$$LL(\theta) = \prod_{i=1}^{m} [h_\theta(x)]^{y} [1 - h_\theta(m)]^{1-y}$$

log likelihood

$$\log LL(\theta) = \log \left(\prod_{i=1}^{m} [h_\theta(x)]^{y} [1 - h_\theta(m)]^{1-y} \right)$$

$$= \sum_{i=1}^{m} \left(\log \underbrace{[h_\theta(x^{(i)})]^{y^{(i)}}}_{a} \underbrace{[1 - h_\theta(m^{i})]^{1-y^{(i)}}}_{b} \right)$$

$\log(a \cdot b \cdot c \cdot d) = \log a + \log b + \log c + \log d$

$$= \sum_{i=1}^{m} \left(\log [h_\theta(x^{(i)})]^{y^{(i)}} + \log [1 - h_\theta(m^{i})]^{1-y^{(i)}} \right)$$

$$\log LL(\theta) = \sum_{i=1}^{m} \left(y^{(i)} \log [h_\theta(x^{(i)})] + (1-y^{(i)}) \log [1 - h_\theta(m^{i})] \right)$$

$$\log LL(\theta) = \sum_{i=1}^{m} \left(y^{(i)} \log \frac{1}{1+e^{-\theta^T x^{(i)}}} + (1-y^{(i)}) \log \left(1 - \frac{1}{1+e^{-\theta^T x^{(i)}}} \right) \right)$$

$$\log LL(\theta) = \sum_{i=1}^{m} \left(y^{(i)} \log \frac{1}{1+e^{-\theta^T x^{(i)}}} + (1-y^{(i)}) \log \left(\frac{e^{-\theta^T x^{(i)}}}{1+e^{-\theta^T x^{(i)}}} \right) \right)$$

↳ **log likelihood**

↳ optimize

hypo
error / objec
gradient

$$= \frac{\partial}{\partial\theta}\left[\sum_{i=1}^{m}\left(y^{(i)}\log\frac{1}{1+e^{-\theta^T x^{(i)}}} + (1-y^{(i)})\log\left(\frac{e^{-\theta^T x^{(i)}}}{1+e^{-\theta^T x^{(i)}}}\right)\right)\right]$$

$$= \sum_{i=1}^{m}\frac{\partial}{\partial\theta}\left(y^{(i)}\log\frac{1}{1+e^{-\theta^T x^{(i)}}} + (1-y^{(i)})\log\left(\underbrace{\frac{\overbrace{e^{-\theta^T x^{(i)}}}^{b}}{1+e^{-\theta^T x^{(i)}}}}_{a}\right)\right)$$

$$= \sum_{i=1}^{m}\frac{\partial}{\partial\theta}\left(y^{(i)}\log\frac{1}{1+e^{-\theta^T x^{(i)}}} + (1-y^{(i)})\log\left(\frac{1}{1+e^{-\theta^T x^{(i)}}}\right) + (1-y^{(i)})\log\left(e^{-\theta^T x^{(i)}}\right)\right)$$

$$= \sum_{i=1}^{m}\frac{\partial}{\partial\theta}\left(\log\left(\frac{1}{1+e^{-\theta^T x^{(i)}}}\right) + (1-y^{(i)})\log\left(e^{-\theta^T x^{(i)}}\right)\right)$$

$$= \sum_{i=1}^{m}\frac{\partial}{\partial\theta}\left(\log\left(1+e^{-\theta^T x^{(i)}}\right)^{-1} + (1-y^{(i)})\log\left(e^{-\theta^T x^{(i)}}\right)\right)$$

$$= \sum_{i=1}^{m}\frac{\partial}{\partial\theta}\left(-\log\left(1+e^{-\theta^T x^{(i)}}\right) + (1-y^{(i)})\left(-\theta^T x^{(i)}\right)\right)$$

$$= \sum_{i=1}^{m} \left[\frac{\partial}{\partial \theta} \left(-\log \left(1 + e^{-\theta^T x^{(i)}} \right) \right) + \frac{\partial}{\partial \theta} (1 - y^{(i)}) (-\theta^T x^{(i)}) \right]$$

$$= \sum_{i=1}^{m} \left[\frac{-1}{1+e^{-\theta^T x^{(i)}}} \left(e^{-\theta^T x^{(i)}} \right) \underline{\underline{(-x^{(i)})}} + (1 - y^{(i)}) \underline{\underline{(-x^{(i)})}} \right]$$

$$= \sum_{i=1}^{m} \left[(1 - y^{(i)}) - \frac{e^{-\theta^T x^{(i)}}}{1+e^{-\theta^T x^{(i)}}} \right] (-x^{(i)})$$

$$= \sum_{i=1}^{m} \left[1 - \frac{e^{-\theta^T x^{(i)}}}{1+e^{-\theta^T x^{(i)}}} - y^{(i)} \right] (-x^{(i)})$$

$$= \sum_{i=1}^{m} \left[\frac{1}{1+e^{-\theta^T x^{(i)}}} - y^{(i)} \right] (\underline{\underline{-}}x^{(i)})$$

$$\frac{\partial}{\partial \theta} (\log L(\theta)) = \sum_{i=1}^{m} \left[y^{(i)} - \frac{1}{1+e^{-\theta^T x^{(i)}}} \right] (x^{(i)})$$

$$\longrightarrow \frac{1}{m} \sum_{i=1}^{m} \left[y^{(i)} - \frac{1}{1+e^{-\theta^T x^{(i)}}} \right] (x^{(i)})$$

Instead of optimizing LL(θ)
optimizing average LL(θ)

$$\frac{\partial}{\partial\theta}\left(\log LL(\theta)\right) = \frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - h_\theta(x^{(i)})\right)x^{(i)}$$

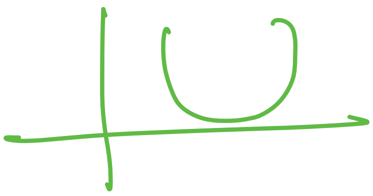
final Gradient

Linear Regression

$\theta^T x^{(i)}$

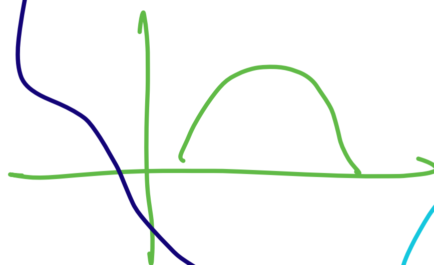
Logistic Regression

$$\frac{1}{1+e^{-\theta^T x^{(i)}}}$$



$$\theta = \theta - \eta \frac{\partial J(\theta)}{\partial\theta}$$

Gradient Descent



$$\theta = \theta + \eta \frac{\partial LL(\theta)}{\partial\theta}$$

Gradient Ascent

0.67

≥ 0.5 1

0.33

Start with random values of $\theta_0, \theta_1 \ldots\ldots \theta_n$

do

{

how good your θ is?

update θ's

} while (convergence)

$$\sum_{i=1}^{m}\left(y^{(i)}\log\frac{1}{1+e^{-\theta^T x^{(i)}}}+(1-y^{(i)})\log\left(1-\frac{1}{1+e^{-\theta^T x^{(i)}}}\right)\right)$$

$$\left[\sum_{i=1}^{m}\left(y^{(i)}\log h_\theta(x^{(i)})+(1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right)\right]$$

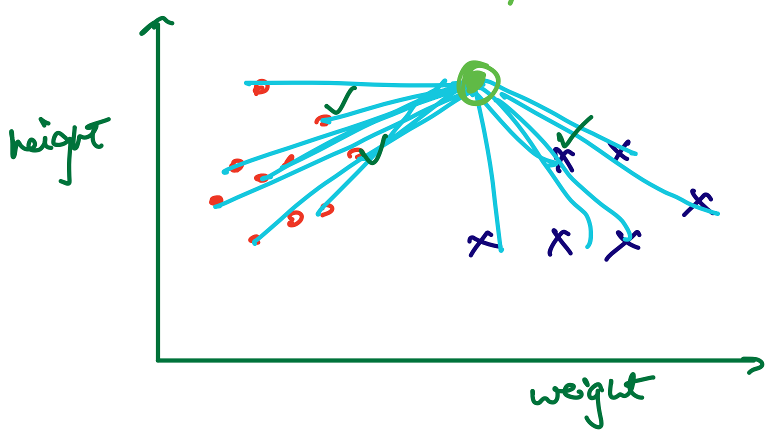
Binary Cross Entropy

Linear Reg.	Logistic Reg.
- Regression	- Classification
- MSE (Minimise)	- Binary Cross Entropy (maximise)
- Gradient Descent $\theta = \theta - \eta \frac{\partial L}{\partial \theta}$	- Gradient Ascent $\theta = \theta + \eta \frac{\partial LL(\theta)}{\partial(\theta)}$

KNN (K Nearest Neighbours)

→ Supervised ML Algo → y value given

→ Classification & Regression



① Distance b/w test point and all the training points

—, —, —, —, — . . .

(2) Distance sort inc order

— small
—
—
—
—
— large

(3) k given
↳ 3

a small
b
c
—
—
— large

} 3 data points closest to your test data point

(4) a → 1
b → 1
c → 0
} majority vote

(5) test data point class 1.

Training Time: $O(1)$

Test Time:

1 test data point : $M + \underbrace{M\log M}_{\text{sorting}} + k$ ✓
M → Distance

OR

$M + M\log k$ ✓

q test datapoint : $q(M + M\log k)$

eg.

N

2, 28, 10, 12, 3, 6, 20, 24

k largest elements

↳ 3

→ N log N sorting Algo

$2, 3, 6, 10, 12, \underbrace{20, 24, 28}_{3}$

N log N + k

min heap

$\cancel{2}, 28, \cancel{10}$

$\cancel{12}, 20, 24$

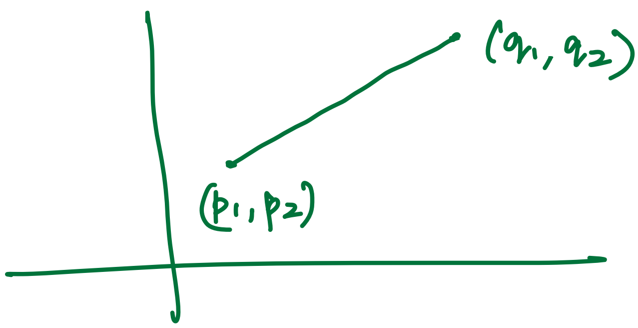
heap size k

1 deletion: log k

N deletion: $\underline{N \log k}$

Distance:

1 Euclidean Distance

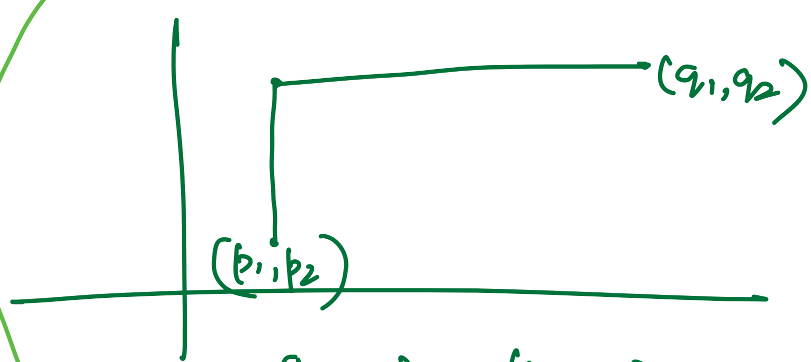


$$\sqrt{(q_1-p_1)^2 + (q_2-p_2)^2}$$

k=2

$$\left(\sum_{i=1}^{n} (q_i - p_i)^2\right)^{1/2}$$

2 Manhattan Distance



$$(p_1-q_1) + (p_2-q_2)$$

$$\sum_{i=1}^{n} (p_i - q_i)$$

Minkowski Distance

$$\left(\sum_{i=1}^{n} |p_i - q_i|^k \right)^{1/k}$$

$k = 1$