LOGISTIC REGRESSION

A REGRESSION WHICH IS NOT REGRESSION

What is Logistic Regression?

Form of regression that allows the prediction of discrete variables by a mix of continuous and discrete predictors.

Logistic regression is often used bec
relationship between the DV (a dis
and a predictor is non-linear

Discrete data can only take on certain individual values.

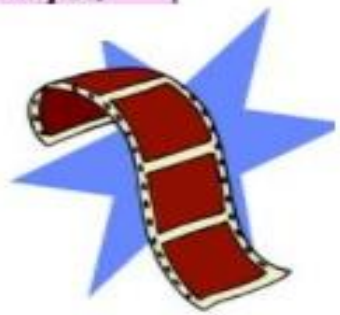
Continuous data can take on any value in a certain range.

Example 1

Number of pages in a book is a **discrete variable**.

Example 2

Length of a film is a **continuous variable**.



Example 3

Shoe size is a **Discrete variable**. E.g. $5, 5\frac{1}{2}, 6, 6\frac{1}{2}$ etc. <u>Not</u> in between.

Example 4

Temperature is a **continuous variable**.

Example 5

Number of people in a race is a **discrete variable**.

Example 6

Time taken to run a race is a **continuous variable**

Discrete Variables

Whether or not a Person Smokes

$$Y = \begin{cases} \text{Non} - \text{smoker} \\ \text{Smoker} \end{cases}$$

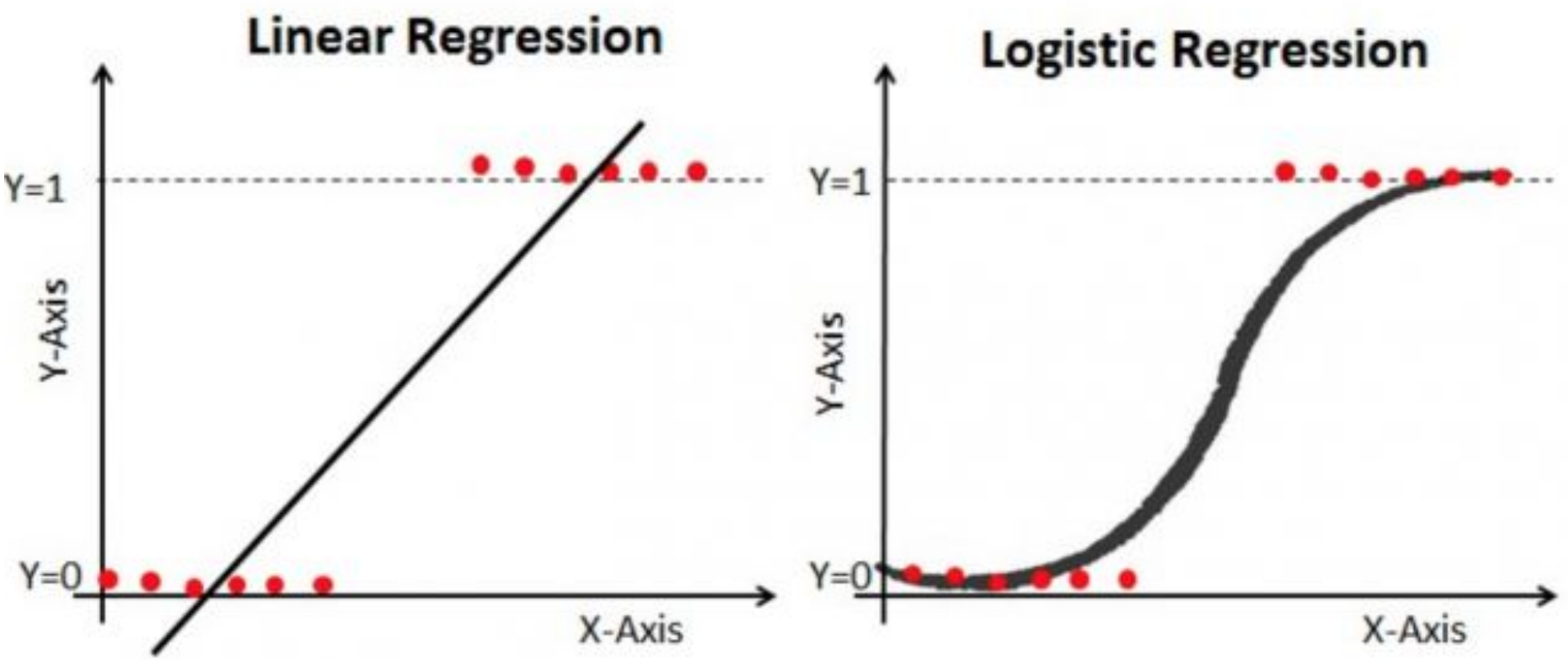
Success of a Medical Treatment

$$Y = \begin{cases} \text{Survives} \\ \text{Dies} \end{cases}$$

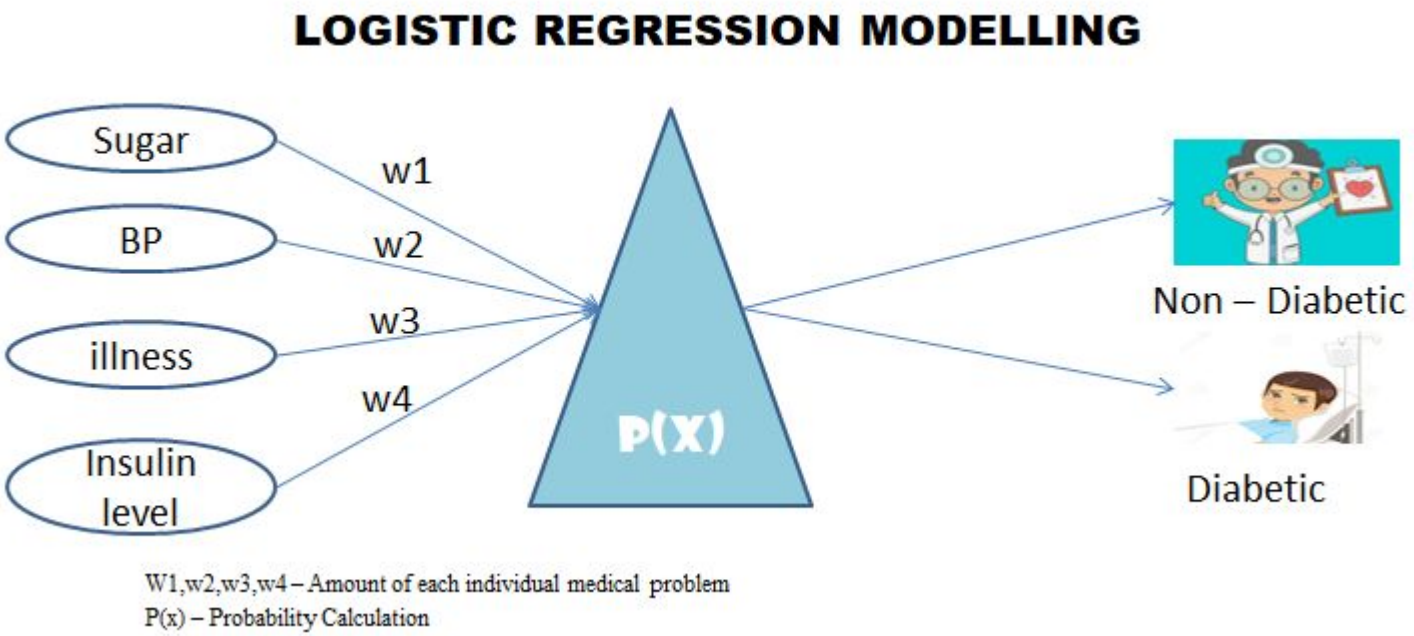
Opinion Poll Responses

$$Y = \begin{cases} \text{Agree} \\ \text{Neutral} \\ \text{Disagree} \end{cases}$$

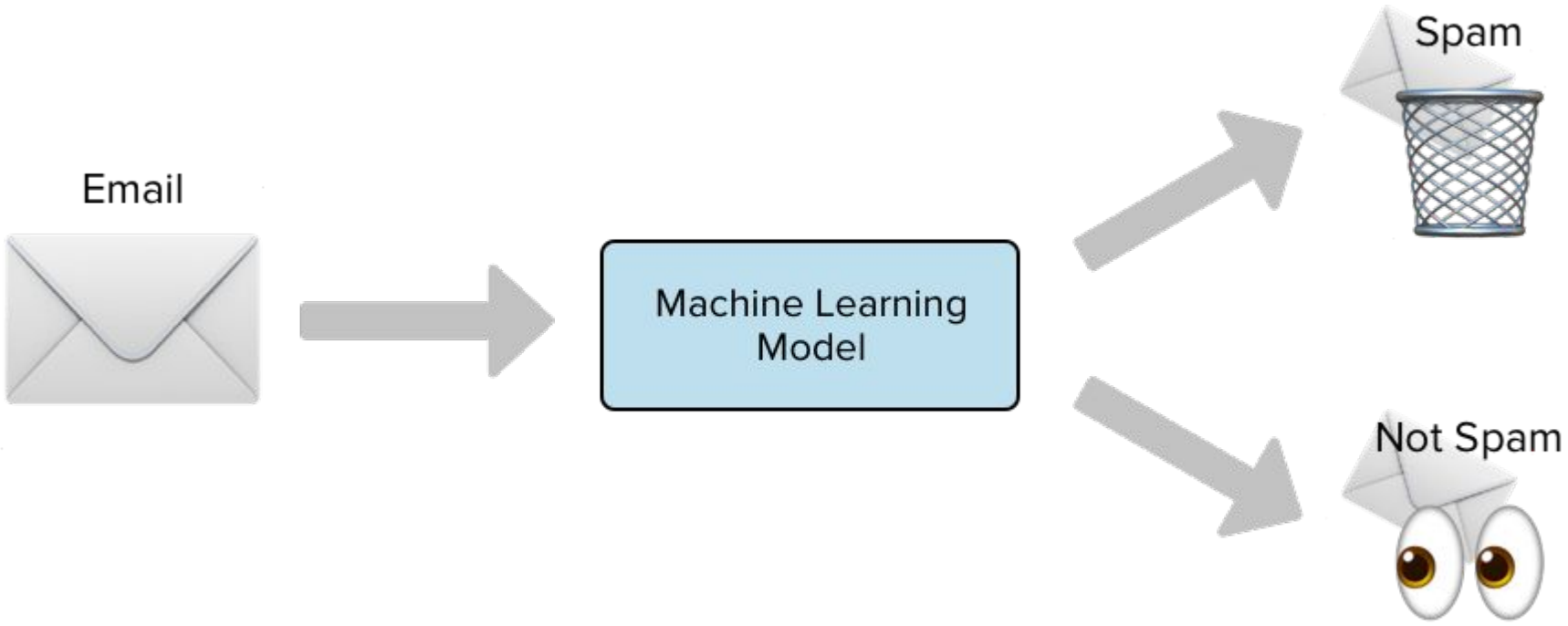
Linear vs Logistic Regression



Use Case 1 : Diabetic Prediction



Use Case 2: Spam Detection

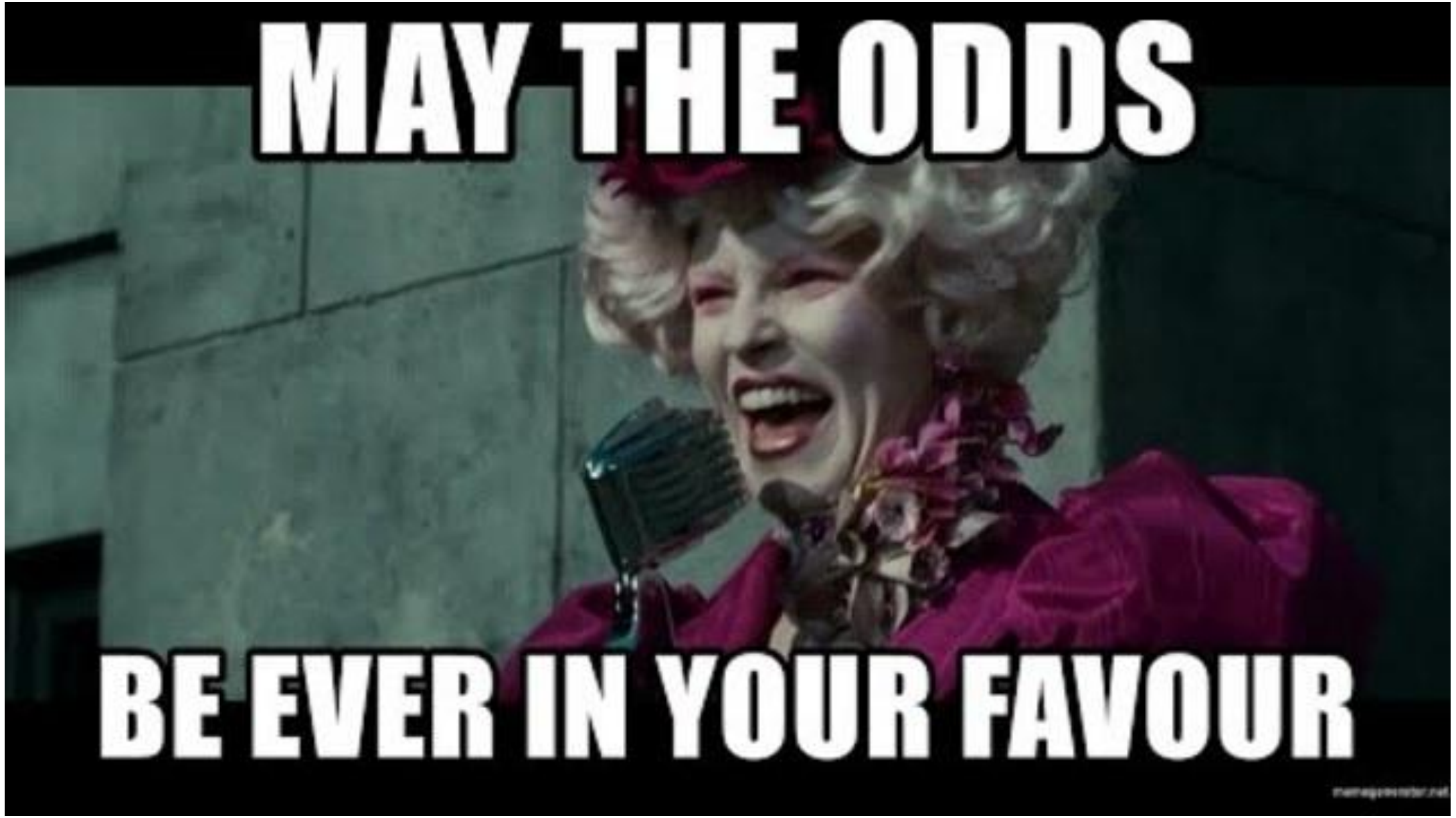


Background

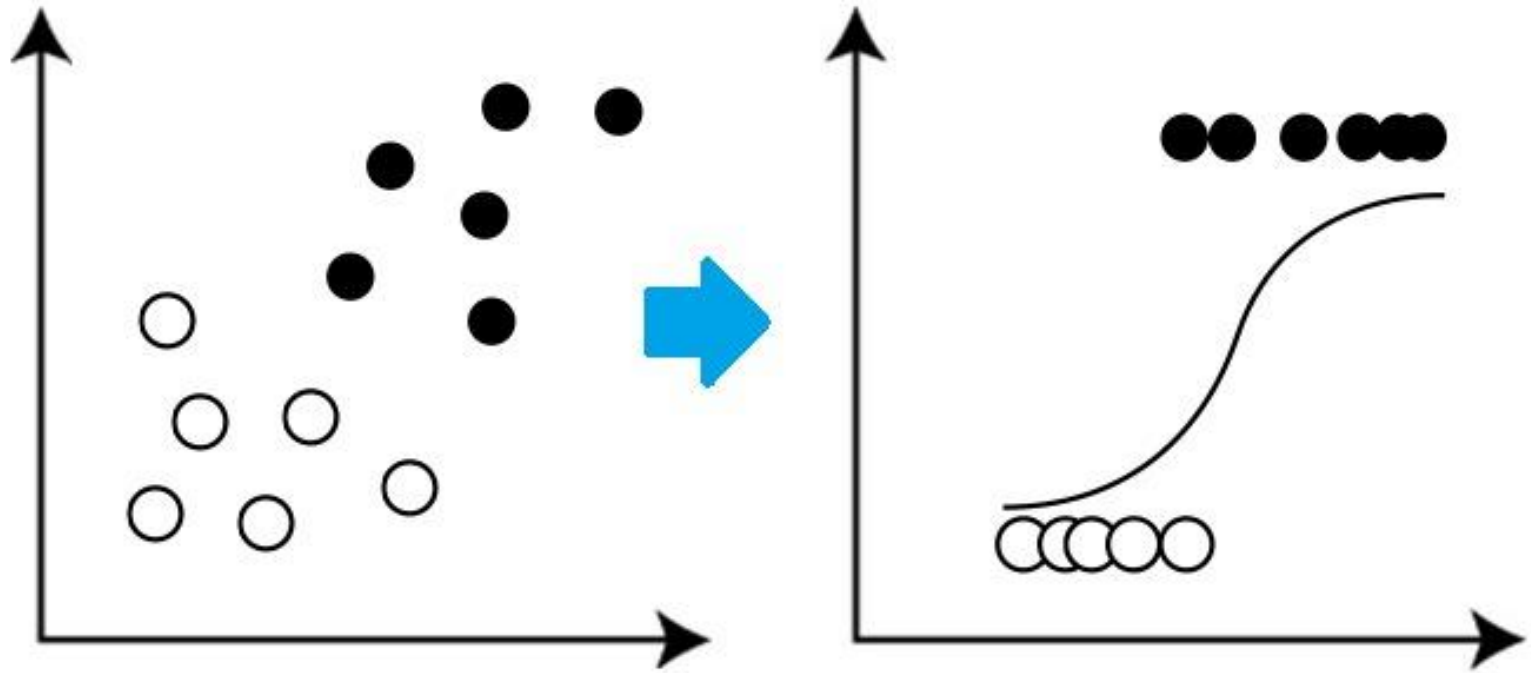
Y = A BINARY RESPONSE (DV)

1 POSITIVE RESPONSE (Success)
- P

0 NEGATIVE RESPONSE (Failure)
- Q = (1-P)

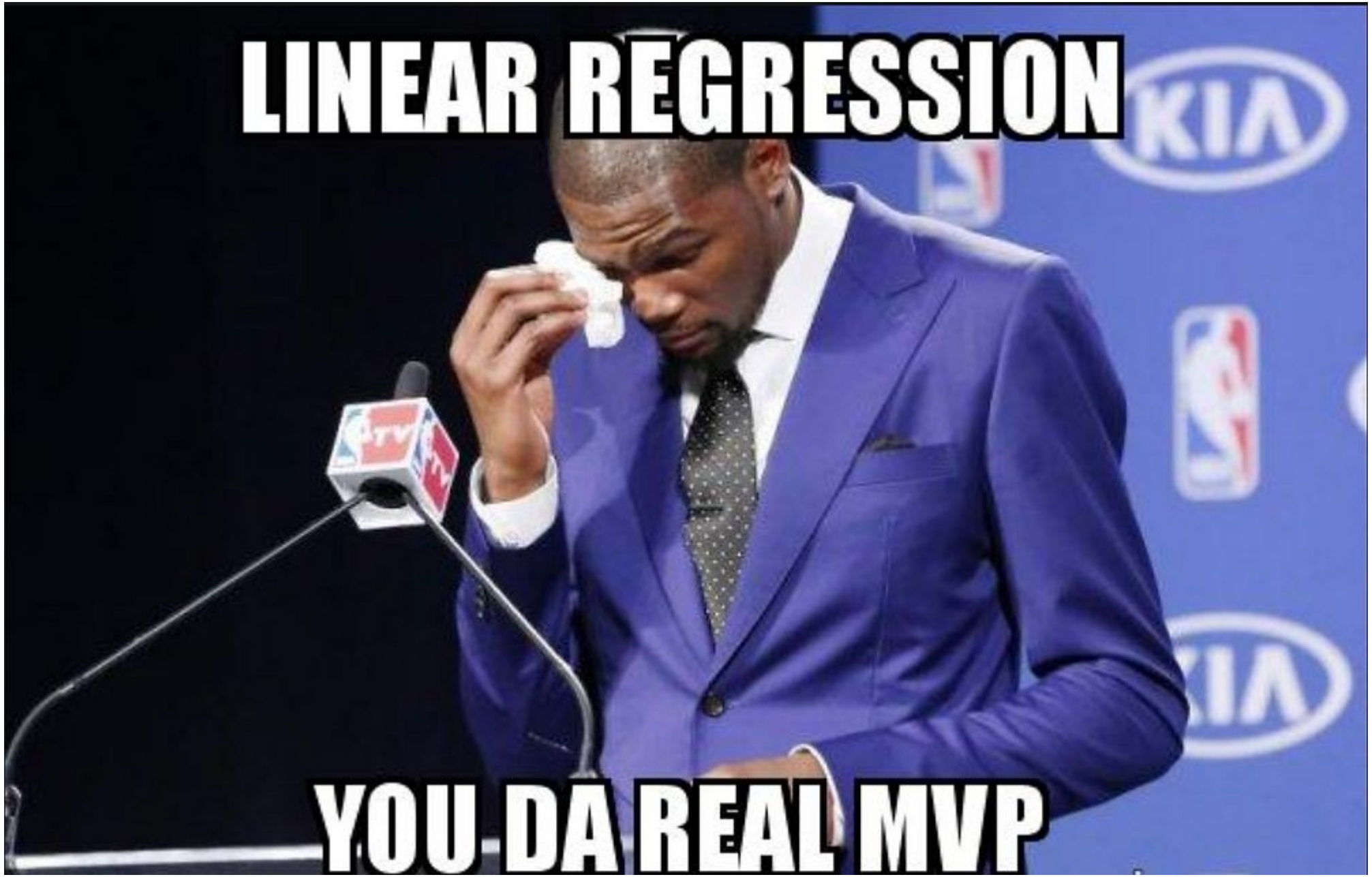

MAY THE ODDS
BE EVER IN YOUR FAVOUR

LOGISTIC REGRESSION

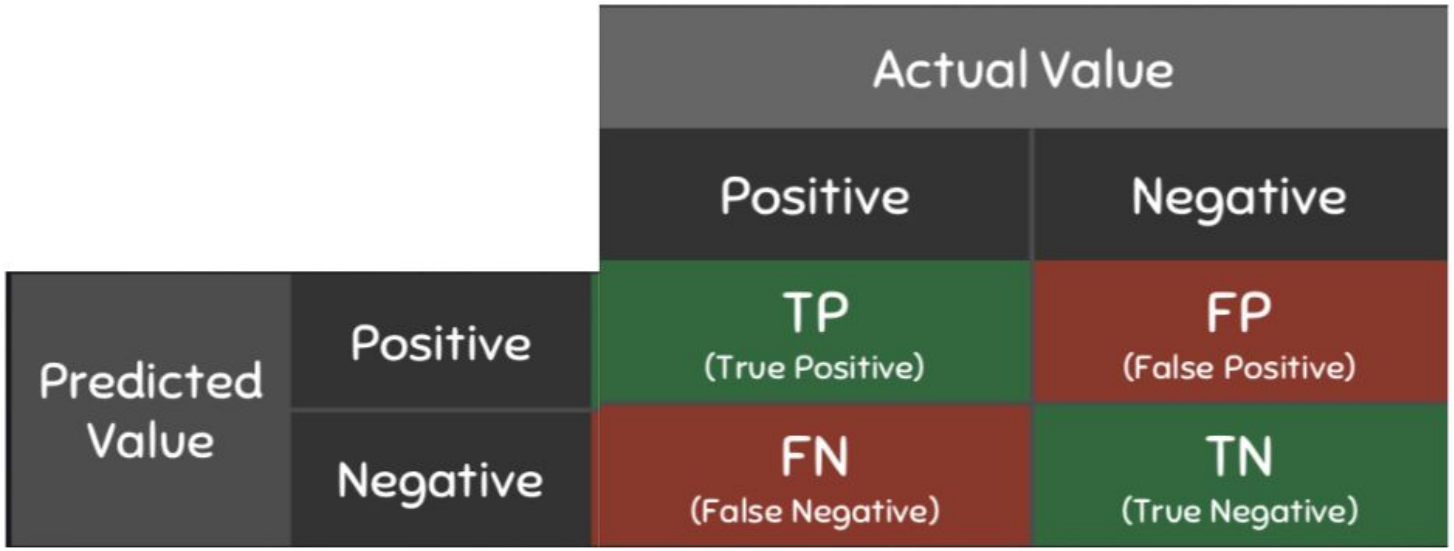


Assumptions

- The only "real" limitation on logistic regression is that the outcome must be discrete.
- If the outcome is continuous then multiple regression is more powerful given that the assumptions are met.
- Ratio of cases to variables – using discrete variables requires that there are enough responses in every given category
- Linearity in the logit – the regression equation should have a linear relationship with the logit form of the DV. There is no assumption about the predictors being linearly related to each other.
- Absence of multicollinearity
- No outliers


LINEAR REGRESSION
KIA
YOU DA REAL MVP

Confusion Matrix



- True Positive (TP) : Observation is positive, and is predicted to be positive.
- False Negative (FN) : Observation is positive, but is predicted negative.
- True Negative (TN) : Observation is negative, and is predicted to be negative.
- False Positive (FP) : Observation is negative, but is predicted positive.

Classification Metrics

Accuracy:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

Recall:

$$Recall = \frac{TP}{TP + FN}$$

Precision:

$$Precision = \frac{TP}{TP + FP}$$

F_1 score:

$$F_1 = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}}$$

Classification or Regression?

- No of eggs in a basket =
- No of kids in a class =
- Volts of electricity=
- No of Facebook likes =
- Wind Speed =
- Water temperature. =


I DON'T KNOW HOW TO USE LOGISTIC REGRESSION
AND AT THIS POINT I'M TOO AFRAID TO ASK
memegenerator.net