Aayush Gupta 2017125

Page no.

Date / /

# Machine Learning Assignment 1

(3) (i) Given, $P(y=1|x,w) = g(w_0 + w_1 x)$
where $g(z)$ is a logistic function.
Also, it is an increasing function

i.e. $$g(z) = \frac{1}{1+e^{-z}}$$

As ~~it~~ it is a function of $x$
$w_0 + w_1 x$ acts as a linear equation, so it extends from $-\infty$ to $+\infty$ when ~~x~~ extends from $-\infty$ to $+\infty$ and ~~w~~ $w_1 \neq 0$

So $$g(z) = \frac{1}{1+e^{-z}}$$

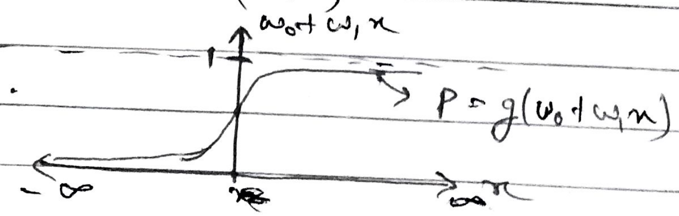
~~[illegible]~~

$$-\infty < z = w_0 + w_1 x < \infty$$

$$g(z) \text{ at } z = -\infty = \frac{1}{1+e^{\infty}} \Rightarrow 0$$

$$g(z) \text{ at } z = \infty = \frac{1}{1+e^{-\infty}} \rightarrow 1$$

So, for ~~[illegible]~~ range of $P(y=1|x,w)$ is $(0,1)$



Date / /

Page no.

(ii) logit function : $l(x) = \log\left(\frac{x}{1-x}\right)$

$$\frac{d}{dx} l(x) = \left(\frac{\cancel{1-x}}{x}\right) \frac{(1-x) - x(-1)}{(1-x)^{\cancel{2}}}$$

$$= \frac{1}{x(1-x)}$$

which is always positive for $0 < x < 1$

Hence, it is an increasing function.

For $0^+$ & $1^-$, logit tends to $-\infty$ and $+\infty$ respectively.

So, logit goes from $(-\infty$ to $\infty)$ for $0 < x < 1$

(4) (a) As, MSE takes the square of error, hence the higher errors are given more weights and are ~~easier to~~ easier to reduce using the gradient descent as higher errors are undesirable and reduced significantly due to higher weights.

(b) In a data, where noise is very far away from the actual data, MSE would give higher weight to reduce the loss, which is infact not good for the performance of the actual data

My NoteBook

Date / / Page no.

~~R~~MSE would increase the error in the actual data. In such cases MAE is considered as a better option, as it is more robust to the noise and ~~focu~~ focuses more on actual data.

(c) Quantile loss function is more useful over MAE when we need to predict intervals rather than discrete points which have variable variance. It is basically a modification of MAE

$$L_\gamma(y, y^p) = \sum_{i=y_i<y_i^p} (\gamma-1).|y_i - y_i^p| + \sum_{i=y_i \geq y_i^p} \gamma|y_i - y_i^p|$$

where $\gamma$ is a parameter which is called quantile. By changing it we can adjust the weight of errors which can help us to adjust the ~~amou~~ amount of overestimation or underestimation. For $\gamma = 0.5$, it turns into MAE.