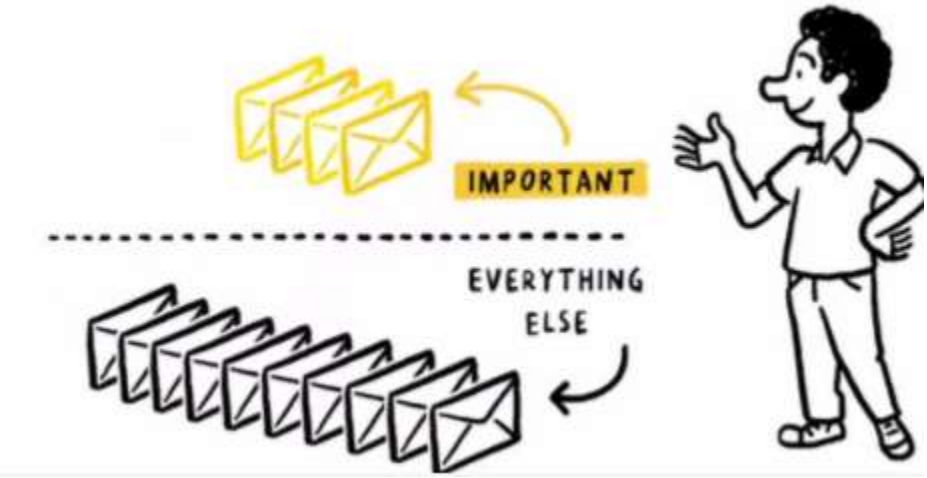
Classification- Summary

Classification predictive modeling is the task of approximating a mapping function (f) from input variables (X) to discrete output variables (y).



The output variables are often called labels or categories. The mapping function predicts the class or category for a given observation.

For example, an email of text can be classified as belonging to one of two classes: “spam“ and “*not spam*“.

- A classification problem requires that examples be classified into one of two or more classes.
- A classification can have real-valued or discrete input variables.
- A problem with two classes is often called a two-class or binary classification problem.
- A problem with more than two classes is often called a multi-class classification problem.

It is common for classification models to predict a continuous value as the probability of a given example belonging to each output class. The probabilities can be interpreted as the likelihood or confidence of a given example belonging to each class. A predicted probability can be converted into a class value by selecting the class label that has the highest probability.

For example, a specific email of text may be assigned the probabilities of 0.1 as being “spam” and 0.9 as being “not spam”. We can convert these probabilities to a class label by selecting the “not spam” label as it has the highest predicted likelihood.

There are many ways to estimate the skill of a classification predictive model, but perhaps the most common is to calculate the classification accuracy.

The classification accuracy is the percentage of correctly classified examples out of all predictions made.

For example, if a classification predictive model made 5 predictions and 3 of them were correct and 2 of them were incorrect, then the classification accuracy of the model based on just these predictions would be:

accuracy = correct predictions / total predictions * 100
accuracy = 3 / 5 * 100
accuracy = 60%

In some cases, it is possible to convert a regression problem to a classification problem. For example, the quantity to be predicted could be converted into discrete buckets.

For example, amounts in a continuous range between $0 and $100 could be converted into 2 buckets:

- Class 0: $0 to $49
- Class 1: $50 to $100

https://machinelearningmastery.com/classification-versus-regression-in-machine-learning/

Assumptions of Logistic Regression

Assumptions in Logistic Regression

Logistic Regression does not make many of the key assumptions of linear regression that are based on ordinary least squares algorithms – particularly regarding linearity, normality, homoscedasticity, and measurement level.

First, logistic regression does not require a linear relationship between the dependent and independent variables. Second, the error terms (residuals) do not need to be normally distributed. Third, homoscedasticity is not required.

However, some other assumptions still apply.

First, logistic regression requires the variables to be independent of each other. In other words, *there should not be any Multicollinearity*. This means that the independent variables should not be too highly correlated with each other.

Second, logistic regression assumes *linearity of independent variables and log odds*. although this analysis does not require the dependent and independent variables to be related linearly, it requires that the independent variables are linearly related to the log odds. This can be deduced from the equation-

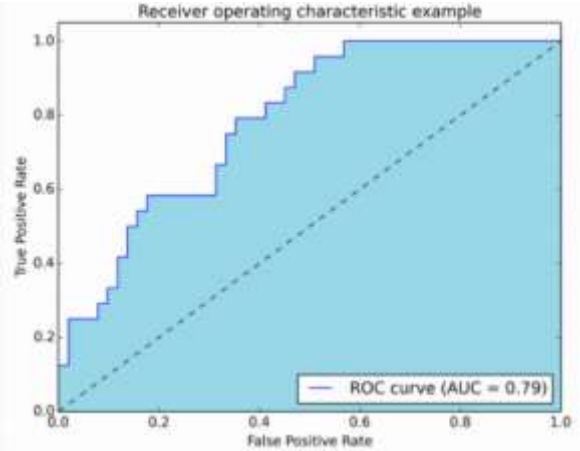
$\log(odds) = b_0 + b_1x_1 + b_2x_2 + ... + b_nx_n$

So, $\log(odds) \propto X_1$ and so on.

Finally, logistic regression typically requires a large sample size. A general guideline is that you need at minimum of a 10 cases with the least frequent outcome (the classes of your interest, usually 1's) for each independent variable in your model.

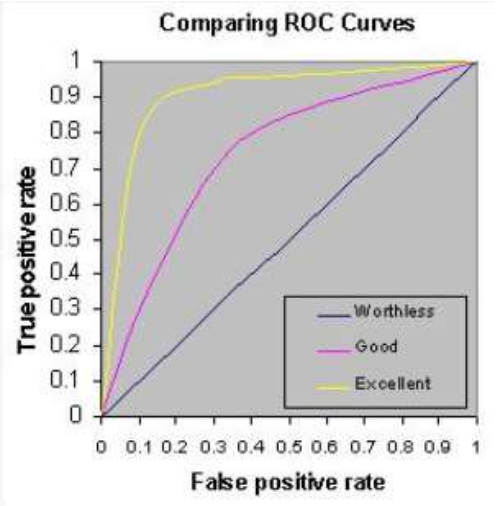
Area Under Curve

Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. The area under the ROC curve (AUC) is a measure of how well a parameter can distinguish between two diagnostic groups (1's/0's). A test with perfect discrimination has a ROC curve that passes through the upper left corner (100% sensitivity, 100% specificity). Therefore the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test.



The best possible prediction method would yield a point in the upper left corner or coordinate (0,1) of the ROC space, representing 100% sensitivity (no false negatives) and 100% specificity (no false positives). The (0,1) point is also called a *perfect classification*.

A random guess would give a point along a diagonal line (the so-called *line of no-discrimination*) from the left bottom to the top right corners (regardless of the positive and negative base rates). An intuitive example of random guessing is a decision by flipping coins. As the size of the sample increases, a random classifier's ROC point tends towards the diagonal line. In the case of a balanced coin, it will tend to the point (0.5, 0.5).



Type 1 & Type 2 Errors