

Due at **11:59PM February 7, 2024**

What to submit: Submit a PDF to CatCourses. You can use the provided .tex and put your answers in solution sections bellow. Select all choices that apply for multi-choices problems.

Student Name:

---

## Part I: Linear Algebra

1. Which of the following vectors are in the span of the vectors  $\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$ ?

(a)  $\begin{bmatrix} 5 \\ 5 \\ 0 \end{bmatrix}$

(b)  $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

(c)  $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$

(d)  $\begin{bmatrix} 7 \\ -1 \\ 8 \end{bmatrix}$

*Solution:*

2. Let

$$A = \begin{bmatrix} 3 & -1 & 1 \\ 2 & 0 & 2 \end{bmatrix}, B = \begin{bmatrix} 2 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$

Find the inverse of  $AB^T$ .

*Solution:*

3. Which of the following statement is always true?

- (a) If  $A$  is a  $3 \times 5$  matrix and  $B$  is a  $5 \times 4$  matrix, then  $(AB)^T$  is a  $3 \times 4$  matrix.
- (b) If  $A = A^T$ , then the diagonal entries of  $A$  must be either 0 or 1's.
- (c) If  $AB = A^T B^T$ , then  $A$  and  $B$  must be of the same size.
- (d)  $AA^T = A^T A$

*Solution:*

4. Show that the following vectors form a linearly dependent set in  $\mathbb{R}^4$  by expressing  $\mathbf{v}_2$  as a linear combination of the other two.

$$\mathbf{v}_1 = \begin{bmatrix} 6 \\ 0 \\ 5 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 0 \\ 3 \\ 1 \\ -1 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 4 \\ -7 \\ 1 \\ 3 \end{bmatrix}$$

*Solution:*

## Part II: Probability

5. Suppose that  $P(A \cap B) = 0.4$  and  $P(B) = 0.9$ . Find  $P(A|B)$ .

*Solution:*

6. A random variable,  $X$ , has the probability distribution table as shown.

x	-2	-1	0	1	2
$P(X = x)$			0.4	0.1	0.1

Assume that  $P(X = -2) = P(X = -1)$ . Compute the expectation and variance of  $X$ .

*Solution:*

7. A motor insurance company insures drivers in age group A, B and C. 40% of the customers are in group A, 25% are in B, and 35% are in group C. The company's record shows that each year, 2% of customers in age group A, 1% in group B and 1.5% in group C made a claim. Given that a driver made a claim, what is the probability that the driver is from age group C?

*Solution:*

## Part III: Neural Network

8. Which of the followings would you consider to be valid activation functions?

(a)  $f(x) = -\min(2, x)$

(b)  $f(x) = 0.9x + 1$

(c)  $f(x) = \begin{cases} \min(x, 0.1x) & \text{if } x \geq 0 \\ \min(x, 0.1x) & \text{if } x < 0 \end{cases}$

(d)  $f(x) = \begin{cases} \max(x, 0.1x) & \text{if } x \geq 0 \\ \min(x, 0.1x) & \text{if } x < 0 \end{cases}$

*Solution:*

9. Which of the following indicates overfitting?

- (a) High training error, high test error
- (b) Low training error, low test error
- (c) Low training error, high test error
- (d) High training error, low test error

*Solution:*

10. Suppose we are training a simple neural network with two layers for regression. The network takes two-dimensional input  $[x_1, x_2]$  and gives a scalar  $\bar{y}$ . The first layer is a linear layer followed by ReLU. The second layer is also a linear layer followed by ReLU.

First layer:  $h_1 = \text{ReLU}(ax_1 + bx_2 + c)$

Second layer:  $\bar{y} = \text{ReLU}(dh_1 + e)$

Loss:  $L = (y - \bar{y})^2$

The network parameters are initialized as follows,

$$a = 2, b = 3, c = 1, d = 2, e = -5 \quad (1)$$

- (a) Given one training data point  $[x_1, x_2] = [1, 0]$  and its ground truth  $y = 3$ , compute  $h_1, \bar{y}, L$  in the forward pass.
- (b) Compute the gradient of the loss w.r.t. network parameters  $a, b, c, d, e$  respectively.
- (c) If the learning rate is set as 0.1, compute the updated value of  $a, b, c, d$ , and  $e$  after one iteration of gradient descent.

*Solution:*