Cost function

Logistic Regression

$-\log(0) = \infty$

Log loss

Error function

$$\text{cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & y = 1 \\ -\log(1-h_\theta(x)) & y = 0 \end{cases}$$

$h_\theta(x)$: Prediction

y: actual

$-\log(1) \approx 0$

$-\log(1-1) = \infty$

$-\log(1-0) \approx 0$

Diabetic → 1 — y

Diabetic → 0 — y

Best cases

Actual Case → $y = 1$
Prediction → $\hat{y} = 1$
Error = 0

Actual case → $y = 0$
Prediction → $\hat{y} = 0$
Error = 0

$\hat{y}$: $h_\theta(x)$

1. $\hat{y} = mx + c$ → sigmoid function

2. $h_\theta(x) = \theta_0 + \theta_1 x$

θ_0: intercept, c

θ_1: slope, m

Worst case

Actual case $\rightarrow y = 1$

Prediction $\rightarrow \hat{y} = 0$

Error $= \infty$

Actual $\rightarrow y = 0$

Prediction $\rightarrow \hat{y} = 1$

Error $= \infty$

Goal $\rightarrow$ Reduce the error value

combined above two equation

$$\boxed{cost(h_\theta(x), y) = -y_i \log(h_\theta(x_i)) - (1-y_i)\log(1-h_\theta(x_i))}$$

Actual label/ data of one class

Prediction

Actual data/label of another class

$\theta_0 + \theta_1 x$

$mx + c$

$$MSE = \frac{1}{2m}\sum_{i=1}^{m}\left(\sigma(i) - y_i\right)^2$$

$\sigma(i)$ → sigmoid function

Why MSE not suitable for Logistic Regression??

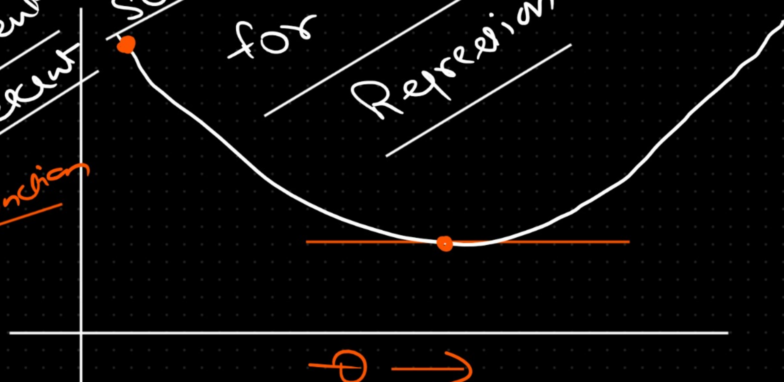
$$\Rightarrow \frac{1}{2m}\sum_{i=1}^{m}\left(\frac{1}{1+e^{-(\theta_0+\theta_1 x)}} - y_i\right)^2$$

Multiple local minima → difficulty to optimize

Plot → $\boxed{\text{Non-Convex curve}}$

Convex curve → single local & global minima

Traditional gradient descent suitable for linear Regression



Not suitable traditional gradient descent