2017147 - Divyam Anshumaan.

ML - Assignment - 1.

Q3. i) ~~$P \in [0,1]$~~ $P \in (0,1)$ as

$g(z) = 1/1+\exp(-z)$ which is bound between $(0,1)$ as it is parameterized by w.

We can change w_0, w_1 as needed to produce $e^{-w_0'} \cdot e^{-w_1' x} = e^{-k}$ where $P(y=1 \mid x; k) \in [0,1)$

ii) logit function $l(x) = \log(x/1-x)$

1. $l(x)$ is continuous in domain $[0,1]$

~~proof~~: for any $c \in [0,1]$

i) $\log(c/1-c)$ is defined and

$\lim_{c \to 0} \log(c/1-c) = -\infty$ and

$\lim_{c \to 1} \log(c/1-c) = \infty$ —(1)

ii) $\lim_{c \to 1} \log(c/1-c) = l(c)$ and

$\lim_{c \to 0} l(c)$ are defined right and left limit.

Hence $l(x)$ is continuous in domain $[0,1]$

from (1) we conclude that range $l(x) \in (-\infty, \infty)$

Q4.

a) $J(\theta) = \frac{1}{m} \sum_{i=1}^{m} |h_\theta(x^{(i)}) - y^{(i)}|$

Gradient update $= \theta_i \leftarrow \theta_i - \alpha \cdot \frac{\partial J(\theta)}{\partial \theta_i}$

where,

$$\frac{\partial J(\theta)}{\partial \theta_j} = \begin{cases} \frac{1}{m} \sum_i (-1) \cdot x_j^{(i)}, & \text{if } h_\theta(x^{(i)}) - y^{(i)} < 0 \\ 0, & h_\theta(x^{(i)}) - y^{(i)} = 0 \\ \frac{1}{m} \sum_i (1) \cdot x_j^{(i)}, & \text{if } h_\theta(x^{(i)}) - y^{(i)} > 0 \end{cases}$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m} \sum_i (-1) \cdot x_j^{(i)}, \quad h_\theta(x^{(i)}) - y^{(i)} < 0$$

$$= 0, \quad h_\theta(x^{(i)}) - y^{(i)} = 0$$

$$= \frac{1}{m} \sum_i (1) \cdot x_j^{(i)}, \quad h_\theta(x^{(i)}) - y^{(i)} > 0$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{\alpha}{m} \sum \text{sgn}(h_\theta(x^{(i)}) - y^{(i)}) \cdot x_j^{(i)}$$

where $\text{sgn}(\cdot)$ is the signum function.

- this is difficult to optimize because of the fixed size parameter update. As it is not looking at the cost incurred, the fixed update will keep overshooting the optimal unless α is small.

In case of MSE, step size decreased w/ decrease in cost, so steps sizes decreased as the optimum was approached.

b) Used for datasets w/ high variance or many outliers, since the parameter update gives the same weight to all costs of the same sign.

Thus noisy examples are ignored.

CLASSTIME | Page No.
Date / /

c) We can specifically penalize over-estimation or under-estimation more than the other.

This also provides a range or interval of predictions as compared to exact values. This becomes relevant for models-data with non-constant variance, which cannot be solved properly with MAE like losses, which will give bad estimates.