

Logistic Regression-2

What if we want to predict the odds of $y_i = 1$ vs $y_i = 0$?

Log-odds: Shows how the model is similar to a linear model which is predicting log odds of $y_i = 1$ vs $y_i = 0$, defined as :

$$\log_e(\text{odds}) = \log\left[\frac{p}{1-p}\right]; p = \frac{1}{1+e^{-z_i}} = \frac{e^{z_i}}{e^{z_i}+1} \text{ and } 1-p = \frac{1}{e^{z_i}+1}$$

On substituting the value of p and $1-p$, and solving it we get

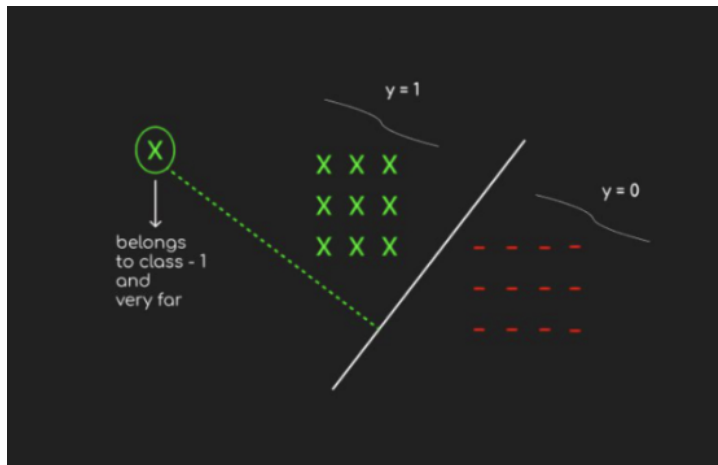
$$\log_e(\text{odds}) = \log_e[e^{z_i}] = z_i = w^T x_i + w_0$$

The image shows a handwritten derivation on a black background. At the top, it states $\text{odds} = \frac{p}{1-p} = e^{w^T x + w_0}$. To the right, it notes $\|w\|=1$. Below this, the equation $\log_e(\text{odds}) = \log_e\left(\frac{p}{1-p}\right)$ is written, with the fraction inside a red box. An arrow points from this box to the underlined text "log-odds-ratio". To the right of the box, the expression $w^T x + w_0$ is shown with checkmarks above the w and x terms. A red bracket under $w^T x + w_0$ is labeled "linear fn in x".

Note: Hence the name Regression in Logistic Regression.

Does Outlier Impact Logistic Regression?

Case 1: When an outlier is on the correct side



The loss value comes out to be very small.

$$L = -[y \log \hat{y} + (1 - y) \log(1 - \hat{y})]$$

$y = 1$ 0

Since the outlier belongs to class 1, L becomes

$$L = -\log \hat{y}$$

Now,

$$\hat{y} = \sigma(z^i)$$

(z^i) is significantly large

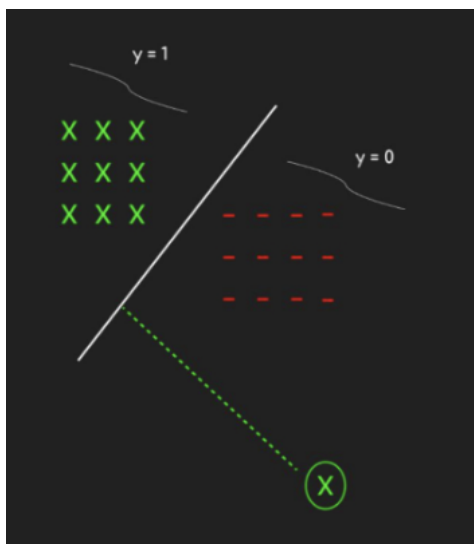
$$\sigma(z^i) \rightarrow 1$$

$$\log \hat{y} \rightarrow 0$$

Very low impact !

Conclusion: The impact of the outlier is very low on the hyperplane.

Case 2: When the outlier is on the opposite side.



$$L = -[y \log \hat{y} + (1 - y) \log(1 - \hat{y})]$$

$y = 1$

Since the outlier belongs to class 1, L becomes

$$L = -\log \hat{y}$$

Now,

$$\hat{y} = \sigma(z^i)$$

(z^i) is large in negative i.e., -4.3

$$\sigma(z^i) \rightarrow 0$$

$$\log \hat{y} \rightarrow \text{large}$$

Very high impact !

Loss values come out to be large.

Conclusion: Hyperplane will shift its position to minimize its loss => Huge impact on hyperplane.

Is there any way to use Logistic Regression for multiclass classification?

- **One vs Rest Method:**

- For $y_i = \{1, 2, 3, \dots, K\}$, generate k-binary logistic Regression models
- We train a total of K models i.e. class 1 vs rest, class 2 vs rest
 - Such that i^{th} class is given label = 1
 - Other classes are given label = 0 when training the i^{th} model
- Final prediction -> Argmax of all of the predictions made by each model