Assignment 2

Math for NN

Solve the following sheet

Questions

Linear Algebra (Questions 1–20)

Q1. Let

$$\mathbf{x} = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} 4 \\ 0 \\ -1 \end{pmatrix} \in \mathbb{R}^3.$$

Compute:

(a) The dot product $\mathbf{x}^\top \mathbf{y}$.

(b) The Euclidean norms $\|\mathbf{x}\|_2$ and $\|\mathbf{y}\|_2$.

(c) The cosine similarity $\cos\theta = \dfrac{\mathbf{x}^\top \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}$.

Q2. Consider the vectors in $\mathbb{R}^3$

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix}.$$

(a) Are $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ linearly independent?

(b) If they are linearly dependent, express one vector as a linear combination of the other two.

(c) What is the rank of the matrix $\mathbf{A} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3]$?

Q3. Let

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 1 & -1 \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

(a) Compute $\mathbf{y} = \mathbf{A}\mathbf{x}$.

(b) Write $\mathbf{y}$ explicitly as a linear combination of the columns of $\mathbf{A}$.

Q4. Let

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

(a) Compute the products $\mathbf{AB}$ and $\mathbf{BA}$.

(b) Are $\mathbf{A}$ and $\mathbf{B}$ commuting matrices (i.e. does $\mathbf{AB} = \mathbf{BA}$)?

Q5. For the 2×2 matrix

$$\mathbf{A} = \begin{pmatrix} 3 & -1 \\ 2 & 4 \end{pmatrix},$$

compute $\det(\mathbf{A})$ and interpret its absolute value as an area-scaling factor of the linear transformation $\mathbf{x} \mapsto \mathbf{Ax}$.

Q6. Compute the determinant of the 3×3 matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & -1 \\ 2 & 3 & 4 \end{pmatrix}.$$

Is $\mathbf{A}$ invertible?

Q7. For the matrix

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix},$$

(a) Compute $\det(\mathbf{A})$.

(b) Find $\mathbf{A}^{-1}$ explicitly.

Q8. Let

$$= \mathrm{diag}(1, 3, 5) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}.$$

(a) Find the eigenvalues of .

(b) For each eigenvalue, give a corresponding eigenvector.

Q9. Consider the symmetric matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Write the quadratic form $\mathbf{x}^\top \mathbf{Ax}$ as an explicit function of x_1 and x_2.

Q10. Let

$$\mathbf{A} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}.$$

(a) Compute $\mathbf{x}^\top \mathbf{Ax}$ for a general vector $\mathbf{x} = (x_1, x_2)^\top$.

(b) Show that $\mathbf{A}$ is positive semidefinite (psd) and decide whether it is positive definite.

Q11. Let

$$\mathbf{B} = \begin{pmatrix} 2 & -1 \\ 4 & -2 \end{pmatrix}.$$

(a) Determine the rank of $\mathbf{B}$.

(b) Find a basis for the column space of $\mathbf{B}$.

Q12. For the same matrix

$$\mathbf{B} = \begin{pmatrix} 2 & -1 \\ 4 & -2 \end{pmatrix},$$

find all solutions of $\mathbf{B}\mathbf{x} = \mathbf{0}$ and give a basis for the null space $\mathcal{N}(\mathbf{B})$.

Q13. Let

$$\mathbf{a} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

Compute the orthogonal projection of $\mathbf{x}$ onto the line spanned by $\mathbf{a}$, i.e.

$$\operatorname{proj}_{\mathbf{a}}(\mathbf{x}) = \frac{\mathbf{a}^\top \mathbf{x}}{\mathbf{a}^\top \mathbf{a}} \mathbf{a}.$$

Q14. Let

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 0 & 1 \\ 3 & 4 \end{pmatrix}.$$

Compute $\operatorname{tr}(\mathbf{AB})$ and $\operatorname{tr}(\mathbf{BA})$ and verify the property $\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA})$.

Q15. Let

$$\mathbf{x} = \begin{pmatrix} 1 \\ -2 \\ 3 \\ -4 \end{pmatrix}.$$

Compute the following norms:

$$\|\mathbf{x}\|_1, \qquad \|\mathbf{x}\|_2, \qquad \|\mathbf{x}\|_\infty.$$

Q16. Consider the matrix

$$= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

(a) Show that $^\top = \mathbf{I}_2$, i.e. is orthogonal.

(b) Let $\mathbf{x} = (1, 2)^\top$. Compute $\|\mathbf{x}\|_2$ and $\|\mathbf{x}\|_2$ and verify that the norm is preserved.

Q17. Let

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ -1 & 0 \end{pmatrix} \in \mathbb{R}^{3\times 2}.$$

(a) Compute $\mathbf{A}^\top \mathbf{A}$.

(b) Compute $\det(\mathbf{A}^\top \mathbf{A})$ and comment on its positive definiteness.

Q18. Consider the linear classifier in $\mathbb{R}^2$ with weight vector $\mathbf{w} = (2, -1)^\top$ and bias $b = -3$. The decision function is

$$f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b.$$

Classify the following points as lying on the positive side, negative side, or exactly on the decision boundary $f(\mathbf{x}) = 0$:

$$\mathbf{x}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \qquad \mathbf{x}_2 = \begin{pmatrix} 3 \\ 0 \end{pmatrix}.$$

Q19. Let

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

Compute the outer product $\mathbf{x}\mathbf{y}^\top$.

Q20. Consider the tall matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \in \mathbb{R}^{3\times 2}.$$

(a) Compute $\mathbf{A}^\top \mathbf{A}$ and $(\mathbf{A}^\top \mathbf{A})^{-1}$.

(b) Compute the Moore–Penrose pseudoinverse

$$\mathbf{A}^\dagger = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top.$$

Calculus (Questions 21–35)

Q21. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = 3x^3 - 5x^2 + 2x - 7.$$

Compute $f'(x)$.

Q22. Let

$$f(x) = e^{2x^2+1}.$$

Compute $f'(x)$ using the chain rule.

Q23. Let

$$f(x) = \ln(x^2 + 1).$$

(a) Compute $f'(x)$.

(b) Compute the second derivative $f''(x)$.

Q24. The ReLU activation is defined as

$$\mathrm{ReLU}(x) = \max(0, x).$$

(a) Write its derivative for $x < 0$ and for $x > 0$.

(b) Explain why the derivative does not exist at $x = 0$.

Q25. Consider the function of two variables

$$f(x_1, x_2) = x_1^2 x_2 + 4x_2.$$

$2x_1x_2$

Compute the partial derivatives $\dfrac{\partial f}{\partial x_1}$ and $\dfrac{\partial f}{\partial x_2}$.

$x_1 + 4$

Q26. Let

$$f(x_1, x_2, x_3) = x_1 + 2x_2^2 - 3x_3^3.$$

Compute the gradient $\nabla f(x_1, x_2, x_3)$ and evaluate it at the point $(1, -1, 1)$.

Q27. Consider the quadratic function

$$f(x_1, x_2) = x_1^2 + 4x_1x_2 + 3x_2^2.$$

(a) Compute the gradient $\nabla f(x_1, x_2)$.

(b) Compute the Hessian matrix $\mathbf{H}_f$.

(c) Determine whether $\mathbf{H}_f$ is positive definite, negative definite, or indefinite.

Q28. Let the vector-valued function $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^3$ be

$$\mathbf{f}(x_1, x_2) = \begin{pmatrix} x_1^2 \\ x_1 x_2 \\ \sin x_2 \end{pmatrix}.$$

(a) Compute the Jacobian matrix $\mathbf{J}(x_1, x_2) = \dfrac{\partial \mathbf{f}}{\partial (x_1, x_2)}$.

(b) Evaluate $\mathbf{J}(1, 0)$.

Q29. Let $f : \mathbb{R}^d \to \mathbb{R}$ be the linear function

$$f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b,$$

where $\mathbf{w} \in \mathbb{R}^d$ and $b \in \mathbb{R}$ are constants. Compute the gradient $\nabla_{\mathbf{x}} f(\mathbf{x})$.

Q30. Let $f : \mathbb{R}^d \to \mathbb{R}$ be the quadratic form

$$f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x},$$

where $\mathbf{A} \in \mathbb{R}^{d \times d}$ is a constant matrix. Compute the gradient $\nabla_{\mathbf{x}} f(\mathbf{x})$ in terms of $\mathbf{A}$ and $\mathbf{x}$. (No need to assume symmetry, but comment on the symmetric case.)

Q31. Consider the scalar parameter $\theta \in \mathbb{R}$ and the loss

$$L(\theta) = (\theta - 3)^2.$$

(a) Compute $\dfrac{dL}{d\theta}$.

(b) Starting from $\theta_0 = 0$ and using gradient descent with learning rate $\eta = 0.1$, compute the updated parameter θ_1.

Q32. In logistic regression or a neural unit, consider

$$z = \sigma(\mathbf{w}^\top \mathbf{x}), \qquad \sigma(t) = \frac{1}{1 + e^{-t}},$$

where $\mathbf{x} \in \mathbb{R}^d$ is the input and $\mathbf{w} \in \mathbb{R}^d$ are the weights. Compute the gradient $\dfrac{\partial z}{\partial \mathbf{w}}$.

Q33. Let

$$f(x) = x^4 - 2x^2.$$

(a) Compute $f'(x)$ and $f''(x)$.

(b) Find all critical points (where $f'(x) = 0$).

(c) Determine for which values of x the function is convex (i.e. $f''(x) \geq 0$).

Q34. Let

$$y = g(u), \qquad g(t) = t^2, \qquad u = \mathbf{a}^\top \mathbf{x},$$

with $\mathbf{a}, \mathbf{x} \in \mathbb{R}^d$. Use the multivariate chain rule to compute the gradient $\nabla_{\mathbf{x}} y$.

Q35. Let $\in \mathbb{R}^{n \times d}$, $\mathbf{y} \in \mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^d$. Consider the least-squares loss

$$L(\mathbf{w}) = \frac{1}{2}\|\mathbf{w} - \mathbf{y}\|_2^2 = \frac{1}{2}(\mathbf{w} - \mathbf{y})^\top (\mathbf{w} - \mathbf{y}).$$

Compute the gradient $\nabla_{\mathbf{w}} L(\mathbf{w})$.

Probability (Questions 36–50)

Q36. Two fair six-sided dice are rolled.

(a) Describe the sample space.

(b) Compute the probability that the sum of the two dice is exactly 10.

Q37. With the same experiment (two fair dice), compute the probability that the sum of the two dice is at least 9.

Q38. Let A be an event with $P(A) = 0.3$.

(a) Compute $P(A^c)$.

(b) Let B be another event with $P(B) = 0.5$, and assume A and B are disjoint. Compute $P(A \cup B)$ and $P(A \cap B)$.

Q39. There are 10 coins in a box. One coin is "bad" and always lands on tails. The other 9 coins are fair. You pick a coin uniformly at random and flip it once.

(a) What is the probability that the outcome of this flip is tails?

(b) Given that the outcome is tails, what is the probability that you picked the bad coin? (Use Bayes' rule.)

Q40. A disease has prevalence $P(D) = 0.01$ in a population. A test has

$$P(\text{positive} \mid D) = 0.99, \qquad P(\text{positive} \mid D^c) = 0.05.$$

(a) Compute $P(\text{positive})$.

(b) Compute $P(D \mid \text{positive})$ using Bayes' theorem.

Q41. At a university, 60% of students are in Computer Science (CS) and 40% in Electrical Engineering (EE). The probability of passing a certain course is $P(\text{pass} \mid \text{CS}) = 0.9$ and $P(\text{pass} \mid \text{EE}) = 0.7$.

(a) Use the law of total probability to compute $P(\text{pass})$.

(b) Compute $P(\text{CS} \mid \text{pass})$.

Q42. Let X be the result of rolling a fair six-sided die ($X \in \{1, 2, 3, 4, 5, 6\}$).

(a) Write the probability mass function (pmf) of X.

(b) Compute $\mathbb{E}[X]$.

(c) Compute $\text{Var}(X)$.

Q43. On a single fair die, define a random variable X by

$$X = \begin{cases} 0, & \text{if the outcome is 1 or 2,} \\ 1, & \text{if the outcome is } 3, 4, 5, 6. \end{cases}$$

(a) Find the pmf of X.

(b) Compute $\mathbb{E}[X]$ and $\mathrm{Var}(X)$.

Q44. A discrete random variable X takes values in $\{1, 2, 4\}$ with

$$P(X = 1) = 0.2, \quad P(X = 2) = 0.5, \quad P(X = 4) = 0.3.$$

Compute $\mathbb{E}[X]$ and $\mathrm{Var}(X)$.

Q45. Let X and Y be independent random variables with

$$\mathbb{E}[X] = 1, \quad \mathrm{Var}(X) = 2, \qquad \mathbb{E}[Y] = 3, \quad \mathrm{Var}(Y) = 4.$$

Compute:

(a) $\mathbb{E}[X + Y]$,

(b) $\mathrm{Var}(X + Y)$,

(c) $\mathrm{Cov}(X, Y)$,

(d) the correlation coefficient $\rho_{X,Y}$.

Q46. The joint pmf of (X, Y) is given by

$$P(X = x, Y = y) = \begin{cases} 0.2, & (x, y) = (0, 0), \\ 0.3, & (x, y) = (0, 1), \\ 0.1, & (x, y) = (1, 0), \\ 0.4, & (x, y) = (1, 1), \\ 0, & \text{otherwise.} \end{cases}$$

(a) Compute $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

(b) Compute $\mathbb{E}[XY]$.

(c) Compute $\mathrm{Cov}(X, Y)$.

Q47. Let $X \sim \mathcal{N}(\mu, \sigma^2)$ be a normal (Gaussian) random variable.

(a) Write the probability density function (pdf) of X.

(b) State its mean and variance.

Q48. Let $X \sim \mathrm{Bernoulli}(p)$.

(a) Write the pmf of X.

(b) Compute $\mathbb{E}[X]$ and $\mathrm{Var}(X)$.

Q49. For two random variables X and Y, suppose

$$\mathrm{Cov}(X, Y) = 3, \qquad \mathrm{std}(X) = 2, \qquad \mathrm{std}(Y) = \sqrt{5}.$$

Compute the correlation coefficient

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\mathrm{std}(X)\,\mathrm{std}(Y)}.$$

Q50. The chain rule of probability for three variables states that

$$P(X_1, X_2, X_3) = P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1, X_2).$$

(a) Write this factorization explicitly.

(b) If we additionally assume that X_3 is conditionally independent of X_1 given X_2, simplify $P(X_1, X_2, X_3)$ using this conditional independence.