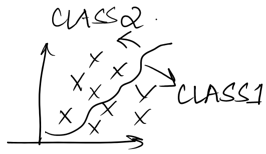
Logistic Regression

26 March 2023 21:33

1. Sigmoid Function
2. Hypothesis
3. Decision Boundary
4. log loss Function



$n \rightarrow$ number of features

$m \rightarrow$ no of data pts.

$X \rightarrow$ input data $(m \times n)$

$y \rightarrow$ target/pred/class

$X(i), y(i) \rightarrow i^{th}$ training ex.

$W \rightarrow$ weight (parameter) of $n \times 1$

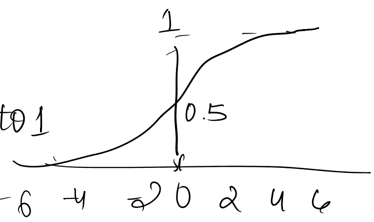
$b \rightarrow$ bias (parameter)

yhat $\rightarrow$ hypothesis (o/p value b/w 0 & 1)

BINARY CLASSIFICATION

$n = 2$

Sigmoid $f^n = \dfrac{1}{1+e^{-z}}$ $\rightarrow$ all i/p in 0 to 1



hypothesis θ_1 θ_2

$$\text{y-hat} = W.X + b \longrightarrow \text{Linear}$$

$$\text{y-hat} = \text{sigmoid}(W.X+b)$$

$$= \frac{1}{1+e^{-(WX+b)}}$$

log loss f^n

Error $\rightarrow$ minimized.

$LR \rightarrow MSE$

$$P(y=1 \mid X; W,b) = \text{y-hat}$$

$$P(y=0 \mid X; W,b) = 1-\text{y-hat}.$$

$$J(W,b) = \frac{1}{m}\sum_{i=1}^{m} L(\hat{y}^{(i)}, y^{(i)}) = -\frac{1}{M}\sum_{i=1}^{m}\left[y^{(i)}\log \hat{y}^{(i)}\right] + \left[(1-y^{(i)})\log(1-\hat{y}^{(i)})\right]$$

* Binary cross entropy loss f^n

Gradient Descent

$$W = W - lr * dW$$
$$b = b - lr * db$$

$\gg$ partial derivative w.r.t W & b

$W = W - lr * dW$

$b = b - lr * db$ >> partial derivative of loss f'n wrt W & b

$$dW = \frac{1}{m} * (y\text{-hat} - y) X$$

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

Decision Boundary (2 classes)

$y = 1 \Rightarrow$ when $y\text{-hat} >= 0.5 \Rightarrow W \cdot X + b \geq 0$

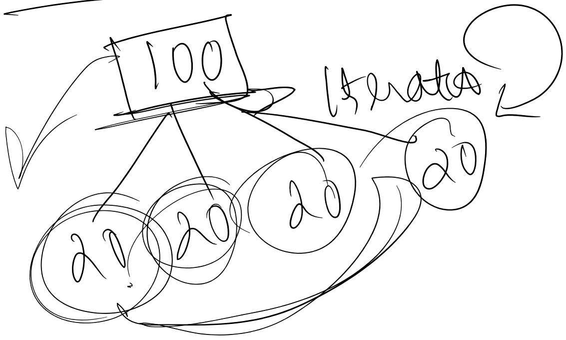
$y = 0 \Rightarrow$ when $y\text{-hat} < 0.5 \Rightarrow W \cdot X + b < 0$

This means (sigmoid graph)

$y = 1$ when $W \cdot X + b \geq 0$ → Decision Boundary

$y = 0$ when $W \cdot X + b < 0$.

Mini Batch GD



for all data pts