LIVE: Interactive problem solving session

29-Sept-2019, 10 AM

AppliedAICourse.com

1. Which one of them are True statements

a. Gaussian naive Bayes assumes that the feature-values are generated from Gaussian distribution.

b. Linear Regression assumes features are normal distributed.

c. KNN assumes the data is distributed in circular shapes.

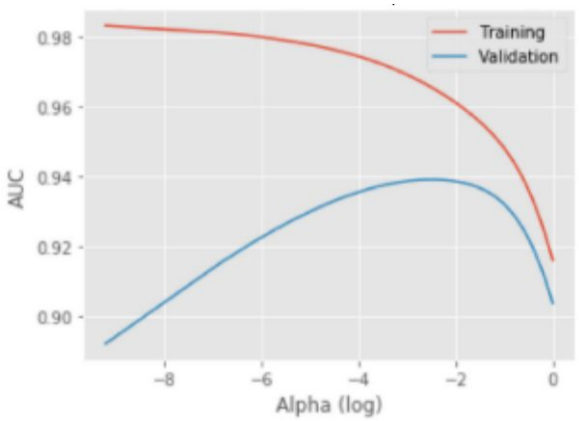
d. In Gaussian Naive Bayes the conditional probability of the features i.e., (P(X_i | y_k) is assumed to be Gaussian

A. a, b B. b, c, d C. c D. d

2. Given an input image of size 28X28 , what will be the output dimensions when 40 convolutional filters of size 5,5 with padding = 1 and stride = 1 are applied ?

a. **25 * 25**
b. **25 * 25 * 40**
c. **26 * 26**
d. **26 * 26 * 40**

3. What can be deduce from this plot ?



1. **The train and test data are coming from different distribution**
2. **It's better to search for hyper parameter within the range (10^{-10} to 10^{-8}), and then choose the best hyper parameter**
3. **It's better to search for hyper parameter within the range (1 to 10^{3}), and then choose the best hyper parameter**
4. **Model is overfitting the training data, we better change the model**
5. **AUC might not be best metric for this problem**
6. **Choose 0.001 as the best hyper parameter**
7. **Choose 1 as the best hyper parameter**

The correct statement(s) is/are _____

4. Choose all the correct statements

a. IDF value of some words will always change if a document was removed from the corpus

b. TF of a word in a document will change after removal of a document from the corpus

c. TF-IDF can be used to provide a vector representation for each document in a corpus

d. TF-IDF can provide a vector representation for each word in a document corpus

e. TF-IDF representation of a word can vary from document to document

5. Consider following statements

i. TSNE is a non-convex optimization problem
j. TSNE can be used as a dimensionality reduction technique.
k. We can produce the similar results as PCA using TSNE with perplexity=number of data-points

A. Only one of them is correct
B. Only two of them are correct
C. All three of them correct
D. None of them are correct

6. Consider you have a dataset with 90% of negative points, and 10% of the positive points, and it is given that predicting a positive point as negative points will cost a lot more than predicting negative points as positive data points. Which of the error metric will be more appropriate for this data set?

a. **Balance the data and then use the Accuracy**
b. **F1 Score**
c. **A custom error metric like, a*|type1 errors| + b*|type2 errors|, a>>b**
d. **AUC**

7. Consider data set with one independent feature(X) and one dependent feature (Y) written as (x_i,y_i) pairs as follows: {(2,0), (1.414, 1.414), (0,2) , (-1.414,1.414) , (-2,0), (-1.414,-1.414), (0,-2), (1.414.1.414) , (10,10), (0,0) }

And assume you are fitting linear regression model on this dataset with Linear_regression.fit(X,Y) to minimize MSE(Mean Squared Error) value on the train data itself. Then, the minimum MSE(Y, Linear_regression.predict(X)) = _____ (round your answer to 3 decimals)

8. Consider the following statements and choose the correct options which are correct.

a. Larger “k” value makes the KNN model overfit.

b. Standardization affects the rank-correlation between variables.

c. Scaling of high dimensional vectors changes the cosine similarity between them.

d. The dimensionality of the bias term will always be a constant in every epoch of mini batch SGD on Linear Regression.

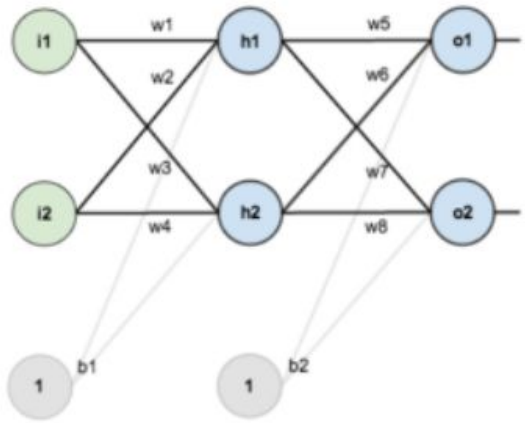
e. Zero correlation implies that the two variables are not related.

A. **a, b, c** **B. d** **C. d, e** **D. all of them are wrong**

9. Consider the neural network architecture. [LENGTHY]

It is given that, the weight vector [w1, w2, w3, w4, w5, w6, w7, w8]=[0.15, 0.20, 0.25, 0.30, 0.40, 0.45, 0.50, 0.55], bias terms [b1, b2]=[1,1] and the true outputs are [o1, o2]=[0.01, 0.99]. The value of w1 is _____ after back propagation for the inputs [i1, i2]=[0.05, 0.1].

It is given that the learning rate α=0.5. The error for each output neuron is computed using the squared error/loss function and sum of errors across all outputs is taken as the total error/loss. We use logistic functions as the activation function in both hidden and output layers.



Solution: https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/

10. Given an input RGB image of size 28X28, what will be the number of tunable parameters in the conv-layer with 64 convolutionals and each filter of size 5 pixels wide and 5 pixels long with padding = 1 & stride =(1,2)?
HINT: don’t miss the bias terms

a.1600
b.1664
c.4800
d.4864

11. Consider a neural network with a single RNN layer and a SoftMax layer for doing a 10-class classification. Let's just say the hidden vector (number of neurons in the RNN-layer) is of length 200 and the input embedded word vector is of length 32. The number of trainable parameters in the network(excluding the bias terms) _____.

a.46600
b.466000
c.48400
d.46400

12. You have a weight matrix W i.e., the set of weights to be learned in a neural network. You want the final learnt matrix to be nearly an orthonormal matrix (W_T. W = I). How can you achieve this while training the network?

1. **Initialize W as an orthogonal matrix**
2. **Add $((P^{-1}WR) - I)$ to the cost function, where I is identity matrix and P & R are invertible matrices.**
3. **It can't be achieved, gradients are not controlled by users.**
4. **None of the above**

13. Which of the parameters listed below are learnable using BPTT(back propagation through time) in a RNN?

1. **Hidden state at each time step**
2. **Weights in the computation of hidden state**
3. **Number of timesteps to unroll for BPTT**
4. **All of them**

A. **2** **B. 4** **C. 2, 1** **D. 2,3**

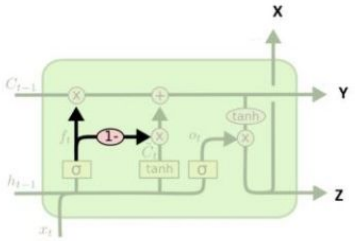
14. Choose the correct statements

a. Two different ROC curves can have same AUC value

b. If $A = V\sum U^T$ i.e. SVD Decomposition, then rank of A = 1 + number of non-zero Rows in $\sum$

C. The difference between the Gini index and Entropy (in a binary classification setting) at each split while building a Decision-Tree is always in the interval [0.5, 1].

15. We are performing a sentiment analysis task on reviews using an LSTM with the following model parameters: Embedded layer is of dimensionality 10,000. The number of words in each review(with padding) is 100. The number of LSTM neurons is 460. Assume that the model uses the below given LSTM architecture:



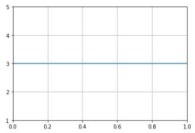
X represents the output at each time step -> O/P(t). Y represents the cell state at each time step -> C(t). Z represents the hidden state at each time step -> h(t). Given the above information what could be the possible sizes of vector X, Y and Z . NOTE: "??" represents any integer

a. **X: (460,??) , Y: (350,??), Z: (460,??)**
b. **X: (350,??) , Y: (460,??), Z: (350,??)**
c. **X: (460,??) , Y: (460,??), Z: (350,??)**
d. **None of the above**

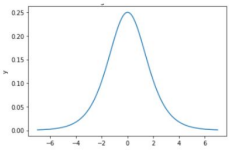
i. ReLU

ii. SoftPlus

A.

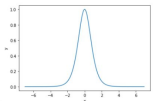


B.



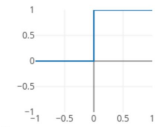
iii. Sigmoid function

C.



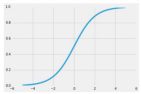
iv. Linear Activation function

D.



v. tanh

E.



16. Consider the above table where the first column represents the activation functions and the right column represents the gradient curves of the activation function, please choose the correct mapping from the below options.

a. **i<->D, ii<->B, iii<->E, iv<->A, v<->C**
b. **i<->E, ii<->C, iii<->B, iv<->A, v<->D**
c. **i<->E, ii<->B, iii<->D, iv<->C, v<->A**
d. **i<->E, ii<->B, iii<->D, iv<->C, v<->A**

17. Number of tunable parameters at max pooling layer in the below model are _____

Input(28, 28, 3) ⇒ Conv(filters=64, stride=(1,1), kernel=(4,4)) ⇒ maxpool(stride=(2,2)) ⇒ flatten() ⇒ dense(10, softmax)

18. In a CNN consisting of max pooling layer, the multiplication factor for a gradient passing through a neuron which is non-maximal in the pooling window is _____? (float value, round it one decimal) (e.g if answer is 1/10 then write 0.1; not 0.10)

19.

```
x = keras.layers.Input(batch_shape = (None, 4096))
hidden = keras.layers.Dense(512, activation = 'relu')(x)
hidden = keras.layers.BatchNormalization()(hidden)
hidden = keras.layers.Dropout(0.5)(hidden)
predictions = keras.layers.Dense(80, activation = 'sigmoid')(hidden)
mlp_model = keras.models.Model(input = [x], output = [predictions])
```

In the above code snippet, the number of tunable parameters at BatchNormalization() layer are:

A. 2 B. 2048 C. 1024 D. 4

20. Consider the data sets with features

1. **Data 1 : (X, Y, X)**
2. **Data 2: (X, Y, X, X)**
3. **Data 3: (X, Y, X, X, 2*X)**
4. **Data 4: (X, Y, X, X, 2*X, X)**

Assume we have applied PCA to reduce the dimensions of all 3 datasets to 2d, which of the

following statement is true

A. PCA(Data 1) = PCA(Data 2) = PCA(Data 3) = PCA(Data 4)

B. PCA(Data 1) ≠ PCA(Data 2) ≠ PCA(Data 3) ≠ PCA(Data 3)
C. PCA(Data 1) ≠ PCA(Data 2) ≠ PCA(Data 3) = PCA(Data 4)
D. PCA(Data 1) = PCA(Data 2) ≠ PCA(Data 3) ≠ PCA(Data 4)
E. PCA(Data 1) = PCA(Data 2) ≠ PCA(Data 3) = PCA(Data 4)