Quiz 2: Neural Networks and Deep Learning

Instructions

Select the best answer for each question. Questions 1-8 cover core concepts, while questions 9-10 are more challenging.

Questions

1. What is the key limitation of linear models that neural networks overcome?

A) Linear models are too slow to train B) Linear models cannot handle large datasets C) Linear models cannot represent non-linear relationships in data D) Linear models require too much memory

2. In a neural network, what is the purpose of an activation function?

A) To speed up the training process B) To introduce non-linearity into the model C) To normalize the input data D) To reduce the number of parameters

3. Which of the following is a common activation function used in neural networks?

A) Logarithmic function B) Polynomial function C) ReLU (Rectified Linear Unit) D) Fourier transform

4. Which loss function is typically used for multi-class classification problems where the output layer uses a Softmax activation?

A) Mean Squared Error (MSE) B) Mean Absolute Error (MAE) C) Binary Cross-Entropy (BCE) D) Categorical Cross-Entropy

5. What does the Universal Approximation Theorem primarily suggest about neural networks?

A) A sufficiently large neural network with a non-linear activation can approximate any continuous function. B) Neural networks can only approximate linear functions efficiently. C) Neural networks require an infinite number of neurons for practical tasks. D) The theorem guarantees that training will always converge to the global minimum.

6. When is a Sigmoid activation function commonly used in the output layer of a neural network?

A) For binary classification problems. B) For multi-class classification problems with more than two classes. C) For regression problems predicting continuous values. D) For generating feature embeddings.

7. Why is the XOR problem significant in the context of neural networks?

A) It demonstrates the need for complex feature engineering even for simple problems. B) It illustrates a simple case where data is not linearly separable, requiring non-linear models. C) It proves that linear models are always faster to train. D) It shows the limitations of using too many hidden layers.

8. In the mathematical model of an artificial neuron, what calculation is performed *before* applying the activation function?

A) The final output prediction of the network. B) The backpropagation error gradient. C) The normalization of the input features. D) The weighted sum of the inputs plus the bias term.

9. Which technique can help linear models handle non-linear patterns, although it might require domain knowledge and doesn't scale well?

A) Manual feature engineering (e.g., adding polynomial or interaction terms). B) Using the ReLU activation function. C) Applying dropout regularization. D) Increasing the number of data samples.

10. What is the primary goal of using the gradient descent update rule when training a model like linear regression or a neural network?

A) To find a closed-form analytical solution for the parameters. B) To iteratively adjust the model's parameters to minimize a defined loss function. C) To determine the optimal number of features or neurons. D) To randomly initialize the weights and biases.

Answers

1. C
2. B
3. C
4. D
5. A
6. A
7. B
8. D
9. A

10. B

Explanations

1. Neural networks can model complex, non-linear relationships that linear models cannot represent.
2. Activation functions introduce non-linearity into neural networks, allowing them to learn complex patterns. Without activation functions, neural networks would just be stacked linear models.
3. ReLU (Rectified Linear Unit) is one of the most commonly used activation functions in modern neural networks.
4. Categorical Cross-Entropy is the standard loss function for multi-class classification when the final layer uses Softmax to output a probability distribution across classes.
5. The Universal Approximation Theorem states that a neural network with at least one hidden layer and a non-linear activation function can approximate any continuous function, highlighting their theoretical power.
6. A Sigmoid function outputs a value between 0 and 1, making it suitable for the output layer in binary classification tasks where the output represents the probability of belonging to the positive class.
7. The XOR problem is a classic example of a non-linearly separable dataset, demonstrating the need for models capable of learning non-linear decision boundaries, like neural networks.
8. A neuron first computes the weighted sum of its inputs and adds a bias term (often called the pre-activation or logit), and then applies the activation function to this sum.
9. Manual feature engineering, such as creating polynomial features (x^2, x^3) or interaction terms ($x_1 imes x_2$), allows linear models to capture some non-linearities, but it's often difficult and problem-specific.
10. Gradient descent is an iterative optimization algorithm used to update the model's parameters (weights and biases) in the direction that reduces the loss function, aiming to find the parameters that minimize the error.