General Neural Network Questions

1. **Why sigmoid in the output layer for binary classification?**
   Sigmoid outputs probabilities (0 to 1) for binary classification.

2. **What does input_dim=X.shape[1] mean in the Dense layer?**
   It sets the input layer to match the number of features in X.

3. **Purpose of StandardScaler in preprocessing?**
   Scales features to zero mean and unit variance for faster convergence.

4. **What does batch_size=32 mean in model.fit?**
   Processes 32 samples per gradient update, balancing speed and stability.

5. **Why binary_crossentropy as loss? What for multiclass?**
   Measures binary classification error; use categorical_crossentropy for multiclass.

6. **Role of adam optimizer?**
   Adapts learning rate for efficient weight updates.

7. **Why relu in hidden layers?**
   Adds non-linearity, avoids vanishing gradients, and is computationally efficient.

8. **Why split data with train_test_split?**
   Trains on one set, tests on unseen data to evaluate generalization.

Convolutional Neural Network (CNN) Questions

9. **Purpose of Conv2D layer?**
   Extracts spatial features like edges from images using filters.

10. **Why use MaxPooling2D after Conv2D?**
    Reduces dimensions, lowers computation, and prevents overfitting.

11. **What does Flatten layer do?**
    Converts 2D feature maps to 1D for dense layers.

12. **Significance of rescale=1./255 in ImageDataGenerator?**
    Normalizes pixel values to [0, 1] for stable training.

13. **Why input shape (64, 64, 3) in CNN?**
    Represents 64x64 RGB images (height, width, 3 channels).

14. **Purpose of Dropout(0.5) in custom CNN?**
    Randomly drops 50% of neurons to prevent overfitting.

15. **Why more layers in custom CNN?**
    Learns complex features for better pattern recognition.

16. **Changes for multiclass classification in CNN?**
    Use softmax in output layer and categorical_crossentropy loss.

Transfer Learning Questions

17. **What is transfer learning, and why used?**
    Uses pre-trained model features to save time and improve performance.

18. **Why set base model layers to trainable = False?**
    Freezes pre-trained layers to retain learned features.

19. **Role of GlobalAveragePooling2D?**
    Averages feature maps to a vector for classification.

20. **Why input shape (224, 224, 3) for VGG16?**
    Matches pre-trained model's expected 224x224 RGB input.

21. **Difference between categorical_crossentropy and binary_crossentropy?**
    categorical for multiclass, binary for two-class problems.

22. **How does Adam differ from SGD?**
    Adam uses adaptive learning rates for faster convergence.

Recurrent Neural Network (RNN, LSTM, GRU) Questions

23. **Why LSTM for stock price prediction?**
    Captures long-term dependencies in time-series data.

24. **What is sequence_length=60 in create_sequences?**
    Uses 60 past time steps to predict the next value.

25. **Why use MinMaxScaler for stock prices?**
    Scales data to [0, 1] for stable training.

26. **Difference between SimpleRNN, LSTM, GRU?**
    SimpleRNN has vanishing gradients; LSTM and GRU handle long-term dependencies, with GRU simpler.

27. **Why mean_squared_error for stock prediction?**
    Minimizes error for continuous regression predictions.

28. **How does Dense(1) produce predictions in LSTM?**
    Outputs one value for the next stock price.

Next Word Prediction Questions

29. **Role of Embedding layer?**
    Converts words to dense vectors for semantic representation.

30. **Why softmax in output layer?**

   Outputs probabilities over vocabulary for next word.

31. **What does to_categorical do?**

   Converts word indices to one-hot encoded vectors.

32. **Why calculate max_seq_len?**

   Pads sequences to fixed length for consistent input.

33. **How does LSTM help in next word prediction?**

   Captures contextual dependencies in text sequences.

Autoencoder Questions

34. **Purpose of autoencoder in MNIST code?**

   Learns compressed data representations for reconstruction.

35. **Why input/output shape (784,) in autoencoders?**

   Represents flattened 28x28 MNIST images.

36. **Difference between vanilla and deep autoencoder?**

   Vanilla has one encoding/decoding layer; deep has multiple.

37. **Why binary_crossentropy in autoencoders?**

   Measures pixel-wise error for normalized images.

38. **Role of UpSampling2D in convolutional autoencoder?**

   Increases feature map size for image reconstruction.

39. **How does convolutional autoencoder differ?**

   Uses convolutional layers for spatial data, unlike dense layers.

40. **Why normalize MNIST to [0, 1]?**

   Stabilizes training and matches loss function.

Q-Learning Questions

41. **Purpose of Q-matrix?**

   Stores expected rewards for state-action pairs.

42. **What does gamma represent?**

   Balances immediate vs. future rewards (0.8).

43. **Why -1 for invalid actions in R?**

   Prevents agent from choosing invalid actions.

44. **Significance of Bellman equation?**

   Updates Q-values with immediate and future rewards.

45. **Why normalize Q-matrix?**
    Scales values to [0, 100] for interpretability.

46. **What happens at terminal states (2 or 3)?**
    Episode ends as goals are reached.

47. **How does learning rate (α) affect Q-learning?**
    Controls how much new rewards update Q-values.

General Deep Learning Concepts

48. **Supervised vs. unsupervised learning in notebook?**
    Supervised (CNNs) uses labels; unsupervised (autoencoders) learns from unlabeled data.

49. **Why use epochs? Effect of number?**
    Epochs repeat dataset passes; too few underfit, too many overfit.

50. **What is overfitting? How do dropout/augmentation help?**
    Overfitting is poor generalization; dropout deactivates neurons, augmentation adds data variety.