Logistic Regression Cost Function

$\rightarrow \hat{y}^{(i)} = \sigma(w^T x^{(i)} + b)$, 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:

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

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

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

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

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

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

Expanded Observations part 1

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

Expanded Observations part 2

Wednesday, August 23, 2017 9:03 AM

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

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

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

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

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

Expanded Observations part 3

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

(minimize)