**Editorial-W5A2: Understanding Neural Network Learning: From Perceptron to Overfitting**

**Question 1 (MSQ)**

A neural network model is being trained to classify emails as spam or not-spam. The engineer initializes all weights and bias to zero and sets a learning rate of 0.1.

**Question:** (Select all that apply) What is the primary role of the learning rate in this scenario?
A) It determines the number of layers in the network
B) It decides the activation function to use
C) It sets the initial values of weights
D) It controls how much weights are updated during training

**Correct Answer:** D) It controls how much weights are updated during training

**Explanation:**
The learning rate specifies the step size for updating weights after each prediction error, impacting how quickly or slowly the model learns.

**Question 2 (MSQ)**

An engineer uses perceptron to learn the AND logic gate. After several epochs, the error becomes zero and the weights stop changing.

**Question:** (Select all that apply) What does this indicate about the AND gate problem?
A) It is not linearly separable
B) The perceptron failed to converge
C) It is linearly separable and the perceptron has converged
D) More epochs are needed

**Correct Answer:** C) It is linearly separable and the perceptron has converged

**Explanation:**
The perceptron algorithm converges with zero error only for linearly separable problems like the AND gate.

**Question 3 (MCQ)**

A student tries to train a single-layer perceptron to solve the XOR logic gate problem. Even after 10,000 epochs, the error does not approach zero.

**Question:** Why does the perceptron fail to solve the XOR problem?
A) The learning rate is too high
B) XOR is not linearly separable
C) The bias is missing
D) The dataset is too small

**Correct Answer:** B) XOR is not linearly separable

**Explanation:**
A single-layer perceptron cannot solve problems that are not linearly separable, such as XOR.

**Question 4 (MSQ)**

A neural network uses the sigmoid activation function defined as
σ(z) = ( 1 + e^−z ) ^(-1)

**Question:** (Select all that apply) What is the main advantage of using the sigmoid function in neural networks?
A) It outputs only integer values
B) It guarantees zero error after training
C) It provides non-linearity and outputs values between 0 and 1
D) It is faster than all other activations

**Correct Answer:** C) It provides non-linearity and outputs values between 0 and 1

**Explanation:**
The sigmoid introduces non-linearity, allowing the network to handle complex patterns and output probabilities.

**Question 5 (MCQ)**

During training, a neural network's weights and bias are now changing by very magnitude, and the error remains very close to zero.

**Question:**
What has most likely occurred in the training process?
A) The network is underfitting
B) The model has converged
C) The learning rate is too high
D) The data is not linearly separable

**Correct Answer:** B) The model has converged

**Explanation:**
Convergence is achieved when weights stabilize and the error is minimized, indicating successful training.

**Question 6 (MCQ)**

An engineer wants to classify images of cats using three features: ear, whisker, and fur. She uses a neural network with twenty hidden neurons, each having different weights and thresholds.

**Question:**
What is the main benefit of using multiple neurons with different weights and thresholds in the hidden layer?

A) It reduces the number of epochs needed
B) It allows the network to capture more complex patterns
C) It eliminates the need for an activation function
D) It guarantees zero error for all datasets

**Correct Answer:** B) It allows the network to capture more complex patterns

**Explanation:**
Multiple neurons with diverse weights and thresholds enable the network to model and learn more intricate relationships in the data.

**Question 7 (MSQ)**

A neural network is trained using the following update rule for weights:

new weight = old weight + (learning rate × error × input)

**Question:** (Select all that apply) What is the purpose of this update rule?
A) To randomly initialize weights
B) To reduce error by adjusting weights
C) To increase the number of neurons
D) To set the bias value

**Correct Answer:** B) To reduce error by adjusting weights

**Explanation:**
This rule incrementally updates weights in direction that reduces error to minimize prediction error, improving the model's accuracy over time.

**Question 8 (MSQ)**

While training a neural network for handwritten digit recognition, the training accuracy reaches 99%, but validation accuracy plateaus at 72%.

**Question:** (Select all that apply) What does this performance gap most likely indicate?
A) Underfitting
B) High Bias
C) Overfitting
D) Low Variance

**Correct Answer:** C) Overfitting

**Explanation:**
A large discrepancy between training and validation accuracy suggests the model memorized training data patterns (noise, outliers) but fails to generalize to unseen data, a hallmark of overfitting.

**Question 9 (MSQ)**

A team is trying to solve a non-linearly separable problem with a neural network.

**Question:** (Select all that apply) What should they do to improve their model's performance?
A) Use a single-layer perceptron
B) Reduce the dataset size
C) Remove the bias term
D) Add more hidden layers and use non-linear activation functions

**Correct Answer:** D) Add more hidden layers and use non-linear activation functions

**Explanation:**
Multi-layer networks with non-linear activations can model complex, non-linear relationships that single-layer perceptrons cannot.

**Question 10 (MSQ)**

Which of the following best describes the analogy between neural network learning and human learning? (Select all that apply)
A) Both rely solely on memorization
B) Both involve trial, error, correction, and adaptation
C) Both require non-linear activation functions
D) Both always achieve zero error

**Correct Answer:** B) Both involve trial, error, correction, and adaptation

**Explanation:**
Neural networks, like human learning, improve through repeated feedback and adjustment.

**Question 11 (MCQ)**

A company is developing a simple AI system to automatically approve or reject access requests based on two criteria:

- **Criterion 1:** Importance of the request (input 1)
- **Criterion 2:** Security clearance of the requester (input 2)

They use a perceptron model with the following configuration:

- **Weights:** [2, −1] (importance weight = 2, clearance weight = −1)
- **Inputs:** [1, 1]
- **Bias:** −0.5
- **Activation:** Step function (output is 1 if z>0, otherwise 0)

**Question:**
Based on this model, will the AI system approve the current access request?
A) Yes (Output = 1)
B) No (Output = 0)

C) Output = −1
D) Output = 2

**Correct Answer:** A) Yes (Output = 1)

**Explanation:**

- Calculate the weighted sum:
  $z = (2\times1)+(-1\times1)+(-0.5) = 0.5$
- Apply the step function:
  Since $z = 0.5>0$, the output is **1** (approval).
- Therefore, the AI system will approve the request.

**Question 12 (MSQ)**

A team at MediHealth Analytics is training a neural network to predict patient readmission risks. They use backpropagation to optimize the model.

**Question:**
Which of the following statements about backpropagation are **correct**? (Select all that apply)
A) It updates network weights based on error gradients
B) It computes the gradient of the loss function with respect to each weight
C) It is exclusive to recurrent neural networks (RNNs)
D) It requires differentiable activation functions to work effectively

**Correct Answers:**
A) It updates network weights based on error gradients
B) It computes the gradient of the loss function with respect to each weight
D) It requires differentiable activation functions to work effectively

**Explanation:**

- **A** and **B** are correct because backpropagation calculates gradients of the loss relative to weights (**B**) and uses these gradients to update weights via optimization methods like gradient descent.
- **D** is correct: Backpropagation relies on differentiable activation functions (e.g., sigmoid, ReLU) to compute gradients through the chain rule.
- **C** is incorrect: Backpropagation is a universal algorithm applicable to **all neural network architectures** (CNNs, RNNs, etc.), not just RNNs.

**Question 13 (MSQ)**

A smart home security system uses a simple perceptron to decide whether to trigger an alert. The system considers three factors:

- **Motion detected in the living room** (Input 1: 1 if motion is detected, 0 otherwise)
- **Window sensor triggered** (Input 2: 1 if a window is opened, 0 otherwise)

- **Time is between 11 PM and 6 AM** (Input 3: 1 if true, 0 otherwise)

The system is designed to be more sensitive to motion and nighttime activity, and less sensitive to the window sensor alone.

The perceptron is configured as follows:

- **Inputs:** [1, 0, 1] (ordering same as weights)
- **Weights:** [0.5, 0.1, 0.6] (motion = 0.5, window = 0.1, night = 0.6)
- **Threshold:** 0.85
- **Activation:** Step function (outputs 1 if weighted sum ≥ threshold, otherwise 0)

**Question:** Choose correct options A) Yes (Output = 1)
B) No (Output = 0)
C) z = 1.1
D) z = 0.85
**Correct Answer:** A, C
**Explanation:**

- **Calculate the weighted sum:**
  z= ( 1× 0.5) + (0 × 0.1) + (1 × 0.6) = 0.5 + 0 + 0.6 = 1.1
- **Apply the step function:**
  Since z = 1.1 ≥ 0.85 (threshold), the output is **1** (alert triggered).

The smart home system will trigger an alert because the combined evidence meets the activation threshold.

**Question 14 (MSQ)**

A teacher at an educational institution uses PyTorch to store student assessment scores in the following tensor:

```
import torch

scores = torch.tensor([[[8, 5, 0, 1],
                        [5, 9, 9, 2]],
                       [[18, 15, 1, -1],
                        [25, 19, 0, 12]]])
```

**Question:** (Select all correct options) Which indexing method will retrieve the value 25 from the tensor?
A) scores[2, 2, 1]
B) scores[1, 1, 0]
C) scores[1, -1, 0]
D)scores[1, 1, -4]

**Correct Answer:** B, C, D

**Explanation:** Python uses 0 based indexing, also can use backward indexing to access the same.

**Question 15 (MSQ)**

A financial institution, SecureBank, is training a deep learning model to detect fraudulent transactions in real time. Initially, they used a CPU-based system, but as transaction volumes grew exponentially, training times became unmanageable-taking 72 hours per epoch. After switching to a GPU, training time dropped to 4 hours per epoch, enabling faster model iteration.

**Question:** (Select all correct options) What is the primary reason GPUs drastically reduced training time compared to CPUs?
A) GPUs eliminate the need for labeled training data
B) GPUs automatically correct prediction errors in the model
C) GPUs enhance model accuracy without additional training
D) GPUs execute thousands of parallel operations simultaneously

**Correct Answer:** D) GPUs execute thousands of parallel operations simultaneously

**Explanation:**

- **Key Architectural Difference**:
    - **CPUs** (e.g., SecureBank's initial setup) process tasks sequentially, struggling with large-scale parallel computations.
    - **GPUs** (e.g., NVIDIA CUDA cores) have thousands of smaller cores optimized for parallel matrix operations, critical for deep learning.
- **Impact on Training**:
    - Fraud detection models require processing millions of transactions. GPUs parallelize these tasks, reducing epoch time from **72 hours** to **4 hours**.
- **Why Other Options Fail**:
    - **A**: GPUs do not generate or label data-they accelerate existing computations.
    - **B/C**: Accuracy improvements depend on model architecture and training data, not hardware alone.