Logistic Regression

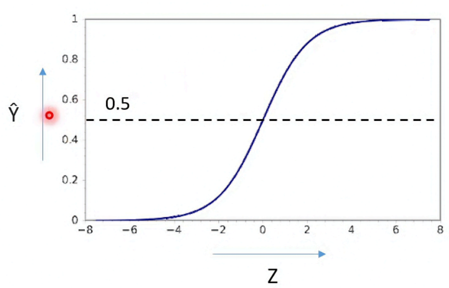
Created	@August 3, 2025
Edited	@August 5, 2025 6:55 AM
Archive	☐

Intuition:

Features:

1. It is a supervised learning model
2. Classification model
3. Best for binary classification problem
4. Uses Sigmoid function

Sigmoid Function:



$$\hat{Y} = \frac{1}{1+e^{-Z}} \qquad Z = w.X + b$$

Sigmoid Function

$\hat{Y}$ - Probability that (y = 1)

$\hat{Y}$ = P(Y=1 | X)

X - input features

w – weights
(number of weights is equal to the number of input features in a dataset)

b - bias

$\hat{Y} = \sigma(Z)$

`Y^ is probability of Y being 1 for given value of X`

Advantages of LR:

1. Easy to implement
2. Performs well on data with linear relationship
3. Less prone to overfitting for low dimensional dataset

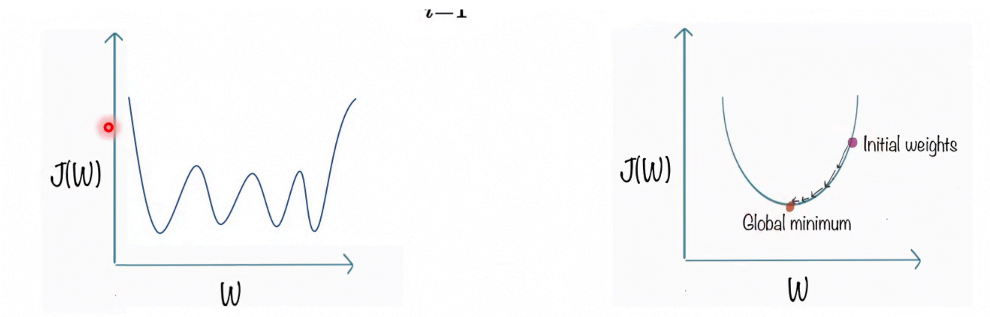
Disadvantages:

1. High dimensional data causes overfitting
2. Difficult to capture complex relationships in a dataset
3. Sensitive to Outliers
4. Needs a larger dataset

Math Behind Logistic Regression:

Sigmoid function

Loss Function:



Gradient Descent
With Local minima

Gradient Descent
With Global minima

Loss function of logistic Regression:

Binary Cross Entropy Loss Function or Log Loss:

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

The value of y can be either zero or one and y^ lies between zero and one.

When y = 1

$$L(1, \hat{y}) = -(1 \underline{\log \hat{y}} + \underline{(1-1) \log (1 - \hat{y})}) \Rightarrow L(1, \hat{y}) = -\log \hat{y}$$

we always want a smaller Loss function value, hence y^ should be very large(Closer to 1), so that (-log y^) will be a large negative number.;

When y = 0

$$L(0, \hat{y}) = -(0 \log \hat{y} + (1-0)\log(1-\hat{y})) \quad \Rightarrow \quad L(0, \hat{y}) = -\log(1-\hat{y})$$

In this case y^ should be very small(Closer to zero), so that -log(1-y^) will be a large negative number.

Loss function mainly applies for a single training set as compared to the cost function which deals with a penalty for a number of training sets or the complete batch

Cost function for Logistic Regression:

$$J(w, b) = \frac{1}{m}\Sigma(L(y^{(i)}, \hat{y}^{(i)})) = -\frac{1}{m}\Sigma(y^{(i)}\log\hat{y}^{(i)} + (1-y^{(i)})\log(1-\hat{y}^{(i)}))$$

('m' denotes the number of data points in the training set)

It is just average of all the loss function.

It changes with weight and bias of Model.

Gradient Descent for Logistic Regression:

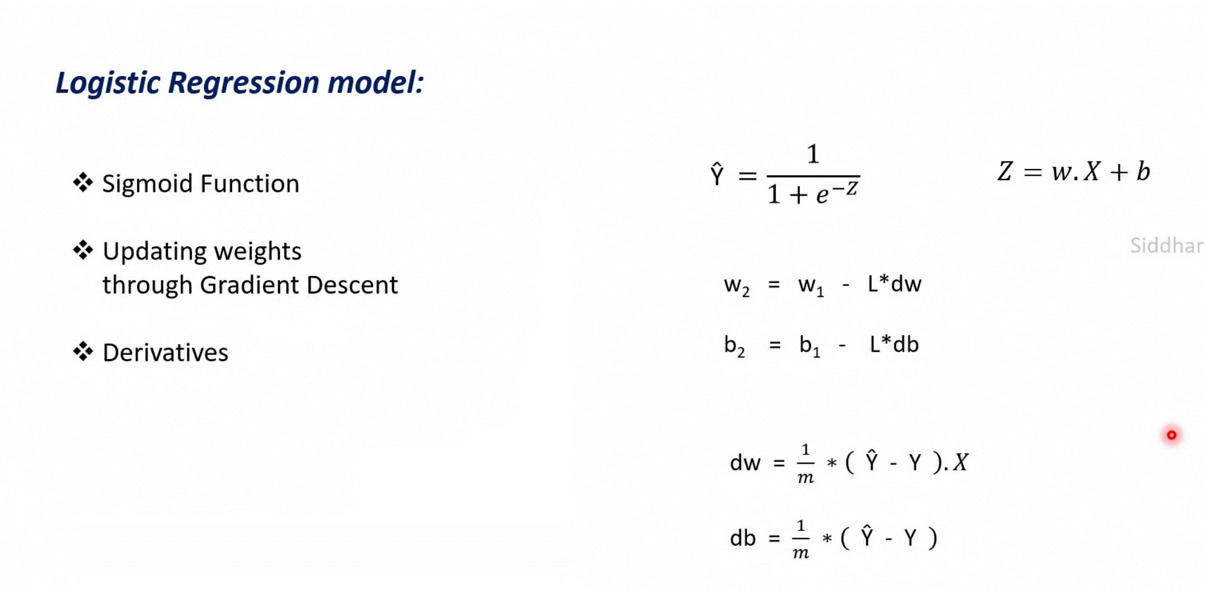
$$dw = \frac{1}{m} * (\hat{Y} - Y).X$$

$$db = \frac{1}{m} * (\hat{Y} - Y)$$

These equations are for Logistic Regression

m is the number of data points in the dataset

Summary:



Workflow of the logistic regression model:

1. Set learning rate and number of iterations; initiate random weight and bias value.
2. Build logistic regression function (Sigmoid Function)
3. Update the parameters using gradient descent —> we will get the best model with minimum cost function
4. Build the predict function to determine the class of the data point