# Logistic regression

Logistic regression is a statistical method used for modelling the probability of binary outcome. It is commonly employed in situations where the dependent variable is categorical and represents two classes, often coded as 0 and 1. Logistic regression models the relationship between the independent variables and the probability of belonging particular category

## Key points

* Binary Outcome:
  + It is used when the dependent variable is binary or dichotomous indicating two possible outcomes (e.g. yes/no, success/failure etc)
* Sigmoid function:
  + The logistic regression model used the sigmoid (logistic) function to transform a linear combination of the independent variables into a probability between 0 and 1.
* Log-Odds Transformation:
  + The linear combination of independent variables is transformed into log-odds using logistic function. The log-odds represent the logarithm of the odds of the event occurring.
* Equation:
  + The logistic regression equation of single predictor is given by:
    - Where p is the probability of the event, is the intercept, is the coefficient for the predictor x.
* Odds Ratio:
  + The odds ratio represents the multiplicative change in odds for a one-unit change in the predictor variable.
* Maximum Likelihood Estimation (MLE):
  + Logistic regression estimates the parameters (coefficients) using the maximum likelihood estimation method, maximising the likelihood of observing the given outcomes.
* Assumptions:
  + The data should be or almost linearly separable.

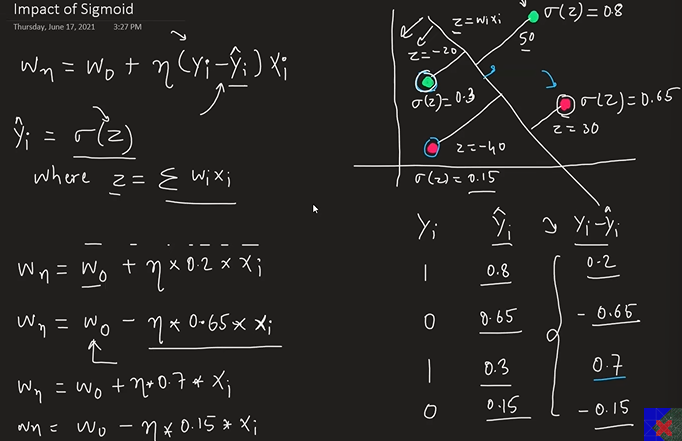
### Logistic Regression Training Process:

1. **Initialization:**
   * Initial weights are assigned randomly.
2. **Sigmoid Function:**
   * The sigmoid function (σ) is used to transform the weighted sum (z) into a probability value between 0 and 1.
3. **Weight Adjustment:**
   * The weights are adjusted based on the difference between the predicted (​) and actual () values using the gradient descent algorithm.
4. **Decision Boundary:**
   * The decision boundary is determined by the weights, and the sigmoid function ensures the output falls between 0 and 1.
5. **Loss Function:**
   * The loss function measures the difference between the predicted and actual values. Commonly used is the binary cross-entropy loss function for binary classification.
6. **Gradient Descent:**
   * Gradient descent is employed to minimize the loss function, finding the optimal weights that result in the best fit line.
7. **Batch or Stochastic Gradient Descent:**
   * The optimization process can involve batch gradient descent (updating weights after evaluating the entire dataset) or stochastic gradient descent (updating weights after each data point).

### Key Differences:

* **Sigmoid Activation:**
  + Logistic regression uses the sigmoid function to produce probabilities, transforming the linear combination of weights and features into a range suitable for classification.
* **Loss Function:**
  + The loss function for logistic regression is based on the negative log-likelihood of the predicted probabilities, penalizing the model more for confidently incorrect predictions.
* **Gradient Descent:**
  + The weights are updated using the gradient of the loss function with respect to the weights, ensuring a systematic approach to finding the optimal weights.

In above function we are using step function where the output is either 0 or 1. In this case to make the transformation possible we have to use the sigmoid function. The sigmoid converts – infinity to +infinity between the range 0 and 1. With this value we can update in above equation



With the help of loss function we can find the best fit line which will separate the classifications correctly. There might be multiple lines to find the best fit line then we have to use loss function.

Random weights are given initially and then the loop is run to find the best weight. Based on misclassification points it will adjust the weights and draw the line which correctly separates the data points correctly either number of loops provided or convergence is defined where all the class is classified correctly. Whatever, outputs falls or greater than 0 then its positive region and if its less than 0 then it falls into negative region. Transformation of line comes after adjusting the weights (coefficients) of independent variable. Based on the co-ordinates of the independent variable, its adjusted (added/subtracted) to random weights assigned in first instance. However, in machine learning, the weights are adjust with specific learning rate which doesn’t make drastically changes in each iterations.