STAT 714 fa 2023 Lec 01

Least squares estimation in linear models

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

$A^2 = AA$ $\qquad AA = A$

Idempotent matrix

A square matrix $\mathbf{A}$ is called *idempotent* if $\mathbf{A}^2 = \mathbf{A}$.

Exercise: Verify that these are idempotent matrices:

$$\begin{bmatrix} 3 & -2 \\ 3 & -2 \end{bmatrix} \qquad \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{I}$$

$$\begin{pmatrix} 3 & -2 \\ 3 & -2 \end{pmatrix} \begin{pmatrix} 3 & -2 \\ 3 & -2 \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ 3 & -2 \end{pmatrix}$$

Always have P projects onto Col P.

Projection matrix

A square matrix $\mathbf{P}$ is called a *projection matrix* onto the space V if

1. $\mathbf{P}$ is idempotent
2. for any $\mathbf{x}$, $\mathbf{P}\mathbf{x} \in V$
3. for any $\mathbf{z} \in V$, $\mathbf{P}\mathbf{z} = \mathbf{z}$

Sometimes we call projection matrices simply "projections".

Theorem (Every idempotent matrix is a projection)

Every idempotent matrix is a projection onto its own column space.

Prove the result.

Let A be idempotent.
Show that A is a proj. matrix onto $\text{Col}\, A$.

(i) idempotent $\boxed{\checkmark}$

(ii) For any $\underset{\sim}{x}$, we have $A\underset{\sim}{x} \in \text{Col}\, A$.

(iii) Let $\underset{\sim}{x} \in \text{Col}\, A$. Then $\underset{\sim}{x} = A\underset{\sim}{z}$ for some $\underset{\sim}{z}$.

Then $A\underset{\sim}{x} = A\underbrace{A\underset{\sim}{x}}_{\underset{\sim}{z}} = A\underset{\sim}{z} = \underset{\sim}{x}$.

$$P \text{ projects onto } \text{Col } P = \text{Span}\left\{\begin{pmatrix}3\\3\end{pmatrix}, \begin{pmatrix}-2\\-2\end{pmatrix}\right\} = \text{Span}\left\{\begin{pmatrix}3\\3\end{pmatrix}\right\}$$

symmetric

Exercise: Let

$$\mathbf{P} = \begin{bmatrix} 3 & -2 \\ 3 & -2 \end{bmatrix}, \quad \tilde{\mathbf{P}} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}, \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

1. Find $\mathbf{Pv}$, $\tilde{\mathbf{P}}\mathbf{v}$, $(\mathbf{I} - \mathbf{P})\mathbf{v}$, and $(\mathbf{I} - \tilde{\mathbf{P}})\mathbf{v}$.
2. Give the spaces onto which $\mathbf{P}$, $\tilde{\mathbf{P}}$, $(\mathbf{I} - \mathbf{P})$, and $(\mathbf{I} - \tilde{\mathbf{P}})$ project.
3. Draw pictures.

$$(\text{Col } \tilde{P})^\perp$$

$$P\underset{\sim}{v} = \begin{pmatrix}3 & -2\\3 & -2\end{pmatrix}\begin{pmatrix}1\\2\end{pmatrix} = \begin{bmatrix}-1\\-1\end{bmatrix}, \quad (I - P)\underset{\sim}{v} = \underset{\sim}{v} - P\underset{\sim}{v} = \begin{pmatrix}1\\2\end{pmatrix} - \begin{pmatrix}-1\\-1\end{pmatrix} = \begin{pmatrix}2\\3\end{pmatrix}$$

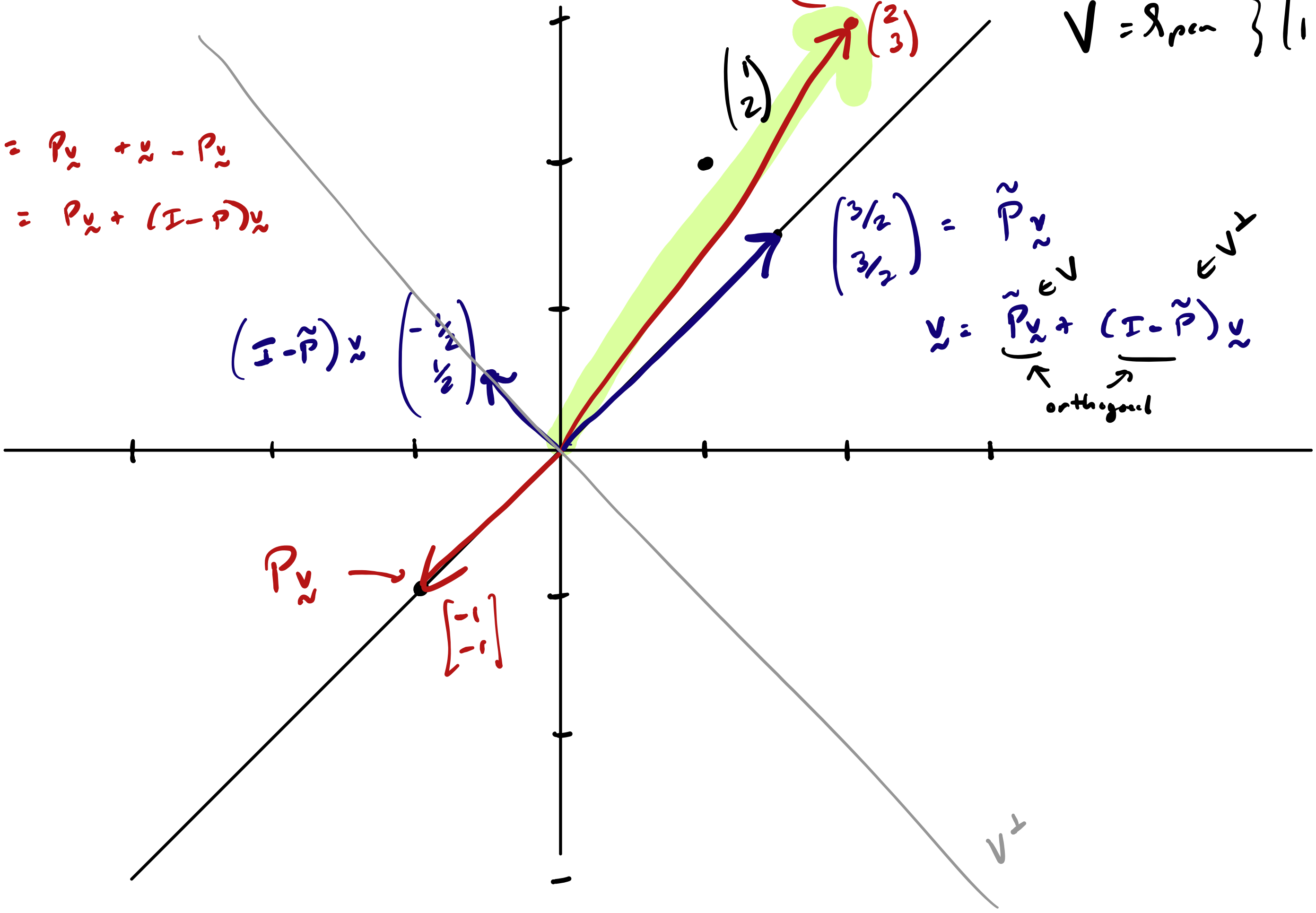
$$\tilde{P}\underset{\sim}{v} = \begin{pmatrix}1/2 & 1/2\\1/2 & 1/2\end{pmatrix}\begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}3/2\\3/2\end{pmatrix} \quad (I - \tilde{P})\underset{\sim}{v} = \underset{\sim}{v} - \tilde{P}\underset{\sim}{v} = \begin{pmatrix}1\\2\end{pmatrix} - \begin{pmatrix}3/2\\3/2\end{pmatrix} = \begin{pmatrix}-1/2\\1/2\end{pmatrix}$$

$$V = \text{span}\left\{\begin{pmatrix}1\\1\end{pmatrix}\right\}$$

$$\underset{\sim}{v} = P\underset{\sim}{v} + \underset{\sim}{v} - P\underset{\sim}{v}$$
$$= P\underset{\sim}{v} + (I - P)\underset{\sim}{v}$$

$$\underset{\sim}{v} = \underbrace{\tilde{P}\underset{\sim}{v}}_{\in V} + \underbrace{(I - \tilde{P})\underset{\sim}{v}}_{\in V^\perp}$$

orthogonal



We like projections that let us orthogonally decompose any vector $\mathbf{x}$ as

$$\mathbf{x} = \mathbf{P}\mathbf{x} + (\mathbf{I} - \mathbf{P})\mathbf{x}, \quad \text{where} \quad \mathbf{P}\mathbf{x} \cdot (\mathbf{I} - \mathbf{P})\mathbf{x} = 0.$$

Orthogonal projection

Let $\mathbf{P}$ be a projection matrix onto a subspace V. The projection is an *orthogonal projection* if $\underline{(\mathbf{I} - \mathbf{P})}$ is the projection matrix onto $V^{\perp}$.

Discuss: Which projection matrix corresponds to an orthogonal projection?

$$\mathbf{P} = \begin{bmatrix} 3 & -2 \\ 3 & -2 \end{bmatrix}, \quad \tilde{\mathbf{P}} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}.$$

Symmetry & P makes P an orthogonal projection.

Theorem (Uniqueness of symmetric projection)

Let $\mathbf{P}$ *be symmetric and a projection onto a vector space* V. *Then*

1. $\mathbf{P}$ *the unique symmetric projection onto* V.
2. $(\mathbf{I} - \mathbf{P})$ *is the unique symmetric projection onto* $V^{\perp}$.

$$\boxed{c^T c = 0 \iff c = 0}$$

See results A.16 and Cor A.4 of Monahan (2008).

Prove the result.

① Let P_1 and P_2 be symmetric and projection matrices onto V.

Want to show $P_1 = P_2$.

Sufficient to show $(P_1 - P_2)^T (P_1 - P_2) = 0$

We have

$$(P_1-P_2)^T(P_1-P_2) = P_1^T P_1 - P_1^T P_2 - P_2^T P_1 + P_2^T P_2$$

$$= P_1 P_1 - P_1 P_2 - P_2 P_1 + P_2 P_2$$

$$= P_1 - \underbrace{P_1 P_2}_{P_2} - \underbrace{P_2 P_1}_{P_1} + P_2$$

Take any $\underset{\sim}{z}$. Then $P_1 \underset{\sim}{z} \in V$, so that $P_2 \underbrace{P_1 \underset{\sim}{z}}_{\in V} = P_1 \underset{\sim}{z}$.

true for all $\underset{\sim}{z}$

$$\Rightarrow \quad P_2 P_1 = P_1 .$$

Take any $\underset{\sim}{z}$. Then $P_2 \underset{\sim}{z} \in V$. Then $P_1 \underbrace{P_2 \underset{\sim}{z}}_{\in V} = P_2 \underset{\sim}{z} \quad \forall \underset{\sim}{z}$.

$$= P_1 - P_2 - P_1 + P_2$$

$$= 0.$$

(2) Just need to show that $I-P$ projects onto $V^\perp$.

$V = \text{Col } P$. $\quad V^\perp = (\text{Col } P)^\perp = \text{Nul } P^T = \text{Nul } P$.

So show that $I-P$ projects onto $\text{Nul } P$.

(i) $(I-P)(I-P) = I - P - P + P = I - P$.

(ii) For $\underline{\underline{\text{any}}}$ $\underset{\sim}{z}$, we have $P(I-P)\underset{\sim}{z} = P\underset{\sim}{z} - PP\underset{\sim}{z} = P\underset{\sim}{z} - P\underset{\sim}{z} = \underset{\sim}{0}$

$\Rightarrow (I-P)\underset{\sim}{z} \in \text{Nul } P$

(iii) Take $\underset{\sim}{v} \in \text{Nul } P$. Then $P\underset{\sim}{v} = 0$.

And $(I - P)\underset{\sim}{v} = \underset{\sim}{v} - \underbrace{P\underset{\sim}{v}}_{0} = \underset{\sim}{v}$.

Generalized inverse of a matrix

A matrix **G** which satisfies $\mathbf{AGA} = \mathbf{A}$ is called a *generalized inverse* of **A**.

$A\underset{\sim}{x} = b \iff \underset{\sim}{x} = A^{-1}\underset{\sim}{b}$ if A is invertible.

Theorem (Generalized inverses for solving systems of equations)

Suppose $\boxed{\mathbf{Ax} = \mathbf{b}}$ *is consistent and let* **G** *be a generalized inverse of* **A**. *Then*

1. $\mathbf{Gb}$ *is a solution to* $\mathbf{Ax} = \mathbf{b}$. — Says that $A(G\underset{\sim}{b}) = \underset{\sim}{b}$
2. $\hat{\mathbf{x}}$ *is a solution to* $\mathbf{Ax} = \mathbf{b}$ *iff there exists* $\mathbf{z}$ *such that* $\hat{\mathbf{x}} = \mathbf{Gb} + (\mathbf{I} - \mathbf{GA})\mathbf{z}$.

See Res A.12 and A.13 of Monahan (2008).

Prove the results.

① If $A\underset{\sim}{x} = \underset{\sim}{b}$ is consistent, then there is at least one solution $\hat{\underset{\sim}{x}}$ such that $A\hat{\underset{\sim}{x}} = \underset{\sim}{b}$. Then

$$A(G\underset{\sim}{b}) = AGA\hat{\underset{\sim}{x}} = A\hat{\underset{\sim}{x}} = \underset{\sim}{b}.$$

claim:

(2) $\hat{\underset{\sim}{x}}$ is a solution to $A\underset{\sim}{x} = \underset{\sim}{b}$ $\Rightarrow$ $\exists\, \underset{\sim}{z}$ s.t. $\hat{\underset{\sim}{x}} = Gb + (I - GA)\underset{\sim}{z}$.

proof:

Write $$\hat{\underset{\sim}{x}} = G\underset{\sim}{b} + \hat{\underset{\sim}{x}} - G\underset{\sim}{b}$$

$$\left(A\hat{\underset{\sim}{x}} = \underset{\sim}{b} \Rightarrow GA\hat{\underset{\sim}{x}} = G\underset{\sim}{b}\right)$$

$$= G\underset{\sim}{b} + \hat{\underset{\sim}{x}} - GA\hat{\underset{\sim}{x}}$$

$$= G\underset{\sim}{b} + (I - GA)\hat{\underset{\sim}{x}}$$

So $\hat{\underset{\sim}{x}} = Gb + (I - GA)\underset{\sim}{z}$ with $\underset{\sim}{z} = \hat{\underset{\sim}{x}}$.

For some $\underset{\sim}{z}$, set $\hat{\underset{\sim}{x}} = Gb + (I - GA)\underset{\sim}{z}$. $\Rightarrow$ $\hat{\underset{\sim}{x}}$ is a solution to $A\underset{\sim}{x} = \underset{\sim}{b}$

$$A\hat{x} = A\left(Gb + (I - GA)\underset{\sim}{z}\right)$$

$$= AGb + A(I - GA)\underset{\sim}{z}$$

$$= AGb + \underbrace{(A - \overbrace{AGA}^{A})\underset{\sim}{z}}_{\underset{\sim}{0}}$$

$$= AGb$$

since Gb is a solution to $A\underset{\sim}{x} = \underset{\sim}{b}$.

$$= \underset{\sim}{b}$$

$$\{\underset{\sim}{x} : A\underset{\sim}{x} = \underset{\sim}{b}\} = \{G\underset{\sim}{b} + (I - GA)\underset{\sim}{z} \text{ for all } \underset{\sim}{z}\},$$

where G is a gen. inverse of A.

Theorem (Generalized inverse recipe using block structure)

Let $\mathbf{A}$ *be an* $m \times n$ *matrix with rank* r*. If we can partition* $\mathbf{A}$ *as*

$$\mathbf{A} = \begin{bmatrix} \mathbf{C} & \mathbf{D} \\ \mathbf{E} & \mathbf{F} \end{bmatrix}, \quad \text{with } \mathbf{C}\ r \times r \text{ invertible, then} \quad \mathbf{G} = \begin{bmatrix} \mathbf{C}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}$$

is a generalized inverse of $\mathbf{A}$*, where the* $\mathbf{O}$ *matrices have the necessary dimensions.*

Similarly, if $\mathbf{F}$ *is* $r \times r$ *invertible, then* $\mathbf{G} = \begin{bmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{F}^{-1} \end{bmatrix}$ *is a gen. inverse of* $\mathbf{A}$*.*

See Res A.10 and Cor A.3 of Monahan (2008).

Make gen. inv. of *any* matrix by permuting rows/columns to get such a partition. See Res A.11 of Monahan (2008).

Prove the first result.

If $A = \begin{pmatrix} \underset{r\times r}{C} & D \\ E & F \end{pmatrix}$ invertible. rank $A = r$.

Write E, F as linear comb of rows C D

$$[E \;\; F] = K[C \;\; D]$$

$$\Rightarrow \; E = KC$$

$$F = KD$$

Let $G = \begin{pmatrix} C^{-1} & 0 \\ 0 & 0 \end{pmatrix}$.

Then

$$AGA = \begin{pmatrix} C & D \\ E & F \end{pmatrix}\begin{pmatrix} C^{-1} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} C & D \\ E & F \end{pmatrix}$$

$$= \begin{pmatrix} I & 0 \\ EC^{-1} & 0 \end{pmatrix}\begin{pmatrix} C & D \\ E & F \end{pmatrix}$$

$$= \begin{pmatrix} C & D \\ E & EC^{-1}D \end{pmatrix}$$

$$= \begin{pmatrix} C & D \\ E & K\underbrace{CC^{-1}}_{I}D \end{pmatrix}$$

$$= \begin{pmatrix} C & D \\ E & F \end{pmatrix}.$$

We sometimes denote the generalized inverse of a matrix $\mathbf{A}$ by $\mathbf{A}^{-}$.

Theorem (Projections constructed with a generalized inverse)

Let $\mathbf{A}^{-}$ be a generalized inverse of $\mathbf{A}$. Then

1. $\mathbf{A}\mathbf{A}^{-}$ *is a projection onto* $\operatorname{Col}\mathbf{A}$. ✓
2. $(\mathbf{I} - \mathbf{A}^{-}\mathbf{A})$ *is a projection onto* $\operatorname{Nul}\mathbf{A}$.

See Res A.14 and A.15 of Monahan (2008).

Prove the result.

① (i) $\underbrace{AA^{-}A}_{A}A^{-} = AA^{-}$ ✓

$Ax = v$

$\uparrow A^{-}v$

(ii) For any z, $AA^{-}z \in \operatorname{Col} A$

$AA^{-}v = v$

(iii) For any $v \in \operatorname{Col} A$, $v = Ab$ for some b.

Then $AA^{-}v = \overbrace{AA^{-}A}b = Ab = v$.

Write $A\underset{\sim}{x} = \underset{\sim}{v}$. The solution is $A^-\underset{\sim}{v}$, from earlier result.
So $AA^-\underset{\sim}{v} = \underset{\sim}{v}$.

② $(I - A^-A)$ proj. onto $\mathrm{Nul}\, A$.

(i) $$(I - A^-A)(I - A^-A) = I - A^-A - A^-A + \underbrace{A^-\overbrace{AA^-A}^{A}}_{A^-A}$$
$$= I - A^-A$$

(ii) For any $\underset{\sim}{z}$, $$A(I - A^-A)\underset{\sim}{z} = (A - \overbrace{AA^-A}^{A})\underset{\sim}{z} = \underset{\sim}{0}$$
so $(I - A^-A)\underset{\sim}{z} \in \mathrm{Nul}\, A$.

(iii) For $\underset{\sim}{v} \in \mathrm{Nul}\, A$.

$$(I - A^-A)\underset{\sim}{v} = \underset{\sim}{v} - A^-A\underset{\sim}{v} = \underset{\sim}{v}.$$

For the remainder of the lecture, let $\mathbf{X}$ be an $n \times p$ matrix and $\mathbf{y}$ be a vector in $\mathbb{R}^n$.

We have in mind data coming from a model like

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \varepsilon$$

...but we are not thinking yet about the distribution of ε.

We consider least-squares "estimation" of $\mathbf{b}$, but no statistics yet—only geometry.

$$\hat{\underset{\sim}{b}} = \underset{\underset{\sim}{b}}{\operatorname{argmin}} \; \| \underset{\sim}{y} - X\underset{\sim}{b} \|^2$$

Least-squares solution

A *least-squares solution* to $\mathbf{Xb} = \mathbf{y}$ is a vector $\hat{\mathbf{b}} \in \mathbb{R}^p$ such that

$$\|\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}\| \leq \|\mathbf{y} - \mathbf{Xb}\| \quad \text{for all } \mathbf{b} \in \mathbb{R}^p.$$

Theorem (Least-squares solution iff solution to normal equations)

1. *The equation* $\mathbf{X}^T\mathbf{Xb} = \mathbf{X}^T\mathbf{y}$ *is consistent.*
2. $\|\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}\| \leq \|\mathbf{y} - \mathbf{Xb}\|$ *for all* $\mathbf{b} \in \mathbb{R}^p$ *if and only if* $\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}$.

See Cor 2.1 and Res 2.3 of Monahan (2008).

We call the set of equations $\mathbf{X}^T\mathbf{Xb} = \mathbf{X}^T\mathbf{y}$ the *normal equations*.

Prove the result.

① $X^T X \underset{\sim}{b} = X^T \underset{\sim}{y}$ is always consistent.

$\underbrace{X^T X \underset{\sim}{b}}_{\in \text{Col } X^T X}$ $\underbrace{X^T \underset{\sim}{y}}_{\in \text{Col } X^T}$

$$\boxed{(\text{Col } A)^\perp = \text{Nul } A^T}$$

Sufficient to show $\text{Col } X^T X = \text{Col } X^T$.

$$(\text{Col } X^T X)^\perp = \text{Nul } X^T X$$
$$(\text{Col } X^T)^\perp = \text{Nul } X$$

same.

② claim: $X^T X \hat{\underset{\sim}{b}} = X^T \underset{\sim}{y} \Rightarrow \|\underset{\sim}{y} - X\hat{\underset{\sim}{b}}\| \le \|\underset{\sim}{y} - X\underset{\sim}{b}\| \quad \forall \underset{\sim}{b} \in \mathbb{R}^p$

proof: Write, for any $\underset{\sim}{b} \in \mathbb{R}^p$

$$\|\underset{\sim}{y} - X\underset{\sim}{b}\|^2 = \|\underset{\sim}{y} - X\hat{\underset{\sim}{b}} - (X\underset{\sim}{b} - X\hat{\underset{\sim}{b}})\|^2$$

$$= \left((\underline{\underset{\sim}{y} - X\hat{\underset{\sim}{b}}}) - (\underline{X\underset{\sim}{b} - X\hat{\underset{\sim}{b}}})\right)^T \left((\underline{\underset{\sim}{y} - X\hat{\underset{\sim}{b}}}) - (\overbrace{X\underset{\sim}{b} - X\hat{\underset{\sim}{b}}}^{X(\underset{\sim}{b} - \hat{\underset{\sim}{b}})})\right)$$

$$= (\underset{\sim}{y} - X\hat{\underset{\sim}{b}})^T (\underset{\sim}{y} - X\hat{\underset{\sim}{b}}) + (X\underset{\sim}{b} - X\hat{\underset{\sim}{b}})^T (X\underset{\sim}{b} - X\hat{\underset{\sim}{b}})$$
$$- 2(\underset{\sim}{y} - X\hat{\underset{\sim}{b}})^T (X\underset{\sim}{b} - X\hat{\underset{\sim}{b}})$$

$$= \|\underset{\sim}{y} - X\hat{\underset{\sim}{b}}\|^2 + \|X(\underset{\sim}{b} - \hat{\underset{\sim}{b}})\|^2$$
$$- 2 \underbrace{(\underset{\sim}{y} - X\hat{\underset{\sim}{b}})^T}_{[\underbrace{X^T(\underset{\sim}{y} - X\hat{\underset{\sim}{b}})}_{X^T\underset{\sim}{y} - X^T X\hat{\underset{\sim}{b}} = 0}]^T} X(\underset{\sim}{b} - \hat{\underset{\sim}{b}})$$

$$= \|\underset{\sim}{y} - X\hat{\underset{\sim}{b}}\|^2 + \underbrace{\|X(\underset{\sim}{b} - \hat{\underset{\sim}{b}})\|^2}_{\ge 0}$$

$$\Rightarrow \|\underset{\sim}{y} - X\underset{\sim}{b}\|^2 \ge \|\underset{\sim}{y} - X\hat{\underset{\sim}{b}}\|^2.$$

claim: $\|\underset{\sim}{y} - \underbrace{X\hat{\underset{\sim}{b}}}_{\in \text{Col } X}\| \leq \|\underset{\sim}{y} - \underbrace{X\underset{\sim}{b}}_{\in \text{Col } X}\| \quad \forall\ \underset{\sim}{b} \in \mathbb{R}^p \quad \Rightarrow \quad X^T X \hat{\underset{\sim}{b}} = X^T \underset{\sim}{y}$

proof: Best Approx. theorem says $X\hat{\underset{\sim}{b}} = \text{Proj}_{\text{Col } X}\, \underset{\sim}{y}$.

Now write

$$\underset{\sim}{y} = \underbrace{X\hat{\underset{\sim}{b}}}_{\in \text{Col } X} + \underbrace{(\underset{\sim}{y} - X\hat{\underset{\sim}{b}})}_{\in (\text{Col } X)^\perp = \text{Nul } X^T}$$

By the orthogonal decomp. theorem, this decomp. is unique and

$X\hat{b}$ is orthogonal to $\underset{\sim}{y} - X\hat{b}$.

Since $\underset{\sim}{y} - X\hat{\underset{\sim}{b}} \in \text{Nul } X^T$ then

$$X^T(\underset{\sim}{y} - X\hat{\underset{\sim}{b}}) = \underset{\sim}{0}$$

$$\iff X^T X \hat{\underset{\sim}{b}} = X^T \underset{\sim}{y}.$$

Can also use calculus to obtain the normal equations...

For a real-valued function $Q(\mathbf{x})$ taking vectors in $\mathbb{R}^n$, define

$$\frac{\partial}{\partial \mathbf{x}} Q(\mathbf{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1} Q(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial x_n} Q(\mathbf{x}) \end{bmatrix}.$$

Theorem (Derivative of linear and quadratic forms)

For a vector **a** *and a matrix* **A**, *we have*

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{a}^T \mathbf{x} = \mathbf{a} \quad \textit{and} \quad \frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{x} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x}.$$

A least-squares solution of $\mathbf{X}\mathbf{b} = \mathbf{y}$ is a minimizer of $Q(\mathbf{b}) = \|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2$.

Exercise: Use fact that $\hat{\mathbf{b}}$ minimizes $Q(\mathbf{b})$ iff $\frac{\partial}{\partial \mathbf{b}} Q(\mathbf{b})\Big|_{\mathbf{b}=\hat{\mathbf{b}}} = \mathbf{0}$ to get normal eqs.

$$Q(\underset{\sim}{b}) = \| \underset{\sim}{y} - X \underset{\sim}{b} \|^2$$

$$= (\underset{\sim}{y} - X\underset{\sim}{b})^T (\underset{\sim}{y} - X\underset{\sim}{b})$$

$$= \underset{\sim}{y}^T \underset{\sim}{y} - 2\underset{\sim}{y}^T X \underset{\sim}{b} + \underset{\sim}{b}^T X^T X \underset{\sim}{b}$$

$$\frac{\partial}{\partial \underset{\sim}{b}} Q(\underset{\sim}{b}) = \underline{-2 X^T \underset{\sim}{y}} + \underline{\left(X^T X + (X^T X)^T \right)} \underset{\sim}{b}$$

$$= -2 X^T \underset{\sim}{y} + 2 X^T X \underset{\sim}{b}$$

$$\overset{\text{set}}{=} \underset{\sim}{0}$$

$$\Rightarrow \quad X^T X \underset{\sim}{b} = X^T \underset{\sim}{y} .$$

gen inv. of $\mathbf{A}$

2. $\hat{\mathbf{x}}$ *is a solution to* $\mathbf{Ax} = \mathbf{b}$ *iff there exists* $\mathbf{z}$ *such that* $\hat{\mathbf{x}} = \mathbf{Gb} + (\mathbf{I} - \mathbf{GA})\mathbf{z}$.

Theorem (Characterization of solutions to the normal equations)

The vector $\hat{\mathbf{b}}$ *is a solution to* $\boxed{\mathbf{X}^T\mathbf{X}}\hat{\mathbf{b}} = \boxed{\mathbf{X}^T\mathbf{y}}$ *iff there exists a vector* $\mathbf{z}$ *such that*

$$\hat{\mathbf{b}} = \boxed{(\mathbf{X}^T\mathbf{X})^-}\mathbf{X}^T\mathbf{y} + (\mathbf{I} - (\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X})\mathbf{z}.$$

If $\mathbf{X}$ *has full-column rank, then* $\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ *is the unique solution.*

rank X^TX = rank X

Col X^TX = Col X^T

=> dim Col X^TX = dim Col X^T

rank X^TX = rank X^T = rank X

Prove the result.

$$\hat{\mathbf{b}} = \boxed{(\mathbf{X}^T\mathbf{X})^-}\mathbf{X}^T\mathbf{y} + (\mathbf{I} - (\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X})\mathbf{z}. \leftarrow$$

Exercise: Characterize the set of solutions to the normal equations when

$$\mathbf{y} = \begin{bmatrix} 1 \\ -1 \\ 2 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

To find $\hat{b}$ such that $\|y - X\hat{b}\| \leq \|y - Xb\| \quad \forall\, b \in \mathbb{R}^p$,

take $\hat{b}$ satisfying $X^TX\hat{b} = X^Ty$.

$$X^TX = \begin{bmatrix} 3 & 1 & 2 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 0 & 3 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \leftarrow \text{has rank } 3. \qquad X^Ty = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

F

Theorem (Generalized inverse recipe using block structure)

Let $\mathbf{A}$ be an $m \times n$ matrix with rank r. If we can partition $\mathbf{A}$ as

$$\mathbf{A} = \begin{bmatrix} \mathbf{C} & \mathbf{D} \\ \mathbf{E} & \mathbf{F} \end{bmatrix}, \quad \text{with } \mathbf{C} \ r \times r \text{ invertible, then} \quad \mathbf{G} = \begin{bmatrix} \mathbf{C}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}$$

is a generalized inverse of $\mathbf{A}$, where the $\mathbf{O}$ matrices have the necessary dimensions.

Similarly, if $\mathbf{F}$ is $r \times r$ invertible, then $\mathbf{G} = \begin{bmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{F}^{-1} \end{bmatrix}$ is a gen. inverse of $\mathbf{A}$.

Find gen. inv. of X^TX.

$$(X^TX)^- = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix}^{-1}}_{\frac{1}{2}\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 3 \end{bmatrix}} & \\ 0 & & & \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & -1/2 \\ 0 & 0 & -1/2 & 3/2 \end{bmatrix}$$

$$\hat{\underset{\sim}{b}} = (X^TX)^- X^T \underset{\sim}{y} + \left(I - (X^TX)^- X^TX\right) \underset{\sim}{z} \qquad \underset{\sim}{z} \in \mathbb{R}^4$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & -1/2 \\ 0 & 0 & -1/2 & 3/2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \\ 0 \end{bmatrix} + \underbrace{\left(I - \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & -1/2 \\ 0 & 0 & -1/2 & 3/2 \end{bmatrix} \begin{bmatrix} 3 & 1 & 2 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 0 & 3 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \right)} \underset{\sim}{z}$$

$$= \begin{bmatrix} 0 \\ 1 \\ 1/2 \\ -1/2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \underset{\sim}{z} \quad , \quad \underset{\sim}{z} \in \mathbb{R}^4$$

$$= \begin{bmatrix} 0 \\ 1 \\ 1/2 \\ -1/2 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} w \quad , \quad w \in \mathbb{R}.$$

G is a gen inv of A if $AGA = A$

These are helper results for constructing the orthogonal projection onto Col $\mathbf{X}$.

Theorem ("Cool result" and generalized inverse of $\mathbf{X}$)

1. $\mathbf{X}^T\mathbf{X}\mathbf{A} = \mathbf{X}^T\mathbf{X}\mathbf{B} \iff \mathbf{X}\mathbf{A} = \mathbf{X}\mathbf{B}$.
2. $(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T$ *is a generalized inverse of* $\mathbf{X}$. ✓

See Res 2.4 and 2.5 of Monahan (2008).

Prove the results.

(2) Since $(X^TX)^-$ is a gen. inv. of X^TX, we can write

$$X^TX(X^TX)^-X^TX = X^TX(I)$$

$$\iff X\boxed{(X^TX)^-X^T}X = X(I)$$

gen. inv. of X

$X\hat{b}$ is the orthogonal proj. of $\underset{\sim}{y}$ onto Col X.

Theorem (Orthogonal projection onto Col $\mathbf{X}$)

The matrix $\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T$ *and the matrix* $\mathbf{I} - \mathbf{P}_\mathbf{X}$ *are*

1. *projections onto* Col $\mathbf{X}$ *and* Nul $\mathbf{X}^T$, *respectively*
2. *invariant to the choice of generalized inverse*
3. *symmetric (therefore unique)*

See Thm 2.4 and Res 2.6 of Monahan (2008).

Prove the results.

① Show that P_X is a projection onto Col X.

(i) idempotent? $P_X P_X = \underbrace{X \underset{\uparrow \text{ g-inv. of } X}{\boxed{(X^TX)^- X^T}} X}(X^TX)^- X^T$

$= X(X^TX)^-X^T$

(ii) For any $\underset{\sim}{y} \in \mathbb{R}^n$, $\quad P_X \underset{\sim}{y} = X(X^TX)X^T\underset{\sim}{y} \in \text{Col}X$

(iii) For any $\underset{\sim}{v} \in \text{Col}\,X$, $\quad \exists\ \underset{\sim}{u} : \underset{\sim}{v} = X\underset{\sim}{u}$.

$$
\begin{aligned}
P_X \underset{\sim}{v} &= X(X^TX)X^T\underset{\sim}{v} \\
&= X\boxed{(X^TX)X^T}(X\underset{\sim}{u}) \quad \text{some } \underset{\sim}{u} \\
&= X\underset{\sim}{u} \\
&= \underset{\sim}{v}
\end{aligned}
$$

② Let G_1 and G_2 be gen. inverses of X^TX.

Then

$$\cancel{X^T}X G_1 X^TX = \cancel{X^T}X G_2 X^TX = \cancel{X^T}X$$

same?

Cool next
$\Leftrightarrow$ $\boxed{XG_1X^T}X = \boxed{XG_2X^T}X$

$\Leftrightarrow$ $\cancel{X^T}X G_1^T X^T = \cancel{X^T}X G_2^T X^T$

Cool result
$\Leftrightarrow$ $X G_1^T X^T = X G_2^T X^T$ $\qquad AGA$

$\Leftrightarrow$ $X G_1 X^T = X G_2 X^T = P_X$

③ Let G be a gen inv of A, where A is a symmetric matrix.
Then $\underbrace{\frac{1}{2}(G + G^T)}_{\text{symmetric}}$ is also a gen inverse of A.

Choose a symmetric g-inv of X^TX, say $(X^TX)^-$.

Then $(P_X)^T = X((X^TX)^-)^T X^T = X(X^TX)^- X^T.$

Let $\mathbf{A}^-$ *be a generalized inverse of* $\mathbf{A}$*. Then*

1. $\mathbf{AA}^-$ *is a projection onto* Col $\mathbf{A}$. ✓

From here

$X(X^TX)^- X^T \underset{\sim}{y}$

Result

We have $\mathbf{X}^T\mathbf{Xb} = \mathbf{X}^T\mathbf{y}$ *if and only if* $\mathbf{Xb} = \mathbf{P}_X\mathbf{y}$.

See Res 2.7 of Monahan (2008).

is a projection onto Col X^T

Prove the result.

$\Rightarrow$ Let $\hat{\underset{\sim}{b}}$ satisfy $X^TX\hat{\underset{\sim}{b}} = X^T\underset{\sim}{y} = \boxed{X^T(X^T)^-} X^T\underset{\sim}{y}$

$\in$ Col X^T

$= X^T X(X^TX)^- X^T\underset{\sim}{y}$

col result $\iff$ $X\hat{\underset{\sim}{b}} = X(X^TX)^- X^T\underset{\sim}{y}$

$= P_X\underset{\sim}{y}$

Find $(X^T)^-$:

Use another version of the "Col result":
$AX^TX = BX^TX \iff AX^T = BX^T$.

Then we have
$X^TX(X^TX)^- X^TX = X^TX$

$\iff$

$X^T\underbrace{X(X^TX)^-}X^T = X^T$

So $X(X^TX)^-$ is a g.inv of X^T.

$\Leftarrow$ let $X\hat{\underset{\sim}{b}} = P_X \underset{\sim}{y}$. Then $X\hat{\underset{\sim}{b}} = X(X^TX)^- X^T \underset{\sim}{y}$

g-inv of X^T

$$\Rightarrow X^TX\hat{\underset{\sim}{b}} = X^T \boxed{X(X^TX)^-} X^T \underset{\sim}{y}$$

$$= X^T \underset{\sim}{y}.$$

Sums of squares

For $\hat{\mathbf{b}}$ satisfying $\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}$ we define the

1. *fitted values* as $\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}} = P_X \underset{\sim}{y}$
2. *residuals* as $\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}}$
3. *total sum of squares (SST)* as $\|\mathbf{y}\|^2$
4. *regression sum of squares (SSR)* as $\|\hat{\mathbf{y}}\|^2$
5. *error sum of squares (SSE)* as $\|\hat{\mathbf{e}}\|^2$.

Theorem (Sum of squares decomposition)

We have SST $=$ SSR $+$ SSE, *or* $\|\mathbf{y}\|^2 = \|\hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{e}}\|^2$.

Prove the result.

$$\|\underset{\sim}{y}\|^2 = \|\underset{\sim}{y} - \hat{\underset{\sim}{y}} + \hat{\underset{\sim}{y}}\|^2 = \|\hat{\underset{\sim}{e}} + \hat{\underset{\sim}{y}}\|^2$$

$$= \|\hat{\underset{\sim}{e}}\|^2 + \|\hat{\underset{\sim}{y}}\|^2 + \underbrace{2\,\hat{\underset{\sim}{e}}^T\hat{\underset{\sim}{y}}}_{=0}$$

$$= \|\hat{\underset{\sim}{e}}\|^2 + \|\hat{\underset{\sim}{y}}\|^2 + 2\underbrace{(\underset{\sim}{y} - X\hat{b})^T}_{\in \mathcal{C}(X)^\perp}\underbrace{\hat{\underset{\sim}{y}}}_{\in \mathcal{C}(X)}$$

$$= \|\hat{\underset{\sim}{e}}\|^2 + \|\hat{\underset{\sim}{y}}\|^2.$$

Monahan, J. F. (2008). *A primer on linear models*. CRC Press.