

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

log likelihood

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

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

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

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

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

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

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

↳ log likelihood
↳ optimize

hypo
error / loss
gradient

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

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

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

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

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

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

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

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

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

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

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

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

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

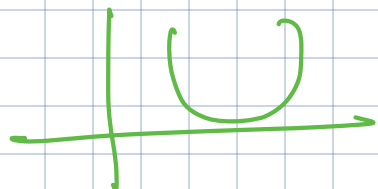
Instead of minimizing $J(\theta)$
 optimizing average $U(\theta)$

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

first Gradient

Linear Regression

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

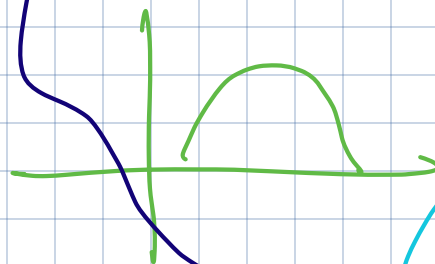


$$\theta = \theta - \eta \frac{\partial J(\theta)}{\partial \theta}$$

Gradient Descent

Logistic Regression

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



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

Gradient Ascent

$$\frac{0.67}{70.5} \approx 0.0095 \approx 0.01$$

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

do

{

how good your θ is?

update θ 's

} while (convergence)

↓

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

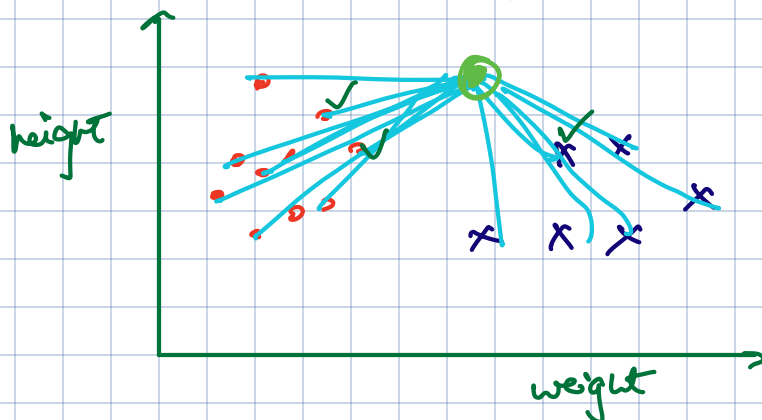
$$\left[\sum_{i=1}^m \left(y^{(i)} \log \sigma(x^{(i)}) + (1 - y^{(i)}) \log (1 - \sigma(x^{(i)})) \right) \right]$$

Binary Cross Entropy

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

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
 $\hookrightarrow 3$

$\left. \begin{array}{c} a \\ b \\ c \end{array} \right\}$ small } 2 data points closest to your test data point

— large

④ $\left. \begin{array}{l} a \rightarrow 1 \\ b \rightarrow 1 \\ c \rightarrow 0 \end{array} \right\}$ majority vote

⑤ test data point class 1.

Training Time: $O(1)$

Test Time:

Test data point: $M + \frac{M \log M + k}{2}$ ✓

Distance Sorting

OR

$M + M \log k$ ✓

q test data point: $q(M + M \log k)$