**AI Interview / assignment Solutions**

**Part A: MCQ**

Q1. There are 60 data points which are randomly split into 3 classes of equal size. All partitions are equally likely. A and B are two data points among them. What is the probability that two data points A and B will end up in the same class?

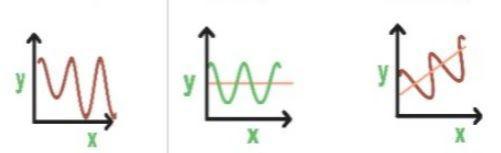
1. 1/3
2. 19/59 **(correct answer)**
3. 18/58
4. 19/60

Q2. A report is stating that a person is suffering from COVID-19 though he is actually not affected by coronavirus. Which of the following case is correct

1. True positive
2. False positive (**correct)**
3. True negative
4. False negative

Q3. Which of the following time series data can be declared as stationary?

(a) **(b)** **(correct)** (c)



Q4.Consider the following training set of m=4 training examples:

* y

1 0.5

2 1

4 2

0 0

Consider the linear regression model hθ(x)=θ0+θ1x. What are the values of θ0 and θ1 that you

​​ ​​ ​​ ​​ ​​

would expect to obtain upon running gradient descent on this model? (Linear regression will be

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| able to fit this data perfectly.) | |  |  |  |
| ​​ | ​​ |  |  |  |
| **(A)** θ0=0.5,θ1=0 **(correct)** | |  |  |  |
| ​​ | ​​ |  |  |  |
| (B) θ0=0.5,θ1=0.5 | |  |  |  |
| ​​ | ​​ |  |  |  |
| (C) θ0=1,θ1=0.5 | |  |  |  |
| ​​ | ​​ |  |  |  |
| (D) θ0=0,θ1=0.5 | |  |  |  |
| ​​ | ​​ |  |  |  |
| (E) θ0=1,θ1=1 | |  |  |  |
| Q5. Let f be some function so that | |  |  |  |
| ​​ ​​ |  |  |  |  |
| f(θ0,θ1) outputs a number. For this problem, | |  |  |  |
| f is some arbitrary/unknown smooth function (not necessarily the | |  |  |  |
| cost function of linear regression, so f may have local optima). | | ​​ | ​​ |  |
|  | ​​ ​​ |  |

Suppose we use gradient descent to try to minimize f(θ0,θ1) as a function of θ0 and θ1. Which of the following statements are true? (Check all that apply)



Statement1: If  and  are initialized at the global minimum, then one iteration will not change their values.

Statement2: Setting the learning rate  to be very small is not harmful, and can only speed up the convergence of gradient descent.



Statement3: No matter how  and  are initialized, so long as  is sufficiently small, we can safely expect gradient descent to converge to the same solution.



Statement4: If the first few iterations of gradient descent cause  to increase rather than decrease, then the most likely cause is that we have set the learning rate  to too large a value.

1. Statement 1 and Statement 2 are correct
2. Statement 2 and Statement 3 are correct
3. Statement 1 and Statement 4 are correct **(correct)**
4. Statement 2 and Statement 4 are correct

Q6. What are the effects of pooling operation on a CNN based model

Statement1: Increase feature / dimensionality

Statement2: Reduce feature / dimensionality

Statement3: Performs down sampling operations

Statement4: Performs up sampling operations

1. Statement 1 and Statement 3 are correct
2. Statement 1 and Statement 4 are correct
3. Statement 2 and Statement 4 are correct **(correct)**
4. Statement 2 and Statement 3 are correct

Q7. Role of activation function

Statement1:Introduce non-linearity

Statement2: Delivers output based on linear combinations of inputs

Statement3: To learn complex patterns in data

1. Statement 1 and Statement 3 are correct **(correct)**
2. Statement 2 and Statement 3 are correct
3. Statement 1 and Statement 2 are correct
4. Only statement 1 is correct

Q8. Which of the following is a widely used and effective machine learning algorithm based on the idea of bagging?

1. Decision tree
2. Regression
3. Classification
4. Random forest **(correct)**

Q9. How do you handle missing or corrupted data in a dataset?

1. Drop missing rows or columns
2. Replace missing values with mean/median/mode
3. Assign a unique category to missing values
4. All of the above **(Correct)**

Q10. Why is second order differencing in time series needed?

1. To remove stationarity
2. To find the maxima or minima at the local point
3. Both A and B **(correct)**
4. None of the above

Q11. When performing regression or classification, which of the following is the correct way to preprocess the data?

1. Normalize the data -> PCA -> training **(correct)**
2. PCA -> normalize PCA output -> training
3. Normalize the data -> PCA -> normalize PCA output -> training
4. None

Q12. Which of the following is an example of feature extraction?

1. Constructing bag of words vector from an email
2. Applying PCA projects to a large high-dimensional data
3. Removing stopwords in a sentence
4. All of the above **(correct)**

Q13. Which of the following is true about Naive Bayes ?

(A) Assumes that all the features in a dataset are equally important

1. Assumes that all the features in a dataset are independent
2. Both A and B **(correct)**
3. None of the above options

Q14. Which of the following statements about regularization is not correct?

1. Using too large a value of lambda can cause your hypothesis to underfit the data
2. Using too large a value of lambda can cause your hypothesis to overfit the data
3. Using a very large value of lambda cannot hurt the performance of your hypothesis
4. None of the above **(correct)**

Q15. How can you prevent a clustering algorithm from getting stuck in bad local optima?

1. Set the same seed value for each run
2. Use multiple radom initializations **(correct)**
3. Both A and B
4. None of the above

Q16. In which of the following cases will K-means clustering fail to give good results? 1) Data points with outliers 2) Data points with different densities 3) Data points with nonconvex shapes

1. 1 and 2
2. 2 and 3
3. 1,2 and 3 **(correct)**
4. 1 and 3

Q17. Which of the following is a reasonable way to select the number of principal components "k"?

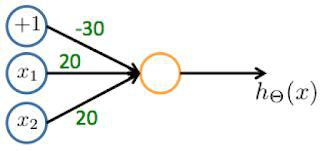
1. Choose k to be the smallest value so that at least 99% of the varinace is retained **(Correct)**
2. Choose k to be 99% of m (k = 0.99\*m, rounded to the nearest integer)
3. Choose k to be the largest value so that 99% of the variance is retained
4. Use the elbow method.

Q18. You run gradient descent for 15 iterations with a=0.3 and compute J(theta) after each iteration. You find that the value of J(Theta) decreases quickly and then levels off. Based on this, which of the following conclusions seems most plausible?

1. Rather than using the current value of a, use a larger value of a (say a=1.0)
2. Rather than using the current value of a, use a smaller value of a (say a=0.1)
3. a=0.3 is an effective choice of learning rate **(correct)**
4. None of the above

Q19. Consider​ the following neural network which takes two binary-valued inputs



 and outputs . Which of the following logical functions does it (approximately) compute

1. NAND (meaning “NOT AND”)
2. OR
3. XOR (exclusive OR)
4. AND **(correct)**

Q20. 3 people Alice, Bob and Charlie. Alice can shoot with probability 0.2, Bob with 0.5 and Charlie with 1. What is the probability of Bob surviving if they all were shooting in cyclic order.

1. 1/10
2. 13/30 **(correct)**
3. 8/10
4. None

Q21. Histogram equalization is not used for

1. Image enhancement
2. Image translation
3. Contrast adjustment **(correct)**
4. All

Q22. Which filter is advisable to remove salt and pepper noise

1. Gaussian filter
2. Median filter (**correct)**
3. Averaging filter
4. Contraharmonic mean filter

Q23. Which of the following color space doesn’t have a channel corresponding to luminosity of image

1. RGB
2. LAB
3. HSV **(correct)**
4. all

Q24. What is padding in image processing

1. The process of removing noise from an image
2. The process of modifying pixel intensity value of an image
3. The process of adding layers of zeros to an image **correct**
4. All

Q25. Second derivative of approximation says that value at the end of ramp must be

1. Zero
2. Nonzero **correct**
3. Positive
4. negative

**Part B: Fill in the blanks with correct option**

Q1. Information gain leads to ​……………… **decrease** ……….. ​(increase/decrease) in Entropy?

Q2. A decision tree is a ​…… **nonlinea**r ……………….​(linear/nonlinear) algorithm for

……………… **classificationn** …….. ​(classificationn /regression) which works by trying to

………… **decrease** …………...​(increase/decrease) entropy.

Q3. Logistic regression is used for**​… classification** …………………… ​(regression/classification)

Q4. K-Nearest Neighbours take more time for ​………… **testing** …………… ​(training/testing) less

time for ​………… **training** …………… ​(training/testing)

Q5. A table has 3 columns. First column has values ranging from -1 to +1. The second column has values ranging from 1 to 100. The third column has values ranging from 0 to +256. What should be the idle range of output **​………-1 to 256**……………… ( ​from 0 to +1, from 0 to 256, from -1 to 256)