Math Exercises

CSCI 567 Machine Learning

Spring 2025

Instructor: Haipeng Luo

Important notes: These exercises are all related to the **required prerequisites** of this course. A **complete** understanding of these exercises is a **"must"** if you want to succeed in this course. If you have no ideas what (some of) these exercises are talking about, then you do not have the required background and should drop the course. If you have learned these before but forget some of them already, ask yourself if you can pick them up quickly and, perhaps more importantly, whether you feel comfortable with these contents. If not, you should drop the course as well.

MULTIPLE-CHOICE QUESTIONS: one or more correct choices for each question.

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) $(I + A)^{-1} = I - A$

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

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

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.

(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.

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

(a) A^{-1}

(b) $A + B$

(c) $AA^\top$

(d) $A - B$

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: $\|\mathbf{w}\|_2 = \sqrt{\sum_i w_i^2}$.)

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

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

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

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

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

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

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 Alice gets a red ball and Bob gets a blue ball? 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 $\frac{2}{5}$

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.

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

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

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

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 $f(\mathbf{x}) \propto \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^\top \Sigma^{-1}(\mathbf{x}-\mu)\right)$.

Q5 Which of the following statements are true?

(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 $\mathbb{E}[XX^\top] - \mathbb{E}[X]\mathbb{E}[X]^\top$.

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

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

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

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

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 $\frac{\partial f}{\partial \mathbf{w}}$?

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

Q6 For a differentiable 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.