Template for generation of Theory Questions

Date: 30-05-2025

Question generated by: Ms. Sajida Shaikh

Topic / Module: Mathematics for Artificial Intelligence

Session 1 & 2

Q1. What is the formula for cosine similarity between two vectors A and B?

A. (A · B) / ||A + B||

B. ||A - B|| / (||A|| \* ||B||)

C. (A · B) / (||A|| \* ||B||)

D. A · B

Answer: C. (A · B) / (||A|| \* ||B||)

---------------------------------------------------------------------------------------------------

Q2. Which of the following best describes orthogonal vectors in terms of dot product?

A. Dot product is 1

B. Dot product is -1

C. Dot product is 0

D. Dot product is undefined

Answer: C. Dot product is 0

--------------------------------------------------------------------------------------------------------

Q3. Which of the following is true about an orthogonal matrix Q?

A. Q⁻¹ = Qᵀ

B. Q has all positive eigenvalues

C. Q is always symmetric

D. QᵀQ ≠ I

Answer: A. Q⁻¹ = Qᵀ  
Solution: Orthogonal matrix satisfies QᵀQ = I

-------------------------------------------------------------------------------------------------------------

Session 3 & 4

Q4: The span of a set of vectors is:

A. The set of all scalar multiples of one vector

B. A linear combination of unit vectors

C. The set of all linear combinations of those vectors

D. A masure of vector length

Answer: C. The set of all linear combinations of those vectors

Solution: The span is the vector subspace formed by all possible linear combinations.

-------------------------------------------------------------------------------------------------------------

Q5: A vector is said to be normalized if:

A. Its norm is zero

B. Its magnitude is 1

C. Its angle with x-axis is 0

D. It has all equal components

Answer: B. Its magnitude is 1

Solution: A normalized (unit) vector has magnitude (norm) = 1.

---------------------------------------------------------------------------------------------------------

#### **Q6. What problem does L1 regularization (Lasso) help address in machine learning models?**

A) Overfitting by reducing model complexity  
B) Underfitting by adding more features  
C) Increasing computation time  
D) Converting categorical data into numerical data

**Answer:** A) Overfitting by reducing model complexity

**Explanation:** L1 regularization encourages sparsity by shrinking some coefficients to zero, effectively selecting important features.

---------------------------------------------------------------------------------------------------

Session 5

**Q7.** The **rank** of a matrix is:

1. The number of zero rows in its echelon form
2. The number of pivot columns in its row-reduced form
3. Always equal to the number of columns
4. Always less than the number of rows

**Answer:** B  
 **Solution:**  
Rank = number of **linearly independent rows or columns**, which equals the number of pivot positions.

-------------------------------------------------------------------------------------------------------

Q8. A **basis** for a vector space must:

A. Be orthogonal  
B. Contain exactly one vector  
C. Be linearly independent and span the space  
D. Be formed by unit vectors

**Answer:** C  
 **Solution:**

A basis spans the vector space and consists of **linearly independent** vectors.

-------------------------------------------------------------------------------------------------------

Session 6 & 7

#### **Q9. What type of matrix represents a reflection transformation?**

A) Diagonal matrix  
B) Orthogonal matrix with determinant -1  
C) Symmetric matrix  
D) Upper triangular matrix

**Answer:** B) Orthogonal matrix with determinant -1

**Explanation:** A reflection matrix is an orthogonal matrix whose determinant is -1, flipping the orientation of vectors.

----------------------------------------------------------------------------------------------------

#### **Q10.What is a linear transformation in the context of matrices?**

A) A transformation that preserves vector addition and scalar multiplication  
B) A transformation that changes the rank of a matrix  
C) A transformation that increases the determinant of a matrix  
D) A transformation that only applies to square matrices

**Answer:** A) A transformation that preserves vector addition and scalar multiplication

**Explanation:** A linear transformation TT satisfies T(a+b)=T(a)+T(b)T(a + b) = T(a) + T(b) and T(ca)=cT(a)T(ca) = cT(a), preserving linear structure.

-------------------------------------------------------------------------------------------------------

Q11: The Hadamard product of two matrices is:

A. The dot product of vectors

B. The element-wise multiplication of matrices

C. The matrix product

D. The product of transpose and original

Answer: B

Solution: Hadamard product multiplies corresponding elements of two matrices of the same size.

-------------------------------------------------------------------------------------------------------

Q12. If A⁻¹ exists, then:

A. AA⁻¹ = A

B. AA⁻¹ = 0

C. AA⁻¹ = I

D. A⁻¹ = Aᵀ

Answer: C

Solution: If A⁻¹ exists, then AA⁻¹ = A⁻¹A = I.

------------------------------------------------------------------------------------------------------

Q13. The determinant of a matrix is used to:

A. Find eigenvalues

B. Determine if a matrix is invertible

C. Compute the matrix trace

D. Normalize the matrix

Answer: B

Solution: A matrix is invertible only if its determinant is non-zero.

-------------------------------------------------------------------------------------------------------

Session 8 & 9

Q14: Which of the following is not a characteristic of a positive definite matrix?

* A. All eigenvalues are positive
* B. It is symmetric
* C. It has zero determinant
* D. xᵀAx > 0 for any non-zero x

Answer: C

Solution: Positive definite matrix has non-zero determinant and positive eigenvalues.

------------------------------------------------------------------------------------------------------

Q15. Low-rank matrix approximation is useful in:

* A. Increasing the storage size
* B. Reducing the dimensionality of data
* C. Amplifying noise in data
* D. Randomizing matrix elements

Answer: B

Solution: Low-rank approximation reduces dimensionality, preserves structure, and improves computational efficiency.

-------------------------------------------------------------------------------------------------------

Session 10 & 11

#### **Q16. How are eigenvalues of a matrix related to its determinant?**

A) The determinant is the sum of the eigenvalues  
B) The determinant is the product of the eigenvalues  
C) The determinant is always equal to one of the eigenvalues  
D) The determinant is unrelated to the eigenvalues

**Answer:** B) The determinant is the product of the eigenvalues

----------------------------------------------------------------------------------------------

#### **Q17. What is the primary use of Singular Value Decomposition (SVD) in data science?**

1. Matrix inversion only  
   B) Dimensionality reduction, noise filtering, and principal component analysis (PCA)  
   C) Solving linear equations only  
   D) None of the above

**Answer:** B) Dimensionality reduction, noise filtering, and principal component analysis (PCA)

**Explanation:** SVD is widely used in PCA for reducing the dimensionality of datasets and in applications like image compression.

---------------------------------------------------------------------------------------------------

Q18. Which of the following is a key use of eigenvalues in AI/ML?

A. Encrypting datasets

B. Measuring statistical variance

C. Data dimensionality reduction

D. Sorting datasets

Answer: C

Solution: Eigenvalues are used in PCA for dimensionality reduction.

------------------------------------------------------------------------------------------------------------

Q19. Eigenvectors represent:

A. Directions that remain unchanged after a linear transformation

B. The magnitude of transformation

C. The determinant of the matrix

D. The inverse of a matrix

Answer: A

Solution: Eigenvectors point in directions that are scaled by the transformation matrix.

-------------------------------------------------------------------------------------------------------

Q20: The first step in computing PCA for a dataset is:

A. Apply a convolutional layer

B. Center the data by subtracting the mean

C. Calculate the determinant

D. Maximize eigenvalues

Answer: B

Solution: PCA starts by centering the data to ensure the mean is zero.

-----------------------------------------------------------------------------------------------------------Session 12 & 13

**Q21.** The chain rule is used when:

A. Functions are added  
B. A function is composed within another function  
C. A function is constant  
D. Derivatives are not defined

**Answer:** B  
**Explanation:** Chain Rule is applied to differentiate composite functions, i.e., f(g(x))f(g(x))f(g(x)).

---------------------------------------------------------------------------------------------------

Q22.Different with respect to z h(z)=sin(z6)+sin6(z)

1. cos(z6)+ cos6(z)
2. 6z5cos(z6)+6sin5(z)cos(z)
3. 6z5cos(z)+6sin5(z)cos(z)
4. 6z5cos(z6)+6sin5(z)

Answer : B

----------------------------------------------------------------------------------------------------

Session 14 & 15

### **Q23. What does the directional derivative of a function measure?**

A) The rate of change of the function in a specific direction  
B) The second derivative of the function  
C) The gradient of the function at a point  
D) The minimum value of the function

**Answer:** **A) The rate of change of the function in a specific direction**

**Explanation:** The **directional derivative** tells us how a function changes **along a given direction** at a particular point.

-------------------------------------------------------------------------------------------------------

### **Q24. What happens when we take the directional derivative in the opposite direction of the gradient?**

A) The function increases the fastest  
B) The function remains constant  
C) The function decreases the fastest  
D) The function diverges

**Answer:** **C) The function decreases the fastest**

**Explanation:** Moving **opposite** to the gradient vector leads to the **steepest descent**, which is the basis of **Gradient Descent**.

-------------------------------------------------------------------------------------------------------

Session 16 & 17

**Q25.** The second-order terms in the multivariate Taylor series expansion of a function f(x)f(x)f(x) are used to capture:  
A. The linear change of the function  
B. The curvature of the function  
C. The exact function value  
D. The direction of steepest ascent

**Answer:** B  
 **Explanation:** The second-order terms in the Taylor series describe the curvature of the function (i.e., how the function changes in a nonlinear manner near a point).

-------------------------------------------------------------------------------------------------------

**Q26.** In the context of the gradient of a matrix, the derivative of a function f(X)f(X)f(X) with respect to XXX is known as the:  
A. Jacobian matrix  
B. Hessian matrix  
C. Gradient matrix  
D. Identity matrix

**Answer:** A  
**Explanation:** The Jacobian matrix is the matrix of first-order partial derivatives of a vector-valued function with respect to the components of the input vector or matrix.

-------------------------------------------------------------------------------------------------------

**Q27.** In machine learning, backpropagation is used to:  
A. Minimize the loss function by updating the weights of the network  
B. Compute the gradients of the loss function with respect to the weights  
C. Apply activation functions to hidden layers  
D. All of the above

**Answer:** D  
**Explanation:** Backpropagation is a process used to compute gradients of the loss function, update the weights, and apply activation functions in neural networks.

------------------------------------------------------------------------------------------------------

Session 18

**Q28.** In convex optimization, which of the following is always true?  
A. The feasible set is non-convex  
B. The problem has no solution  
C. The cost function must be non-differentiable  
D. Any stationary point is a global minimum

**Answer:** D  
**Explanation:** In convex optimization, any stationary point (where gradient is zero) is guaranteed to be the global minimum.

-----------------------------------------------------------------------------------------------------

**Q29.** Lagrange multipliers are used to:  
A. Solve unconstrained optimization problems  
B. Convert nonlinear equations to linear equations  
C. Find the local maxima and minima of a function subject to constraints  
D. Eliminate variables from a function

**Answer:** C  
**Explanation:** Lagrange multipliers help optimize a function subject to equality constraints.

----------------------------------------------------------------------------------------------------

Session 19

**Q30.** Which of the following is true about the **Hessian matrix** in optimization?  
A. It determines the direction of steepest ascent  
B. It is used to compute the gradient of the function  
C. It helps determine whether a point is a local minimum, maximum, or saddle point  
D. It is only used for non-differentiable functions

**Answer:** C  
**Explanation:** The Hessian matrix helps determine whether a point is a local minimum, maximum, or saddle point by analyzing the curvature of the function.

-------------------------------------------------------------------------------------------------------

**Q31.** In neural networks, the **Jacobian matrix** is used in:  
A. Backpropagation for computing the gradients  
B. Finding the local minima using second-order derivatives  
C. Convergence rate analysis of gradient descent  
D. Determining the function’s convexity

**Answer:** A  
**Explanation:** In neural networks, the Jacobian matrix is used during backpropagation to compute the gradients of the loss function with respect to the weights.

------------------------------------------------------------------------------------------------------

Session 20

**Q32.** A **saddle point** in optimization refers to:  
A. A point where the gradient is zero and is a local minimum  
B. A point where the gradient is zero and is neither a maximum nor a minimum  
C. A point with a maximum gradient  
D. A point where the function is concave

**Answer:** B  
**Explanation:** A saddle point occurs when the gradient is zero, but the point is neither a local maximum nor a local minimum, i.e., the function curves upwards in one direction and downwards in another.

------------------------------------------------------------------------------------------------------

### Q33. What is typically the effect of a **high learning rate** on Gradient Descent algorithms?

a) Faster convergence  
b) Slow convergence  
c) Risk of overshooting the minimum  
d) No effect

**Answer**: c) **Risk of overshooting the minimum**

**Explanation**: A high learning rate can cause the updates to overshoot the optimal point, resulting in poor convergence or divergence of the algorithm.

-------------------------------------------------------------------------------------------------------

**Q34.** **Mini-batch Gradient Descent** has the advantage of:  
A. Having faster convergence and reduced variance compared to SGD  
B. Always finding the global minimum  
C. Having no randomness in the updates  
D. Converging slower than Batch Gradient Descent

**Answer:** A  
**Explanation:** Mini-batch Gradient Descent offers a trade-off between the high variance of SGD and the computational cost of Batch Gradient Descent, providing faster convergence.

-------------------------------------------------------------------------------------------------------

Q35. If A is an m×n real matrix, which of the following correctly represents its Singular Value Decomposition?

A. A=UΣV   
B. A=UΣVT  
C. A=VΣUT  
D. A=UTΣV

**Correct Answer: B.** A=UΣVT

Q36. A 2×2 matrix A satisfies tr(A2)=5 tr(A)=5 and tr(A)=3. Find det(A)

1. 2
2. 4
3. 5
4. 7

Ans: A

Q37. Find Trace of matrix

A=​2 4 0

1 3 5

7 0 6​

1. 11
2. 10
3. 7
4. 9

#### Q38.**How are eigenvalues of a matrix related to its determinant?**

1. The determinant is the sum of the eigenvalues  
   B) The determinant is the product of the eigenvalues  
   C) The determinant is always equal to one of the eigenvalues  
   D) The determinant is unrelated to the eigenvalues

**Answer:** B) The determinant is the product of the eigenvalues

#### **Q39. What type of matrix represents a reflection transformation?**

A) Diagonal matrix  
B) Orthogonal matrix with determinant -1  
C) Symmetric matrix  
D) Upper triangular matrix

**Answer:** B) Orthogonal matrix with determinant -1

**Explanation:** A reflection matrix is an orthogonal matrix whose determinant is -1, flipping the orientation of vectors.

#### **Q40. What is the effect of applying an identity matrix as a transformation?**

A) It scales the vector  
B) It reflects the vector  
C) It rotates the vector  
D) It leaves the vector unchanged

✅ **Answer:** D) It leaves the vector unchanged