Logistic Regression

Sid

Classification

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

0: "Negative Class"

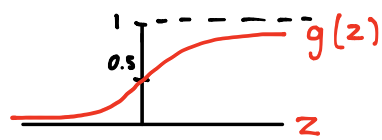
1: "Positive Class"

Logistic Regression : $0 \le h_\theta(x) \le 1$

Hypothesis :

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

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



(sigmoid)

$h_\theta(x)$ = probability that $y = 1$ on input x

$p(y=1 \mid x ; \theta)$

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

Decision Boundary

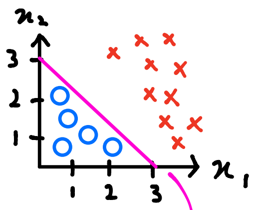
$h_\theta(x) \ge 0.5$: predict $y = 1$

$g(z) \ge 0.5$ when $z \ge 0$

$h_\theta(x) = g(\theta^T x) \ge 0.5$ when $\theta^T x \ge 0$

$h_\theta(x) < 0.5$: predict $y = 0$

$h_\theta(x) = g(\theta^T x) \quad g(z) < 0.5 \quad \theta^T x < 0$



$x_1 + x_2 = 3$

Decision Boundary

$h_\theta(x) = g(\theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2)$

$$\theta = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \qquad \theta^T x = -3 + x_1 + x_2$$

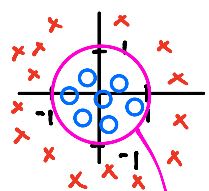
predict $y = 1$

$\rightarrow -3 + x_1 + x_2 \geq 0$

$x_1 + x_2 \geq 3$

property of hypothesis/parameters not data set

Non-linear decision boundary



$x_1^2 + x_2^2 = 1$

Decision boundary

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

$$\theta = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \qquad y = 1 \; : \; -1 + x_1^2 + x_2^2 \geq 0$$

$x_1^2 + x_2^2 \geq 1$

Training set: $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$

m examples

$$x \in \begin{bmatrix} x_0 \\ x_1 \\ \dots \\ x_n \end{bmatrix} \quad x_0 = 1, \; y \in \{0, 1\}$$

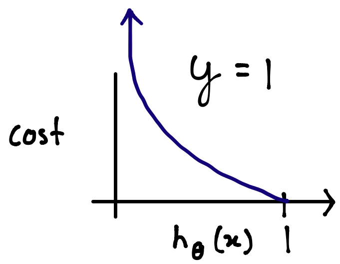
$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

Cost function:

Linear: $J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \underbrace{\frac{1}{2}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2}_{\mathrm{Cost}(h_\theta(x^{(i)}),\, y^{(i)})}$

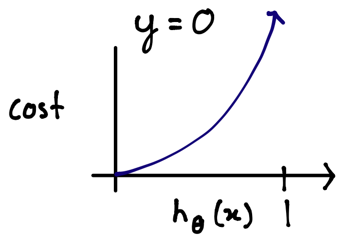
Logistic :

$$\text{Cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & , y = 1 \\ -\log(1-h_\theta(x)) & , y = 0 \end{cases}$$



Cost = 0 if $y = 1$, $h_\theta(x) = 1$

as $h_\theta(x) \to 0$, cost $\to \infty$



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

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

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

Gradient descent

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

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

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

Vectorized

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

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

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

Advanced optimization

```
function [jVal, gradient] = costFunction(theta)

    jVal     = J(θ)
    gradient = ∂/∂θ J(θ)

end
```

```
options = optimset('GradObj', 'on', 'MaxIter', '100');
initialTheta = zeros(2, 1);
[optTheta, functionVal, exitFlag] =
    fminunc(@costFunction, initialTheta, options);
```

Multiclass classification

- One-vs-all : create separate 2-class classification problems
  - ↳ $h_\theta^{(i)}(x) = P(y = i \mid x; \theta) \qquad (i = 1, 2, \dots)$
- $\max_i h_\theta^{(i)}(x)$ on new input x pick class i