Linear Algebra and Calculus Review

⇒ Little Calculus Review

$$X_1, \dots, X_n \overset{iid}{\sim} U(a,b) \Leftrightarrow \underset{\sim}{\overset{pdf}{f(x)}} = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

The maximum likelihood estimators for a and b are

$$\hat{a}_{MLE} = X_{(1)}, \quad \hat{b}_{MLE} = X_{(n)}$$

"order statistics" (no proof at this moment)

We want to check whether the estimators are unbiased or not.
So we need to compute $E[\hat{a}_{MLE}]$ and $E[\hat{b}_{MLE}]$

- $E[\hat{a}_{MLE}] = \int_a^b x\,(\text{pdf of } X_{(1)})\,dx$

how to find pdf of $X_{(1)}$? We can find cdf and derive it!

$$P(X_{(1)} \le x) = 1 - P(X_{(1)} > x) = 1 - P(\text{all } X_i\text{'s} > x) = 1 - \prod_{i=1}^{n} P(X_i > x)$$

$$F(x_{(1)}) = 1 - \prod_{i=1}^{n}\Big(1 - \underbrace{\frac{x-a}{b-a}}_{\text{cdf of } U(a,b)}\Big) = 1 - \left(1 - \frac{x-a}{b-a}\right)^n$$

$$f(x_{(1)}) = \frac{\partial F(x_{(1)})}{\partial x} = -n\left(1 - \frac{x-a}{b-a}\right)^{n-1} \cdot -\frac{1}{(b-a)}$$

chain rule ∴ $(f \circ g)' = (f' \circ g)\, g'$ — composition

$$= \frac{n}{b-a}\left(\frac{b-\cancel{a}-x+\cancel{a}}{b-a}\right)^{n-1} = \frac{n}{(b-a)^n}(b-x)^{n-1} \quad //$$

Now, the expectations:

- $E[X_{(1)}] = \int_a^b x\left[\frac{n}{(b-a)^n}(b-x)^{n-1}\right]dx$

by parts: $\int f(x)\,g'(x)\,dx \Rightarrow \int_a^b \underbrace{u}_{\text{derive}}\,\overbrace{dv}^{\text{integrate}} = uv\Big|_a^b - \int_a^b v\,du$

$$u = x \Rightarrow \frac{du}{dx} = 1 \Rightarrow du = dx$$

$$dv = (b-x)^{n-1}dx \Rightarrow v = \int (b-x)^{n-1}dx = -\frac{(b-x)^n}{n} \quad \text{(indefinite integral)}$$

Rewriting...

$$E[X_{(1)}] = \frac{n}{(b-a)^n}\left[\overset{u}{x}\overbrace{\left(-\frac{(b-x)^n}{n}\right)}^{v}\Big|_a^b - \int_a^b \overbrace{-\frac{(b-x)^n}{n}}^{v}\,\overset{du}{dx}\right] = \frac{n}{(b-a)^n}\left[-\frac{b\cancel{(b-b)^n}}{n} + \frac{a(b-a)^n}{n} - \left[\frac{(b-x)^{n+1}}{n(n+1)}\right]\Big|_a^b\right]$$

with $\cancel{(b-b)^n} \to 0$

$$= \frac{n}{(b-a)^n}\left[\frac{a(b-a)^n}{n} - \frac{(b-b)^n}{n(n+1)} + \frac{(b-a)^{n+1}}{n(n+1)}\right] = a + \frac{b-a}{n+1}$$

with $(b-b)^n \to 0$

($\neq a \Rightarrow X_{(1)}$ is biased MLE of a.)

Same thing for $E[X_{(n)}]$

pdf of $X_{(n)}$ (easier !)

$$P(X_{(n)} \leq x) = P(\text{all } x_i\text{'s} \leq x) = \prod_{i=1}^{n} P(X_i \leq x) = \prod_{i=1}^{n} \left(\frac{x-a}{b-a}\right) = \left(\frac{x-a}{b-a}\right)^n$$

$$f(x_{(n)}) = \frac{\partial F(x_{(n)})}{\partial x} = n\left(\frac{x-a}{b-a}\right)^{n-1} \cdot \frac{1}{(b-a)} = \frac{n}{(b-a)^n}(x-a)^{n-1}$$

$$E[X_{(n)}] = \int_a^b x\left[\frac{n}{(b-a)^n}(x-a)^{n-1}\right]dx$$

by parts (again)

$$\int u\,dv = uv\Big|_a^b - \int_a^b v\,du$$

$$\left[\begin{array}{l} u = x \Rightarrow \frac{du}{dx} = 1 \Rightarrow du = dx \\ dv = (x-a)^{n-1} \Rightarrow v = \int (x-a)^{n-1}dx = \frac{(x-a)^n}{n} \end{array}\right]$$

Rewriting ...

$$E[X_{(n)}] = \frac{n}{(b-a)^n}\left[x\frac{(x-a)^n}{n}\Big|_a^b - \int_a^b \frac{(x-a)^n}{n}dx\right] = \frac{n}{(b-a)^n}\left[\frac{b(b-a)^n}{n} - \frac{a\cancel{(a-a)^n}^{\to 0}}{n} - \frac{(x-a)^{n+1}}{n(n+1)}\Big|_a^b\right]$$

$$= \frac{n}{(b-a)^n}\left[\frac{b(b-a)^n}{n} - \frac{(b-a)^{n+1}}{n(n+1)} + \frac{\cancel{(a-a)^{n+1}}^{\to 0}}{n(n+1)}\right] = b - \frac{b-a}{n+1}$$

$\left(\neq b \Rightarrow X_{(n)} \text{ is biased est. of } b\right)$

This is a result for $U(a,b)$, so by replacing a and b in the above results, you can find Expectations for different Uniforms!

Let's try something a bit harder...

Compute $var(X_{(1)})$ and $var(X_{(n)})$.

Let's use the following formula:

$$var(X) = E[X^2] - (E[X])^2$$

(Proof: $var(X) = E(X-\mu)^2 = E(X^2 - 2X\mu + \mu^2) = E[X^2] - 2\mu E[X] + \mu^2 =$
$= E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \mu^2 = E[X^2] - (E[X])^2$ //)

Scanned by CamScanner

Let's start with $var(X_{(1)}) = E[X_{(1)}^2] - (E[X_{(1)}])^2$.

we already have $E[X_{(1)}]$, so $(E[X_{(1)}])^2 = \left(a + \frac{b-a}{n+1}\right)^2$

$$E[X_{(1)}^2] = \int_a^b x^2 \left[\frac{n}{(b-a)^n}(b-x)^{n-1}\right] dx \quad (1)$$ { by parts again

$$\left[\begin{array}{l} u = x^2 \Rightarrow \frac{du}{dx} = 2x \Rightarrow du = 2x\,dx \\ dv = (b-x)^{n-1}dx \Rightarrow v = -\frac{(b-x)^n}{n} \end{array}\right]$$

(we'll need to do $\int$ by parts twice)

Rewriting (1) ..

$$\frac{n}{(b-a)^n}\int_a^b x^2(b-x)^{n-1}dx = \frac{n}{(b-a)^n}\left[-\frac{x^2(b-x)^n}{n}\Bigg|_a^b + \int_a^b \frac{(b-x)^n}{n}2x\,dx\right]$$

$$= \frac{n}{(b-a)^n}\left[-\frac{b^2\cancel{(b-b)^n}^{\,0}}{n} + \frac{a^2(b-a)^n}{n} + \frac{2}{n}\int_a^b \underbrace{x(b-x)^n dx}_{\text{by parts}}\right]$$

$\left\{\begin{array}{l} u = x \Rightarrow du = d \\ dv = (b-x)^n \\ V = -\frac{(b-x)^{n+1}}{n+1}\end{array}\right.$

$$= \frac{n}{(b-a)^n}\left[\frac{a^2(b-a)^n}{n} + \frac{2}{n}\left(-\frac{x(b-x)^{n+1}}{n+1}\Bigg|_a^b + \int_a^b \frac{(b-x)^{n+1}}{n+1}dx\right)\right]$$

$$= \frac{n}{(b-a)^n}\left[\frac{a^2(b-a)^n}{n} + \frac{2}{n}\left(\frac{-b\cancel{(b-b)^{n+1}}^{\,0}}{n+1} + \frac{a(b-a)^{n+1}}{n+1} - \frac{(b-x)^{n+2}}{(n+1)(n+2)}\Bigg|_a^b\right)\right]$$

$$= \frac{n}{(b-a)^n}\left[\frac{a^2(b-a)^n}{n} + \frac{2}{n}\left(\frac{a(b-a)^{n+1}}{n+1} - \frac{\cancel{(b-b)^{n+2}}^{\,0}}{(n+1)(n+2)} + \frac{(b-a)^{n+2}}{(n+1)(n+2)}\right)\right]$$

$$= \frac{n}{(b-a)^n}\left[\frac{a^2(b-a)^n}{n} + \frac{2a(b-a)^{n+1}}{n(n+1)} + \frac{2(b-a)^{n+2}}{n(n+1)(n+2)}\right]$$

$$= a^2 + \frac{2a(b-a)}{n+1} + \frac{2(b-a)^2}{(n+1)(n+2)}$$ // (you can simplify if you want)

So, $var(X_{(1)}) = E[X_{(1)}^2] - (E[X_{(1)}])^2$

$$= a^2 + \frac{2a(b-a)}{n+1} + \frac{2(b-a)^2}{(n+1)(n+2)} - \left(a + \frac{b-a}{n+1}\right)^2$$

Now, let's compute $var(X_{(n)}) = E[X_{(n)}^2] - (E[X_{(n)}])^2$

We already have $E[X_{(n)}] = b - \frac{b-a}{n+1}$, so $(E[X_{(n)}])^2 = \left(b - \frac{b-a}{n+1}\right)^2$

$$E[X_{(n)}^2] = \int_a^b x^2 \left[\frac{n}{(b-a)^n}(x-a)^{n-1}\right] dx \quad (1) \Big\} \text{ by parts}$$

$$\left[\begin{array}{l} u = x^2 \Rightarrow \frac{du}{dx} = 2x \Rightarrow du = 2x\,dx \\ dv = (x-a)^{n-1} \Rightarrow v = \frac{(x-a)^n}{n} \end{array}\right]$$

Rewriting (1) ...

$$E[X_{(n)}^2] = \frac{n}{(b-a)^n}\left[\frac{x^2(x-a)^n}{n}\Big|_a^b - \frac{2}{n}\underbrace{\int_a^b x(x-a)^n dx}_{\text{by parts}}\right] \quad \left\{\begin{array}{l} u = x \Rightarrow \frac{du}{dx} = 1 \Rightarrow du = dx \\ dv = (x-a)^n \Rightarrow v = \frac{(x-a)^{n+1}}{n+1} \end{array}\right.$$

$$= \frac{n}{(b-a)^n}\left[\frac{b^2(b-a)^n}{n} - \cancelto{0}{\frac{a^2(a-a)^n}{n}} - \frac{2}{n}\left(\frac{x(x-a)^{n+1}}{n+1}\Big|_a^b - \int_a^b \frac{(x-a)^{n+1}}{n+1}dx\right)\right]$$

$$= \frac{n}{(b-a)^n}\left[\frac{b^2(b-a)^n}{n} - \frac{2}{n}\left(\frac{b(b-a)^{n+1}}{n+1} - \cancelto{0}{\frac{a(a-a)^{n+1}}{n+1}} - \frac{(x-a)^{n+2}}{(n+1)(n+2)}\Big|_a^b\right)\right]$$

$$= \frac{n}{(b-a)^n}\left[\frac{b^2(b-a)^n}{n} - \frac{2}{n}\left(\frac{b(b-a)^{n+1}}{n+1} - \frac{(b-a)^{n+2}}{(n+1)(n+2)} + \cancelto{0}{\frac{(a-a)^{n+2}}{(n+1)(n+2)}}\right)\right]$$

$$= \frac{n}{(b-a)^n}\left[\frac{b^2(b-a)^n}{n} - \frac{2b(b-a)^{n+1}}{n(n+1)} + \frac{2(b-a)^{n+2}}{n(n+1)(n+2)}\right]$$

$$= b^2 - \frac{2b(b-a)}{n+1} + \frac{2(b-a)^2}{(n+1)(n+2)}$$

So, $var(X_{(n)}) = E[X_{(n)}^2] - (E[X_{(n)}])^2 = b^2 - \frac{2b(b-a)}{n+1} + \frac{2(b-a)^2}{(n+1)(n+2)} - \left(b - \frac{b-a}{n+1}\right)^2$

Linear Algebra Review

Some important concepts

① Vector space or linear space (over the reals) is composed of 3 objects, a set V and 2 operations ('+','x'), and V must be a nonempty set. The objects in V can be anything, although we call them vectors. Vector spaces must satisfy:

addition '+'
1. $x + y = y + x$, $\forall x, y \in V$ (+ is commutative)
2. $(x+y)+z = x+(y+z)$ $\forall x, y, z \in V$ (+ is associative)
3. $0 + x = x$, $\forall x \in V$ (0 is additive identity)
4. $\forall x \in V$ $\exists (-x) \in V$ s.t. $x + (-x) = 0$ (existence of additive inverse)

scalar multiplication 'x'
5. $(\alpha\beta)x = \alpha(\beta x)$, $\forall \alpha, \beta \in \mathbb{R}$, $\forall x \in V$ (scalar mult is associative)
6. $\alpha(x+y) = \alpha x + \alpha y$, $\forall \alpha \in \mathbb{R}$, $\forall x, y \in V$ (right distributive rule)
7. $(\alpha+\beta)x = \alpha x + \beta x$, $\forall \alpha, \beta \in \mathbb{R}$, $\forall x \in V$ (left distributive rule)
8. $1x = x$, $\forall x \in V$

Examples:
1. The set $\mathbb{R}$ of real numbers is a vector space over $\mathbb{R}$.
2. The set $\mathbb{R}^2$ of all ordered pairs of real numbers is a vector space over $\mathbb{R}$.
3. The set $C(\mathbb{R})$ of all continuous functions defined on the real number line
4. The set $C([a,b])$ of all continuous functions defined on $[a,b]$ is a vector space over $\mathbb{R}$.
5. The set P of all polynomials with real coefficients ——— " ———
6. The set P_n of all polynomials of degree $\leq n$ with $\mathbb{R}$ coeff. ——— " ———
7. The set $M_{m,n}$ of all $m \times n$ matrices, with real entries ——— " ———
8. $V = \{0\}$ (where $0 \in \mathbb{R}^n$)

• Question 1: Let V be the set of all fifth-degree polynomials with standard operations. Is it a vector space? Answer: No! Because it's not closed under addition

Proof by counterexample
$f = x^5 + x - 1$ and $g = -x^5 \in V \Rightarrow f + g = x - 1 \notin V$

• Question 2: Let $V = \{(x,y) . x \geq 0, y \geq 0\}$ with standard operations. Is it a vector space?
Answer: No! Not every element in V has an additive inverse (e.g. $-(1,1) = (-1,-1) \notin V$

• Question 3: Is $V = \{(x, \frac{1}{2}x) : x \in \mathbb{R}\}$ a vector space? Yes! Check properties..

Proof . For any $x, y \in \mathbb{R}$

Addition
(1,2) $(x, \frac{1}{2}x) + (y, \frac{1}{2}y) = (x+y, \frac{1}{2}(x+y))$
(3) $0 = (0,0)$ (contains zero element)
(4) $-(x, \frac{1}{2}x) = (-x, \frac{1}{2}(-x))$ (every element has additive inverse)

. For 2 scalars α and β and $x, y \in \mathbb{R}$.
(5) $\beta\alpha(x, \frac{1}{2}x) = \beta(\alpha x, \alpha\frac{1}{2}x)$
(6) $\alpha(x+y, \frac{1}{2}(x+y)) = (\alpha x + \alpha y, \frac{1}{2}\alpha x + \frac{1}{2}\alpha y)$
(7) $(\alpha+\beta)(x, \frac{1}{2}x) = \alpha(x, \frac{1}{2}x) + \beta(x, \frac{1}{2}x)$
(8) $1(x, \frac{1}{2}x) = (x, \frac{1}{2}x)$

② Subspace: <u>subset of vectors in $\mathbb{R}^n$</u> (could also be all $\mathbb{R}^n$) satisfying 3 conditions.
subset of vector space

1. contains $\vec{0}$
2. closed under addition (for $\vec{v}_1, \vec{v}_2 \in S \to \vec{v}_1 + \vec{v}_2 \in S$)
3. closed under scalar multiplication ($\alpha \in \mathbb{R}, \vec{v}_i \in S \to c\vec{v}_i \in S$)

Closure means that even when you perform '+' or 'x', you will still get a result that belongs to your subspace.

Examples:

① $V = \{\vec{0}\} = \left\{\begin{bmatrix}0\\0\\0\end{bmatrix}\right\}$

Is it a subspace of $\mathbb{R}^3$?

Check.

① contains $\vec{0}$ (yes)

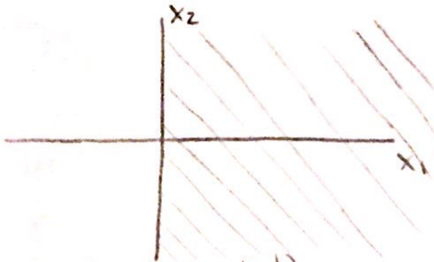
② closed under addition

$\begin{bmatrix}0\\0\\0\end{bmatrix} + \begin{bmatrix}0\\0\\0\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$ if I add any member of the set, I end up in the set ∴ closed under addition

③ closed under multiplication

$\alpha\begin{bmatrix}0\\0\\0\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$ (ended up in the zero-vector set) ∴ closed under multiplication

② $S = \left\{\begin{bmatrix}x_1\\x_2\end{bmatrix} \in \mathbb{R}^2 \mid x_1 \geq 0\right\}$ ∴ Let's see on the Cartesian Plane



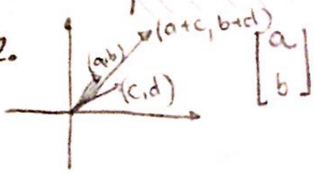
→ x_1 is constrained to ≥ 0

→ x_2 can be chosen arbitrarily

Is S a subspace of $\mathbb{R}^2$?

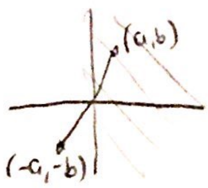
Check:

1. Contains $\vec{0} = \begin{bmatrix}0\\0\end{bmatrix}$? Yes

2. 

$\begin{bmatrix}a\\b\end{bmatrix} + \begin{bmatrix}c\\d\end{bmatrix} = \begin{bmatrix}a+c\\b+d\end{bmatrix}$, since $a \geq 0$ and $c \geq 0 \to (a+c) \geq 0$ always falls in the right quadrants.

3. $-1\begin{bmatrix}a\\b\end{bmatrix} = \begin{bmatrix}-a\\-b\end{bmatrix}$



⇒ falls out of the subspace.
Not closed under scalar multiplication.

③ Linear Combination and independence of vectors

• Linear Combination: given $\underline{n}$ vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ and $\underline{n}$ scalars $\alpha_1, \alpha_2, \dots, \alpha_n$, their L.C. is the vector $\vec{v}$ s.t.

$$\vec{v} = \alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2 + \dots + \alpha_n \vec{v}_n$$

Example. $\vec{v}_1 = \begin{bmatrix}1\\2\end{bmatrix}$ $\vec{v}_2 = \begin{bmatrix}0\\5\end{bmatrix}$ Can you express $\vec{v}_3 = \begin{bmatrix}3\\11\end{bmatrix}$ as a L.C of $\vec{v}_1$ and $\vec{v}_2$?

Answer: YES! $\vec{v}_3 = a\begin{bmatrix}1\\2\end{bmatrix} + b\begin{bmatrix}0\\5\end{bmatrix} = \begin{bmatrix}a\\2a+5b\end{bmatrix}$ $\quad a = 3$, $2a+5b = 11 \Rightarrow b = 1$

Therefore $\vec{v}_3 = \begin{bmatrix}3\\11\end{bmatrix}$ is a L.C of $\vec{v}_1$ and $\vec{v}_2$.

• Linear dependence and independence: a set of elements (vectors) $S = \{\vec{v}_1, \dots, \vec{v}_n\}$ is said to be linearly independent if the equation

$\alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2 + \dots + \alpha_n \vec{v}_n = 0$ has only one trivial solution $\alpha_1 = 0, \alpha_2 = 0, \dots, \alpha_n = 0$
(i.e. coefficients uniquely determined)

$\Rightarrow$ no vector can be expressed as a linear combination of the others.

S is linearly dependent if it's not linearly independent (at least one nontrivial, i.e. nonzero, solution to the above equation).

④ Span: set of all vectors that can be generated by taking L.C. of the given vectors. Let V be a vector space over $\mathbb{R}$ and $S = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ be a subset of V. We say that S is a spanning set of V if every vector $\vec{v}$ of V can be written as a L.C. of vectors in S. (S spans V).
Then the span of S is the set of all L.C. of vectors in S

$$\text{span}(S) = \{\alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2 + \dots + \alpha_n \vec{v}_n : \alpha_1, \alpha_2, \dots, \alpha_n \text{ are scalars}\}$$

S spans V $\Longleftrightarrow$ V is spanned by S.

Also, span is the smallest subspace of V that contains S.

Example $S = \left\{\begin{bmatrix}1\\-1\\2\end{bmatrix}, \begin{bmatrix}2\\1\\3\end{bmatrix}, \begin{bmatrix}-1\\0\\2\end{bmatrix}\right\}$ Is $\text{span}(S) = \mathbb{R}^3$?

Let's take L.C of the 3 vectors $c_1\begin{bmatrix}1\\-1\\2\end{bmatrix} + c_2\begin{bmatrix}2\\1\\3\end{bmatrix} + c_3\begin{bmatrix}-1\\0\\2\end{bmatrix} = \begin{bmatrix}a\\b\\c\end{bmatrix}$

$$\begin{cases}1c_1 + 2c_2 - 1c_3 = a\\ -1c_1 + 1c_2 + 0c_3 = b\\ 2c_1 + 3c_2 + 2c_3 = c\end{cases} \xrightarrow{+} \begin{cases}1c_1 + 2c_2 - 1c_3 = a\\ 3c_2 - 1c_3 = a+b\\ 2c_1 + 3c_2 + 2c_3 = c\end{cases} \xrightarrow{\times(-2)\,+} \begin{cases}1c_1 + 2c_2 - 1c_3 = a\\ 3c_2 - 1c_3 = a+b\\ -1c_2 + 4c_3 = a+c\end{cases} \xrightarrow{\times(3)\,+} \begin{cases}1c_1 + 2c_2 - 1c_3 = a\\ 3c_2 - 1c_3 = a+b\\ 11c_3 = -5a+b+3c\end{cases}$$

$$\begin{cases}c_3 = \frac{1}{11}(3c - 5a + b)\\ c_2 = \frac{1}{3}(b + a + c_3)\\ c_1 = a - 2c_2 + c_3\end{cases}$$

$\therefore$ for any vector $\begin{bmatrix}a\\b\\c\end{bmatrix}$ in $\mathbb{R}^3$, I can always find some L.C. of vectors in S that add up to $\begin{bmatrix}a\\b\\c\end{bmatrix}$.

So, $\text{span}(S) = \mathbb{R}^3$

Are they L.I.? Yes $c_1\begin{bmatrix}1\\-1\\2\end{bmatrix} + c_2\begin{bmatrix}2\\1\\3\end{bmatrix} + c_3\begin{bmatrix}-1\\0\\2\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$
Only solution: $c_1 = c_2 = c_3 = 0$

Another example of span:

$S = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \end{bmatrix} \right\}$ $\begin{cases} 1c_1 + 0c_2 = a \\ 2c_1 + 5c_2 = b \end{cases}$ $\Rightarrow c_1 = a$, $\Rightarrow c_2 = \dfrac{b-2a}{5}$

$\text{span}(S) = \mathbb{R}^2$ and the vectors in S are L.I.

(If you have exactly 2 vectors and they span $\mathbb{R}^2$ → they must be L.I.)
3 " $\mathbb{R}^3$ → "
⋮ ⋮ ⋮

⑤ Basis of a subspace: set of vectors $\{\vec{v}_1, \dots, \vec{v}_n\}$ is a basis for the vector space V if

① $V = \text{span}(\vec{v}_1, \dots, \vec{v}_n)$
② $\{\vec{v}_1, \dots, \vec{v}_n\}$ are L.I.

In fact, any L.C. of the orthogonal standard basis vectors $e_1 = (1,0,0,\dots,0)$, $e_2 = (0,1,0,0,\dots,0)$, …, $e_n = (0,0,0,\dots,0,1)$ spans $\mathbb{R}^n$ and, therefore, forms a basis for $\mathbb{R}^n$ (L.I.).

Fact: for a given vector space V, the number of vectors in any basis is the same → basis is the "minimum" set of vectors that spans the subspace $= \dim(V)$.

We assign $\dim\{0\} = 0$ and $\dim V = \infty$ if there is no basis.

⑥ Column space $(C(A))$ is the set of all L.C.'s of $\vec{v}_1, \dots, \vec{v}_n \Rightarrow C(A) = \text{span}\{\vec{v}_1, \dots, \vec{v}_n\}$

⑦ Nullspace $N(A) = \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \vec{0}\}$: set of vectors $\vec{x}$ orthogonal to all rows of A. If $N(A) = \vec{0}$ we have L.I. vectors.

Equivalently, Nullspace $N(A)$ is the set of vectors mapped to 0 by $y = Ax$

To find Nullspace: reduce to rref (reduced row echelon form)

$$A_{3\times4} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & 4 & 0 \\ 4 & 3 & 2 & 1 & 0 \end{array}\right] \xrightarrow{\times 4} \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 3 & 0 \end{array}\right] (-1)$$

$x \in \mathbb{R}^4$

$A\vec{x} = \vec{0}$ ↳ if I take the set of all x's that satisfy the above equation, I get the null space (subspace)

$$\begin{cases} x_1 + x_2 + x_3 + x_4 = 0 \\ x_1 + 2x_2 + 3x_3 + 4x_4 = 0 \\ 4x_1 + 3x_2 + 2x_3 + x_4 = 0 \end{cases}$$

$$\text{rref}(A) \Rightarrow \left[\begin{array}{cccc|c} 1 & 0 & -1 & -2 & 0 \\ 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \Rightarrow \begin{cases} x_1 - x_3 - 2x_4 = 0 \\ x_2 + 2x_3 + 3x_4 = 0 \end{cases}$$

$$\begin{cases} x_1 = x_3 + 2x_4 \\ x_2 = -2x_3 - 3x_4 \end{cases}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \end{bmatrix}$$

i.e. the solution set $(N(A))$ of the equation $A\vec{x} = 0$ can be represented as a L.C. of the vectors

$N(A) = \text{span}\left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \end{bmatrix} \right\}$

⑧ One-to-one and onto matrices

- One-to-one : $f(x) = f(y) \overset{\text{if and only if}}{\Longleftrightarrow} x = y$
- onto(Y): if and only if for every $y \in Y$ there is, at least, one $x \in X$ s.t. $f(x) = y$

The above definitions apply to functions, linear transformations ($T: \mathbb{R}^n \to \mathbb{R}^n$) and to matrices by extension (since a $m \times n$ matrix A can be identified with the linear transformation $L_A: \mathbb{R}^n \to \mathbb{R}^m$ given by $L_A(x) = Ax$)

More on onto: A is called onto if $R(A) = \mathbb{R}^m$
↳ Range $R(A) = \{Ax : x \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$

$\Leftrightarrow$
- $Ax = y$ can be solved in x for any y
- columns of A span $\mathbb{R}^m$
- A has right inverse (ie. $\exists B \in \mathbb{R}^{n \times m}$ s.t. $AB = I$)
- rows of A are independent
- $N(A^T) = \{0\}$
- $\det(AA^T) \neq 0$

⑨ Inverse of a matrix: $A \in \mathbb{R}^{n \times n}$ → important — is invertible or nonsingular if one of the equivalent conditions is true

- $\det A \neq 0$
- $\det A^TA = \det AA^T \neq 0$
- A is full rank
- Columns of A are a basis for $\mathbb{R}^n$
- rows of A are a basis for $\mathbb{R}^n$
- $y = Ax$ has a unique solution x, $\forall y \in \mathbb{R}^n$
- $AA^{-1} = A^{-1}A = I$ (right and left inverse)
- $N(A) = \{0\}$ (L.I. vectors)

→ Rank of a matrix $A_{m \times n} = \dim C(A)$ is the maximum number of independent columns/rows of the matrix. Properties:

- $\text{rank}(A) \leq \min(n, m)$
- $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$
- $\text{rank}(AB) \geq \text{rank}\, A_{m \times n} + \text{rank}\, B_{n \times p} - n$
- $\text{rank}(A) = \text{rank}(A^T)$
- $\text{rank}(A) + \dim N(A) = n$
- ↳ A is full rank when $\text{rank}(A) = \min(m, n)$

$$\text{rank}(A^TA) \begin{cases} p, & n \geq p \rightsquigarrow \text{may be invertible} \\ n, & n < p \rightsquigarrow \text{not invertible} \end{cases}$$

Full rank matrices:

- for square matrices: full rank means nonsingular
- for skinny matrices ($m \geq n$): full rank means indep columns
- for fat matrices ($m \leq n$): full rank means indep rows

Matrix Calculus

Some General tips/properties to get started:

① $AB \neq BA$

② Factoring: keep factors on the same side (right or left)

$AX + BX = (A+B)X$ or $XA + XB = X(A+B)$

$AX + XB \rightarrow$ does not factor out

③ No matrix division: you must multiply by the inverse of the matrix

$(X^TX)\beta = X^Ty \Rightarrow \beta = (X^TX)^{-1}X^Ty$

④ Dimensions must match to allow for matrix multiplication

$A_{m\times n} \times B_{n\times p} = C_{m\times p}$

⑤ It's possible to switch inverse with transpose and vice-versa:

$[(X^TX)^{-1}]^T = [(X^TY)^T]^{-1} = [(X^TX)]^{-1}$ $(A^TB^T) = (BA)^T$

or $(A^{-1})^T = (A^T)^{-1}$ since $\begin{cases} A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I \\ (A^{-1})^TA^T = (AA^{-1})^T = I^T = I \end{cases}$

⑥ $(ABC)^T = C^TB^TA^T$

⑦ Multiplication always same side (right or left):

$A = B \Rightarrow AC = BC$ (right) or $CA = CB$ (left)

⑧ $X^{-1}X = I$ or $XX^{-1} = I$

⑨ Associativity of Multiplication $A(BC) = (AB)C$

⑩ scalar associativity: $\alpha(AB) = (\alpha A)B = A(\alpha B) = (AB)\alpha$

* Inner Product

$\langle x,y\rangle = x_1y_1 + x_2y_2 + \dots + x_ny_n = x^Ty = \sum_{i=1}^{n} x_iy_i$

Properties

- $\langle \alpha x, y\rangle = \alpha\langle x,y\rangle$
- $\langle x+y, z\rangle = \langle x,z\rangle + \langle y,z\rangle$
- $\langle x,y\rangle = \langle y,x\rangle$
- $\langle x,x\rangle = \|\vec{x}\|_2^2$ (or $\|\vec{x}\|^2$) $= \vec{x}\cdot\vec{x}$
  $= x_1^2 + x_2^2 + x_3^2 + \dots + x_n^2$
- $\langle x,x\rangle = 0 \overset{\text{"iff"}}{\Longleftrightarrow} x = 0$

ex: $x = \begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix}$ $y = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$ $\langle x,y\rangle = 4.1 + 1.2 + 0.4 = 6$

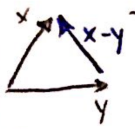
$\Rightarrow$ a scalar

* Norms

- Euclidean norm (ℓ^2 norm): for $x \in \mathbb{R}^n$, the

$\|x\| = \|x\|_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} = \sqrt{x^Tx} = \sqrt{\langle x,x\rangle}$

measures length of a vector (from origin)



$\rightarrow$ norm defines distance between vectors

$\underbrace{dist(x,y)}_{\text{length, no direction}} = \|x-y\|_2 = \sqrt{(x-y)(x-y)}$

$= \sqrt{\|x\|_2^2 + \|y\|_2^2 - 2xy}$

Properties:

- $\|\alpha x\| = |\alpha|\,\|x\|$ (homogeneity)
- $\|x+y\| \leq \|x\| + \|y\|$ (triangle inequality)
- $\|x\| \geq 0$ (nonnegativity)
- $\|x\| = 0 \Longleftrightarrow x = 0$ (definiteness)

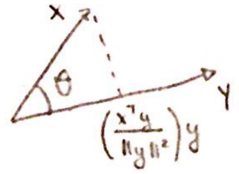
- $L1$-norm: $\|x\|_1 = \sum_{i=1}^{n} |x_i|$
- Lp-norm: $\|x\|_p = \left[\sum_{i=1}^{n} |x_i|^p\right]^{1/p}$
- ∞-norm: $\|x\|_\infty = \max_i |x_i|$

- **Cauchy-Schwarz inequality and angle between vectors**

For any $x, y \in \mathbb{R}^n$ $\quad |x^T y| \leq \|x\|\|y\|$ $\quad$ ($|\langle x, y\rangle|$) $\qquad$ Recall: $\|x\| = \sqrt{x^T x} = \sqrt{\langle x, x\rangle}$

the following special cases also hold:

- $|\langle x,y\rangle|^2 \leq \langle x,x\rangle \langle y,y\rangle$
- $(|x_1| + \cdots + |x_n|)^2 \leq n \sum_{i=1}^{n} x_i$
- (unsigned) angle between vectors in $\mathbb{R}^n$



$$\theta = \angle(x,y) = \cos^{-1} \frac{x^T y}{\|x\|\|y\|}$$

$$\Rightarrow x^T y = \|x\|\|y\| \cos\theta$$

- $x^T y > 0$: $\angle(x,y)$ is acute
- $x^T y < 0$: $\angle(x,y)$ is obtuse

- **Derivatives**:

$x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}_{p\times 1}$

- $\dfrac{\partial(\,)}{\partial x} = \begin{pmatrix} \frac{\partial(\cdot)}{\partial x_1} \\ \vdots \\ \frac{\partial(\cdot)}{\partial x_p} \end{pmatrix}$

- $\dfrac{\overbrace{\partial(a^T x)}^{\text{scalar}}}{\partial x_{p\times1}} = \begin{pmatrix} \frac{\partial(a^T x)}{\partial x_1} \\ \vdots \\ \frac{\partial(a^T x)}{\partial x_p} \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix}_{p\times1} = \vec{a}_{p\times1}$ $\quad$ where $a^T_{1\times p}$

- $\dfrac{\overbrace{\partial(x^T a)}^{\text{scalar}}}{\partial x_{p\times1}} = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} = \vec{a}_{p\times1}$

$\Rightarrow$ vectors' dimensions must match! $\dfrac{\partial(\text{scalar})}{\partial(\text{vector})}$

- $U^T = (u_1, \ldots, u_K)_{1\times K}$

$$\frac{\partial U^T}{\partial x_{p\times1}} = \begin{bmatrix} \frac{\partial u_1}{\partial x_1} & \cdots & \frac{\partial u_K}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial u_1}{\partial x_p} & \cdots & \frac{\partial u_K}{\partial x_p} \end{bmatrix}_{p\times K}$$

- $$\frac{\partial x^T}{\partial x} = \begin{bmatrix} \frac{\partial x_1}{\partial x_1} & \cdots & \frac{\partial x_p}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial x_p} & \cdots & \frac{\partial x_p}{\partial x_p} \end{bmatrix}_{p\times p} = I$$

- $$\frac{\partial(x^T A x)}{\partial x} = \frac{\partial(x^T)}{\partial x} Ax + \frac{\partial(\overbrace{x^T A^T}^{(x^T A x)^T})}{\partial x} x = Ax + A^T x \quad \left(= 2Ax \text{ if } A = A^T \Rightarrow A \text{ is symmetric}\right)$$

- $$\frac{\partial(U^T V)}{\partial x} = \frac{\partial U^T}{\partial x} V + \frac{\partial \overbrace{V^T}^{(U^T V)^T}}{\partial x} U \quad \left(\frac{\partial(U^T U)}{\partial x} = 2 \frac{\partial U^T}{\partial x} U \right)$$

- **Gradient** of $f(x,y,z)$ (1st order derivative)

$$\nabla f(x,y,z) = \left[\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y} \quad \frac{\partial f}{\partial z} \right]$$

- **Hessian** of $f(x,y,z)$ (2nd order derivative)

$$\nabla^2 f(x,y,z) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} \end{bmatrix}$$

Ex: $f(x,y,z) = x^2 - y^2 + 4z^2 - 2xy + 4y^2$

$$\nabla_f(x,y,z) = \left[2x - 2y \quad -2y - 2x + 4z \quad 8z + 4y \right]$$

$$\nabla^2_f(x,y,z) = \begin{bmatrix} 2 & -2 & 0 \\ -2 & -2 & 4 \\ 0 & 4 & 8 \end{bmatrix}$$