Deriving Logistic Regression

→ using "Logistic Loss"

Data:

$$D = \{x, y\}$$

$x \rightarrow \in R^d$, Data points

$y \rightarrow$ Target Variable (Binary) $\in \{+1, -1\}$

Line eqn:

W.K.T, eqn of Line $\Rightarrow$ $\boxed{y = mx + b}$

Goal:

Find slope 'm' & intercept 'b', such that it should predict the target variable correctly.

Cases: Let, $\boxed{z = y_{true} * (mx+b)}$ — (a)

1) If $y_{true} > 0$ & $(mx+b) > 0$
 then, $z > 0$.
2) If $y_{true} < 0$ & $(mx+b) < 0$
 then, $z > 0$.

} Correctly classified

3) If $y_{true} > 0$ & $(mx+b) < 0$
 (or)
4) $y_{true} < 0$ & $(mx+b) > 0$
 then, $z < 0$

} misclassified.

"We need to find 'm' and 'b' such that maximum no. of points should be correctly classified."

Scanned By Scanner Go

So, we can write the optimization function as,

$$\boxed{f(w,b) = \underset{w,b}{\operatorname{argmax}}\left(\sum_{i=1}^{n} y_i * (mx+b)\right)}$$

$$\Rightarrow f(w,b) = \underset{w,b}{\operatorname{argmax}}\left(\sum_{i=1}^{n} z_i\right) \quad \text{--- } (1)$$

As the values ~~will~~ of z will depend on the magnitude of 'x' [from eq (a)]. we will apply "Squashing" using "Sigmoid" function.

eq Sigmoid, $\sigma(x) = \dfrac{1}{1+e^{-x}}$

eq^n (1) will become,

$$f(w,b) = \underset{w,b}{\operatorname{argmax}}\left(\sum_{i=1}^{n} \sigma(z_i)\right)$$

$$\Rightarrow \boxed{f(w,b) = \underset{w,b}{\operatorname{argmax}}\left(\sum_{i=1}^{n} \frac{1}{1+e^{-z_i}}\right)} \quad \text{--- } (2)$$

Note: If two functions ($f(x)$ & $g(x)$) are either monotonically increasing or decreasing, then we can write them as $f(g(x))$ and the results won't change.

→ As "log(x)" and sigmoid are both monotonically increasing functions, we can write eq^n (2) as.

$$\boxed{f(w,b) = \underset{w,b}{\operatorname{argmax}}\left(\sum_{i=1}^{n} \log\left(\frac{1}{1+e^{-z_i}}\right)\right)}$$

$$\Rightarrow f(w,b) = \underset{w,b}{\operatorname{argmax}} \sum_{i=1}^{n} \left[\log(1) - \log(1+e^{-z_i})\right] \quad \because \log(a/b) = \log(a) - \log(b)$$

(4)

Scanned By Scanner Go

$$= \underset{w,b}{\operatorname{argmax}} \sum_{i=1}^{n} 0 - \log(1+e^{-z_i}) \quad \because \log(1) = 0.$$

$$\boxed{f(w,b) = -\underset{w,b}{\operatorname{argmax}} \sum_{i=1}^{n} \log(1+e^{-z_i})} \text{ — eq}^n \ 5$$

w.r.t, $f(x) = \operatorname{argmax}(x)$

$= -\operatorname{argmin}(-x)$

Now, eq^n 5 can be written as,

$$\boxed{f(w,b) = \underset{w,b}{\operatorname{argmin}} \sum_{i=1}^{n} \log(1+e^{-z_i})} \text{ — eq}^n \ 6$$

"eq^n 6 is the optimization eq^n used to find the optimal slope 'm' and intercept 'b'"

Gradient Descent for Slope 'm':

$$\Rightarrow \frac{\partial f(m,b)}{\partial m} = \frac{\partial}{\partial m}\left(\sum_{i=1}^{n} \log(1+e^{-z_i})\right)$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial m}\left[\log\left[1+e^{-[y_i * (mx_i+b)]}\right]\right]$$

$$= \sum_{i=1}^{n} \frac{1}{1+e^{-(y_i * (mx_i+b))}} \times e^{-(y_i * (mx_i+b))} \times (-y_i * x_i)$$

$$\boxed{\frac{\partial f}{\partial m} = \sum_{i=1}^{n} -y_i * x_i * \left(\frac{e^{-z_i}}{1+e^{-z_i}}\right)} \rightarrow \text{eq}^n \ 7$$

w.k.T, $\boxed{\sigma(z) = \frac{1}{1+e^{-z}}}$ → eq^n (b)

$$\Rightarrow 1+e^{-z} = \frac{1}{\sigma(z)}$$

$$\Rightarrow e^{-z} = \frac{1}{\sigma(z)} - 1$$

$$\Rightarrow \boxed{e^{-z} = \frac{1-\sigma(z)}{\sigma(z)}} \rightarrow eq^n\ (c)$$

we can write eq^n (7) as,

$$\frac{\partial f}{\partial m} = \sum_{i=1}^{n} -y_i * x_i * (e^{-z_i}) * \left(\frac{1}{1+e^{-z_i}}\right)$$

substituting eq^n (b) & (c) in above eq^n,

$$= \sum_{i=1}^{n} -y_i * x_i * \frac{1-\sigma(z)}{\cancel{\sigma(z)}} * \cancel{\sigma(z)}$$

$$\boxed{\frac{\partial f}{\partial m} = -\sum_{i=1}^{n} x_i * y_i * \left(1-\sigma(mx+b)\right)} - eq^n\ (8)$$

eq^n (8) is the gradient descent of slope 'm' for logistic loss func^n.

Gradient Descent for ~~slo~~ Intercept 'b':

$$\frac{\partial f}{\partial b} = \frac{\partial}{\partial b}\left(\sum_{i=1}^{n} \log(1+e^{-z_i})\right)$$

$$= \sum_{i=1}^{n} \frac{1}{1+e^{-(y_i * (mx+b))}} \times e^{-(y_i * (mx+b))} \times (-y_i)$$

$$= \sum_{i=1}^{n} -y_i * \frac{e^{-z_i}}{1+e^{-z_i}}$$

$$= \sum_{i=1}^{n} -y_i * e^{-z_i} * \frac{1}{1+e^{-z_i}}$$

from eq^n (b) & (c)

$$= \sum_{i=1}^{n} -y_i * \frac{1-\sigma(z)}{\cancel{\sigma(z)}} * \cancel{\sigma(z)}$$

$$\Rightarrow \boxed{\frac{\partial f}{\partial b} = -\sum_{i=1}^{n} y_i * \left(1 - \sigma\left(y_i * (mx+b)\right)\right)}$$

↳ eq^n (9)

eq^n (9) is the gradient descent of Intercept for logistic loss function.