**Skill test Questions and Answers**

**1) True-False: Is Logistic regression a supervised machine learning algorithm?**

A) TRUE  
B) FALSE

**Solution:** A

True, Logistic regression is a supervised learning algorithm because it uses true labels for training. Supervised learning algorithm should have input variables (x) and an target variable (Y) when you train the model .

**2) True-False: Is Logistic regression mainly used for Regression?**

A) TRUE  
B) FALSE

**Solution: B**

Logistic regression is a classification algorithm, don’t confuse with the name regression.

**3) True-False: Is it possible to design a logistic regression algorithm using a Neural Network Algorithm?**

A) TRUE  
B) FALSE

**Solution: A**

True, Neural network is a is a *universal* approximator so it can implement linear regression algorithm.

**4) True-False: Is it possible to apply a logistic regression algorithm on a 3-class Classification problem?**

A) TRUE  
B) FALSE

**Solution: A**

Yes, we can apply logistic regression on 3 classification problem, We can use One Vs all method for 3 class classification in logistic regression.

**5) Which of the following methods do we use to best fit the data in Logistic Regression?**

A) Least Square Error  
B) Maximum Likelihood  
C) Jaccard distance  
D) Both A and B

**Solution: B**

Logistic regression uses maximum likely hood estimate for training a logistic regression.

**6) Which of the following evaluation metrics can not be applied in case of logistic regression output to compare with target?**

A) AUC-ROC  
B) Accuracy  
C) Logloss  
D) Mean-Squared-Error

**Solution: D**

Since, Logistic Regression is a classification algorithm so it’s output can not be real time value so mean squared error can not use for evaluating it

**7) One of the very good methods to analyze the performance of Logistic Regression is AIC, which is similar to R-Squared in Linear Regression. Which of the following is true about AIC?**

A) We prefer a model with minimum AIC value  
B) We prefer a model with maximum AIC value  
C) Both but depend on the situation  
D) None of these

**Solution: A**

We select the best model in logistic regression which can least AIC. For more information refer this source: <http://www4.ncsu.edu/~shu3/Presentation/AIC.pdf>

**8) [True-False] Standardisation of features is required before training a Logistic Regression.**

A) TRUE  
B) FALSE

**Solution: B**

Standardization isn’t required for logistic regression. The main goal of standardizing features is to help convergence of the technique used for optimization.

**9) Which of the following algorithms do we use for Variable Selection?**

A) LASSO  
B) Ridge  
C) Both  
D) None of these

**Solution: A**

In case of lasso we apply a absolute penality, after increasing the penality in lasso some of the coefficient of variables may become zero.

**Context: 10-11**

Consider a following model for logistic regression: P (y =1|x, w)= g(w0 + w1x)  
where g(z) is the logistic function.

In the above equation the P (y =1|x; w) , viewed as a function of x, that we can get by changing the parameters w.

**10) What would be the range of p in such case?**

A) (0, inf)  
B) (-inf, 0 )  
C) (0, 1)  
D) (-inf, inf)

**Solution: C**

For values of *x* in the range of  real number from −∞ to +∞ Logistic function will give the output between (0,1)

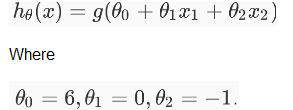
**11) In above question what do you think which function would make p between (0,1)?**  
  
A) logistic function  
B) Log likelihood function  
C) Mixture of both  
D) None of them

**Solution: A**

Explanation is same as question number 10

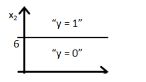
**Context: 12-13**

Suppose you train a logistic regression classifier and your hypothesis function H is

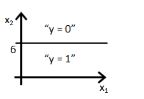


**12) Which of the following figure will represent the decision boundary as given by above classifier?**

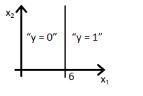
A)



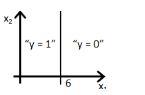
B)



C)



D)

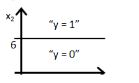


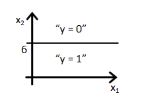
**Solution: B**

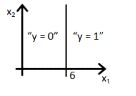
Option B would be the right answer. Since our line will be represented by y = g(-6+x2) which is shown in the option A and option B. But option B is the right answer because when you put the value x2 = 6 in the equation then y = g(0) you will get that means y= 0.5 will be on the line, if you increase the value of x2 greater then 6 you will get negative values so output will be the region y =0.

**13) If you replace coefficient of x1 with x2 what would be the output figure?**

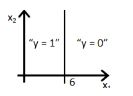
A)

  
B)

  
C)



D)



**Solution: D**

Same explanation as in previous question.

**14) Suppose you have been given a fair coin and you want to find out the odds of getting heads. Which of the following option is true for such a case?**

A) odds will be 0  
B) odds will be 0.5  
C) odds will be 1  
D) None of these

**Solution: C**

Odds are defined as the ratio of the probability of success and the probability of failure. So in case of fair coin probability of success is 1/2 and the probability of failure is 1/2 so odd would be 1

**15) The logit function(given as l(x)) is the log of odds function. What could be the range of logit function in the domain x=[0,1]?**

A) (– ∞ , ∞)  
B) (0,1)  
C) (0, ∞)  
D) (- ∞, 0)

**Solution: A**

For our purposes, the odds function has the advantage of transforming the probability function, which has values from 0 to 1, into an equivalent function with values between 0 and ∞. When we take the natural log of the odds function, we get a range of values from -∞ to ∞*.*

**16) Which of the following option is true?**

A) Linear Regression errors values has to be normally distributed but in case of Logistic Regression it is not the case  
B) Logistic Regression errors values has to be normally distributed but in case of Linear Regression it is not the case  
C) Both Linear Regression and Logistic Regression error values have to be normally distributed  
D) Both Linear Regression and Logistic Regression error values have not to be normally distributed

**Solution:A**

Only A is true. Refer this tutorial <https://czep.net/stat/mlelr.pdf>

**17) Which of the following is true regarding the logistic function for any value “x”?**

**Note:**  
Logistic(x): is a logistic function of any number “x”

Logit(x): is a logit function of any number “x”

Logit\_inv(x): is a inverse logit function of any number “x”

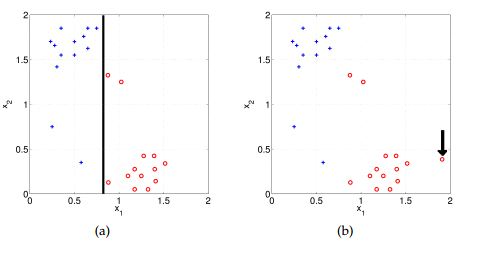
A) Logistic(x) = Logit(x)  
B) Logistic(x) = Logit\_inv(x)  
C) Logit\_inv(x) = Logit(x)  
D) None of these

**Solution: B**

Refer this link for the solution: <https://en.wikipedia.org/wiki/Logit>

**18) How will the bias change on using high(infinite) regularisation?**

Suppose you have given the two scatter plot “a” and “b” for two classes( blue for positive and red for negative class). In scatter plot “a”, you correctly classified all data points using logistic regression ( black line is a decision boundary).



A) Bias will be high  
B) Bias will be low  
C) Can’t say  
D) None of these

**Solution: A**

Model will become very simple so bias will be very high.

**19) Suppose, You applied a Logistic Regression model on a given data and got a training accuracy X and testing accuracy Y. Now, you want to add a few new features in the same data. Select the option(s) which is/are correct in such a case.**

**Note: Consider remaining parameters are same.**

A) Training accuracy increases  
B) Training accuracy increases or remains the same  
C) Testing accuracy decreases  
D) Testing accuracy increases or remains the same

**Solution: A and D**

Adding more features to model will increase the training accuracy because model has to consider more data to fit the logistic regression. But testing accuracy increases if feature is found to be significant

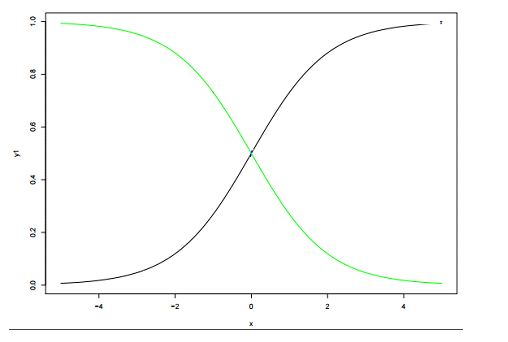
**20) Choose which of the following options is true regarding One-Vs-All method in Logistic Regression.**

A) We need to fit n models in n-class classification problem  
B) We need to fit n-1 models to classify into n classes  
C) We need to fit only 1 model to classify into n classes  
D) None of these

**Solution: A**

If there are n classes, then n separate logistic regression has to fit, where the probability of each category is predicted over the rest of the categories combined.

**21) Below are two different logistic models with different values for β0 and β1.**

**Which of the following statement(s) is true about β0 and β1 values of two logistics models (Green, Black)?**

**Note: consider Y = β0 + β1\*X. Here, β0 is intercept and β1 is coefficient.**

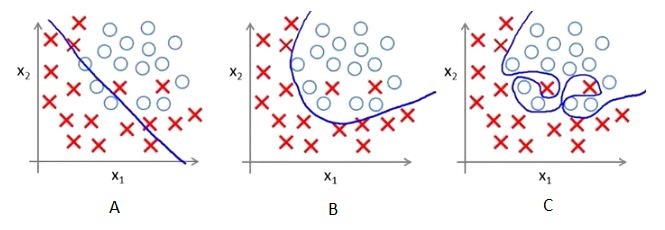
A) β1 for Green is greater than Black  
B) β1 for Green is lower than Black  
C) β1 for both models is same  
D) Can’t Say

**Solution: B**

β0 and β1: β0 = 0, β1 = 1 is in X1 color(black) and β0 = 0, β1 = −1 is in X4 color (green)

**Context 22-24**

Below are the three scatter plot(A,B,C left to right) and hand drawn decision boundaries for logistic regression.



**22) Which of the following above figure shows that the decision boundary is overfitting the training data?**

A) A  
B) B  
C) C  
D)None of these

**Solution: C**

Since in figure 3, Decision boundary is not smooth that means it will over-fitting the data.

**23) What do you conclude after seeing this visualization?**

1. The training error in first plot is maximum as compare to second and third plot.
2. The best model for this regression problem is the last (third) plot because it has minimum training error (zero).
3. The second model is more robust than first and third because it will perform best on unseen data.
4. The third model is overfitting more as compare to first and second.
5. All will perform same because we have not seen the testing data.

A) 1 and 3  
B) 1 and 3  
C) 1, 3 and 4  
D) 5

**Solution: C**

The trend in the graphs looks like a quadratic trend over independent variable X. A higher degree(Right graph) polynomial might have a very high accuracy on the train population but is expected to fail badly on test dataset. But if you see in left graph we will have training error maximum because it underfits the training data

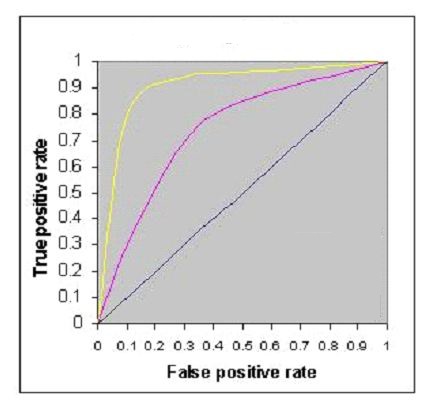
**24) Suppose, above decision boundaries were generated for the different value of regularization. Which of the above decision boundary shows the maximum regularization?**

A) A  
B) B  
C) C  
D) All have equal regularization

**Solution: A**

Since, more regularization means more penality means less complex decision boundry that shows in first figure A.

**25) The below figure shows AUC-ROC curves for three logistic regression models. Different colors show curves for different hyper parameters values. Which of the following AUC-ROC will give best result?**

  
A) Yellow  
B) Pink  
C) Black  
D) All are same

**Solution: A**

The best classification is the largest area under the curve so yellow line has largest area under the curve.

**26) What would do if you want to train logistic regression on same data that will take less time as well as give the comparatively similar accuracy(may not be same)?**

Suppose you are using a Logistic Regression model on a huge dataset. One of the problem you may face on such huge data is that Logistic regression will take very long time to train.

A) Decrease the learning rate and decrease the number of iteration  
B) Decrease the learning rate and increase the number of iteration  
C) Increase the learning rate and increase the number of iteration  
D) Increase the learning rate and decrease the number of iteration

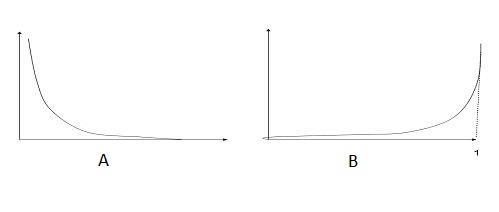
**Solution: D**

If you decrease the number of iteration while training it will take less time for surly but will not give the same accuracy for getting the similar accuracy but not exact you need to increase the learning rate.

**27) Which of the following image is showing the cost function for y =1.**

**Following is the loss function in logistic regression(Y-axis loss function and x axis log probability) for two class classification problem.**

**Note: Y is the target class**

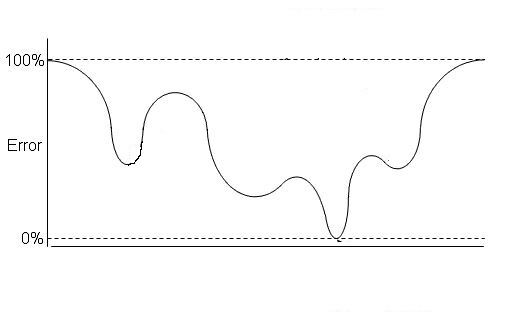


A) A  
B) B  
C) Both  
D) None of these

**Solution: A**

A is the true answer as loss function decreases as the log probability increases

**28) Suppose, Following graph is a cost function for logistic regression.**

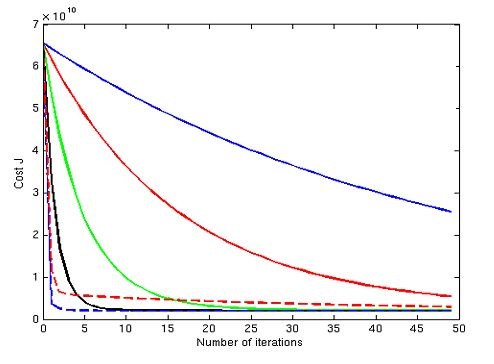
  
**Now, How many local minimas are present in the graph?**

A) 1  
B) 2  
C) 3  
D) 4

**Solution: C**

There are three local minima present in the graph

**29) Imagine, you have given the below graph of logistic regression  which is shows the relationships between cost function and number of iteration for 3 different learning rate values (different colors are showing different curves at different learning rates ).**

Suppose, you save the graph for future reference but you forgot to save the value of different learning rates for this graph. Now, you want to find out the relation between the leaning rate values of these curve. Which of the following will be the true relation?

**Note:**

1. The learning rate for blue is l1
2. The learning rate for red is l2
3. The learning rate for green is l3

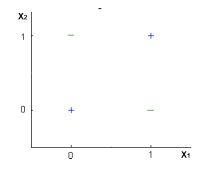
A) l1>l2>l3  
B) l1 = l2 = l3  
C) l1 < l2 < l3

D) None of these

**Solution: C**

If you have low learning rate means your cost function will decrease slowly but in case of large learning rate cost function will decrease very fast.

**30) Can a Logistic Regression classifier do a perfect classification on the below data?**



Note: You can use only X1 and X2 variables where X1 and X2 can take only two binary values(0,1).

A) TRUE  
B) FALSE  
C) Can’t say  
D) None of these

**Solution: B**

No, logistic regression only forms linear decision surface, but the examples in the figure are not linearly separable.

https://www.cs.cmu.edu/~tom/10701\_sp11/midterm\_sol.pdf

### 1. What is a logistic function? What is the range of values of a logistic function?

f(z) = 1/(1+e -z )  
The values of a logistic function will range from 0 to 1. The values of Z will vary from -infinity to +infinity.

### 2. Why is logistic regression very popular?

Logistic regression is famous because it can convert the values of logits (logodds), which can range from -infinity to +infinity 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.

### 3. What is the formula for the logistic regression function?

f(z) = 1/(1+e-(α+1X1+2X2+….+kXk))  
[The Difference between Data Science, Machine Learning and Big Data!](https://upgrad.com/blog/difference-between-data-science-machine-learning-and-big-data/)

### 4. How can the probability of a logistic regression model be expressed as conditional probability?

P(Discrete value of Target variable | X1, X2, X3….Xk).  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.

### 5. 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.  
The odds of winning the lottery = (Probability of winning)/(probability of not winning)  
The odds of winning the lottery = 0.01/0.99  
The odds of winning the lottery is 1 to 99, and the odds of not winning the lottery is 99 to 1.

### 6. 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.  
   Logistic model = α+1X1+2X2+….+kXk. The output of the same will be logits.  
   Logistic function = f(z) = 1/(1+e-(α+1X1+2X2+….+kXk)). The output, in this case, will be the probabilities.

### 7. How to interpret the results of a logistic regression model? Or, what are the meanings of alpha and beta in a logistic regression model?

Alpha is the baseline in a logistic regression model. It is the log odds for an instance when all the attributes (X1, X2,………….Xk) are zero. In practical scenarios, the probability of all the attributes being zero is very low. In another interpretation, Alpha is the log odds for an instance when none of the attributes is taken into consideration.  
   Beta is the value by which the log odds change by a unit change in a particular attribute by keeping all other attributes fixed or unchanged (control variables).

### ***8.  What is odds ratio?***

Odds ratio is the ratio of odds between two groups. For example, let’s assume that we are trying to ascertain the effectiveness of a medicine. We administered this medicine to the ‘intervention’ group and a placebo to the ‘control’ group.  
   Odds ratio (OR) = (odds of the intervention group)/(odds of the control group)  
**Interpretation**  
If odds ratio = 1, then there is no difference between the intervention group and the control group  
If odds ratio is greater than 1, then the control group is better than the intervention group  
If odds ratio is less than 1, then the intervention group is better than the control group.  
[5 Breakthrough Applications of Machine Learning](https://upgrad.com/blog/machine-learning-and-its-breakthrough-applications/)

### 9. What is the formula for calculating odds ratio?

In the formula above, X1 and X0 stand for two different groups for which odds ratio needs to be calculated. X1i stands for the instance ‘i’ in group X1. Xoi stands for the instance ‘i’ in group X0. stands for the coefficient of the logistic regression model. Note that the baseline is not included in this formula.

### 10. Why can’t linear regression be used in place of logistic regression for binary classification?

The reasons why linear regressions cannot be used in case of binary classification are as follows:  
**Distribution of error terms**: The distribution of data in case of linear and logistic regression is different. Linear regression assumes that error terms are normally distributed. In case of binary classification, this assumption does not hold true.  
**Model output**: In linear regression, the output is continuous. In 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 case of logistic regression.

### 11. Is the decision boundary linear or nonlinear in the case of a logistic regression model?

The decision boundary is a line that separates the target variables into different classes. The decision boundary can either be linear or nonlinear. In case of a logistic regression model, the decision boundary is a straight line.  
Logistic regression model formula = α+1X1+2X2+….+kXk. This clearly represents a straight line. Logistic regression is only suitable in such cases where a straight line is able to separate the different classes. If a straight line is not able to do it, then nonlinear algorithms should be used to achieve better results.

### 12. 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 we 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 = 30 (the number of tails)  
Pr(X=60 |n = 100, p)  
The likelihood function is the probability that the number of heads received is 60 in a trail of 100 coin tosses, where the probability of heads received in each coin toss is p. Here the coin toss result follows a binomial distribution.  
This can be reframed as follows:  
       Pr(X=60|n=100,p) = c x p60x(1-p)100-60  
       c = constant  
       p = unknown parameter  
The likelihood function gives the probability of observing the results using unknown parameters.

### 13. What is the Maximum Likelihood Estimator (MLE)?

 The MLE chooses those sets of unknown parameters (estimator) that maximise the likelihood function. The method to find the MLE is to use calculus and setting the derivative of the logistic function with respect to an unknown parameter to zero, and solving it will give the MLE. For a binomial model, this will be easy, but for a logistic model, the calculations are complex. Computer programs are used for deriving MLE for logistic models.  
(Here’s another approach to answering the question.)  
MLE is a statistical approach to estimating the parameters of a mathematical model. MLE and ordinary square estimation give the same results for linear regression if the dependent variable is assumed to be normally distributed. MLE does not assume anything about independent variables.

### 14. What are the different methods of MLE and when is each method preferred?

In case of logistics regression, there are two approaches of MLE. They are conditional and unconditional methods. Conditional and unconditional methods are algorithms that use different likelihood functions. The unconditional formula employs joint probability of positives (for example, churn) and negatives (for example, non-churn). The conditional formula is the ratio of the probability of observed data to the probability of all possible configurations.  
The unconditional method is preferred if the number of parameters is lower compared to the number of instances. If the number of parameters is high compared to the number of instances, then conditional MLE is to be preferred. Statisticians suggest that conditional MLE is to be used when in doubt. Conditional MLE will always provide unbiased results.  
[These 6 Machine Learning Techniques are Improving Healthcare](https://upgrad.com/blog/machine-learning-applications-in-healthcare-2018/)

### 15. What are the advantages and disadvantages of conditional and unconditional methods of MLE?

Conditional methods do not estimate unwanted parameters. Unconditional methods estimate the values of unwanted parameters also. Unconditional formulas can directly be developed with joint probabilities. This cannot be done with conditional probability. If the number of parameters is high relative to the number of instances, then the unconditional method will give biased results. Conditional results will be unbiased in such cases.

### 16. What is the output of a standard MLE program?

The output of a standard MLE program is as follows:  
**Maximised likelihood value**: This is the numerical value obtained by replacing the unknown parameter values in the likelihood function with the MLE parameter estimator.  
**Estimated variance-covariance matrix**: The diagonal of this matrix consists of estimated variances of the ML estimates. The off-diagonal consists of the covariances of the pairs of the ML estimates.

### 17. Why can’t we use Mean Square Error (MSE) as a cost function for logistic regression?

In logistic regression, we use the sigmoid function and perform a non-linear transformation to obtain the probabilities. Squaring this non-linear transformation will lead to non-convexity with local minimums. Finding the global minimum in such cases using gradient descent is not possible. Due to this reason, MSE is not suitable for logistic regression. Cross-entropy or log loss is used as a cost function for logistic regression. In the cost function for logistic regression, the confident wrong predictions are penalised heavily. The confident right predictions are rewarded less. By optimising this cost function, convergence is achieved.

### 18. Why is accuracy not a good measure for classification problems?

Accuracy is not a good measure for classification problems because it gives equal importance to both false positives and false negatives. However, this may not be the case in most business problems. For example, in case of cancer prediction, declaring cancer as benign is more serious than wrongly informing the patient that he is suffering from cancer. Accuracy gives equal importance to both cases and cannot differentiate between them.

### 19. What is the importance of a baseline in a classification problem?

Most classification problems deal with imbalanced datasets. Examples include telecom churn, employee attrition, cancer prediction, fraud detection, online advertisement targeting, and so on. In all these problems, the number of the positive classes will be very low when compared to the negative classes. In some cases, it is common to have positive classes that are less than 1% of the total sample. In such cases, an accuracy of 99% may sound very good but, in reality, it may not be.  
Here, the negatives are 99%, and hence, the baseline will remain the same. If the algorithms predict all the instances as negative, then also the accuracy will be 99%. In this case, all the positives will be predicted wrongly, which is very important for any business. Even though all the positives are predicted wrongly, an accuracy of 99% is achieved. So, the baseline is very important, and the algorithm needs to be evaluated relative to the baseline.

### 20. What are false positives and false negatives?

False positives are those cases in which the negatives are wrongly predicted as positives. For example, predicting that a customer will churn when, in fact, he is not churning.  
False negatives are those cases in which the positives are wrongly predicted as negatives. For example, predicting that a customer will not churn when, in fact, he churns.

### 21. What are the true positive rate (TPR), true negative rate (TNR), false-positive rate (FPR), and false-negative rate (FNR)?

TPR refers to the ratio of positives correctly predicted from all the true labels. In simple words, it is the frequency of correctly predicted true labels.  
   **TPR = TP/TP+FN**  
   TNR refers to the ratio of negatives correctly predicted from all the false labels. It is the frequency of correctly predicted false labels.  
   **TNR = TN/TN+FP**  
   FPR refers to the ratio of positives incorrectly predicted from all the true labels. It is the frequency of incorrectly predicted false labels.  
   **FPR = FP/TN+FP**  
   FNR refers to the ratio of negatives incorrectly predicted from all the false labels. It is the frequency of incorrectly predicted true labels.  
   **FNR = FN/TP+FN**

### 22. What are precision and recall?

Precision is the proportion of true positives out of predicted positives. To put it in another way, it is the accuracy of the prediction. It is also known as the ‘positive predictive value’.  
   **Precision = TP/TP+FP**  
Recall is same as the true positive rate (TPR).  
[How does Unsupervised Machine Learning Work?](https://upgrad.com/blog/how-does-unsupervised-machine-learning-work/)

### 23. What is F-measure?

It is the harmonic mean of precision and recall. In some cases, there will be a trade-off between the precision and the recall. In such cases, the F-measure will drop. It will be high when both the precision and the recall are high. Depending on the business case at hand and the goal of data analytics, an appropriate metric should be selected.  
F-measure = 2 X (Precision X Recall) / (Precision+Recall)

### 24. What is accuracy?

It is the number of correct predictions out of all predictions made.  
   Accuracy = (TP+TN)/(The total number of Predictions)

### 25. What are sensitivity and specificity?

Specificity is the same as true negative rate, or it is equal to 1 – false-positive rate.  
Specificity = TN/TN + FP.  
Sensitivity is the true positive rate.  
Sensitivity = TP/TP + FN

### 26. How to choose a cutoff point in case of a logistic regression model?

The cutoff point depends on the business objective. Depending on the goals of your business, the cutoff point needs to be selected. For example, let’s consider loan defaults. If the business objective is to reduce the loss, then the specificity needs to be high. If the aim is to increase profits, then it is an entirely different matter. It may not be the case that profits will increase by avoiding giving loans to all predicted default cases. But it may be the case that the business has to disburse loans to default cases that are slightly less risky to increase the profits. In such a case, a different cutoff point, which maximises profit, will be required. In most of the instances, businesses will operate around many constraints. The cutoff point that satisfies the business objective will not be the same with and without limitations. The cutoff point needs to be selected considering all these points. As a thumb rule, choose a cutoff value that is equivalent to the proportion of positives in a dataset.

[What is Machine Learning and Why it matters](https://www.upgrad.com/blog/what-is-machine-learning-and-why-it-matters/)

### 27. How does logistic regression handle categorical variables?

The inputs to a logistic regression model need to be numeric. The algorithm cannot handle categorical variables directly. So, they need to be converted into a format that is suitable for the algorithm to process. The various levels of a categorical variable will be assigned a unique numeric value known as the dummy variable. These dummy variables are handled by the logistic regression model as any other numeric value.

### 28. What is a cumulative response curve (CRV)?

In order to convey the results of an analysis to the management, a ‘cumulative response curve’ is used, which is more intuitive than the ROC curve. A ROC curve is very difficult to understand for someone outside the field of data science. A CRV consists of the true positive rate or the percentage of positives correctly classified on the Y-axis and the percentage of the population targeted on the X-axis. It is important to note that the percentage of the population will be ranked by the model in descending order (either the probabilities or the expected values). If the model is good, then by targeting a top portion of the ranked list, all high percentages of positives will be captured. As with the ROC curve, there will be a diagonal line which represents random performance. Let’s understand this random performance as an example. Assuming that 50% of the list is targeted, it is expected that it will capture 50% of the positives. This expectation is captured by the diagonal line, which is similar to the ROC curve.

### 29. What are the lift curves?

The lift is the improvement in model performance (increase in true positive rate) when compared to random performance. Random performance means if 50% of the instances is targeted, then it is expected that it will detect 50% of the positives. Lift is in comparison to the random performance of a model. If a model’s performance is better than its random performance, then its lift will be greater than 1.  
In a lift curve, lift is plotted on the Y-axis and the percentage of the population (sorted in descending order) on the X-axis. At a given percentage of the target population, a model with a high lift is preferred.

### 30. Which algorithm is better at handling outliers logistic regression or SVM?

Logistic regression will find a linear boundary if it exists to accommodate the outliers. Logistic regression will shift the linear boundary in order to accommodate the outliers. SVM is insensitive to individual samples. There will not be a major shift in the linear boundary to accommodate an outlier. SVM comes with inbuilt complexity controls, which take care of overfitting. This is not true in case of logistic regression.

### 31. How will you deal with the multiclass classification problem using logistic regression?

The most famous method of dealing with multiclass classification using logistic regression is using the one-vs-all approach. Under this approach, a number of models are trained, which is equal to the number of classes. The models work in a specific way. For example, the first model classifies the datapoint depending on whether it belongs to class 1 or some other class; the second model classifies the datapoint into class 2 or some other class. This way, each data point can be checked over all the classes.

### 32. Explain the use of ROC curves and the AUC of an ROC Curve.

An ROC (Receiver Operating Characteristic) curve illustrates the performance of a binary classification model. It is basically a TPR versus FPR (true positive rate versus false-positive rate) curve for all the threshold values ranging from 0 to 1. In a ROC curve, each point in the ROC space will be associated with a different confusion matrix. A diagonal line from the bottom-left to the top-right on the ROC graph represents random guessing. The Area Under the Curve (AUC) signifies how good the classifier model is. If the value for AUC is high (near 1), then the model is working satisfactorily, whereas if the value is low (around 0.5), then the model is not working properly and just guessing randomly.

### 33. How can you use the concept of ROC in a multiclass classification?

The concept of ROC curves can easily be used for multiclass classification by using the one-vs-all approach. For example, let’s say that we have three classes ‘a’, ’b’, and ‘c’. Then, the first class comprises class ‘a’ (true class) and the second class comprises both class ‘b’ and class ‘c’ together (false class). Thus, the ROC curve is plotted. Similarly, for all the three classes, we will plot three ROC curves and perform our analysis of AUC.  
We have so far covered the two most basic ML algorithms, Linear and Logistic Regression, and we hope that you have found these resources helpful.

1. Logistic regression is used to predict \_\_\_\_\_\_\_\_\_ valued output?
   * Continuous
   * Categorical
2. How much marks a strudent can get in a competitive exam based on hours of study can be solved using \_\_\_\_\_\_\_\_\_ regression model
   * Multi-linear
   * Logistic
3. Logistic regression is \_\_\_\_\_\_\_\_\_ when the observed outcome of dependent variable can have only two values such as 0 and 1 or success and failure
   * Binomial
   * Multinomial
   * Ordinal
4. Whether a strudent will pass or fail in the competitive exam based on hours of study can be solved using \_\_\_\_\_\_\_\_\_ regression model
   * Multi-linear
   * Logistic
5. \_\_\_\_\_\_\_\_ regression can be termed as a special case of \_\_\_\_\_\_\_\_\_ regression when the outcome variable is categorical
   * Logistic, Linear
   * Linear, Logistic
6. In logistic regression, the goal is to predict \_\_\_\_\_\_\_\_\_
   * Actual value of outcome dependent variable
   * Odds of outcome dependent variable
7. Which of the following can be used to evaluate the performance of logistic regression model?
   * Adjusted R-Squared
   * AIC
8. Which of the following is link function in logistic regression
   * Identity
   * Logit
9. Logistic regression is \_\_\_\_\_\_\_\_\_ when the observed outcome of dependent variable can have multiple possible types
   * Binomial
   * Multinomial
   * Ordinal
10. In logistic regression, following technique is used to measure the goodness of the fit
    * Sum of squares calculations
    * Deviance calculations
11. Which of the following can be used to evaluate the performance of logistic regression model?
    * AIC
    * Null and Residual Deviance
    * Both of the above
    * None of the above
12. Given two model with different AIC value, which one would be preferred model?
    * One with higher AIC value
    * One with lower AIC value
13. Deviance is a measure of difference between a \_\_\_\_\_\_\_ model and the \_\_\_\_\_\_\_\_\_ model
    * saturated, fitted
    * Fitted, saturated
14. Logistic regression is \_\_\_\_\_\_\_\_\_ when the observed outcome of dependent variable are ordered
    * Binomial
    * Multinomial
    * Ordinal
15. Logit transformation is log of \_\_\_\_\_\_\_\_\_\_\_
    * Odds of the event happening for different levels of each independent variable
    * Ratio of odds of the event happening for different levels of each independent variable
16. Logistic function is \_\_\_\_\_\_\_\_\_
    * Dependent variable equalling a given case
    * Probability that dependent variable equals a case
17. Deviance is is a function of \_\_\_\_\_\_\_\_
    * Exponential function of likelihood ratio
    * Logrithmic function of likelihood ratio
18. The odds of the dependent variable equaling a case (given some linear combination x of the predictors) is equivalent to \_\_\_\_\_\_\_
    * Log function of the linear regression expression
    * Exponential function of the linear regression function
19. Regression coefficients in logistic regression are estimated using \_\_\_\_\_\_\_\_
    * Ordinary least squares method
    * Maximum likelihood estimation method
20. \_\_\_\_\_\_\_\_\_ is analogous to \_\_\_\_\_\_\_\_\_\_ in linear regression
    * Sum of squares calculations, deviance
    * Deviance, sum of squares calculations
21. Deviance can be shown to follow \_\_\_\_\_\_\_\_\_\_
    * t-distribution
    * F-distribution
    * Chi-square distribution
    * None of the above
22. \_\_\_\_\_\_ value of deviance represents the better fit of model
    * Higher
    * Lower
23. If the model deviance is significantly \_\_\_\_\_\_\_\_ than the null deviance then one can conclude that the predictor or set of predictors significantly improved model fit
    * Smaller
    * Larger
24. Which of the following is analogous to R-Squared for logistic regression
    * Likelihood ration R-squared
    * McFadden R-squared
    * Cox and Snell R-Squared
    * All of the above
25. Estimation in logistic regression chooses the parameters that \_\_\_\_\_\_\_\_\_\_\_ the likelihood of observing the sample values
    * Minimizes
    * Maximizes
26. Which of the following tests can be used to assess whether the logistic regression model is well calibrated
    * Hosmer-Lemeshow test
    * ROC Curve
    * Both of the above
27. ROC related with ROC curve stands for \_\_\_\_\_\_\_
    * Regression Optimization Characteristic
    * Regression Operating Characteristic
    * Receiver Operating Characteristic
28. Which of the following is used to identify the best threshold for separating positive and negative classes
    * Hosmer-Lemeshow test
    * ROC Curve
    * Both of the above
29. ROC curve is a plot of \_\_\_\_\_\_\_\_\_\_ vs \_\_\_\_\_\_\_\_\_\_\_
    * Sensitivity, 1-specificity
    * 1-specificity, Sensitivity
30. \_\_\_\_\_\_ the value of AUC, better is the prediction power of the model
    * Lower
    * Higher

### How To Learn Machine Learning Algorithms For Interviews

#### Logistics Regression

Theoretical Understanding:

1. Tutorial 35:Logitic Regression Part 1 <https://www.youtube.com/watch?v=L_xBe7MbPwk>
2. Tutorial 36:Logitic Regression Part 2 <https://www.youtube.com/watch?v=uFfsSgQgerw>
3. Tutorial 39:Logitic Regression Part 3 <https://www.youtube.com/watch?v=V8fS0T_ktn4>
4. Tutorial 42:How To Find Optimal Threshold for Binary classification: <https://www.youtube.com/watch?v=_AjhdXuXEDE>
5. Interview question: <https://www.youtube.com/watch?v=tcaruVHXZwE&t=122s>

In [ ]:

##### *1. What Are the Basic Assumption?*

1. Linear Relation between independent features and the log odds

##### *2. Advantages*

Advantages of Logistics Regression

1. Logistic Regression Are very easy to understand
2. It requires less training
3. Good accuracy for many simple data sets and it performs well when the dataset is linearly separable.
4. It makes no assumptions about distributions of classes in feature space.
5. Logistic regression is less inclined to over-fitting but it can overfit in high dimensional datasets.One may consider Regularization (L1 and L2) techniques to avoid over-fittingin these scenarios.
6. Logistic regression is easier to implement, interpret, and very efficient to train.

##### *3. Disadvantages*

1. Sometimes Lot of Feature Engineering Is required
2. If the independent features are correlated it may affect performance
3. It is often quite prone to noise and overfitting
4. If the number of observations is lesser than the number of features, Logistic Regression should not be used, otherwise, it may lead to overfitting.
5. Non-linear problems can’t be solved with logistic regression because it has a linear decision surface. Linearly separable data is rarely found in real-world scenarios.
6. It is tough to obtain complex relationships using logistic regression. More powerful and compact algorithms such as Neural Networks can easily outperform this algorithm.
7. In Linear Regression independent and dependent variables are related linearly. But Logistic Regression needs that independent variables are linearly related to the log odds (log(p/(1-p)).

##### *4. Whether Feature Scaling is required?*

yes

#### 5. Missing Values

Sensitive to missing values

##### *6. Impact of outliers?*

Like linear regression, estimates of the logistic regression are sensitive to the unusual observations: outliers, high leverage, and influential observations. Numerical examples and analysis are presented to demonstrate the most recent outlier diagnostic methods using data sets from medical domain

##### *Types of Problems it can solve(Supervised)*

1. Classification

##### *Performance Metrics*

##### *Classification*

1. Confusion Matrix
2. Precision,Recall, F1 score
3. Part 1 <https://www.youtube.com/watch?v=aWAnNHXIKww>
4. Part 2 <https://www.youtube.com/watch?v=A_ZKMsZ3f3o>