



Optional

Subjective Questions - I

The following questions related to logistic regression are asked quite frequently in interviews.



Q1. What is a logistic function? What is the range of values of a logistic function?

The logistic function is as defined below:

$$f(z) = \frac{1}{(1+e^{-z})}$$

The values of a logistic function will range from 0 to 1. The values of Z will vary from $-\infty$ to $+\infty$.



Q2. Why is logistic regression very popular/widely used?

Logistic regression is famous because it can convert the values of logits (log-odds), which can range from $-\infty$ to $+\infty$ to a range between 0 and 1. As logistic functions output the probability of occurrence of an event, it can be applied to many real-life scenarios. It is for this reason that the logistic regression model is very popular. Another reason why logistic is in comparison to linear regression is that it is able to handle the categorical variables.



Q3. What is the formula for the logistic regression function?

In general, the formula for logistic regression is given by the following expression:



Q4. How can the probability of a logistic regression model be expressed as a conditional

probability?

The conditional probability can be given as:

$$P(\text{Discrete value of target variable} | X_1, X_2, X_3 \dots X_k)$$

It is the probability of the target variable to take up a discrete value (either 0 or 1 in case of binary classification problems) when the values of independent variables are given. For example, the probability an employee will attrite (target variable) given his attributes such as his age, salary, KRA's, etc.



Q5. What are odds?

It is the ratio of the probability of an event occurring to the probability of the event not occurring. For example, let's assume that the probability of winning a lottery is 0.01.

Then, the probability of not winning is $1 - 0.01 = 0.99$.

Now, as per the definition,

The odds of winning the lottery = (Probability of winning)/(Probability of not winning)

The odds of winning the lottery = $0.01/0.99$

Hence, the odds of winning the lottery is 1 to 99, and the odds of not winning the lottery is 99 to 1



Q6. Why can't linear regression be used in place of logistic regression for binary classification?



regression is different. Linear regression assumes that error terms are normally distributed. In the case of binary classification, this assumption does not hold true.

Model output: In linear regression, the output is continuous. In the case of binary classification, an output of a continuous value does not make sense. For binary classification problems, linear regression may predict values that can go beyond 0 and 1. If we want the output in the form of probabilities, which can be mapped to two different classes, then its range should be restricted to 0 and 1. As the logistic regression model can output probabilities with logistic/sigmoid function, it is preferred over linear regression.

Variance of Residual errors: Linear regression assumes that the variance of random errors is constant. This assumption is also violated in the case of logistic regression



Q7. What is the likelihood function?

The likelihood function is the joint probability of observing the data. For example, let's assume that a coin is tossed 100 times and you want to know the probability of getting 60 heads from the tosses. This example follows the binomial distribution formula.

p = Probability of heads from a single coin toss

n = 100 (the number of coin tosses)

x = 60 (the number of heads – success)

n - x = 40 (the number of tails)

Pr (X=60 | n = 100, p)



This can be reframed as follows:

$$\Pr(X=60|n=100, p) = c \times p^{60} \times (1 - p)^{100-60}$$

c = constant

p = unknown parameter

The likelihood function gives the probability of observing the results using unknown parameters.



Q8. What are the outputs of the logistic model and the logistic function?

The logistic model outputs the logits, i.e. log odds; and the logistic function outputs the probabilities.

$$\text{Logistic model} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$$

The output of the same will be logits.

$$\text{Logistic function} = f(z) = \frac{1}{(1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n)})}$$

The output, in this case, will be the probabilities



Q9. How to interpret the results of a logistic regression model? Or, what are the meanings of the different betas in a logistic regression model?

β_0 is the baseline in a logistic regression model. It is the log odds for an instance when all the attributes $(X_1, X_2, X_3, \dots, X_n)$ are zero. In practical scenarios, the



particular attribute by keeping all other attributes fixed or unchanged (control variables).



Q10. What is odds ratio?

Odds ratio is the ratio of odds between two groups. For example, let's assume that you are trying to ascertain the effectiveness of a medicine. You administered this medicine to the 'intervention' group and a placebo to the 'control' group.

$$\text{Odds Ratio (OR)} = \frac{\text{Odds of the Intervention Group}}{\text{Odds of the Control Group}}$$

Interpretation

- If odds ratio = 1, then there is no difference between the intervention group and the control group.
- If the odds ratio is greater than 1, then the intervention group is better than the control group.
- If the odds ratio is less than 1, then the control group is better than the intervention group.




Q11. What is the formula for calculating odds ratio?


The formula can be given as:


$$OR_{X_1, X_0} = e^{\sum_{i=1}^k \beta_i (X_{1i} - X_{0i})}$$


In the formula above, X_1 and X_0 stand for two different groups for which the odds ratio needs to be calculated. X_{1i} stands for the instance ' i ' in group X_1 . X_{0i} stands for the instance ' i ' in group X_0 . β_0 stands for the coefficient of the logistic regression model. Note that the baseline is not included in this formula.

PG Diploma in
Machine
Learning and AI
- December
2020

 **Learn**

 **Live**

 **Jobs**

 **Discussions**

