

# Amazon ML Summer School

## Previous Year Questions:

**Part A** - Basic ML concepts and math fundamentals on topics such as probability, statistics and linear algebra

1. Two squares are chosen at random on a chessboard. What is the probability that they have a side in common?

a)  $8/13$   
b)  $17/18$   
c)  $5/13$   
d)  $1/18$

2. The police plans to enforce speed limits during morning rush hour on four different routes into the city. The traps on routes A, B, C, and D are operated 40%, 30%, 20%, and 30% of the time, respectively. Biff always speeds to work, and he has probability 0.2, 0.1, 0.5, and 0.2 of using those routes. What is the probability that he'll get a ticket on any one morning?

Select an option:

a) 0.27  
b) 0.93  
c) 0.73  
d) 0.07

3. For random variables X and Y, we have  $\text{Var}(X)=1$ ,  $\text{Var}(Y)=4$ , and  $\text{Var}(2X-3Y)=34$ , then the correlation between X and Y is:

a)  $1/2$   
b)  $1/4$   
c)  $1/3$   
d) None of the above

4. You have two coins. One of them is fair and comes up heads with probability  $1/2$  and the other is biased and comes up heads with probability  $3/4$ . You randomly pick a coin and flip it twice. You get heads both times. What is the probability that you picked the fair coin?

a)  $13/32$   
b)  $4/13$

- c) 2/13
- d) 19/32

5. Which of the following statements is true about PCA?

- (i) We must standardize the data before applying
- (ii) We should select the principal components which explain the highest variance
- (iii) We should select the principal components which explain the lowest variance
- (iv) We can use PCA for visualizing the data in lower dimensions

- a) (i), (ii) and (iv).
- b) (ii) and (iv)
- c) iii) and (iv)
- d) (i) and (iii)

6. The number of solutions for each of the following system of equations is

$$\begin{aligned} 2x+y-z &= 4 \\ x-2y+z &= -2 \\ -x+2y-z &= 2 \end{aligned}$$

Select an option

- a) 0
- b) 1
- c) Inf
- d) Can't be determined

7. A 3-input neuron has weights 1, 4 and 3. The transfer function is linear with the constant of proportionality being equal to 3. The inputs are 4, 8 and 5 respectively. What will be the output?

Select an option

- a) 51
- b) 153
- c) 54
- d) 160

8. In medical screening, it is sometimes more important to avoid false negatives than false positives. (i. e., we want to be sure to detect a disease if present). How would this affect the p-value we might use for a medical test?

Select an option

- a) we would choose a higher p-value
- b) we would choose a lower p-value

- c) we would choose the same p-value
- d) false negatives do not depend on p-value

9. When classifying data with logistic classification, what is the upper bound of the likelihood in the maximum likelihood method? Is this value attainable?

- a) 1, Yes
- b) e, No
- c) 1, No
- d) 0, Yes

10. Let  $M$  and  $S^2$  be the mean and variance of a random sample of size  $n > 1$  respectively from a normal population with an unknown mean  $\mu$  and unknown finite variance  $\sigma^2 > 0$ . Consider the following statements:

- 1)  $S^2$  is an unbiased estimator of  $\sigma^2$ , and  $S$  is an unbiased estimator of  $\sigma$ .
- 2)  $((n-1)/n)M$  is an unbiased estimator of  $\mu$ , and  $((n-1)/n)S^2$  is an unbiased estimator of  $\sigma^2$ .

Which of the following statements is/are true?

- a) 1 only
- b) 2 only
- c) Both 1 and 2
- d) Neither 1 nor 2

11. If  $\text{rank}(A)$  is 2 and  $\text{rank}(AB)$  is 3, then

Select an option

- a)  $\text{rank}(B) = 3$
- b)  $\text{rank}(B) \leq 3$
- c)  $\text{rank}(B) \geq 3$
- d) data insufficient

12. The following set of equations has infinite solutions, if  $a = ?$

$$x + y + z = 1$$

$$ax - ay + 3z = 5$$

$$5x - 3y + az = 6$$

Select an option

- a) -3
- b) 3
- c) -4
- d) 4

13. Suppose you have a dataset with  $m=50$  examples and  $n=200000$  features for each example. You want to use multivariate linear regression to fit the parameters  $\Theta$  to our data. Should you prefer gradient descent or the normal equation?

- a) Gradient descent, since  $\text{inverse}(X^T X)$  will be very slow to compute in the normal equation.
- b) Gradient descent, since it will always converge to the optimal  $\Theta$ .
- c) The normal equation, since it provides an efficient way to directly find the solution.
- d) The normal equation, since gradient descent might be unable to find the optimal  $\Theta$

14. The eigen values of a  $4 \times 4$  square matrix having 0's as the diagonal elements and 1's on the off diagonal elements is

- a) 2, -2, 0, 0
- b) 1, -1, 1, -1
- c) 3, -1, -1, -1
- d) 4, 0, 0, 0

15. Let  $A$  be the  $2 \times 2$  matrix with elements  $a_{11} = a_{12} = a_{21} = +1$  and  $a_{22} = -1$ . Then the eigenvalues of the matrix  $A^{19}$  are:

Select an option

- a) 1024 and -1024
- b)  $1024\sqrt{2}$  and  $-1024\sqrt{2}$
- c)  $4\sqrt{2}$  and  $-4\sqrt{2}$
- d)  $512\sqrt{2}$  and  $-512\sqrt{2}$

16. Given  $y=x^x$ , what is  $dy/dx$  at  $x = 2$ ?

Select an option

- a)  $4(1+\log 2)$
- b)  $4(2+\log 2)$
- c) 4
- d) 8

17. If the characteristic equation of a matrix  $A$  is  $t^2 - t - 1 = 0$  then

Select an option

- a)  $A^{-1}$  does not exist
- b)  $A^{-1}$  exists but cannot be determined from the data
- c)  $A^{-1} = A - I$
- d)  $A^{-1} = A + I$

18. Assume we are trying to fit the data coming from a cubic function which is corrupted by standard Gaussian noise, using a linear and 5th degree polynomial. Let M1 and M5 denote the models corresponding to the linear and 5th degree polynomial. Then,

Select an option

- a)  $\text{Bias}(M1) \leq \text{Bias}(M5)$ ,  $\text{Variance}(M1) \leq \text{Variance}(M5)$ .
- b)  $\text{Bias}(M1) \geq \text{Bias}(M5)$ ,  $\text{Variance}(M1) \leq \text{Variance}(M5)$ .
- c)  $\text{Bias}(M1) \leq \text{Bias}(M5)$ ,  $\text{Variance}(M1) \geq \text{Variance}(M5)$ .
- d)  $\text{Bias}(M1) \geq \text{Bias}(M5)$ ,  $\text{Variance}(M1) \geq \text{Variance}(M5)$ .

## Part B - Programming Questions

1.

There are three robots named Ray, Ben and Kevin. Initially Ray has a string S of length N. while the other two robots have empty strings. We can make either of the following moves:

Move 1: Remove the first character from Ray's string and append it to Ben's string.

Move 2: Remove the last character from Ben's string and append it to Kevin's string.

You must perform either of the two moves mentioned above in such a way that the strings left with Ray and Ben are empty and the string left with Kevin is lexicographically the smallest. Your task is to return this lexicographically smallest string that Kevin has after completing this activity.

Note: For any two given strings, a string is said to be lexicographically smaller than the other if it comes before the other string in the dictionary

### 2. Mean, Median and Mode

Given 'n' integers, find their mean, median and mode.

You are required to fill in a function that takes as inputs an integer 'input1' ( $1 \leq \text{input1} \leq 1000$ ) and an integer array input2[], containing 'input1' integers, and returns output1 as the mean, output2 as the median and output3 as the mode. The mean and median must be correct to six decimal places.

Mean:

Defined as the average of all numbers in the array

Median:

Defined as the middle element of the array.

If  $n$  is even, the median is the average of the two middle elements.

If  $n$  is odd, the median is the middle element of the array

Note: For finding the median, elements in the array have to be listed in numerical order from smallest to largest

Mode:

Defined as the number in the array with the highest frequency.

If many numbers have the same highest frequency, then the mode is calculated by breaking ties in favour of the smallest of the numbers.

Input Specification:

input1: Integer in the range of  $1 \leq \text{input1} \leq 1000$ , denoting length of input array.

input2: Integer input array

Output Specification:

Return output1 (double) variable as the mean

Return output2 (double) variable as the median.

Return output3 (int) variable as the mode.

Example 1:

input1: 3

input2: {1,2,3}

Output: 2.000000, 2.000000, 1

Example 2:

input1: 5

input2: (41,18467,6334,26500,19169)

Output: 14102.200000, 18467.000000, 41

### **3. Minimum Cost to Reach the Destination**

There are  $N$  cities in a country. George is initially at the airport in city 1 and he wants to reach city  $N$ . For any city  $i$ , there is either a flight to city  $(i+1)$  or to  $(i+3)$  if it exists.

You have been given an array  $A$  with the costs of flight tickets for  $N$  cities. To find the cost of a flight ticket between any two cities  $i$  and  $j$ , you take the absolute difference of the costs of those cities in the array  $A$ . You can use the formula  $|a| = |\text{Cost}[i] - \text{Cost}[j]|$  to calculate the cost of a flight ticket, where  $|a|$  represents the absolute value of  $a$ .

Your task is to find and return the minimum possible cost of flight ticket required to reach the city N.

Note:

- The number of cities is always greater than 3.
- Assume 1 based indexing.

Input Specification:

input1: An integer value N, representing the number of cities.

input2: An integer array A, representing the cost of tickets to reach the  $i^{\text{th}}$  a city.

Output Specification:

Return the minimum possible cost of flight ticket required to reach the city N.

Example 1:

input1: 4

input2: (1,4,5,2)

Output: 1

Explanation:

George takes a flight in the below optimal manner:

- From city 1 to city 4, as city 4 is the third next city to city 1 so the cost will be  $|1-2|=1$ .

Hence, 1 is returned as the output.

Example 2:

input1: 6

input2 (4,12,13,18,10,12)

Output: 10

Explanation:

George takes a flight in the below optimal manner: From city 1 to city 2, the cost will be  $|4-12|=8$

- From city 2 to city 3, the cost will be  $|12-13|=1$

. From city 3 to city 6, the cost will be  $|13-12|=1$

Therefore, the total cost is  $8 + 1 + 1 = 10$ . Hence, 10 is returned as the output.

## Selection Test Format:



**The selection test will have two parts - Part A will consist of 20 MCQ on basic ML concepts and math fundamentals on topics such as probability, statistics and linear algebra. Part B will consist of two Programming questions. The overall test duration will be 75 minutes.**

## **FAQs**

1. What is the eligibility criteria for enrolling in ML Summer School?

Engineering students enrolled in Bachelor's/Master's/ PhD degree from any recognized institute of India and are expected to graduate in 2025 or 2026 are eligible to enroll in ML Summer School.

2. What is the selection criteria?

The selection test will have two parts – Part A will consist of 20 MCQ on basic ML concepts and math fundamentals on topics such as probability, statistics and linear algebra. Part B will consist of two Programming questions. The overall test duration will be 75 minutes.

3. Are Amazon's past, current and incoming interns eligible to participate in this program?

Yes, Amazon's past, current and incoming interns are eligible to participate in this program.

4. What is the program schedule for ML Summer School?

The program will be spread over eight days July 6, 7, 13, 14, 20, 21, 27 and 28. Detailed program structure will be shared with selected students.

5. In which mode is the program being hosted in 2024?

Amazon ML Summer School will be held virtually in 2024.

6. Is there a registration fee or course fee?

No, there are no fees associated with this program.