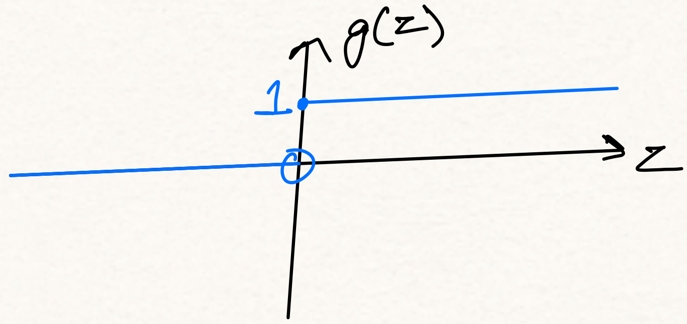
Classification

Perceptron Algorithm

$x^{(i)} \in \mathbb{R}^{d+1} \quad y^{(i)} \in \{0,1\}$

$h_\theta(x) = g(\theta^T x)$

$$g(z) = \begin{cases} 1 ; & z \geq 0 \\ 0 ; & z < 0 \end{cases}$$

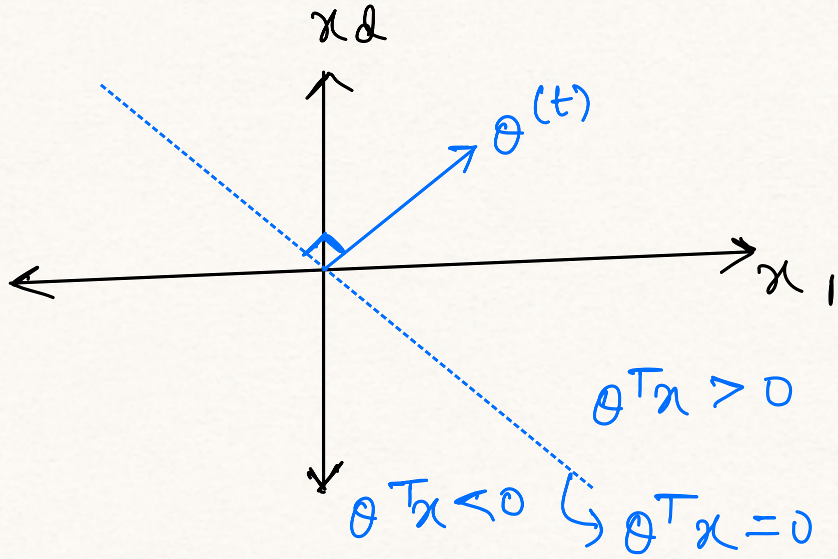


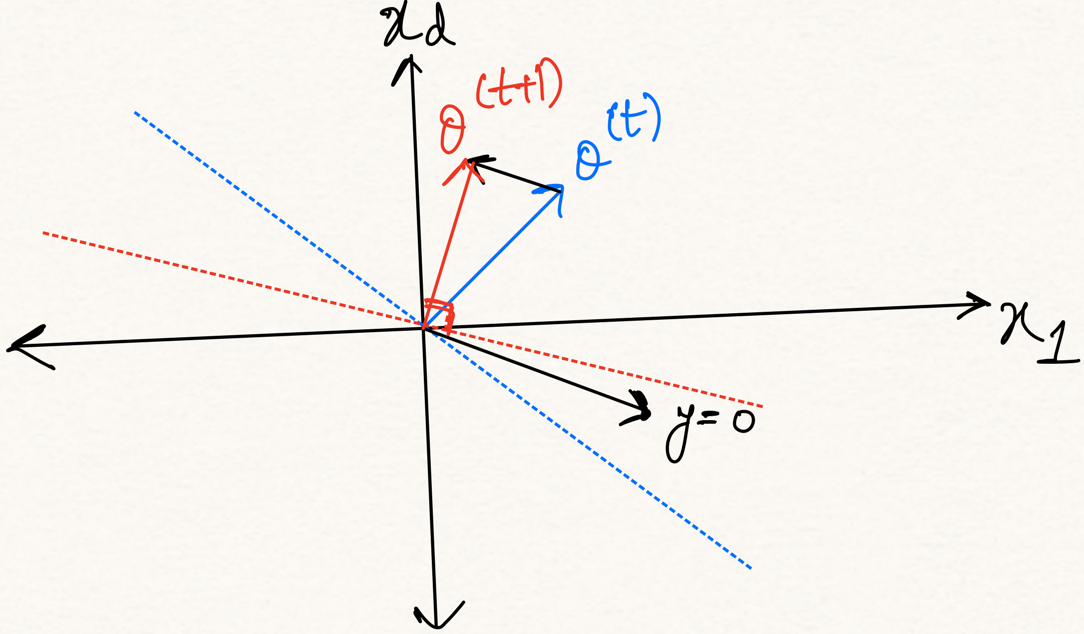
Algo:

$\vec{\theta} := \text{Init.} \ (\vec{0})$

For i in 1, 2, ---

$\theta := \theta + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) \cdot x^{(i)}$





Logistic Regression

$x^{(i)} \in \mathbb{R}^d \quad ; \quad y^{(i)} \in \{0, 1\}$

$y^{(i)} = 1$: positive example

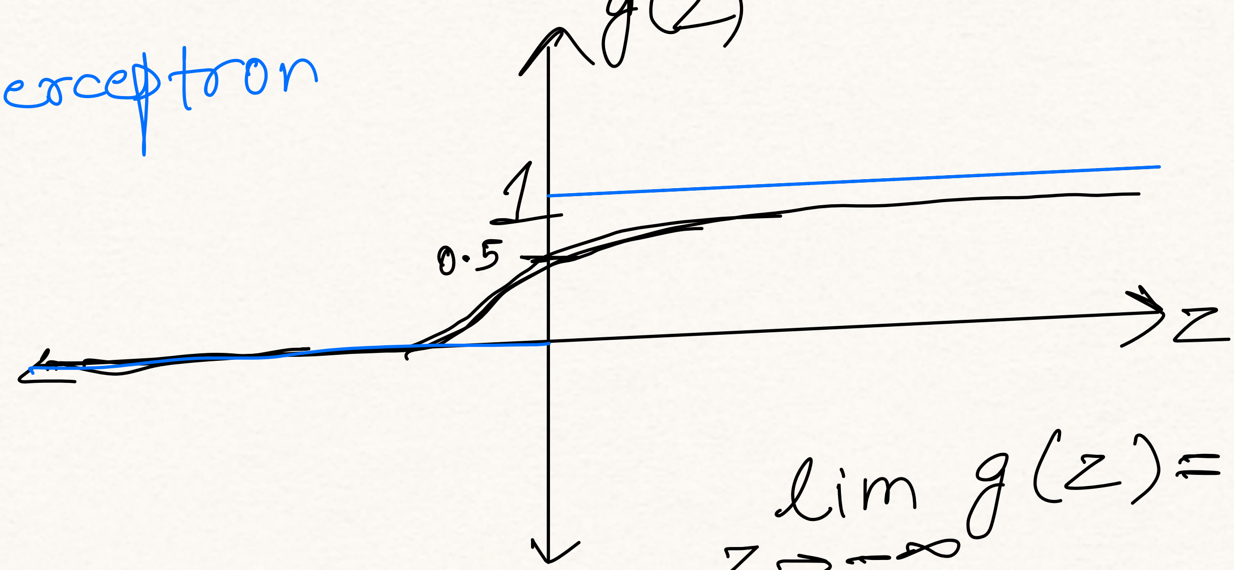
$y^{(i)} = 0$: negative example

$$h_\theta(x) = g(\theta^T x)$$

$$g(z) = \frac{1}{1+e^{-z}} \quad (\text{logistic func}^n)$$

$g(z)$

• → perceptron



$$\lim_{z \to -\infty} g(z) = 0$$

$$\lim_{z \to +\infty} g(z) = 1$$

$$p(y^{(i)} = 1 \mid x^{(i)}; \theta) = h_\theta(x)$$

$$p(y^{(i)} = 0 \mid x^{(i)}; \theta) = 1 - h_\theta(x)$$

→ Probability machine rather than classifier. outputs probability that example is positive or negative. It is on us what we take the threshold probability to be. (Generally taken to be 0.5)

$$p(y|x;\theta) = [h_\theta(x)]^y \times [1-h_\theta(x)]^{1-y}$$

$$L(\theta) = \prod_{i=1}^{n} p(y^{(i)}|x^{(i)};\theta)$$

[I.I.D]

$$\log L(\theta) = \ell(\theta) = \sum_{i=1}^{n} y^{(i)} \log h_\theta(x^{(i)}) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))$$

$$g(z) = \frac{1}{1+e^{-z}}$$

$$g'(z) = \frac{-1}{(1+e^{-z})^2} \times -e^{-z} = \left(\frac{1}{1+e^{-z}}\right)\left(\frac{e^{-z}}{1+e^{-z}}\right)$$

$$= \left(\frac{1}{1+e^{-z}}\right)\left(\frac{1+e^{-z}-1}{1+e^{-z}}\right)$$

$$= \left(\frac{1}{1+e^{-z}}\right)\left(1-\frac{1}{1+e^{-z}}\right)$$

$$= g(z)(1-g(z))$$

$$\ell(\theta) = y\log(g(\theta^T x)) + (1-y)\log(1-g(\theta^T x))$$

$$\nabla_\theta \ell(\theta) = y \cdot \frac{1}{g(\theta^T x)} g'(\theta^T x) \cdot x + (1-y) \cdot \frac{1}{1-g(\theta^T x)} \cdot (-1) g'(\theta^T x) \cdot x$$

$$= y \cdot (1-g(\theta^T x)) \cdot x - (1-y) g(\theta^T x) x$$

$$= [y - g(\theta^T x)] x$$

$$\nabla_\theta \ell(\theta) = [y - h_\theta(x)] x$$

$$\theta := \theta + \alpha [y - h_\theta(x)] x$$

↳ same as perceptoon & G.D.

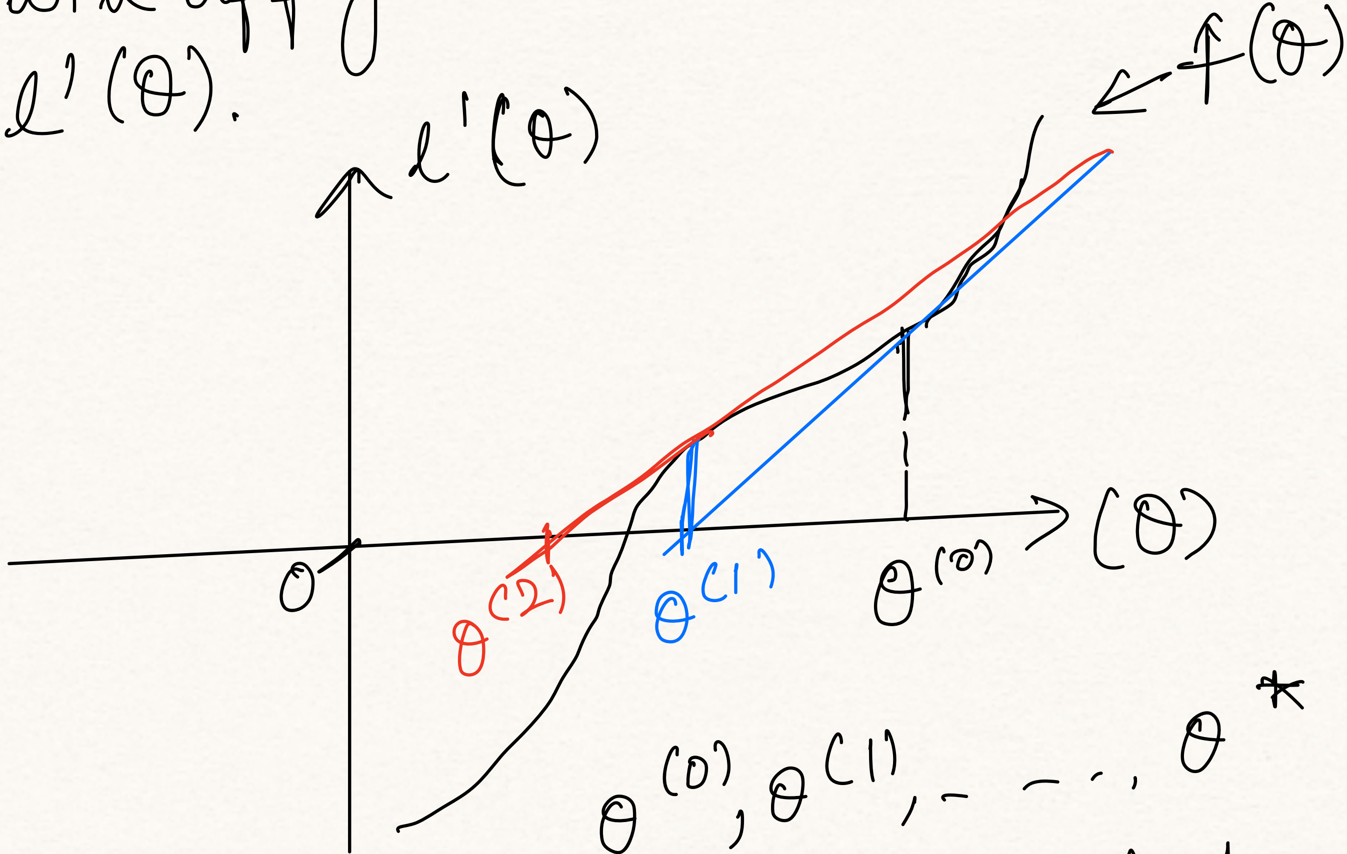
Newton's Method

(Alternative to G.D)

→ Root finding method.

Finds x for which $f(x) = 0$

We will apply Newton's method on $l'(\theta)$.



$\theta^{(0)}, \theta^{(1)}, \ldots, \theta^*$

converges way faster than G.D.

linearly approximate & then jump to that point.

$$\boxed{\theta^{(t+1)} = \theta^{(t)} - \frac{f(\theta^{(t)})}{f'(\theta^{(t)})}}$$

$$\boxed{\theta^{(t+1)} = \theta^{(t)} - \frac{\ell'(\theta^{(t)})}{\ell''(\theta^{(t)})}}$$

Scalar θ.

Vector θ :

$$\boxed{\theta^{(t+1)} = \theta^{(t)} - \alpha . H^{-1} \nabla_\theta \ell(\theta^{(t)})}$$

H = Hessian of loss $\ell(\theta)$

Newton's method $\rightarrow O(d^3)$

G.D $\rightarrow O(d)$