Logistic Regression model

It's a classification algorithm, that is used where the response variable is categorical. The idea of Logistic Regression is to find a relationship between features and probability of particular outcome.

Different types of Logistic Regression Three different types of Logistic Regression are as follows:

$1.Binary Logistic Regression$: In this, the target variable has only two 2 possible outcomes.

For Example, 0 and 1, or pass and fail or true and false.

$2.Multinomial Logistic Regression$: In this, the target variable can have three or more possible values without any order.

For Example, Predicting preference of food i.e. Veg, Non-Veg, Vegan.

$3.Ordinal Logistic Regression$: In this, the target variable can have three or more values with ordering.

For Example, Movie rating from 1 to 5.

Assumptions made in logistic regression

- minimal or no $multicollinearity$ among the independent variables.
- should be a $linearrelationship$ between the logit of the outcome and each predictor variables. The logit function is given as logit(p) = log(p/(1-p)), where p is the probability of the outcome.
- requires a $largesamplesize$ to predict properly.

Advantages of Logistic Regression

- Logistic Regression is very easy to understand.
- It requires less training.
- It performs well for simple datasets as well as when the data set is linearly separable.
- It doesn't make any assumptions about the distributions of classes in feature space.
- A Logistic Regression model is less likely to be over-fitted but it can overfit in high dimensional datasets. To avoid over-fitting these scenarios, One may consider regularization.
- They are easier to implement, interpret, and very efficient to train.

Disadvantages of Logistic Regression

- Sometimes a lot of Feature Engineering is required.
- If the independent features are correlated with each other it may affect the performance of the classifier.

- It is quite sensitive to noise and overfitting.
- Logistic Regression should not be used if the number of observations is lesser than the number of features, otherwise, it may lead to overfitting.
- By using Logistic Regression, non-linear problems can't be solved because it has a linear decision surface. But in real-world scenarios, the linearly separable data is rarely found.
- By using Logistic Regression, it is tough to obtain complex relationships. Some algorithms such as neural networks, which are more powerful, and compact can easily outperform Logistic Regression algorithms.
- In Linear Regression, there is a linear relationship between independent and dependent variables but in Logistic Regression, independent variables are linearly related to the log odds (log(p/(1-p)).

Feature Scaling is required-yes

Missing Values -- Sensitive to missing values

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
```

ODDs

It is the ratio of the probability of an event occurring to the probability of the event not occurring.

Difference between the outputs of the Logistic model and the Logistic function

- The Logistic model outputs the logits, i.e. $log - odds$;
- whereas the Logistic function outputs the $probabilities$.

Can't linear regression be used in place of logistic regression for binary classification

(i)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.

(ii)Model output: In linear regression, the output is continuous. In case of binary classification, an output of a continuous value does not make sense.

- As the logistic regression model can output probabilities with logistic/sigmoid function, it is preferred over linear regression.

(iii)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.

accuracy not a good measure for classification problems it gives equal importance to both false

positives and false negatives

- **False positives**False positives are those cases in which the negatives are wrongly predicted as positives.
- **False negatives** are those cases in which the positives are wrongly predicted as negatives.

- **True Positive Rate** 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
- **True Negative Rate** 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
- **False-Positive Rate** 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
- **False-Negative Rate** 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

Precision is the proportion of true positives out of predicted positives.

- Precision = TP/TP+FP

F-measure: harmonic mean of precision and recall

F-measure = 2 X (Precision X Recall) / (Precision+Recall)

Accuracy the number of correct predictions out of all predictions made.

- Accuracy = (TP+TN)/(The total number of Predictions)

ROC - Receiver Operating Characteristic

Use of ROC curves and the AUC of an ROC Curve

- ROC 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.

Concept of ROC in a multiclass classification

- 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).

logistic regression a generative or a descriptive classifier

- Logistic regression is a $descriptivemodel$.
- Logistic regression learns to classify by knowing what features differentiate two or more classes of objects.
- For example, to classify between an apple and an orange, it will learn that the orange is orange in color and an apple is not.

- On the other hand, a **generative** classifier like a **Naive Bayes** will store all the classes' critical features and then classify based on the features the test case best fits.

can't we use the mean square error cost function used in linear regression for logistic regression

- if we use mean square error in logistic regression, the resultant cost function will be non-convex, i.e., a function with many local minima,

Wald Test useful in logistic regression but not in linear regression

- The Wald test, also known as the Wald Chi-Squared Test,
- is a method to find whether the independent variables in a model are of significance.
- The significance of variables is decided by whether they contribute to the predictions or not.

Maximum Likelihood Estimation to obtain the model coefficients which relate to the predictors and target