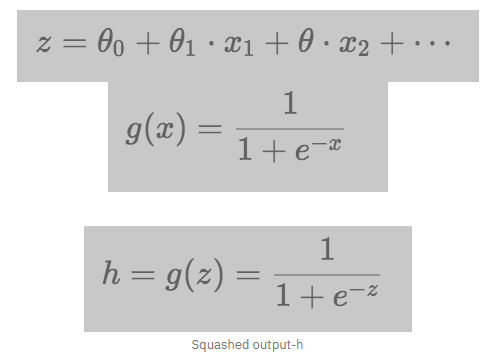
**Introduction:**

Logistic regression is the most famous machine learning algorithm after linear regression. In a lot of ways, linear regression and logistic regression are similar. But the biggest difference lies in what they are used for. Linear regression algorithms are used to predict/forecast values, but logistic regression is used for classification tasks. There are many classification tasks done routinely by people. For example, classifying whether an email is a spam or not, classifying whether a tumor is malignant or benign, classifying whether a website is fraudulent or not, etc. These are typical examples where machine learning algorithms can make our lives a lot easier. A simple, rudimental and useful algorithm for classification is the logistic regression algorithm.



**Dataset:**

"Predict behavior to retain customers. You can analyze all relevant customer data and develop focused customer retention programs."

Each row represents a customer, each column contains customer’s attributes described on the column Metadata.

The data set includes information about:

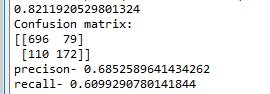
* Customers who left within the last month – the column is called Churn
* Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
* Customer account information – how long they’ve been a customer, contract, payment method, paperless billing, monthly charges, and total charges
* Demographic info about customers – gender, age range, and if they have partners and dependents

**Method:**

We need do some pre-processing before the data can be fed for logistic regression. In the dataset, the values in terms of text like yes or no, electronic payment or credit card or mailed check etc. For logistic regression, we need to encode the values in terms of numbers using the sklearn encode function. The values in all the columns including the column to be predicted is converted to numeric values. Now, the dataset is ready for processing. We consider columns 1 to 16 as factors for predicting the retention rate. The data is split into training (80%) and test (20%) and then perform logistic regression. To randomize we have taken the value of 42. The accuracy of the model is checked by means of a confusion matrix, prediction score, precision score and recall score.

**Result & Interpretation (in terms of Confusion matrix and accuracy metrics):**

The accuracy we got after using logistic regression on the dataset was 82.1%. The confusion matrix and accuracy scores screenshot is attached below.



**Conclusion & Scope:**

Hence, using Logistic Regression we can predict if something like whether customer will stay or not, whether someone will get admission or not. The drawback is that all the data that is used for prediction must be encoded in terms of numbers.