Math Review

CSCI 567 @ USC

Spring 2025

1. Linear Algebra

Q1 Which identities are NOT correct for real-valued matrices A, B, and C? Assume that inverses exist and multiplications are legal.

(a) $(AB)^{-1} = B^{-1}A^{-1}$

(b) $(I + A)^{-1} \neq I - A$

(c) $\mathrm{tr}(AB) = \mathrm{tr}(BA)$

(d) $(AB)^\top \neq A^\top B^\top$

ⓐ $ABB^{-1}A^{-1} = I = B^{-1}A^{-1}AB$

ⓑ $(I-A)(I+A) = I - A^2 \overset{?}{=} I$ iff $A^2 = 0$

(tr(·) only defined on square mat.)

ⓒ $tr(AB) = \underbrace{\sum_i \sum_j A_{ij} B_{ji}}_{(AB)_{ii}} = \underbrace{\sum_j \sum_i B_{ji} A_{ij}}_{(BA)_{jj}} = tr(BA)$

ⓓ $(AB)^T_{ij} = (AB)_{ji} = \sum_k A_{jk} B_{ki}$

$= \sum_k B^T_{ik} A^T_{kj} = (B^T A^T)_{ij} \Rightarrow (AB)^T = B^T A^T$

$(A^T B^T)_{ij} = \sum_k A^T_{ik} B^T_{kj} = \sum_k A_{ki} B_{jk}$.

Def. $X \in \mathbb{R}^{n \times n}$ is PSD if $u^T X u \geq 0 \ \forall u \in \mathbb{R}^n$. X is PD if $u^T X u > 0 \ \forall u \in \mathbb{R}^n \setminus \{0\}$.

Q2 Which of the following statements are true? PSD stands for positive semi-definite.

(a) $XX^\top$ is a PSD matrix if and only if X is PSD. $\Longleftrightarrow$

(b) If X and Y are PSD matrices, then so is $\lambda X + \mu Y$ for any $\lambda, \mu \in \mathbb{R}$.

(c) If $X - Y$ and $X + Y$ are PSD matrices, then so are X and Y.

(d) All eigenvalues of a symmetric PSD matrix are non-negative.

(a) XX^T always PSD : $u^T X X^T u = \|X^T u\|_2^2 \geq 0 \quad \forall u \in \mathbb{R}^n$

- $\Rightarrow$ False: Counterexample

$X = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, $X^T X = I$

not PSD (-ve eigenvalue) PD & PSD

(b) $u^T(\lambda X + \mu Y)u = \lambda \underbrace{(u^T X u)}_{\geq 0} + \mu \underbrace{(u^T Y u)}_{\geq 0} < 0$ if $\lambda = -1, \mu = 0$.

(c) $X = I$ (PD), $Y = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ not PSD

$\Rightarrow X - Y = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$, $X + Y = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$ both PSD

(d) Suppose $u^T X u \geq 0 \ \forall u$.

Let v be eigenvector assoc. with λ.

$\underbrace{v^T X v}_{\geq 0} = v^T(\lambda v) = \lambda \underbrace{\|v\|_2^2}_{> 0} \geq 0$ as $v \neq 0$ (by def.)

$\Rightarrow \lambda \geq 0$

Usually, X PD/PSD $\Rightarrow$ X symmetrical.

Q3 Suppose A and B are two positive definite matrices. Which matrix may NOT be positive definite?

$\forall u \neq 0$: $A^{-1}u \neq 0$ (if $A^{-1}u = 0$, $u = A0 = 0 \ldots$)

(a) A^{-1}

(b) $A + B$

(c) $AA^{\top}$

(d) $A - B$

(a) PD $\Rightarrow (A^{-1}u)^T A (A^{-1}u) > 0 \Rightarrow u^T A^{-T} u > 0 \Rightarrow u^T A^{-1} u > 0$.

(transposing scalar or say $A^T = A$)

(b) $\underbrace{u^TAu}_{>0} + \underbrace{u^TBu}_{>0} > 0$.

(c) $u^TAA^Tu = \|A^Tu\|_2^2 > 0$ ($u^TAu > 0 \Rightarrow Au \neq 0$)

(d) Let $A = B$.

Q4 In a d-dimensional Euclidean space, what is the shortest distance from a point $\mathbf{x_0}$ to a hyperplane $\mathcal{H} = \{\mathbf{x} : \mathbf{w}^\top \mathbf{x} = 0\}$? (Notation: $\|\mathbf{w}\|_2 = \sqrt{\sum_i w_i^2}$.)

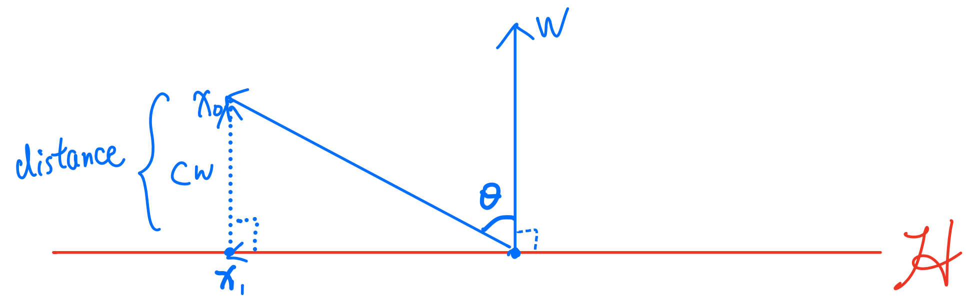
$w \neq 0$.

(a) $|\mathbf{w}^\top \mathbf{x_0}|$

(b) $|\mathbf{w}^\top \mathbf{x_0}| / \|\mathbf{w}\|_2$

(c) $|\mathbf{w}^\top \mathbf{x_0}| / \sqrt{\|\mathbf{w}\|_2^2 + \|\mathbf{x_0}\|_2^2}$

(d) $|\mathbf{w}^\top \mathbf{x_0}| / \|\mathbf{w}\|_2^2$



$$x_1 = x_0 - cw$$

$$\Rightarrow 0 = w^T x_1 = w^T(x_0 - cw)$$

$$= w^T x_0 - c\|w\|_2^2$$

$$\Rightarrow c = \frac{w^T x_0}{\|w\|_2^2} \Rightarrow \text{distance} = \|cw\|_2 = |c|\,\|w\|_2$$

$$= \frac{|w^T x_0|}{\|w\|_2}$$

Q5 Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_N$ are all D-dimensional vectors, and $\underbrace{X \in \mathbb{R}^{N \times D}}_{\text{Design matrix}}$ is a matrix where the n-th row is $\mathbf{x}_n^\top$. Then which of the following identities are correct?

$$X = \begin{bmatrix} - x_1^T - \\ \vdots \\ - x_N^T - \end{bmatrix} \qquad X^T = \begin{bmatrix} | & & | \\ x_1 & \cdots & x_N \\ | & & | \end{bmatrix}$$

(a) $X^\top X = \sum_{n=1}^N \mathbf{x}_n \mathbf{x}_n^\top$ — X^TX: Cov. matrix

(b) $X^\top X = \sum_{n=1}^N \mathbf{x}_n^\top \mathbf{x}_n$ — XX^T: Gram/Kernel matrix.

(c) $XX^\top = \sum_{n=1}^N \mathbf{x}_n \mathbf{x}_n^\top$

(d) $XX^\top = \sum_{n=1}^N \mathbf{x}_n^\top \mathbf{x}_n$

$$(X^TX)_{ij} = \sum_{n=1}^{N} x_{ni} x_{nj}$$

$$(x_n x_n^T)_{ij} = x_{ni} x_{nj}$$

$$(XX^T)_{ij} = x_i^T x_j$$

2. Probability and Statistics

Q1 A bag contains 2 red balls and 3 blue balls. First, Alice draws a ball from the bag randomly (and removes it from the bag). Then, Bob draws a ball randomly too. 1) What is the probability that $\underbrace{\text{Alice gets a red ball}}_{A}$ and $\underbrace{\text{Bob gets a blue ball}}_{B}$? 2) What is the probability that Alice gets a blue ball given that Bob gets a blue ball?

(a) $\frac{3}{10}$ and $\frac{1}{2}$

(b) $\frac{3}{10}$ and $\frac{2}{5}$

(c) $\frac{6}{25}$ and $\frac{1}{2}$

(d) $\frac{6}{25}$ and $\frac{2}{5}$

1) $P(A \cap B) = \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{10}$

2) $P(\bar{A} \mid B) = \frac{P(\bar{A}, B)}{P(B)} = \frac{1}{2}$

$P(B) = P(A \cap B) + P(\bar{A} \cap B)$

$P(\bar{A}, B) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10} = P(A \cap B)$

Q2 For events A, B and C, which of the following identities are correct?

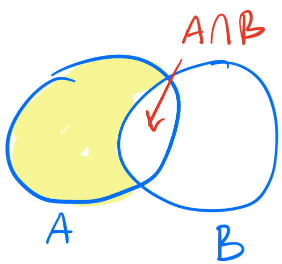
(a) $P(A) - P(A \cap B) = P(A \cup B) - P(B)$

(b) $P(A \cup B) \leq P(A) + P(B) - P(A)P(B)$

(c) $P(A) = P(A \cap C) + P(A \cap \overline{C})$, where $\overline{C}$ denotes the complement of event C.

(d) $P(A) = P(A|C) + P(A|\overline{C})$, where $\overline{C}$ denotes the complement of event C.

(a)

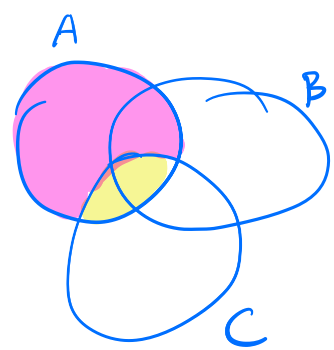


(b) Counterexample:

$\boxed{A|B}$ $P(A) = \frac{1}{2} = P(B)$ and $P(A \cap B) = 0$

$P(A \cup B) = 1 > P(A) + P(B) - P(A)P(B) = \frac{3}{4}$

(c) "Law of total prob."



(d) False. Let $A = C$. $P(A|C) = 1$, $P(A|\bar{C}) = 0$

$P(A) < 1$.

Q3 For events A, B, C and $Z_1, \ldots, Z_T$, which of the following identities are correct?

(a) $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ suppose $P(B) \neq 0$

(b) $\frac{P(A|B,C)}{P(A|C)} = \frac{P(B|A,C)}{P(B|C)}$

(c) $P(\bigcap_{t=1}^{T} Z_t) = \Pi_{t=1}^{T} P(Z_t)$

(d) $P(\bigcap_{t=1}^{T} Z_t) = \Pi_{t=1}^{T} P(Z_t|Z_1, \ldots, Z_{t-1})$

a Bayes' rule.

b $P(A|B,C) = \frac{P(A,B|C)}{P(B|C)}$, $P(B|A,C) = \frac{P(A,B|C)}{P(A|C)}$

c False unless $Z_{1:T}$ (mutually) indep.

d Chain rule (of prob.)

Q4 Which of the following statements on the density function of a Gaussian distribution are true?

(a) The density for a one-dimensional Gaussian distribution with mean μ and variance σ^2 is $f(x) \propto \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$.

(b) The density for a one-dimensional Gaussian distribution with mean μ and variance σ^2 is $f(x) \propto \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$.

(c) The density for a d-dimensional Gaussian distribution with mean $\mu \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$ is $f(\mathbf{x}) \propto \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^\top \Sigma (\mathbf{x}-\mu)\right)$.

(d) The density for a d-dimensional Gaussian distribution with mean $\mu \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$ is $f(\mathbf{x}) \propto \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^\top \Sigma^{-1} (\mathbf{x}-\mu)\right)$.

By def.

Q5 Which of the following statements are true? Def. $\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(\underset{[2\times 1]}{\mu}, \underset{[2\times 2]}{\Sigma})$

(a) Suppose X and Y are two jointly Gaussian random variables. Then $Z = X - 2Y$ is also Gaussian.

(b) Suppose X and Y are two jointly Gaussian random variables. Then the marginal distribution of X is also Gaussian.

(c) Suppose X and Y are two jointly Gaussian random variables. Then the conditional distribution of X given Y is also Gaussian.

(d) For a random vector $X \in \mathbb{R}^n$, its covariance matrix is $\mathbb{E}[XX^\top] - \mathbb{E}[X]\mathbb{E}[X]^\top$.

(a) Any (finite) linear combination of Gaussian rvs is Gaussian.

$$X - 2Y \sim \mathcal{N}\left(E[X-2Y], Var(X-2Y)\right)$$

(b) $X \sim \mathcal{N}(\mu_1, \Sigma_{11})$

(c) $X \mid Y=y \sim \mathcal{N}\left(E[X \mid Y=y], Var(X) - \frac{Cov(X,Y)^2}{Var(Y)}\right)$, $E[X \mid Y=y] = E[X] + \frac{Cov(X,Y)}{Var(Y)}(y - E[Y])$.

(d) $$Cov(X) = E\left[(X - E[X])(X - E[X])^T\right] = E\left[XX^T - X\,E[X]^T - \underbrace{E[X]X^T}_{E[\cdot]} + \underbrace{E[X]E[X]^T}_{=}\right]$$

$$= E[XX^T] - E[X]E[X]^T$$

3. Calculus

Q1 Suppose $\mathbf{a} \in \mathbb{R}^{n\times 1}$ is an arbitrary vector. Which one of the following functions is NOT convex:

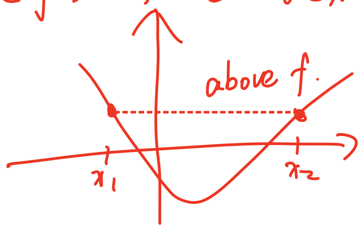
(a) $f(\mathbf{x}) = \sum_{i=1}^{n} |x_i|$

(b) $f(\mathbf{x}) = \sum_{i=1}^{n} a_i x_i$

(c) $f(\mathbf{x}) = \min_{i\in\{1,\ldots,n\}} a_i x_i$

(d) $f(\mathbf{x}) = \sum_{i=1}^{n} \exp(x_i)$

Def. f convex if $f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$ $\forall \lambda \in [0,1]$.



Fact. If f twice differentiable,

Hessian $\nabla^2 f(x)$ PSD $\forall x$ $\Leftrightarrow$ f convex

(a), (b) Check def.

(c) Not convex. Take $a = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ $\Rightarrow$ $f(x) = \min\{x_1, -x_2\}$.

Consider $x = (1,1)$, $y = (-1,-1)$. $f\left(\frac{1}{2}x + \frac{1}{2}y\right) = f(0) = 0$

$f(x) = -1 = f(y)$ but $\frac{1}{2}f(x) + \frac{1}{2}f(y) = -1 < f\left(\frac{1}{2}x + \frac{1}{2}y\right)$.

(d) $\frac{\partial f(x)}{\partial x_i} = e^{x_i}$, $\nabla f(x) = e^{x}$ elementwise.

$\nabla^2 f(x) = \mathrm{diag}(e^{x})$ is PSD $\Rightarrow$ f convex.

Q2 Which of the following are correct chain rules ($g, g_1, \ldots, g_d$ are functions from $\mathbb{R}$ to $\mathbb{R}$)?

(a) For a composite function $f(g(w))$, $\frac{\partial f}{\partial w} = \frac{\partial f}{\partial g}\frac{\partial g}{\partial w}$.

(b) For a composite function $f(g(w))$, $\frac{\partial f}{\partial w} = \frac{\partial f}{\partial g} + \frac{\partial g}{\partial w}$.

(c) For a composite function $f(g_1(w), \ldots, g_d(w))$, $\frac{\partial f}{\partial w} = \left(\frac{\partial f}{\partial g_1}\frac{\partial g_1}{\partial w}, \ldots, \frac{\partial f}{\partial g_d}\frac{\partial g_d}{\partial w}\right)$.

(d) For a composite function $f(g_1(w), \ldots, g_d(w))$, $\frac{\partial f}{\partial w} = \sum_{i=1}^{d} \frac{\partial f}{\partial g_i}\frac{\partial g_i}{\partial w}$.

def. of chain rule.

(c) False. $\frac{\partial f}{\partial w}(\bar{w}) \in \mathbb{R}^{d_w}$ i.e., <u>same shape as w, by def.</u>

Q3 A function $f : \mathbb{R}^{n\times 1} \to \mathbb{R}$ is defined as $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x}$ for some $\mathbf{b} \in \mathbb{R}^{n\times 1}$ and $\mathbf{A} \in \mathbb{R}^{n\times n}$. What is the derivative $\frac{\partial f}{\partial \mathbf{x}}$ (also called the gradient $\nabla f(\mathbf{x})$)?

(a) $(\mathbf{A} + \mathbf{A}^\top)\mathbf{x} + \mathbf{b}$

(b) $2\mathbf{A}^\top \mathbf{x} + \mathbf{b}$

(c) $2\mathbf{A}\mathbf{x} + \mathbf{b}$

(d) $2\mathbf{A}\mathbf{x} + \mathbf{x}$

$x \in \mathbb{R}^{d\times 1}$

$$\nabla_x f(x) \overset{\text{def}}{=} \left[\frac{\partial f(x)}{\partial x_1} \; \cdots \; \frac{\partial f(x)}{\partial x_d} \right]^T$$

Check elementwise.

$$f(x) = \sum_i \sum_j A_{ij} x_i x_j + \sum_i b_i x_i$$

$$\frac{\partial f}{\partial x_i}(x) = \sum_{j\neq i} \left(A_{ij} x_j + A_{ji} x_j \right) + 2A_{ii}x_i + b_i$$

$$\frac{\partial A_{ii} x_i^2}{\partial x_i} = 2A_{ii} x_i$$

$$= \sum_{j=1}^{d} (A_{ij} + A_{ji}) x_j + b_i$$

$$= \left[(A + A^T) x + b \right]_i$$

Q4 A function $f : \mathbb{R}^{n\times n} \to \mathbb{R}$ is defined as $f(\mathbf{A}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^{n\times 1}$. What is the derivative $\frac{\partial f}{\partial \mathbf{A}}$?

(a) $2\mathbf{x}$

(b) $\mathbf{x} + \mathbf{x}^\top$

(c) $\mathbf{x}\mathbf{x}^\top$

(d) $\mathbf{x}^\top\mathbf{x}$

Again, check elementwise.

$$\left[\frac{\partial f}{\partial A}\right]_{ij} \overset{\text{def}}{=} \frac{\partial f}{\partial A_{ij}}$$

$$f(A) = x^T A x = \sum_i \sum_j A_{ij} x_i x_j$$

$$\Rightarrow \frac{\partial f}{\partial A_{ij}} = x_i x_j = [x x^T]_{ij}$$

Q5 A function $f : \mathbb{R}^{n\times 1} \to \mathbb{R}$ is defined as $f(\mathbf{w}) = \ln(1 + e^{-\mathbf{w}^\top \mathbf{x}})$ for some $\mathbf{x} \in \mathbb{R}^{n\times 1}$. What is the derivative $\frac{\partial f}{\partial \mathbf{w}}$?

$$\frac{d \ln x}{dx} = \frac{1}{x}, \frac{de^{ax}}{dx} = ae^{ax}$$

(a) $-\frac{\mathbf{w}}{1+e^{\mathbf{w}^\top \mathbf{x}}}$

(b) $-\frac{\mathbf{x}}{1+e^{\mathbf{w}^\top \mathbf{x}}}$

(c) $-\frac{\mathbf{w}}{1+e^{-\mathbf{w}^\top \mathbf{x}}}$

(d) $-\frac{\mathbf{x}}{1+e^{-\mathbf{w}^\top \mathbf{x}}}$

Let $g(w) = 1 + e^{-w^T x} = 1 + \prod_{i=1}^{n} e^{-w_i x_i}$

$$\frac{\partial g}{\partial w_i} = -x_i e^{-w^T x} \Rightarrow \frac{\partial g}{\partial w} = -x \cdot e^{-w^T x}$$

$$\Rightarrow \frac{\partial f}{\partial g} \frac{\partial g}{\partial w} = \frac{-x \cdot e^{-w^T x}}{1 + e^{-w^T x}} = \frac{-x}{1 + e^{w^T x}}$$

Q6 For a differential function $f : \mathbb{R}^n \to \mathbb{R}$, which of the following statements are correct?

(a) If $\mathbf{x}^\star$ is a minimizer of f, then $\nabla f(\mathbf{x}^\star) = \mathbf{0}$.

(b) If $\mathbf{x}^\star$ is a maximizer of f, then $\nabla f(\mathbf{x}^\star) = \mathbf{0}$.

(c) If $\nabla f(\mathbf{x}^\star) = \mathbf{0}$, then $\mathbf{x}^\star$ is a minimizer of f.

(d) If $\nabla f(\mathbf{x}^\star) = \mathbf{0}$, then $\mathbf{x}^\star$ is a maximizer of f.

Directional derivative: $(w \in \mathbb{R}^n, h \in \mathbb{R})$

$$\vec{d}f(x;w) \overset{\text{def}}{=} \lim_{h \to 0^+} \frac{f(x+hw) - f(x)}{h}$$

Fact: If f continuous differentiable, (i.e., f' exists and cont.)

$$\nabla f(x^*) = 0 \iff \vec{d}f(x;w) \geq 0 \ \forall w \in \mathbb{R}^n.$$

a, b: first-order optimality condition.

Counterexample c, d

e.g., $f(x) = x^3$