Final Exam (NN&DL 5720)

Exam is of 100 marks containing 80 Multiple choice questions. You have 60 minutes to attempt the exam.

[redacted] Switch account

Your email will be recorded when you submit this form

* Required

First Name , Last Name *

Your answer

User ID (SSO) *

Your answer

Find the reasons for Deep Learning recent take off? * 1 point

- [x] We have access to a lot more computational power.
- [ ] Neural Networks are a brand new field.
- [x] We have access to a lot more data.
- [x] Deep learning has resulted in significant improvements in important applications such as online advertising, speech recognition, and image recognition.

If you want to cluster seven data points into three clusters using K-Means. After 1st iteration clusters, C1, C2, C3 has following observations: * 1 point

C1: {(2,2), (4,4), (6,6)}

C2: {(0,4), (4,0)}

C3: {(5,5), (9,9)}

Find the cluster centroids for the second iteration?

- ○ C1: (6,6), C2: (4,4), C3: (9,9)
- ○ C1: (2,2), C2: (0,0), C3: (5,5)
- ○ C1: (4,4), C2: (2,2), C3: (7,7)
- ○ None of the above

An experienced deep learning engineer working on a naïve problem, usually use insight from previous problems to train a good model on the first try, without needing to iterate through different models. * 1 point

- ○ True
- ✓ False

Why the Recurrent Neural Network used for machine translation, like translating English to French? (Mark all that apply.) * 1 point

- ☑ It can be trained as a supervised learning problem.
- ☐ It is strictly more powerful than a Convolutional Neural Network (CNN).
- ☑ It is applicable when the input/output is a sequence (e.g., a sequence of words).
- ☐ RNNs represent the recurrent process of Idea->Code->Experiment->Idea->....

We have a 64x64 input with RGB channel represented as img = (64,64,3). Reshape this array into a column vector. * 1 point

- [x] x= img.reshape((64*64*3,1))
- [ ] x= img.reshape((64*64*3))
- [ ] x= img.reshape((1,64*64*3))
- [ ] x= img.reshape((3*64*64))

Consider the two following random arrays "a" and "b": * 1 point

```
a = np.random.randn(2, 3) # a.shape = (2, 3)
b = np.random.randn(2, 1) # b.shape = (2, 1)
c = a + b
```

What will be the shape of "c"?

- [x] c.shape = (3,2)
- [ ] c.shape = (2,1)
- [ ] Error
- [ ] c.shape = (2,3)

Consider the two following random arrays "a" and "b": * 1 point

```
a = np.random.randn(4, 3) # a.shape = (4, 3)
b = np.random.randn(3, 2) # b.shape = (3, 2)
c = a + b
```

What will be the shape of "c"?

- c.shape = (3,3)
- c.shape = (4,2)
- error
- c.shape = (4,3)

Calculate the precision when the true positive value is 10 and the false positive value is 15. * 1 point

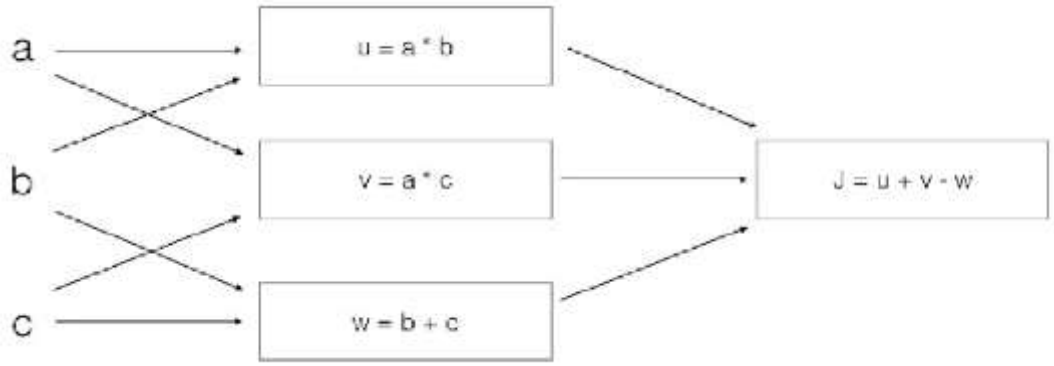
- 0.5
- 0.4
- 0.6
- 0.8

Suppose you have n_x input features per example. Recall that X=[x^(1), x^(2)...x^(m)]. What is the dimension of X? * 1 point

- (1, m)
- (n_x, m)
- (m, n_x)
- (m, 1)

Consider the following computation graph. * 1 point

What is the output J?



- J = (c - 1) * (b + a)
- J = (a - 1) * (b + c)
- J = a *b + b*c + a*c
- J = (b - 1) * (c + a)

The tanh activation usually works better than sigmoid activation function for hidden units because the mean of its output is closer to zero, and so it centers the data better for the next layer. True/False? * 1 point

- True
- False

You are building a binary classifier for recognizing banana (y=1) vs. watermelons (y=0). Which one of these activation functions would you recommend using for the output layer? * 1 point

- ReLU
- Leaky ReLU
- sigmoid
- tanh

Suppose you have built a neural network. You decide to initialize the weights and biases to be zero. Which of the following statements are True? (Check all that apply) * 2 points

- [ ] Each neuron in the first hidden layer will compute the same thing, but neurons in different layers will compute different things, thus we have accomplished “symmetry breaking”.
- [ ] The first hidden layer’s neurons will perform different computations from each other even in the first iteration; their parameters will thus keep evolving in their own way.
- [ ] Each neuron in the first hidden layer will perform the same computation. So even after multiple iterations of gradient descent each neuron in the layer will be computing the same thing as other neurons.
- [ ] Each neuron in the first hidden layer will perform the same computation in the first iteration. But after one iteration of gradient descent they will learn to compute different things because we have “broken symmetry”.

Logistic regression’s weights w should be initialized randomly rather than to all zeros, because if you initialize to all zeros, then logistic regression will fail to learn a useful decision boundary. * 2 points

- True
- False

Which of the following statements is true? * 1 point

- [x] The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers. Correct
- [ ] The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.

During forward propagation, in the forward function for a layer l you need to know what is the activation function in a layer (Sigmoid, tanh, ReLU, etc.). During backpropagation, the corresponding backward function also needs to know what is the activation function for layer l, since the gradient depends on it. True/False? * 1 point

- [x] True
- [ ] False

The effectiveness of an SVM depends upon: 1 point

- [ ] Selection of Kernel
- [ ] Kernel Parameters
- [ ] Soft Margin Parameter C
- [ ] All of the above

Which one is correct? * 1 point

- [ ] The most positively correlated features are good features
- [ ] The most negatively correlated features are good features
- [ ] The features with correlation zero are good features
- [ ] A, B

Which of these about a set is not true? * 1 point

- ◯ Mutable data type
- ◯ Allows duplicate values
- ◯ Data type with unordered values
- ◯ Immutable data type

The validation and test set should * 1 point

- ◯ Come from the different distribution
- ◯ Come from the same distribution
- ◯ be identical to each other
- ◯ Error as display function requires additional argument
- ◯ have same number of instances

Which of these techniques are useful for reducing variance (reducing overfitting)? (Check all that apply.) * 2 points

- ☑ Data augmentation
- ☐ Exploding gradient
- ☑ Dropout
- ☐ vanishing gradient
- ☐ increasing number of layers in model
- ☑ L2 regularization

Why do we normalize the inputs x? * 1 point

- ○ It makes the parameter initialization faster
- ✓ It makes the cost function faster to optimize
- ○ It makes it easier to visualize the data

If searching among a large number of hyperparameters, you should try values in a grid rather than random values, so that you can carry out the search more systematically and not rely on chance. True or False? * 1 point

- ○ True
- ✓ False

Every hyperparameter, if set poorly, can have a huge negative impact on training, and so all hyperparameters are about equally important to tune well. * 1 point

- ○ True
- ✓ False
- ○ Not applicable

It is not necessary to have a target variable for applying dimensionality reduction algorithms. * 1 point

- ○ True
- ○ False

Which of the following techniques would perform better for reducing dimensions of a data set? * 1 point

- ◯ Removing columns which have too many missing values
- ◯ Removing columns which have high variance in data
- ◯ Removing columns with dissimilar data trends
- ◯ None of these

Which of these statements about deep learning programming frameworks are true? (Check all that apply) * 2 points

- ☐ A programming framework allows you to code up deep learning algorithms with typically fewer lines of code than a lower-level language such as Python.
- ☐ Even if a project is currently open source, good governance of the project helps ensure that the it remains open even in the long term, rather than become closed or modified to benefit only one company.
- ☐ Deep learning programming frameworks require cloud-based machines to run.

A company needs an algorithm that can: * 2 points

let us know a bird is flying over a bridge as accurately as possible
takes no more than 10sec to classify a new image
fit in 10MB of memory

If you had three following models with provided performance, which one would you choose?

- ○ test accuracy: 97%, runtime: 1sec, memory size:3MB
- ○ test accuracy: 99%, runtime: 13sec, memory size:9MB
- ○ test accuracy: 98%, runtime: 9sec, memory size:9MB
- ○ test accuracy: 97%, runtime: 3sec, memory size:2MB

Which of the following do you typically see as you move to deeper layers in a ConvNet? * 2 points

- ○ nH and nW increases, while nC decreases
- ○ nH and nW decreases, while nC also decreases
- ○ nH and nW increases, while nC also increases
- ○ nH and nW decrease, while nC increases

The measure of the randomness in the information being processed in the Decision Tree by * 1 point

- ○ Entropy
- ○ Information gain

The idea is to encourage network to learn an encoding and decoding which only relies on activating a small number of neurons. * 1 point

- Convolution Neural Network
- Sparse autoencoder
- Undercomplete autoencoder
- Recurrent Neural architecture

Which of the following do you typically see in a ConvNet? (Check all that apply.) * 2 points

- Multiple CONV layers followed by a POOL layer
- Multiple POOL layers followed by a CONV layer
- FC layers in the last few layers
- FC layers in the first few layers

Select the correct sequence of typical decision tree structure down below: * 2 points

(I) Take the entire data set as input
(II) Divide the input data into two part
(III) Reapply the split to every part recursively
(IV) Stop when meeting desired criteria
(V) Cut the tree when we went too far while doing splits

- (I), (II),(V),(IV),(III).
- (V),(I),(III),(II),(IV).
- (I),(III),(II),(V),(IV).
- (I),(II),(III),(IV),(V).

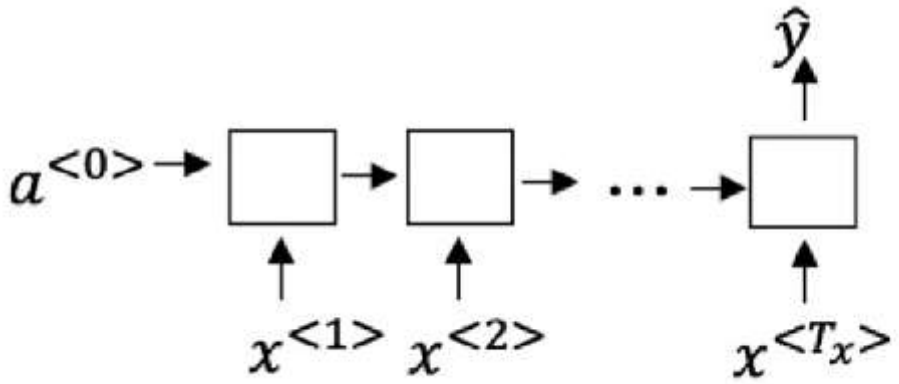
In order to be able to build very deep networks, we usually only use pooling layers to downsize the height/width of the activation volumes while convolutions are used with “valid” padding. Otherwise, we would downsize the input of the model too quickly. * 1 point

- ○ True
- ○ False

Training a deeper network (for example, adding additional layers to the network) allows the network to fit more complex functions and thus almost always results in lower training error. * 1 point

- ○ True
- ○ False
- ○ N/A

To which of these tasks would you apply a many-to-one RNN architecture? * 2 points
(Check all that apply).



- [ ] Image classification (input an image and output a label)
- [ ] Speech recognition (input an audio clip and output a transcript)
- [ ] Sentiment classification (input a piece of text and output a 0/1 to denote positive or negative sentiment)
- [ ] Gender recognition from speech (input an audio clip and output a label indicating the speaker's gender)

Which of the following is not an application of CNN? * 2 points

- ( ) Face Detection and Recognition
- ( ) Image Classification
- ( ) Image Segmentation
- ( ) Edge detection
- ( ) Document Analysis
- ( ) Image Colorization
- ( ) Recommendation Systems
- ( ) Not applicable

Suppose, you have given the following data where x and y are the 2 input variables and Class is the dependent variable. Suppose, you want to predict the class of new data point x=1 and y=1 using Euclidean distance in 7NN. In which class this data point belong to? * 5 points

x	y	Class
−1	1	−
0	1	+
0	2	−
1	−1	−
1	0	+
1	2	+
2	2	−
2	3	+

- + Class
- – Class
- Can't say
- None of these

Primary advantage of using CNN over ANN are: * 2 points

- A CNN can solve classification problems with higher accuracy as compared to ANN.
- A CNN architecture can detect the important features in an image on its own and does not require human intervention.
- None of above
- Both of above

Select option that explains better the Pooling? * 2 points

- Decrease the features size, in order to decrease the computional power that are needed.
- Creates a pool of data in order to improve the accuracy of the alghorithm predicting images.
- It assists in the detection of distorted features, in order to find dominant attributes.
- It assists in the detection of features, even if they are distorted, in addition to decreasing the attribute sizes, resulting in decreased computational need. It is also very useful for extracting dominant attributes.

One way to implement undercomplete autoencoder is to constrain the number of nodes present in hidden layer(s) of the neural network. * 1 point

- True
- False
- N/A

Autoencoders are incapable of learning nonlinear manifolds * 1 point

- True
- False
- N/A

Sparse autoencoders introduces information bottleneck by reducing the number of nodes at hidden layers * 1 point

- ◯ True
- ◯ False
- ◯ Not applicable

Select the correct statement(s) from the following * 1 point

- ☐ The clusters formed by k-means algorithm do not depend on the initial selection of cluster centers.
- ☐ The results of k-means algorithm get impacted by outliers and range of the attributes.
- ☐ K-means clustering automatically selects the most optimum value of k
- ☐ k-means algorithm can be applied to both categorical and numerical variables.

Select the centroid of the following 5 data points * 2 points

X	Y	Z
12	23	45
31	31	31
17	15	25
19	27	45
13	11	27

- ○ 18.4, 21.4, 32.1
- ○ 18.4, 21.4, 34.6
- ○ 34.6, 21.4, 18.4
- ○ 21.4, 32.1, 18.4

The loss is between the original input and the reconstruction from a noisy version of the input. * 1 point

- ○ True
- ○ False

Denoising autoencoders can be used as a tool for feature extraction. * 1 point

- ○ True
- ○ False
- ○ Not relevent

In a classification system, a patient with the disease is classified as not having the disease.
This is an example of a: * 1 point

- True positive
- True negative
- False positive
- False negative

Generative adversarial network is an application of ? * 1 point

- CNN
- LSTM
- Autoencoder
- None of the above

The benefit of Naïve Bayes includes: * 1 point

- Naïve Bayes is one of the fast and easy ML algorithms to predict a class of datasets.
- It is the most popular choice for text classification problems.
- It can be used for Binary as well as Multi-class Classifications.
- All of the above

For what RNN is used and achieve the best results * 1 point

- ◯ Financial predictions
- ◯ Handwriting and images recognition
- ◯ Handwriting and speech recognition
- ◯ Speech and images recognition

What is 'gradient' when we are talking about RNN? * 1 point

- ◯ The most important step of RNN algorithm
- ◯ A parameter that can help you improve the algorhitm´s accuracy
- ◯ A gradient is a partial derivative with respect to its inputs
- ◯ It is how RNN calls it´s features

Which one is correct? select the relevant option/options * 1 point

- ☐ LSTM networks are an extension for recurrent neural networks, which basically extends their memory. Therefore it is not recommended to use it, unless you are using a small Dataset.
- ☐ LSTM networks are an extension for recurrent neural networks, which basically shorten their memory. Therefore it is well suited to learn from important experiences that have very low time lags in between
- ☐ LSTM networks are an extension for recurrent neural networks, which basically extends their memory. Therefore it is well suited to learn from important experiences that have very low time lags in between
- ☐ LSTM networks are an extension for recurrent neural networks, which basically extends their memory. Therefore it is well suited to learn from important experiences that have very long time lags in between

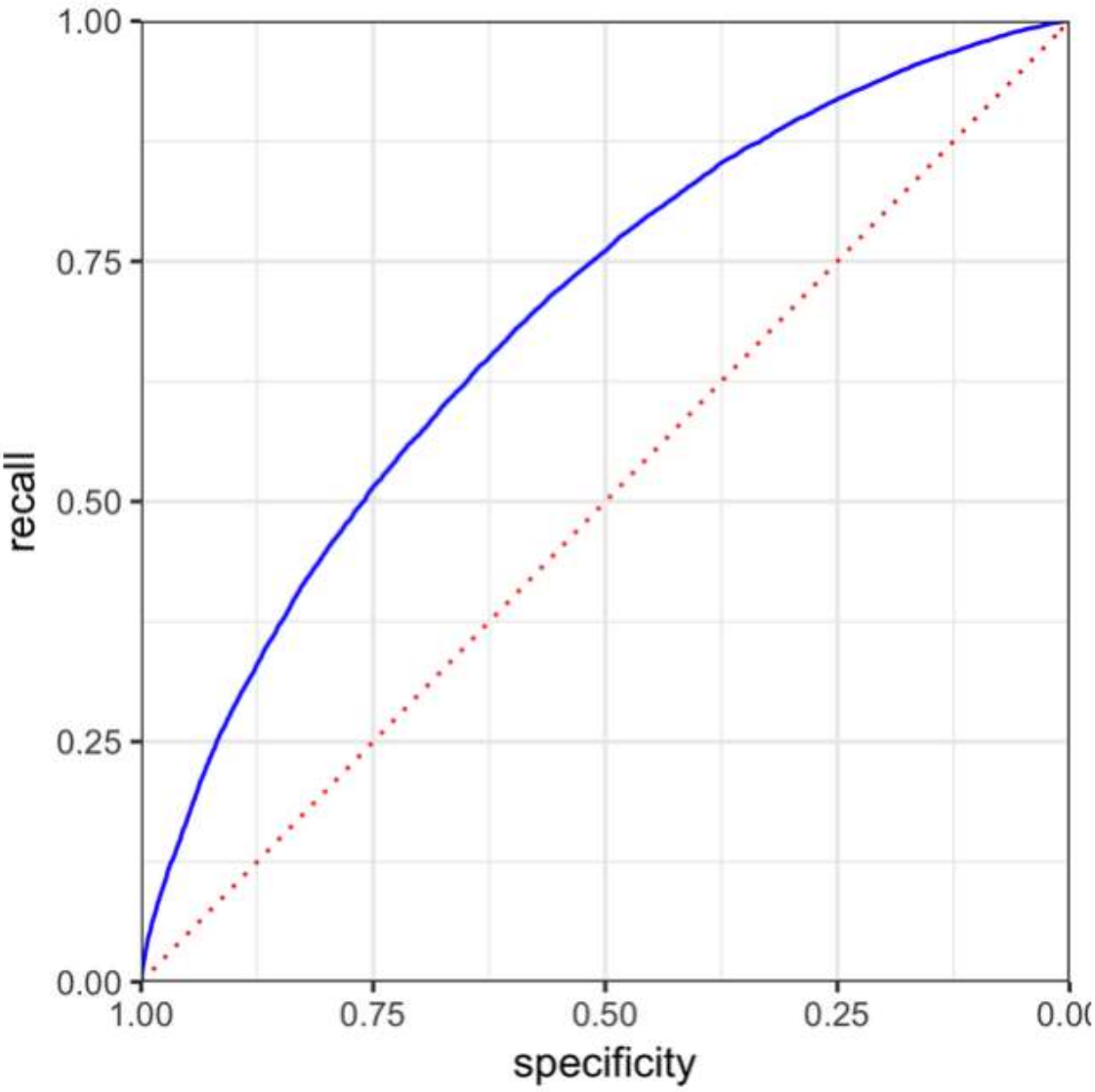
Using the kernel trick, one can get non-linear decision boundaries using algorithms designed originally for linear models. * 1 point

- ◯ True
- ◯ False

Which option is correct about the ROC curve? * 1 point

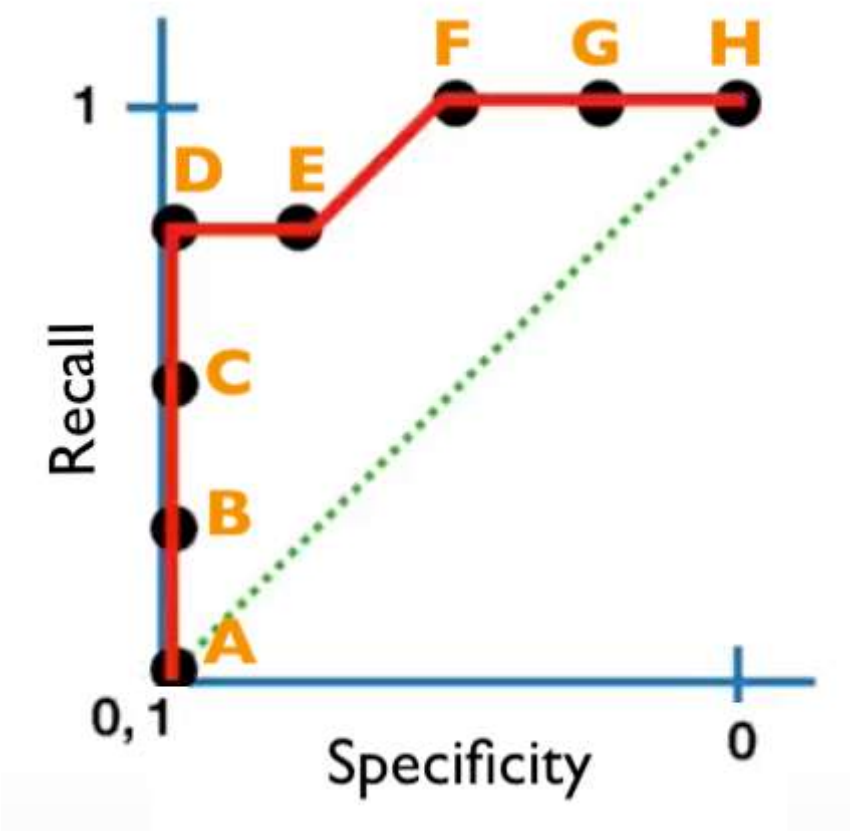
- ◯ You can see that there is a tradeoff between recall and specificity
- ◯ The ROC curve plots sensitivity (recall) on the y-axis against specificity on the x-axis
- ◯ The ideal classifier would classify the 1s without misclassifying more 0s as 1s
- ◯ The ROC curves are especially useful in evaluating data with highly unbalanced outcomes
- ◯ All of the above

The plot below is a ROC curve of a model. What is the recall in this case if we want a classifier with a specificity of at least 50%? * 1 point



- ○ 50%
- ○ 25%
- ○ 100%
- ○ 75%
- ○ 0%

Costco is using a Computer Vision application to detect whether customers wear face masks or not. You as a Data Scientist, got a ROC curve and AUC of the classifier. Our goal is to get as high as possible on successfully detecting customers who do not wear face masks (True Positive) and get as low as possible on False Negative. According to the goal, which threshold on the ROC curve should we take? * 1 point



- A
- B
- C
- D
- E
- F
- G

H

What will be output of following code ? *

2 points

```
import pandas as pd
data = [['Anuj',21],['Rama',25],['Kapil',22]]
df = pd.DataFrame(data,columns=['Name','Age'])
print (df)
```

```
    Name  Age
0   Anuj   21
1   Rama   25
2   Kapil  22
```

Option 1

```
     Name   Age
0   Anuj    21
1   Kapil   22
2  Rama    25
```

Option 2

```
    Name   Age
0   Kapil   22
1   Rama   25
2  Anuj    21
```

Option 3

```
    Name  Age
0   Rama   25
1   Anuj   21
2   Kapil  22
```

Option 4

The execution of following code will result in: * 1 point

```
import pandas as pd
s =pd.Series([1,2,3,4,5],index= ['a','b','c','d','e'])
print(s['f'])
```

- KeyError
- IndexError
- ValueError
- None of the above mentioned

Which rule is followed by the Backpropagation algorithm? * 1 point

- Dynamic Rule
- Chain Rule
- Static Rule
- None

Dataframe object is value mutable. * 1 point

- True
- False

Suppose that you are given two lists: * 1 point

a = [1,2,3]
b = [4,5,6]
Your task is to create a list which contains all the elements of a and b in a single dimension.

Output:
a = [1,2,3,4,5,6]

Which of the following functions will you use?

- a.append(b)
- a.extend(b)
- any one of the above
- none of the above

What will be output for the following code? * 1 point

```
import pandas as pd
import numpy as np
s = pd.Series(np.random.randn(4))
print s.ndim
```

- 0
- 1
- 2
- 3

To import pyplot module we can write * 1 point

- Import matplotlib.pyplot
- Import matplotlib.pyplot as plt
- Import pyplot as plt
- Both (A) and (B)

Which of the following activities are performed by TensorFlow? * 1 point

- Data automation
- Model tracking
- Model retraining
- Performance monitoring
- All of the above

_____ is a regularization technique for neural network models, where randomly selected neurons are ignored during training process. * 1 point

- [ ] Callout
- [ ] Feature Scaling
- [ ] Dropout
- [ ] L2

A _____ requires shape of the input (input_shape) to understand the structure of the input data. * 1 point

- Keras Module
- Keras Model
- Keras layer
- Keras Time

_____ function that propagates errors from nodes of output to input? * 1 point

- Channel Propagation
- gradient
- loss
- dropout
- backpropagation

There is feedback in final stage of backpropagation algorithm. * 1 point

- True
- False

What of the following is true regarding backpropagation rule? * 1 point

- ○ Hidden layers output is not all important, they are only meant for supporting input and output layers
- ○ Actual output is determined by computing the outputs of units for each hidden layer
- ○ It is a feedback neural network
- ○ None of the above

The network that involves backward links from output to the input and hidden layers is * 1 point

- ☐ LSTM
- ☐ RNN
- ☐ Autoencoder
- ☐ Perceptron

Prediction Accuracy of a Neural Network depends on _____ and _____. * 1 point

- ○ Input and Output
- ○ Weight and Bias
- ○ Linear and Logistic Function
- ○ Activation and Threshold

Gradient at a given layer is the sum of all gradients at the previous layers * 1 point

- [ ] True
- [x] False

_____ works best for Image Data. * 1 point

- [ ] AutoEncoders
- [x] Convolution Networks
- [ ] Single Layer Perceptrons
- [ ] Naïve Bayes

The rate at which cost changes with respect to weight or bias is called _____. * 2 points

- [ ] Derivative
- [x] Gradient
- [ ] Rate of Change
- [ ] Loss

Why is the Pooling Layer used in a Convolution Neural Network? * 2 points

- [ ] Padding
- [ ] Dimension Reduction
- [ ] Object Recognition
- [ ] Image Sensing

```
for i in range(1, 10):
    plt.subplot(3, 3, i)
    # Insert ith image with the color map 'grap'
    plt.imshow(trainX[i], cmap=plt.get_cmap('gray'))
plt.show()
```

The output grid would contain? * 1 point

- 6 images
- 7 images
- 9 images
- 10 images

```
model = model_arch()

model.compile(optimizer=Adam(_____=1e-3),
              loss='sparse_categorical_crossentropy',
              metrics=['sparse_categorical_accuracy'])

model.summary()
```

What comes in blank as keyword argument? * 1 point

- lambda
- learning_rate
- epsilon
- value

What is true regarding backpropagation rule? * 1 point

- ○ it is a feedback neural network
- ○ actual output is determined by computing the outputs of units for each hidden layer
- ○ hidden layers output is not all important, they are only meant for supporting input and output layers
- ○ none of the mentioned

By default, Plot() function plots a * 1 point

- ○ Bar chart
- ○ Line chart
- ○ Pie chart
- ○ Horizontal bar chart

Submit Clear form

Never submit passwords through Google Forms.

This form was created inside of University of Central Missouri. Report Abuse

Google Forms