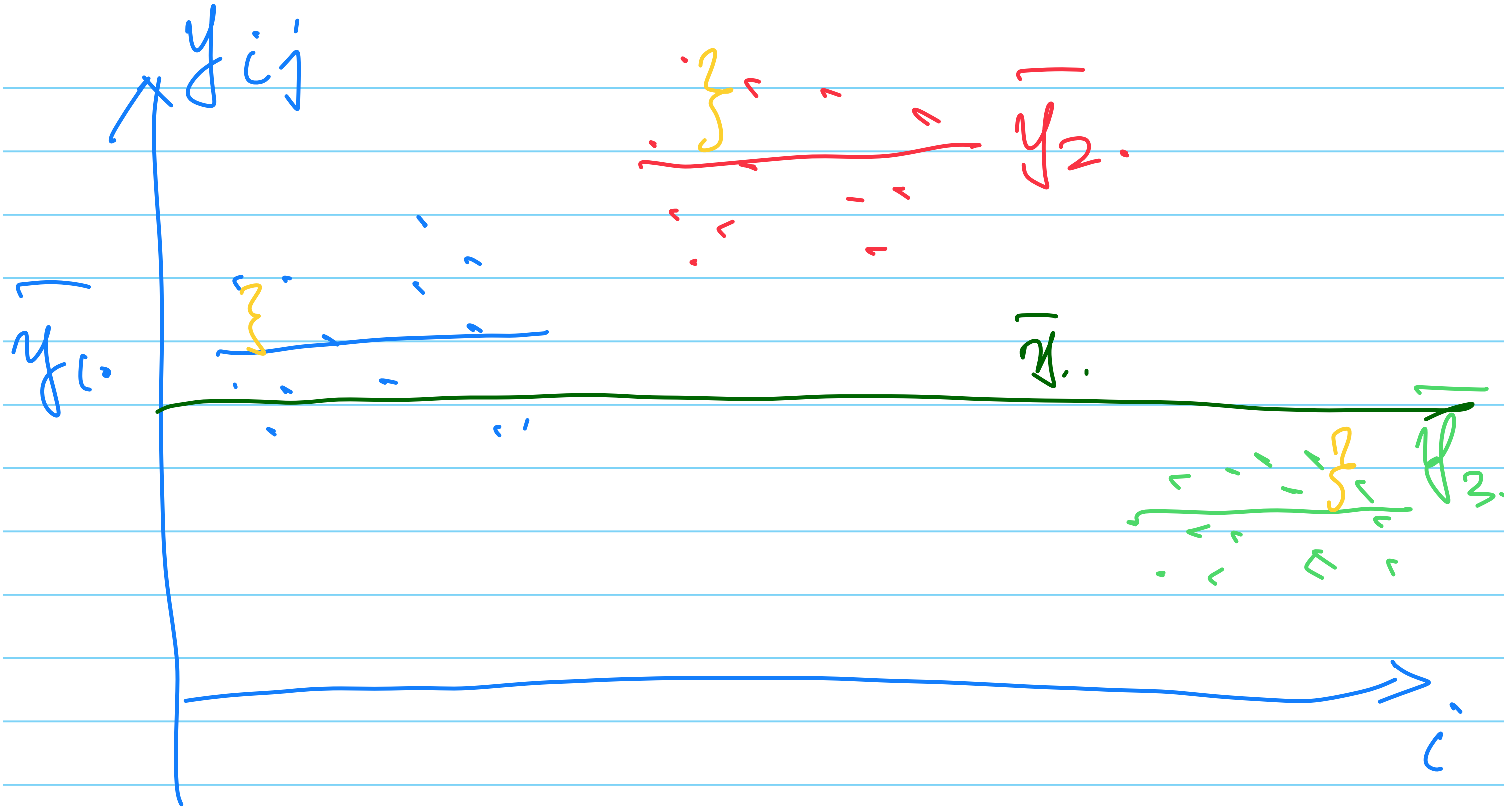
$\bar{y}_{..} \cdot j_n$

$$\hat{y}_0 = P_0 y = (\bar{y}_{..}, \bar{y}_{..}, \cdots, \bar{y}_{..})'$$

$$\hat{y}_1 = P_1 y = (\bar{y}_{1.}, \bar{y}_{1.}, \bar{y}_{2.}, \bar{y}_{2.}, \bar{y}_{3.}, \bar{y}_{3.})$$

$$= \bar{y}_{1.} \cdot x_1 + \bar{y}_{2.} \cdot x_2 + \bar{y}_{3.} \cdot x_3$$





$$\hat{y}_1 = \bar{y}_{1\cdot} \cdot x_1 + \bar{y}_{2\cdot} \cdot x_2 + \bar{y}_{3\cdot} \cdot x_3$$

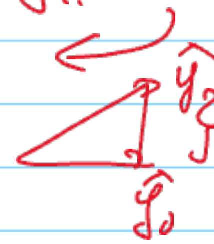
$$= (\underbrace{\bar{y}_{1\cdot}, \ldots, \bar{y}_{1\cdot}}_{n_1}, \underbrace{\bar{y}_{2\cdot}, \ldots, \bar{y}_{2\cdot}}_{n_2},$$

$$\underbrace{\bar{y}_{3\cdot}, \ldots, \bar{y}_{3\cdot}}_{n_3})'$$

$$\hat{y}_0 = \bar{y}_{\cdot\cdot} \mathbf{1}_n = (\bar{y}_{\cdot\cdot}, \ldots, \bar{y}_{\cdot\cdot})'$$

some SS based on $\hat{y}_0$ & $\hat{y}_1$:

$$RSS_0 = \| y - \hat{y}_0 \|^2 = \sum_{i,j} (y_{ij} - \bar{y}_{..})^2$$



$$= \| y \|^2 - \| \hat{y}_0 \|^2$$

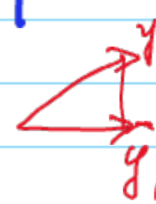
$$= \sum_{i,j} y_{ij}^2 - n \cdot \bar{y}_{..}^2$$

$$\frac{RSS_0}{n-1} = s_y^2 \quad \text{sample variance of } y$$

$$RSS_1 = \| y - \hat{y}_1 \|^2$$

$$= \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 \leftarrow \text{SS within group } i$$

$$= \| y \|^2 - \| \hat{y}_1 \|^2$$



$$= \sum_{ij} y_{ij}^2 - \sum_i n_i \bar{y}_{i.}^2$$

$$RSS_0 - RSS_1$$

$$= \| y - \hat{y}_0 \|^2 - \| y - \hat{y}_1 \|^2$$

$$= \| \hat{y}_0 - \hat{y}_1 \|^2$$

$$= \| \hat{y}_1 \|^2 - \| \hat{y}_0 \|^2 = \sum_i n_i \bar{y}_{i.}^2 - n \bar{y}_{..}^2$$

$$= \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 \cdot n_i \leftarrow \text{SS btw groups}$$

projections in orthogonal spaces

$y = x_1 + x_2 + \cdots + x_k, \quad x_i \perp x_j$

$$\underset{\uparrow\, I_n}{\mathbb{R}^n} = \underset{\uparrow\, P_1}{V_1} \oplus \underset{\uparrow\, P_2}{V_2} \oplus \cdots \oplus \underset{\uparrow\, P_k}{V_k} \qquad (i \neq j)$$

$V_1, V_2, \ldots, V_k$ are orthogonal

$$y = I_n y = P_1 y + P_2 y + \cdots + P_k y$$

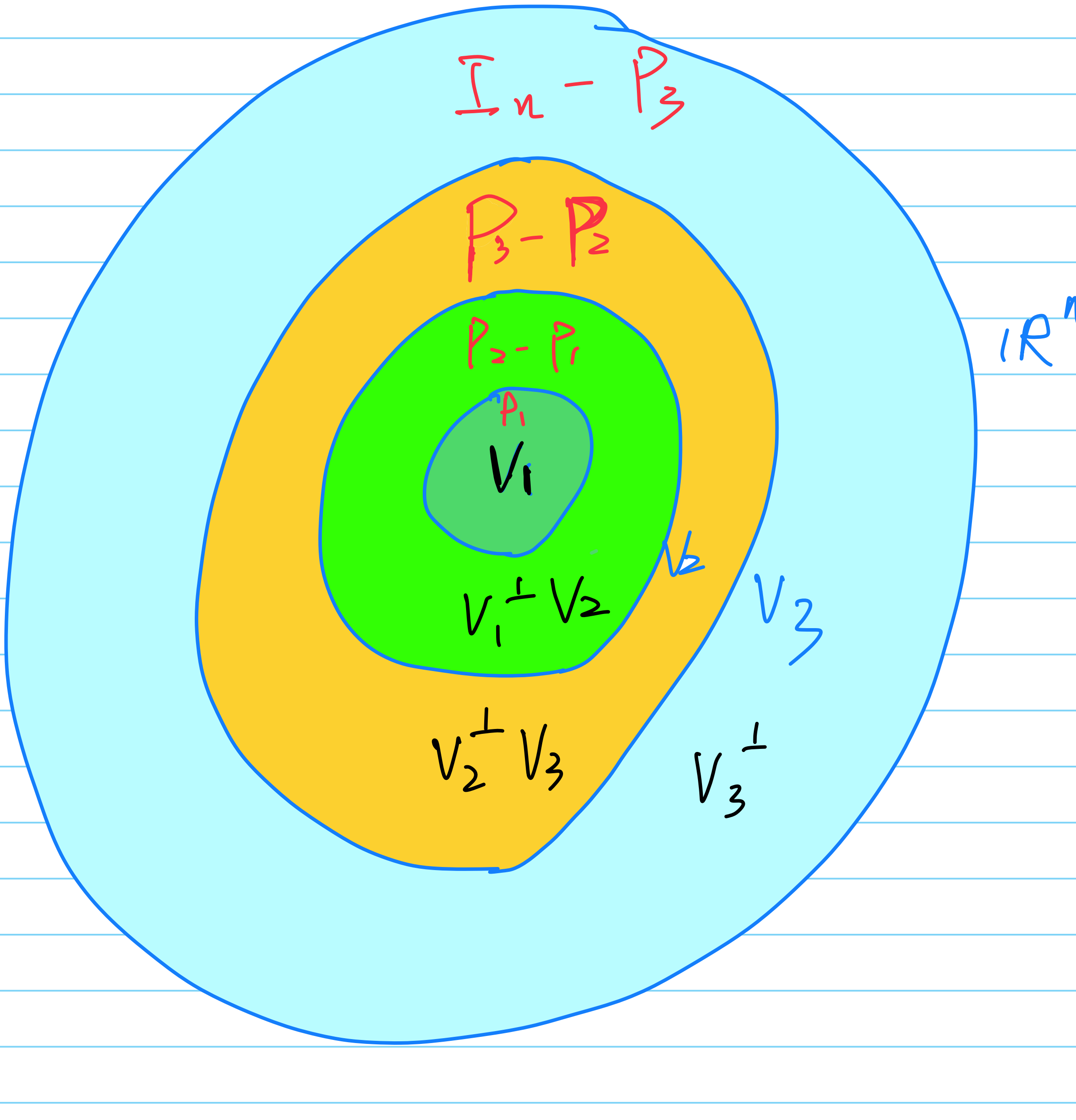
$$\|y\|^2 = \|P_1 y\|^2 + \|P_2 y\|^2 + \cdots + \|P_k y\|^2$$

$P_1 y, \ldots, P_k y$ are all orthogonal.

projection to nested spaces

$$V_1 \subseteq V_2 \subseteq \cdots \subseteq V_k \subseteq \mathbb{R}^n$$

$$\underset{\uparrow\, P_{V_1}}{V_1} \oplus \underset{\uparrow\, P_{V_2} - P_{V_1}}{V_2 V_1^\perp} \oplus \cdots \cdots \oplus \underset{\uparrow\, P_{V_k} - P_{V_{k-1}}}{V_k V_{k-1}^\perp} \oplus \underset{\uparrow\, I_n - P_{V_k}}{V_k^\perp} = \mathbb{R}^n$$


$I_n - P_3$
$P_3 - P_2$
$P_2 - P_1$
P_1
V_1
$V_1^\perp V_2$
V_2
V_3
$V_2^\perp V_3$
$V_3^\perp$
$\mathbb{R}^n$