Linear Algebra and Calculus Exercises: Part I

CSCI 567 Machine Learning

Fall 2022

Instructor: Vatsal Sharan

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

Q3 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 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 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.