Linear Regression: $h_\theta(x) = \theta^T x$

$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \cdots, (x^{(m)}, y^{(m)})\}$

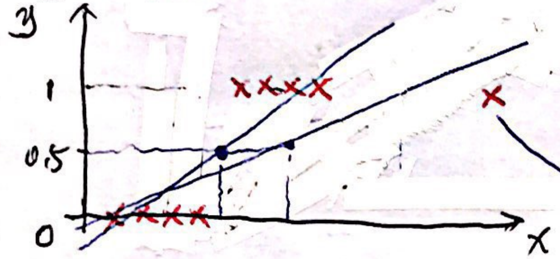
$y^{(i)} \in \mathbb{R}^{1\times1}$ $\quad -\infty < \underset{(h_\theta(x))}{y^{(i)}} < +\infty$ real value

Classification: $y \in \{0, 1\}$ binary classification.

negative class positive class

$y \in \{0, 1, 2, \cdots, k\}$ multiclass classification.

Why not use Linear Regression for classification?



set $h_\theta(x)$ Threshold at 0.5

if $h_\theta(x) \geq 0.5$, predict "y=1"

if $h_\theta(x) < 0.5$, predict "y=0"

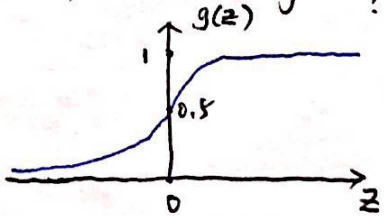
however, add one more data point will change the fitted line a lot.

Logistic Regression

We want $0 \leq h_\theta(x) \leq 1$ instead of $h_\theta(x) \in (-\infty, +\infty)$ in Linear Regression

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

Sigmoid Function
Logistic Function: $g(z) = \dfrac{1}{1+e^{-z}}$



$$\Downarrow$$

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

Interpretation: $h_\theta(x)$ = estimated probability that $\boxed{\text{"y=1"}}$ given "x"

not "y=0"

e.g. $h_\theta(x) = 0.7$ means 70% chance that "y=1" or positive sample.

$h_\theta(x) = P(y=1 \mid x; \theta)$: probability that y=1, given x, parametrized by θ.

$P(y=0 \mid x;\theta) + P(y=1 \mid x;\theta) = 1$

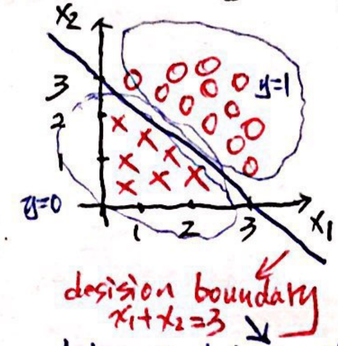
$P(y=0 \mid x;\theta) = 1 - P(y=1 \mid x;\theta)$

the way to compute "y=0" since hypothesis is defined for "y=1"

Decision Boundary

Suppose predict "y=1" if $h_\theta(x) \geq 0.5 \Rightarrow \theta^T x \geq 0$

"y=0" if $h_\theta(x) < 0.5 \Rightarrow \theta^T x < 0$



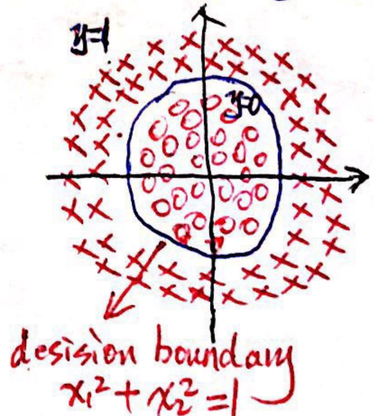
decision boundary
$x_1 + x_2 = 3$

$h_\theta(x) = g(\underset{-3}{\theta_0} + \underset{1}{\theta_1} x_1 + \underset{1}{\theta_2} x_2)$

will predict "y=1" if $-3 + x_1 + x_2 \geq 0$
$\Downarrow$
$x_1 + x_2 \geq 3$

will predict "y=0" if $-3 + x_1 + x_2 < 0$
$\Downarrow$
$x_1 + x_2 < 3$

determinated by the hypothesis, NOT the data



decision boundary
$x_1^2 + x_2^2 = 1$

$h_\theta(x) = g(\underset{-1}{\theta_0} + \underset{0}{\theta_1} x_1 + \underset{0}{\theta_2} x_2 + \underset{1}{\theta_3} x_1^2 + \underset{1}{\theta_4} x_2^2)$

will predict "y=1" if
$-1 + x_1^2 + x_2^2 \geq 0 \Rightarrow x_1^2 + x_2^2 \geq 1$

will predict "y=0" if
$-1 + x_1^2 + x_2^2 < 0 \Rightarrow x_1^2 + x_2^2 < 1$

more complex decision boundary:

$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1^2 x_2 + \theta_6 x_1 x_2^2 + \theta_7 x_1^2 x_2^2$

Oct 4, 2018