$y = f(x) \rightarrow w^T x$

$= \sigma(w^T x)$

X

cabin

$x_1 w_1 + x_2 w_2 + \ldots x_{12} w_{12} = \hat{y}$

mean, median

missing

assumption

$x_1 \times x_3 \rightarrow$ correlated

Transformations (1) Categorical

2) Standardization One-hot encoding

A B C → 1 0 0 / 0 1 0

Logistic Regression (0/1)

MLE → & NLL

$\text{Ber}(y \mid \sigma(w^T x + b))$ → $\hat{y}$

min $\text{Loss} = -\left(y \log \hat{y} + (1-y)\log(1-\hat{y})\right)$

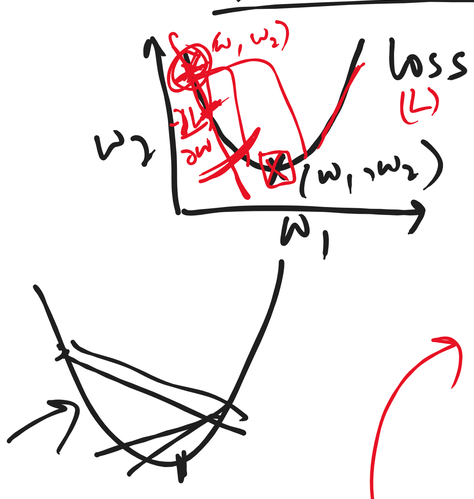
Cross entropy

$\log \hat{y}$ if $\hat{y} \rightarrow 1$

Titanic : X → y

Implement → optimal parameters model

optimal weights.



Gradient Descent

(Algorithm)

1) Initialize parameters (w_1, w_2)

2) $A = \sigma(w^T x + b)$ → $\hat{y}$ calculated → forward pass

$A = [a^{(1)}, a^{(2)} \ldots a^{(n)}]$

$a^{(1)} = w^T x_1 + b$

3) $\frac{\partial L}{\partial w} = \frac{1}{N} \times (A - Y)^T$

$\frac{\partial L}{\partial b} = \frac{1}{N} \sum_{i=1}^{N} (a^{(i)} - y^{(i)})$] optional

4) Update weights (parameters)

$w_{t+1} = w_t - \eta \frac{\partial L}{\partial w}$, η → learning rate

$b_{t+1} = b_t - \eta \frac{\partial L}{\partial b}$

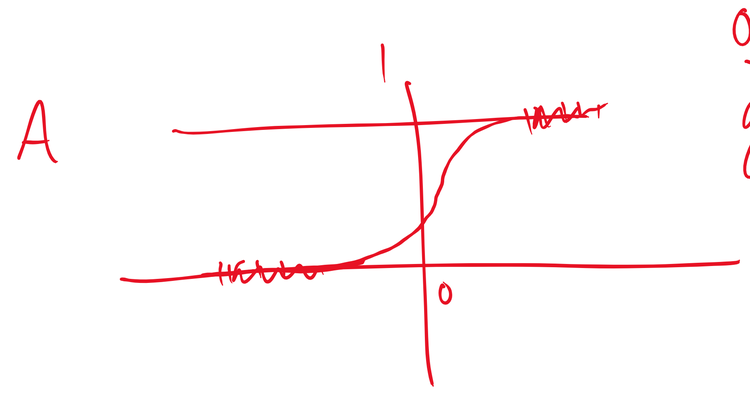
hyperparameters

Repeat epochs

t → epoch

100

t = 0



0.0001

get rid of sigmoid

$\frac{\partial L}{\partial W}$ → Batch : entire dataset

Mini Batch → Batch size

Stochastic : one datapoint