Logistic Regression

Given x, want →

$\hat{y} = P(y = 1 \mid x)$

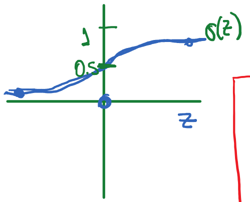
$0 \leq \hat{y} \leq 1x$ $\quad x \in \mathbb{R}^{n_x}$

Parameters: $\boxed{w} \in \mathbb{R}^{n_x}$, $\boxed{b} \in \mathbb{R}$

Output: $\hat{y} = \underbrace{\sigma(w^{T}x + b)}_{z}$

$\hat{y} = \underbrace{\sigma(w_x^T + b)}_{z}$

Sigmoid Function



$\sigma(z) = \dfrac{1}{2 + e^{-z}}$

$X_0 = 1, \quad x \in \mathbb{R}^{n_x+1}$

$\hat{y} = \sigma(\theta^{T}x)$

$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_{n_x} \end{bmatrix}$ θ_0 } b, $\theta_1 \dots \theta_{n_x}$ } w

if z is large → $\sigma(z) \approx \frac{1}{1+0} = 1$

if z is large negative number →

$\sigma(z) = \dfrac{1}{1 + e^{-z}} \approx \dfrac{2}{1 + Big\ Number} \approx 0$

Sigmoid

Cost Function (log reg)

$\rightarrow \hat{y}^{(i)} = \sigma(w^T x^{(i)} + b)$, where $\sigma(z^{(i)}) = \frac{1}{1+e^{-z^{(i)}}}$

Given $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$

want $\rightarrow \hat{y}^{(i)} \approx y^{(i)}$

$z^{(i)} = w^T x^{(i)} + b$

$x^{(i)}$, $y^{(i)}$, $z^{(i)}$ i-th example

Loss (error) function:

$\times$ ~~$\mathcal{L}(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$~~

$$\mathcal{L}(\hat{y}, y) = -\left(y \log(\hat{y}) + (1-y)\log(1-\hat{y})\right)$$

if $y=1$: $\mathcal{L}(\hat{y}, y) = -\log \hat{y}$ ← want $\log \hat{y}$ large, want $\hat{y}$ large

if $y=0$: $\mathcal{L}(\hat{y}, y) = -\log(1-\hat{y})$ ← want $1-\hat{y}$ large ... want $\hat{y}$ small

Cost function: $J(w, b) = \frac{1}{m} \sum_{i=1}^{m} \mathcal{L}(\hat{y}^{(i)}, y^{(i)})$

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

Cost Function Expanded Observations p1

Wednesday, August 23, 2017 8:53 AM

$\hat{y} = \sigma(w^T x + b)$ where $\sigma(z) = \frac{1}{1 + e^{-z}}$

interpret $\hat{y} = P(y = 1 | x)$

$\therefore$ if $y = 1 \longrightarrow P(y/x) = \hat{y}$

if $y = 0 \longrightarrow P(y/x) = 1 - \hat{y}$

Cost Function Expanded Observations p2

Wednesday, August 23, 2017 9:03 AM

$$\left.\begin{cases} \text{if } y=1: & P(x|y) = \hat{y} \\ \text{if } y=0: & P(x|y) = 1-\hat{y} \end{cases}\right\} P(y/x)$$

$$P(y/x) = \hat{y}^{y}\,(1-\hat{y})^{(1-\hat{y})}$$

$$\log(P(y|x)) = \log\left(\hat{y}^{y}\,(1-\hat{y})^{(1-\hat{y})}\right) =$$

$$= y(\log \hat{y}) + (1-\hat{y})\log(1+\hat{y})$$

$$= \ominus \mathcal{L}(\hat{y}, y) \downarrow$$

Cost Function Expanded Observations p3

Wednesday, August 23, 2017 9:19 AM

$$\log P(\text{labels in training set}) = \log \prod_{i=1}^{m} P(y^{(i)} \mid x^{(i)})$$

$$= \sum_{i=1}^{m} \underbrace{\log P(y^{(i)} \mid x^{(i)})}_{-\mathcal{L}(\hat{y}^{(i)} \mid y^{(i)})}$$

Maximum Likelyhood Estimation

$$= -\sum_{i=1}^{m} \mathcal{L}(\hat{y}^{(i)} \mid y^{(i)})$$

Cost: (minimize) $J(w,b) = \frac{1}{m} \sum_{i=1}^{m} \mathcal{L}(\hat{y}^{(i)}, y^{(i)})$