
Sample Midterm Exam – Fall 2025

Q1 – Forward Pass of a Neuron

A neuron takes input $x = 2.0$, weight $w = 0.5$, and bias $b = -0.3$. The activation function is the sigmoid:

$$f(z) = \frac{1}{1 + e^{-z}}, \quad z = wx + b$$

1. Compute z and $\hat{y} = f(z)$.
2. If the target is $y = 0.8$, compute the loss $J = \frac{1}{2}(\hat{y} - y)^2$.

Q2 – Convolution and Pooling Dimensions

An image of size 6×6 is convolved with a 3×3 filter, stride = 2, and no padding.

1. Compute the output feature map size after the convolution.
2. A 2×2 max pooling with stride = 2 is then applied. Compute the final output size after pooling.
3. Compute the total number of learnable parameters (including biases) if 3 filters were used .

Q3 – PCA and Feature Scaling

A dataset contains three features:

- Feature 1: Building height (in meters), range $[0, 100]$
- Feature 2: Energy use (in kWh), range $[0, 10,000]$
- Feature 3: Occupancy rate (in percent), range $[0, 100]$

1. Without scaling, which feature will most strongly influence the first principal component? Explain briefly.

Q4 – SVM Decision Boundary

For a linear SVM, the separating hyperplane is:

$$3x_1 - 4x_2 + 2 = 0$$

For the point $(x_1, x_2) = (2, 1)$, compute the value of $f(x) = 3x_1 - 4x_2 + 2$ and determine whether it lies on the positive or negative side of the boundary.

Q5 – Decision Tree Entropy and Gain

A dataset has the following distribution for the target “Play Tennis”:

Outcome	Count
Play = Yes	6
Play = No	2

1. Compute the entropy before splitting (base 2).

Now, suppose a feature “Wind” divides the data into two groups:

Wind	Play = Yes	Play = No
Weak	4	1
Strong	2	1

2. Compute the entropy for each branch (**Weak** and **Strong**).
3. Based on the entropies, which branch is purer and why?

Q6 – True or False (Concept Check)

Mark each statement as **True (T)** or **False (F)**.

Statement	T/F
1. Increasing the number of filters in a CNN layer decreases the feature map depth.	
2. The sigmoid activation always outputs values between -1 and 1.	
3. In PCA, the first principal component captures the direction of maximum variance in the data.	
4. Gradient descent can converge faster with a larger learning rate, but may overshoot the minimum.	

Q7 – Transformers (Very Short Answer)

Answer each in one short sentence.

1. What is the main purpose of the **attention mechanism** in a Transformer?
2. What is the role of the **Feed-Forward Network (FFN)** layer after attention in each Transformer block?

Q8 – CNN Filter Computation with Stride

Consider a 4×4 grayscale input image and a single 2×2 convolution filter (kernel):

$$X = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 3 \\ 3 & 1 & 0 & 2 \\ 2 & 0 & 1 & 1 \end{bmatrix}, \quad K = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$$

The convolution is performed with: - stride = 2 - no padding - bias $b = 1$

Compute the output feature map values step-by-step.