CS5710 – Machine Learning

Home Assignment –3 PART B
University of Central Missouri
Course: CS5710 – Machine Learning
Fall 2025
Student Name: GANGADHAR NERSU

Q1. From Biological to Artificial

(a) (Biological vs. Artificial)

Neuron: Biological – A nerve cell that receives, processes, and transmits information via electrical impulses.

Artificial – A computational unit that receives weighted inputs, sums them, and applies an activation function.

Synapse: Biological – The junction between two neurons through which signals are transmitted chemically or electrically.

Artificial – The connection (weight) between two artificial neurons that determines the strength and direction of influence.

Activation Function: Biological – The biological mechanism (ion exchange) that determines whether a neuron fires.

Artificial – A mathematical function that introduces nonlinearity to decide the neuron's output based on weighted input.

(b) Two reasons nonlinearity is required in neural networks:

1. Allows modeling of complex, non-linear relationships between input and output.

2. Without nonlinearity, multiple layers collapse into one linear transformation.

(c) DAG assumption in feed-forward NNs:

A Directed Acyclic Graph (DAG) ensures that data flows only forward (input → hidden → output), preventing cycles and ensuring stable learning.

Q2. Architecture & Capacity

(a) Depth and width:

Depth – Number of layers between input and output.

Width – Number of neurons within a layer.

(b) Two roles of hidden units:

1. Feature Extraction – Detects useful patterns.

2. Feature Projection – Maps data into simpler representations.

(c) Trade-offs (D < M vs. D > M):

D < M: Lower capacity, risk of underfitting.

D > M: Higher capacity, risk of overfitting and higher computational cost.

(d) Universal Approximation Idea:

A one-hidden-layer neural network with enough hidden units can approximate any continuous function on a compact domain.

Q3. Perceptron vs. SVM

(a) Perceptron optimization:

Minimizes classification error by adjusting weights but doesn’t optimize a margin.

(b) SVM optimization:

Maximizes the margin between the separating hyperplane and closest data points.

(c) Why SVM generalizes better & real-world scenario:

SVMs generalize better by maximizing margin, reducing overfitting.

Example: Image classification (e.g., handwritten digit recognition).

Q4. Margins and Support Vectors

(a) Margin definition:

Distance between decision boundary and nearest training samples.

(b) Role of support vectors:

Points closest to the boundary that define the optimal hyperplane.

(c) Maximizing margin = robustness:

Larger margin reduces sensitivity to noise and improves stability.

Q5. PCA Concepts

(a) PCA assumption:

Data variance captures structure and information.

(b) Two equivalent definitions:

1. Finds orthogonal directions (principal components) maximizing variance.

2. Computes eigenvectors of covariance matrix with largest eigenvalues.

(c) Centering data:

Ensures variance is measured around mean, not biased by offsets.

(d) Eigenvalue meaning:

Indicates the variance explained by the corresponding principal component.

Q6. PCA vs. Random Projection

(a) Choice of directions:

PCA – data-driven, maximizing variance.

Random Projection – random orthogonal directions.

(b) Advantage & disadvantage:

Advantage (PCA): Optimal, interpretable components.

Disadvantage (PCA): Computationally expensive for large datasets.