Cost function

$$= \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)} - y^{(i)})^2$$

Gradient descent.

$$\theta_j = \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)} - y^{(i)})x_j^{(i)}$$

Classification (Logistic regression)

$$0 \leq h_\theta(x) \leq 1$$

Hypothesis representation
for linear regression.

$$- h_\theta(x) = \theta^T x.$$

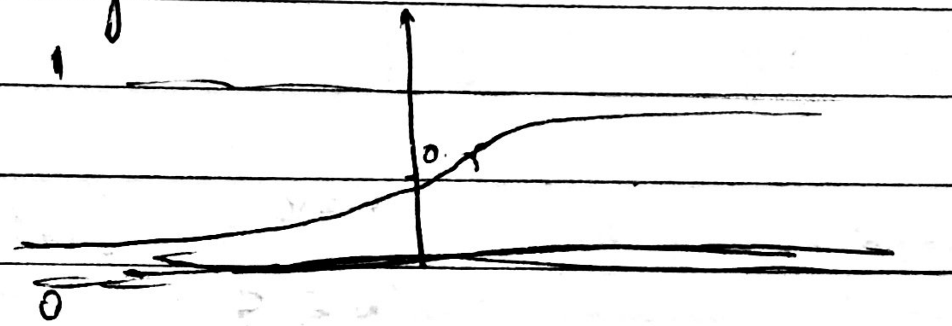
for logistic regression

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

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

$$\boxed{h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}}$$

sigmoid function



fit theta(θ) for our function

for cancer problem,

$y = 1$ malignant

$y = 0$ benign.

if, we get $h_\theta(x) = 0.7$

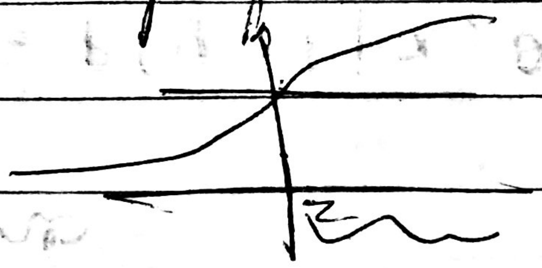
this means 70% chance of being malignant.

$h_\theta(x) = P(y=1 \mid x; \theta)$

{ probability that $y = 1$, given x, parameterized by θ }

if $h_\theta(x) \geq 0.5$, we predict $y = 1$
if $h_\theta(x) < 0.5$, we predict $y = 0$.

in sigmond fungtn



we see th $g(z) \geq 0.5$
when $z \geq 0$

$\therefore h_\theta(x) = g(\theta^T x) \geq 0.5$
when $\boxed{\theta^T x \geq 0}$

$\theta_0 \quad \theta_1 \quad \theta_2$ $\qquad x_1 = 5$

$x_1 = 5$

$5 - x_1 \geq 0$

$5 \geq x_1$

$x \leq 5$

Cost function.

m examples

$$x \in \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \mathbb{R}^{n+1} \qquad x_0 = 1$$

$$y \in \{0,1\}$$

cost function (we could use linear regression cost, however it would be non convex)

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

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \text{cost}(h_\theta(x^{(i)}), y^{(i)})$$

↓ efficiently write

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

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

we get min J for some θ.

$$\therefore h_\theta(x) = \frac{1}{1+e^{-\theta^T x}} \leftarrow x \text{ will be given}$$

Gradient descent

for $\min_\theta J(\theta)$,

$$\theta_j = \theta_j - \alpha \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) x_j^{(i)}$$

; simultaneously update all θ_j

Cost function vectorized implementation

$$h = g(X\theta)$$

$$J(\theta) = \frac{1}{m} \cdot \left(-y^T \log(h) - (1-y)^T \log(1-h)\right)$$

Gradient descent vectorized,

$$\theta := \theta - \frac{\alpha}{m} X^T \left(g(X\theta) - \vec{y}\right)$$

Multiclass classification

weather: sunny, cloudy, rain, snowy.
$\quad\quad\quad\quad$ 1 $\quad\quad$ 2 $\quad\quad$ 3 $\quad\quad$ 4

y = discrete values

we train the model for different classes

$h_\theta^{(i)} x$ for $y = i$