

# Logistic Regression

An Introduction to LR classifier

# Logistic Regression – Binary Classification

- The sentiment analysis task is mainly a binary classification problem to predict whether a given sentence is positive or negative. In our demonstrations we denote '0' as negative and '1' as positive.

# Logistic Regression – The points concept

- Each sentence is mapped to a point.
- If the point is greater than 0.5 then positive else negative.

# Logistic Regression – The points concept

Words/ Documents	going	to	today	i	am	it	is	rain	Points
Document 1	0	0.07	0.07	0	0	0.17	0.17	0.17	0.62
Document 2	0	0	0.07	0.07	0.07	0	0	0	0.41
Document 3	0	0.05	0	0.05	0.05	0	0	0	0.72

# Logistic Regression – Learning algorithm

- A learning algorithm is a specific type of algorithm whose performance increases with time. Logistic regression is a type of learning algorithm. It learns from a training dataset, the pattern of the data and applies the learned logics on new data for prediction.

# Logistic Regression – Linear Equation

Consider the equation:

$$y = a + bx_1 + cx_2 + \dots + dx_{2000}$$

$a, b, c, d$  = coefficients

$X_1, X_2, \dots, X_{2000}$  = independent variables

$y$  = dependent variable

# Logistic Regression – The points concept

Words/ Documents	going	to	today	i	am	it	is	rain	Points
Document 1	0	0.07	0.07	0	0	0.17	0.17	0.17	0.62
Document 2	0	0	0.07	0.07	0.07	0	0	0	0.41
Document 3	0	0.05	0	0.05	0.05	0	0	0	0.72

# Logistic Regression – Linear Equation

Consider the equation:

$$y = a + bx_1 + cx_2 + \dots + dx_{2000}$$

$a, b, c, d$  = coefficients

$X_1, X_2, \dots, X_{2000}$  = independent variables

$y$  = dependent variable



# Logistic Regression – Optimal Coefficients

The algorithm finds the optimal values for the coefficients

# Logistic Regression – Predicting Sentiment

If  $y \geq 0.5 \rightarrow$  Positive sentiment

If  $y < 0.5 \rightarrow$  Negative sentiment

# Logistic Regression – Value range

For some values of the dependent variables, the value of  $y$  can be  $> 1$  or  $< 0$ .

For that, we need some way to restrict the value of  $y$  within the range 0 and 1.

# Logistic Regression – Value range

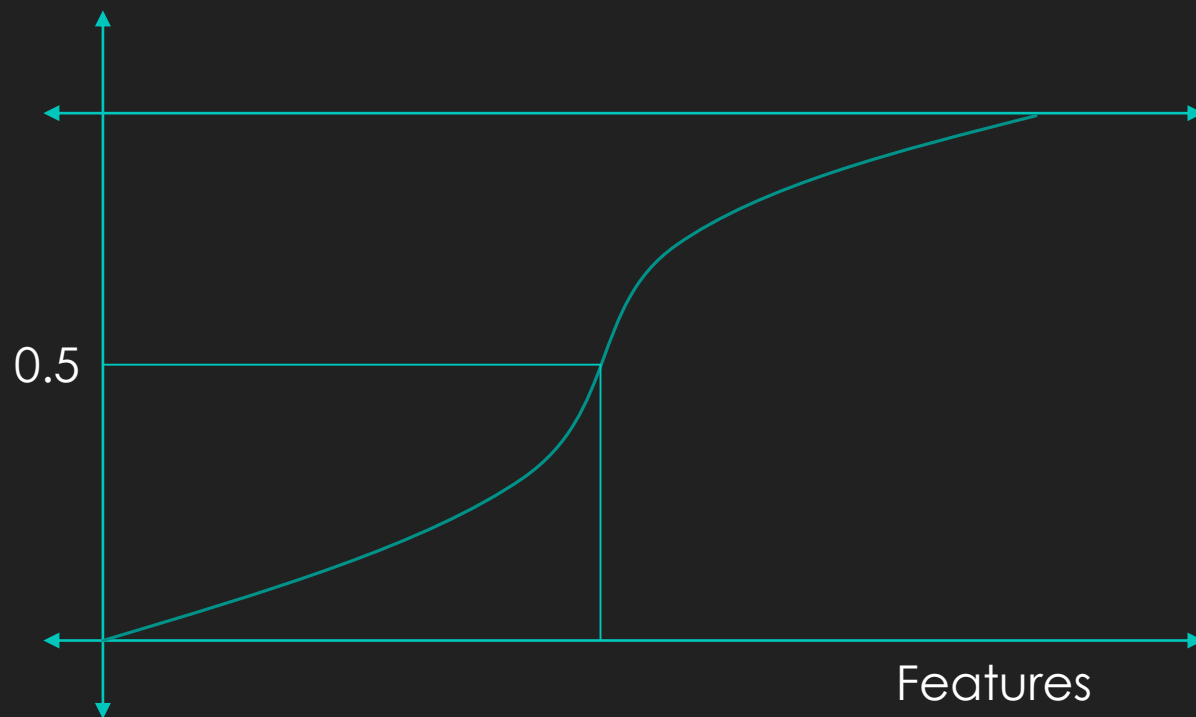
For  $y \geq 0$ ,  $y = e^{(a+bx_1+cx_2+\dots+dx_{2000})}$

$$\text{For } y < 1, \quad y = \frac{e^{(a+bx_1+cx_2+\dots+dx_{2000})}}{e^{(a+bx_1+cx_2+\dots+dx_{2000})} + 1}$$

# Logistic Regression – Value range

$$\ln \left( \frac{y}{y-1} \right) = a + bx_1 + cx_2 + \cdots + dx_{2000}$$

# Logistic Regression – Graph



# Logistic Regression – Summary

- Logistic Regression calculates the values of the coefficients.
- Based on the calculated coefficients, new sentences are given points.
- If the point  $> 0.5$ , the sentence is predicted to be positive and vice versa