Assignment 2

Xunzhi Li

2.Generative methods vs Discriminative methods

Cost of Logistic Regression

$$f(x) = \theta_1 x_1 + \theta_2 x_2 + \cdots \theta_n x_n = \begin{bmatrix} \cdots \theta \cdots \end{bmatrix} \begin{bmatrix} \vdots \\ x \\ \vdots \end{bmatrix}$$

Then we use ~~sigmost~~ sigmoid function.

$$g(x^0) = \frac{1}{1+e^{(-x^0)}} = \frac{e^{x^0}}{e^{x^0}+1}$$

$$x^0 = \theta x$$

$$\Rightarrow \cancel{f(x)=}\ g(x) = \frac{1}{1+e^{(-\theta^T x)}} \Rightarrow \log g(x) = -\log(1+e^{-\theta^T x}) \cdots ①$$

$\Rightarrow$ ~~use Least square Error as loss function~~

$$\text{cost}(g(x), y) = \begin{cases} -\log(g(x)) & \text{if } y=1 \\ -\log(1-g(x)) & \text{if } y=0 \end{cases}$$

$$\Rightarrow \text{cost}(g(x), y) = -y\log(g(x)) - (1-y)\log(1-g(x)) \cdots ②$$

$$① + ② \quad J(w) = -\frac{1}{N}\sum_{n=1}^{N}\left[y_n \log(1+e^{-wx_n}) + (1-y_n)\log\left(1-\frac{1}{1+e^{-wx_n}}\right)\right]$$

$$= -\frac{1}{N}\sum_{n=1}^{N}\left[y_n \log(1+e^{-wx_n}) + (1-y_n)(\log e^{-wx_n} - \log(1+e^{-wx_n}))\right]$$

$$= -\frac{1}{N}\sum_{n=1}^{N}\left[y_n \log(1+e^{-wx_n}) + (1-y_n)(-wx_n - \log(1+e^{-wx_n}))\right]$$

$$= -\frac{1}{N}\sum_{n=1}^{N}\left[y_n w x_n - w x_n - \log(1+e^{-wx_n})\right]$$

$$= -\frac{1}{N}\sum_{n=1}^{N}\left[y_n w x_n - \log(1+e^{wx_n})\right]$$

$$\frac{\partial J(w)}{\partial (w)} = -\frac{1}{N}\sum_{n=1}^{N}\left(y_n x_n - \frac{x_n e^{wx_n}}{1+e^{wx_n}}\right) \longrightarrow \text{sigmoid.}$$

$$= \frac{1}{N}\sum_{n=1}^{N}(f(x_n) - y_n)x_n$$

4. Linear classification

1)

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \propto p(x|y)p(y)$$

$$\Rightarrow \log p(x|y)p(y) = \log \prod_{i=1}^{d} p(x_i|y) + \log p(y)$$

$$= \sum_{i=1}^{d} \log p(x_i|y) + \log(p(y))$$

$p(y)$ is not affected by x. So its constant here, ~~we can skip it~~.

$$\text{So } \mathrm{Map}(y|x) \Rightarrow \mathrm{Max} \sum_{i=1}^{d} \log p(x_i|y) = h(x)$$

because features are multinomial

$$h(x) = \cancel{\underset{y}{\mathrm{argmax}}\, P(y)} \;\; \mathrm{sign}(w^T x + b)$$

$$\begin{cases} w^T x + b > 0 \Rightarrow \text{belongs to class one } (+1) \\ w^T x + b < 0 \Rightarrow \text{belongs to class two } (-1) \end{cases}$$

set $p(x_i | y = +1) \propto p_+^{x_i}$, $p(Y=+1) = \pi_+$

$W_i = \log(p_+) - \log(p_-)$, $\quad b = \log(\pi_+) - \log(\pi_-)$

$$w^T x + b > 0 \Leftrightarrow \sum_{i=1}^{d} x_i (\log(p_+) - \log(p_-)) + \log(\pi_+) - \log(\pi_-) > 0$$

$$\Leftrightarrow e^{\sum_{i=1}^{d} x_i (\log(p_+) - \log(p_-)) + \log(\pi_+) - \log(\pi_-)} > 1$$

$$\Leftrightarrow \prod_{i=1}^{d} \frac{e^{(\log p_+ + \log(\pi_+))}}{e^{\log p_- + \log(\pi_-)}} > 1$$

$$\Leftrightarrow \prod_{i=1}^{d} \frac{(p_+)(\pi_+)}{(p_-)(\pi_-)} > 1 \quad \Leftrightarrow \prod_{i=1}^{d} \frac{P(x_i | y=+1) P(y=+1)}{P(x_i | y=-1) P(y=-1)} > 1$$

$$P(y=+1|x) > P(y=-1|x)$$

$$\Leftrightarrow \mathrm{argmax}\, P(Y=y|x) = +1$$

2)

For logistic Regression

$$P(y=0|x) = \frac{1}{1+e^{w_0+\sum_{i=1}^{d} w_i x_i}}$$

$$P(y=1|x) = 1 - P(y=0|x) = \frac{e^{w_0+\sum_{i=1}^{d} w_i x_i}}{1+e^{w_0+\sum_{i=1}^{d} w_i x_i}}$$

if $P(y=1|x) > P(y=0|x)$, we are gonna say our x belongs to $y=1$ this class.

$$\frac{P(y=1|x)}{P(y=0|x)} > 1 \Rightarrow \log \frac{P(y=1|x)}{P(y=0|x)} > 0$$

$$\Rightarrow \cancel{\log(P(y=1|x)-P(y=0|x)} \log P(y=1|x) - \log(y=0|x) > 0$$

$$\log e^{w_0+\sum_{i=1}^{d} w_i x_i} - \log 1 > 0$$

$$\Rightarrow w_0+\sum_{i=1}^{d} w_i x_i > 0 \quad \text{(its linear classifier)}$$