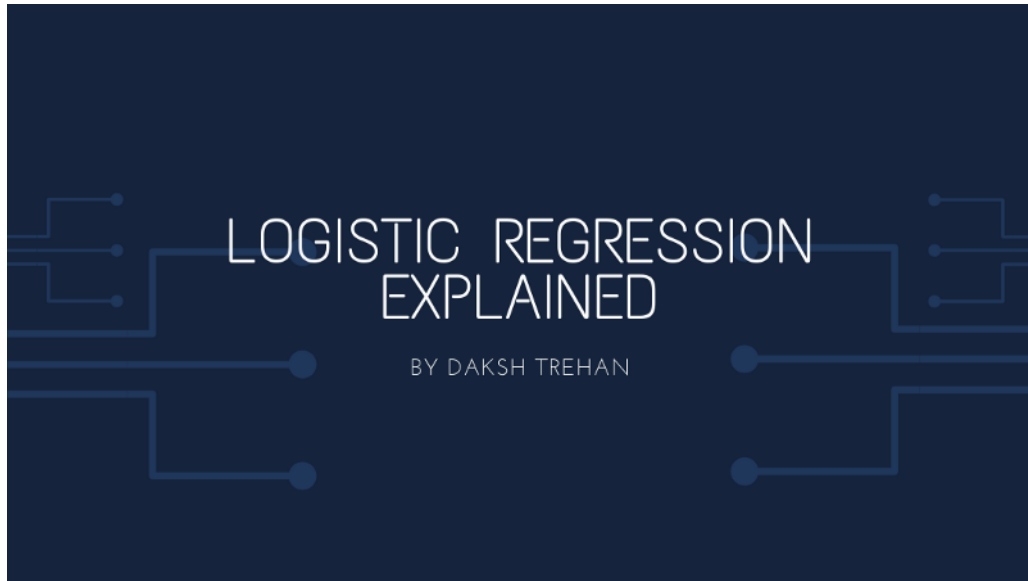

Logistic Regression Explained

Explaining Logistic Regression as easy as it could be.



In linear regression, the Y variable is always continuous. If the Y variable is categorical, you cannot use the [linear regression](#) model.

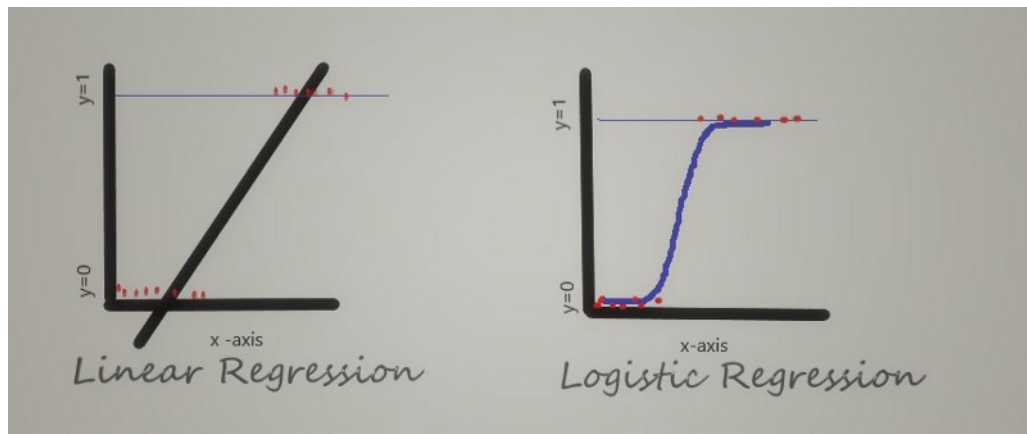
So what would you do when the Y is a categorical variable with 2 classes?

Logistic regression can be used to model and solve such problems, also called as binary classification problems.

Logistic Regression is yet another type of supervised learning algorithm, but its goal is just contrary to its name, rather than regression it aims to classify the data points in two different classes. It is a linear model and produces binary output.

The line differentiating the two classes is known as hyperplane, farther the data point from the hyperplane more will be the confidence of it belonging to that class.

Its goal is to find a legit separating hyperplane that classifies both classes in the best way.



Designed by Daksh Trehan using Paint 3D

A key point to note here is that Y can have 2 classes only and not more than that. If Y has more than 2 classes, it would become a multi-class classification and you can no longer use the vanilla logistic regression for that.

Here are some examples of binary classification problems:

- **Spam Detection:** Predicting if an email is Spam or not
- **Credit Card Fraud:** Predicting if a given credit card transaction is fraud or not
- **Health:** Predicting if a given mass of tissue is benign or malignant

Sigmoid Function

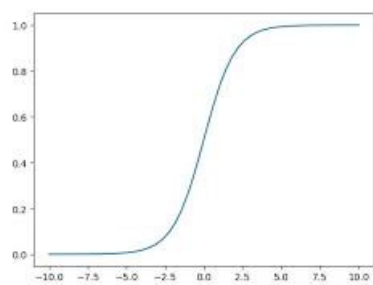
We use the sigmoid function, as it is non-linear and exists between (0 to 1). Therefore, it is especially used for models where we have to **predict the probability** as an output. we use sigmoid to map predictions to probabilities. Since the probability of anything exists only between the range of **0 and 1**, sigmoid is the right choice.

Logistic Regression hypothesis is defined as:

$$h_{\theta}(x) = g(\theta^T x)$$

where function g is the sigmoid function, which is defined as below:

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



Graphical Representation of sigmoid

The function is differentiable and thus we can find the slope of the sigmoid curve at any two points.

$$g'(z) = g(z)(1-g(z))$$

Computing confidence

$$g(z) = \frac{1}{1+e^{-z}} \quad \begin{array}{ll} z \rightarrow \infty & g(z) \rightarrow 1 \\ z \rightarrow -\infty & g(z) \rightarrow 0 \end{array}$$

new hypothesis - $h_\theta(x) = g(\theta^T x)$

$$= \frac{1}{1+e^{-\theta^T x}}$$

if $g(\theta^T x) \geq 0.5$ class 0
 $g(\theta^T x) < 0.5$ class 1

If output computed by sigmoid function ≥ 0.5 then we assume that point belongs to class 0

If output by sigmoid function < 0.5 the point is considered to be in class 1.

Choosing optimal Hyperparameter

Our ultimate goal is to choose the best hyperparameters.

$$= \frac{1}{1+e^{-\theta^T x}}$$

if $g(\theta^T x) \geq 0.5$ class 0
 $g(\theta^T x) < 0.5$ class 1

goal - to get optimal parameters

$$P(y=1 | x; \theta) = h_\theta(x) \quad \# \text{ class 1}$$

$$P(y=0 | x; \theta) = 1 - h_\theta(x) \quad \# \text{ class 0}$$

combining both

$$P(y | x; \theta) = [h_\theta(x)]^y [1 - h_\theta(x)]^{1-y}$$

$$y \in \{0, 1\}$$

If P is probability of object belonging to class 1 then (P-1) would be

probability of object belonging to class 0 as

Probability always lies between 0 and 1.

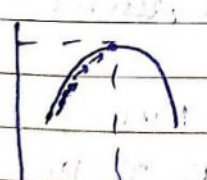
To get optimal θ , we need to maximize likelihood, likelihood is optimal classifier (best hyperplane)

$$L(\theta) = P(y|x; \theta) = \prod P(y_i | x_i; \theta)$$

Taking log both sides

$$\log(L(\theta)) = \log \prod P(y_i | x_i; \theta)$$
$$\log(L(\theta)) = \sum_{i=1}^n y_i \log h x_i + (1 - y_i) \log(1 - h x_i)$$

It is concave g^H & we need to maximize it for best hyperplane
 \therefore we will use gradient ascent



- ① start with random θ
- ② $\theta = \theta + \eta \frac{\partial L(\theta)}{\partial \theta}$
- ③ keep iterating until θ is max or $\frac{\partial L(\theta)}{\partial \theta}$ is max.

for sigmoid μ

$$g(z) = g(z) (1 - g(z))$$

Page 2 of 3

We have combined probability for both classes and derived likelihood that we intend to maximize.

$$\frac{\partial L(\theta)}{\partial \theta} = (y - h\theta(x)) x_j$$
$$\theta = \theta + \eta \sum_{i=1}^n (y_i - h\theta(x_i)) x_j^{(i)}$$

Page 3 of 3

We've successfully derived updated hyperparameters.

Multiclass Logistic Regression

But what if we want to have many outputs using Logistic Regression, for that we can use one v/s rest model.

For illustration, let's suppose our output could be anything belonging to dog, cat and 10 such other classes, but Logistic Regression is a binary model, so our approach would be to implement Vanilla Logistic Regression for dog v/s other classes; if output predicted is dog it's fine, but if test image belongs to some other class we can iterate our former model that is cat v/s other classes and so on.

Logistic Regression using Sci-Kit Learn

```
1 from sklearn.linear_model import LogisticRegression
2 lr=LogisticRegression()
3 lr.fit(x_train,y_train)           #fitting our model
4 lr.predict(x_test)               #predicting on x_test
5 lr.score(x_train,y_train)        #getting accuracy
6 lr.predict_proba(x_train)        #to get likelihood
```

Logistic Regression using sci-kit learn hosted with ❤ by GitHub

[view raw](#)

Advantages of Logistic Regression

1. High efficiency
2. Low Variance
3. Can be easily updated with new data using stochastic Gradient Descent.

Disadvantages of Logistic Regression

1. Doesn't handle large number of categorical variables well.
2. Requires transformation of non-linear features.

Features of Logistic Regression

1. Target is discrete variable
2. Predicted values are probability of targeted values.

Conclusion

Hopefully, this article has not only increased your understanding of **Logistic Regression** but also made you realize, it is not difficult and is already happening in our daily life.

As always, thank you so much for reading, and please share this article if you found it useful! :)

LinkedIn ~ <https://www.linkedin.com/in/dakshtrehan/>

Instagram ~ https://www.instagram.com/daksh_trehan/

Github ~ <https://github.com/dakshtrehan>

Read my other articles :-

[Detecting COVID-19 Using Deep Learning](#)

[Linear Regression Explained](#)

[Determining perfect fit for your ML Model](#)

[Relating Machine Learning Techniques to Real-Life](#)

[Serving Data Science to a Rookie](#)

Follow for further Machine Learning/ Deep Learning blogs.

Medium ~ <https://medium.com/@dakshtrehan>

Cheers.

The cover template was made by me on www.canva.com. The rest pictures are from my notebook.

By [Daksh Trehan](#) on [May 14, 2020](#).

[Canonical link](#)

Exported from [Medium](#) on July 15, 2020.