

# DAY 1 INTERVIEW QUESTIONS - Hypothesis Space, Cost Function

Q1) What is hypothesis space in machine learning?

Ans - Space means all possibilities. Hypothesis is the trained model. Therefore hypothesis space is the set of all possible model for the given training dataset. A specific hypothesis is defined by the parameters that was selected by the cost function. So in turn, the hypothesis space is defined by all the possible parameters that the function can assume.

Q2) Give an example of Hypothesis Space.

Ans - Consider a book seller trying to infer which article user read based on keyword present in the article.

Article	Crime	Music	Academic
$a_1$	true	false	false
$a_2$	true	true	false
$a_3$	false	true	true
$a_4$	true	false	true

Hypothesis space  $H$  could be all Boolean combination of input features.

$2^3 = 8$  possibilities, one for each combination of values for features

Q3) What is Cost Function?

Ans - It is a function that measure the performance of a Machine Learning model for given data. Cost function quantifies the error between predicted value and expected value and present it in the form of single real number.

Purpose of Cost Function is to be either -

- i) Minimized  $\rightarrow$  then returned value is usually cost, loss or error.
- ii) Maximized  $\rightarrow$  then the value it yield is named as reward.

Q4) Give an example of Cost Function?

Ans - In regression, model predicts an output value for each test set. Cost function for regression can be calculated on distance-based error.

$$\text{Distance based error} = \text{Actual} - \text{predicted} = y - y'$$

Types of Distance based error - i) Mean Absolute Error (MAE)

ii) Mean Square Error (MSE) iii) Root Mean Square Error (RMSE)

iv) Mean Absolute Error percentage (MAPE) v) Mean Percentage Error (MPE)

In classification, we use cross entropy for classification problem.



Q5) What is hypothesis testing?

Ans - In hypothesis testing, we evaluate 2 or more exclusive statements on a population using sample data.

There are 2 kinds of hypothesis - i) Null hypothesis ii) Alternate hypothesis

i) Null hypothesis  $\rightarrow$  It is a hypothesis that says there is no statistical significance between two variables in the hypothesis.

ii) Alternate hypothesis  $\rightarrow$  There is a statistically significant relationship between variables. It is opposite of Null hypothesis.

Example - Senior citizen tends to invest in fixed deposit. So 2 hypotheses are  
Null hypothesis  $\rightarrow$  Always tends to be neutral.

There is no relationship between Age of the Customer (especially above 60) & investment in FD.

Alternate hypothesis  $\rightarrow$  There is a relationship between Age of Customer (especially above 60) and investment in FD.

Q6) On what basis, we will reject Null and/or Alternate hypothesis?

Ans - It is based on level of significance and Confidence Interval.

i) Level of Significance  $\rightarrow$  Probability with which we will reject a NULL HYPOTHESIS when it is true is the level of significance.  
- It is denoted by  $\alpha$  (alpha)

ii) Confidence Interval  $\rightarrow$  The probability with which we will accept a NULL HYPOTHESIS when it is true is confidence interval. It is denoted by  $1-\alpha$  (because probability sums to 1)

Level of Significance ( $\alpha$ )	Confidence Interval ( $1-\alpha$ )
0.01 1%	$\rightarrow$ 99% 0.99
0.05 5%	$\rightarrow$ 95% 0.95
0.1 10%	$\rightarrow$ 90% 0.90

iii) P value  $>$  level of significance, NULL HYPOTHESIS is accepted.  
P value  $<$  level of significance, NULL HYPOTHESIS is rejected.