

**Exercise :**

Answer only 16 of the following multiple-choice questions. (PS: +0.75 for correct answer, - 0. 5 for incorrect one.)

Question	Answer
1. Which statement best captures the meaning of the phrase “AI is the new Electricity”? A) AI will soon primarily power electrical grids.      B) AI will transform industries in a way similar to how electricity did in the past. C) AI is a secondary tool used only in specific fields.      D) AI and electricity serve the same purpose in energy production.	B
2. What is the primary purpose of a neural network in a supervised learning setting? A) To cluster similar data points without prior labels.      B) To reduce the dimensionality of data. C) To map inputs to outputs using a set of labeled examples.      D) To generate new data from random noise.	C
3. Which type of neural network is most commonly used for image-related tasks such as photo tagging? A) Convolutional Neural Network (CNN)      B) Recurrent Neural Network (RNN) C) Feed-forward Neural Network      D) Deep Belief Network	A
4. What is a key factor that drives the rapid progress of deep learning? A) Reduction in network complexity      B) Increased scale in data, computation, and algorithmic efficiency C) Minimal use of training data      D) Traditional rule-based programming	B
5. An image is represented as $64 \times 64 \times 3 = 12,288$ . What does 12,288 represent? A) The number of pixels in the image      B) The total number of color channels C) The total number of input features when flattening the image      D) The resolution of the image in kilobytes.	C
6. In gradient descent, how are the weight parameters updated? (Here, $\alpha$ represents the learning rate) A) $w := w + \alpha \cdot dw$ B) $w := w - \alpha \cdot dw$ C) $w := w \times \alpha \cdot dw$ D) $w := w / \alpha \cdot dw$	B
7. What is the derivative of $f(a) = a^2$ with respect to $a$ ? A) 1      B) $a$ C) $a^2$ D) $2a$	D
8. What does “vectorization” mean in the context of neural network programming? A) Using loops to iterate over each element      B) Representing data using binary vectors only C) Expressing operations in terms of matrix and vector computations      D) Converting scalar values into vectors manually	C
9. Which NumPy function is typically used for computing the dot product in a vectorized implementation? A) <code>np.dot()</code> B) <code>np.add()</code> C) <code>np.multiply()</code> D) <code>np.sum()</code>	A
10. In logistic regression, when $z$ is a large negative number, the sigmoid activation output is approximately: A) 1      B) 0      C) 0.5      D) -1	B
11. In the context of gradient descent for logistic regression, why do we compute gradients ( $dw$ and $db$ )? A) To increase the cost function.      B) To measure the accuracy of the model. C) To expand the dataset size.      D) To update the model parameters to minimize the cost function	D

<p><b>12. What are the three main layers in a one-hidden-layer neural network?</b>  convolutional layer, output layer  recurrent layer, output layer</p>	<p>A) Input layer,  C) Input layer,  D) Input layer, pooling layer, output layer</p>	<p><b>B</b></p>
<p><b>13. What is the definition of the ReLU activation function?</b>  A) <math>g(z) = \max(0, z)</math></p>	<p>B) <math>g(z) = \tanh(z)</math>  C) <math>g(z) = 1/(1 + e^{(-z)})</math>  D) <math>g(z) = z^2</math></p>	<p><b>A</b></p>
<p><b>14. Which of the following architectures best illustrates a deep neural network?</b>  A) Input layer → Logistic regression (no hidden layer) → Output  B) Input layer → One hidden layer → Output layer  C) <b>Input layer → Multiple hidden layers → Output layer</b>  D) Input layer → Convolution → Pooling → Output</p>		<p><b>C</b></p>
<p><b>15. What is the primary difference between parameters and hyperparameters in a deep neural network?</b>  A) Parameters are set manually; hyperparameters are learned from data.  B) Hyperparameters include activation functions, while parameters do not.  C) There is no difference; they are synonymous.  D) <b>Parameters are learned from data, while hyperparameters are set before training.</b></p>		<p><b>D</b></p>
<p><b>16. In the context of bias and variance, which scenario is indicative of a model suffering from high variance (i.e., overfitting)?</b>  A) High training error and high development (dev) error  B) Low training error and low development error  C) <b>Low training error but high development error</b>  D) High training error but low development error</p>		<p><b>C</b></p>
<p><b>17. How does dropout regularization help prevent overfitting in neural networks?</b>  A) It adds Gaussian noise to the inputs.  B) <b>It randomly sets a fraction of the hidden layer activations to zero during training, forcing the network to not rely on any single neuron.</b>  C) It decreases the learning rate over iterations.  D) It only regularizes the output layer.</p>		<p><b>B</b></p>
<p><b>18. What is the primary purpose of learning rate decay in training neural networks?</b>  A) <b>To slowly reduce the learning rate over epochs, allowing finer convergence near a minimum.</b>  B) To gradually increase the learning rate over time for faster training.  C) To keep the learning rate constant throughout training.  D) To randomly adjust the learning rate at each iteration.</p>		<p><b>A</b></p>
<p><b>19. Why is it important to initialize neural network weights to small random values instead of zeros?</b>  A) It simplifies the training algorithm.  B) It avoids the use of bias in the network.  C) It guarantees that the network converges in a single iteration.  D) <b>It prevents symmetry, enabling different neurons to learn different features.</b></p>		<p><b>D</b></p>