1. The number of axis of a tensor is called

a. dimensions

b. rank

c. shape

d. attribute

answer - b

1. For which of the following is root mean squared error an appropriate loss function?
2. Binary Classification
3. Multi-class classification
4. Regression
5. Logistic regression

answer – c

1. Which of the following statements is true when you use 1×1 convolution in a CNN?
2. Overfitting is less
3. Can be used for feature pooling
4. Helps in dimensionality reduction
5. All of the above

Answer - d

1. Bag-of-words is more suitable for algorithms like
2. 1D convnet
3. CNN
4. RNN
5. Random forest

Answer – d

1. The input image has been converted into a matrix of size 28 X 28 and a kernel of size 7 X 7 with a stride of 1. What will be the size of the convoluted matrix?
2. 14\*14
3. 22\*22
4. 21\*21
5. 7\*7

Answer – b

1. Which of following activation function cannot be used at output layer to classify an image?
2. Sigmoid
3. Tanh
4. RELU
5. None of the above

Answer – c

1. The vanishing gradient problem was addressed using

a. LSTM

b. ResNets

c. RNN

d. CNN

Answer – a

1. In which neural net architecture, does weight sharing occur?

a. CNN and FCN

b. FCN and RNN

c. CNN and RNN

d. None of the above

Answer – c

1. For a classification task, instead of random weight initializations in a neural network, we set all the weights to zero. Which of the following statements is true?

a. The network will train as learn as usual

b. The neural network will not train as there is no net gradient change

c. The neural network will not train properly

d. The neural network will train but all the neurons will end up recognizing the same thing

Answer - d

1. Which is the most suitable neural network architecture for image classification problem?

a. RNN

b. CNN

c. Perceptron

d. FCN

Answer – b

1. If calculation of reset gate in GRU unit is close to 0, which of the following would occur?

a. Previous hidden state would be ignored

b. Previous hidden state would be not be ignored

c. Forgets the information for future time steps

d. Copies the information through many time steps

Answer - b

1. Batch Normalization is helpful because

a. Normalizes all the input before sending it to the next layer

b. It is a very efficient backpropagation technique

c. It returns back the normalized mean and standard deviation of weights

d. Improves the speed of the network

Answer - a