

# GSOC'19 xgboost loss functions

Avinash Barnwal

May 2019

## Model

Assume the data follow below model:-

$$\begin{aligned}\log y_i &= x_i' \beta + z_i \sigma \\ \log \hat{y}_i &= x_i' \hat{\beta} \\ \eta &= x_i' \hat{\beta} \\ z_i &= \frac{\log y_i - \eta}{\sigma} \sim f\end{aligned}\tag{1}$$

where  $y_i$  is the uncensored response and  $\hat{y}_i$  is the predicted value for  $i$ -th observation and  $\sigma$  is the standard deviation of the error.

## Normal Distribution

$$\begin{aligned}f(z) &= \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} \\ f'(z) &= -zf(z) \\ f''(z) &= -f(z) - zf'(z) \\ F(z) &= \phi(z) \\ F_{Y_i}(y_i) &= F(z_i)\end{aligned}\tag{2}$$

## Logistic Distribution

$$\begin{aligned}f(z) &= \frac{e^z}{(1 + e^z)^2} \\ w &= e^z \\ f'(z) &= f(z) \frac{1 - w}{1 + w} \\ f''(z) &= f(z) \frac{(w^2 - 4w + 1)}{(1 + w)^2} \\ F(z) &= \frac{e^z}{1 + e^z} \\ F_{Y_i}(y_i) &= F(z_i)\end{aligned}\tag{3}$$

## Extreme Distribution

$$\begin{aligned}
w &= e^z \\
f(z) &= we^{-w} \\
F(z) &= 1 - e^{-w} \\
f'(z) &= f(z)(1 - w) \\
f''(z) &= (w^2 - 3w + 1)f(z) \\
F_{Y_i}(y_i) &= F(z_i)
\end{aligned} \tag{4}$$

# 1 Uncensored Data

## 1.1 Loss Function

$$L_i = -\log \text{lik}_i = -\log(f_{Y_i}(y_i|\eta)) \tag{5}$$

where,  $f_\eta(\eta)$  is the probability density function(pdf) of  $\eta$ .

Now using change of variable for probability density function(pdf). We can write the below equations.

### Normal Distribution

$$\begin{aligned}
f_Y(y|\eta) &= f_Z(z) \frac{\partial z}{\partial y} \\
f_Y(y|\eta) &= \frac{\exp\left(-\frac{(\log(y) - \log(\hat{y}_i))^2}{2\sigma^2}\right)}{y\sigma\sqrt{2\pi}} \\
L_i &= -\log\left(\frac{\exp\left(-\frac{(\log(y_i) - \log(\hat{y}_i))^2}{2\sigma^2}\right)}{y_i\sigma\sqrt{2\pi}}\right) \\
L_i &= -\log\left(\frac{\exp\left(-\frac{(\log(y_i) - \eta)^2}{2\sigma^2}\right)}{y_i\sigma\sqrt{2\pi}}\right)
\end{aligned} \tag{6}$$

### Logistic Distribution

$$\begin{aligned}
f_Y(y|\eta) &= f_Z(z) \frac{\partial z}{\partial y} \\
f_Y(y|\eta) &= \frac{\exp\left(\frac{(\log(y) - \eta)}{\sigma}\right)}{\sigma * y * (1 + \exp\left(\frac{(\log(y) - \eta)}{\sigma}\right))^2} \\
L_i &= -\log\left(\frac{\exp\left(\frac{(\log(y) - \eta)}{\sigma}\right)}{\sigma * y * (1 + \exp\left(\frac{(\log(y) - \eta)}{\sigma}\right))^2}\right) \\
&= -\log\left(\frac{\exp\left(\frac{(\log(y) - \eta)}{\sigma}\right)}{\sigma * y * (1 + \exp\left(\frac{(\log(y) - \eta)}{\sigma}\right))^2}\right)
\end{aligned} \tag{7}$$

### Extreme Distribution

$$\begin{aligned}
f_Y(y|\eta) &= f_Z(z) \frac{\partial z}{\partial y} \\
f_Y(y|\eta) &= \frac{e^{\frac{\log(y) - \log(\hat{y}_i)}{\sigma}} e^{-e^{\frac{\log(y) - \log(\hat{y}_i)}{\sigma}}}}{y\sigma} \\
L_i &= -\log \frac{e^{\frac{\log(y_i) - \log(\hat{y}_i)}{\sigma}} e^{-e^{\frac{\log(y_i) - \log(\hat{y}_i)}{\sigma}}}}{y_i\sigma}
\end{aligned} \tag{8}$$

## 1.2 Negative Gradient

In Gradient Boosting and Xgboost, we need to calculate negative gradient of loss function with respect to real-value prediction which is  $\hat{\eta}$ . Here i am changing between  $\hat{\eta}_i$  to  $\hat{\eta}$  to make it general.

$$\begin{aligned}
-\frac{\partial L_i}{\partial \hat{\eta}} &= \frac{\partial \log(f_{Y_i}(y_i|\eta))}{\partial \eta} \\
&= \frac{1}{f_{Y_i}(y_i|\eta)} * \frac{\partial f_{Y_i}(y_i|\eta)}{\partial \eta}
\end{aligned} \tag{9}$$

Using change of variable between  $z$  and  $\eta$ . We can write the below in terms of  $z$ .

$$\frac{\partial f_Y(y|\eta)}{\partial \eta} = \frac{\partial}{\partial \eta} [f_Z(z) \frac{\partial z}{\partial y}] \tag{10}$$

Using product rule of differentiation, we can split  $f_Z(z)$  and  $\frac{\partial z}{\partial \eta}$

$$= \frac{\partial f_Z(z)}{\partial \eta} \frac{\partial z}{\partial y} + f_Z(z) \frac{\partial^2 z}{\partial \eta \partial y} \tag{11}$$

$$= \frac{\partial f_Z(z)}{\partial z} \frac{\partial z}{\partial \eta} \frac{\partial z}{\partial y} + f_Z(z) \frac{\partial^2 z}{\partial \eta \partial y} \tag{12}$$

As we know  $\frac{\partial f_Z(z)}{\partial z} = -zf_Z(z)$

Now, we will calculate the gradient of  $z$  with respect to  $\eta$  as we have  $\frac{\partial z^2}{\partial \eta \partial y}$  in the equation 12.

$$\frac{\partial z}{\partial \eta} = \frac{-1}{\sigma} \tag{13}$$

$$\frac{\partial z}{\partial y} = \frac{1}{y\sigma} \tag{14}$$

We also need double differentiation of  $z$  with respect to  $\eta$  as we have  $\frac{\partial^2 z}{\partial \eta^2}$  in the equation 12.

$$\frac{\partial^2 z}{\partial \eta \partial y} = 0 \tag{15}$$

$$= -\frac{f'_Z(z)}{y\sigma^2} \quad (16)$$

Now going back to original equation of calculating negative gradient of loss function with respect to  $\eta$ . In equation 9, we have calculated the second part of negative gradient of loss function with respect to  $\eta$  which is  $\frac{\partial f_{Y_i}(y_i|\eta)}{\partial \eta}$ . By replacing the 16th equation in the negative gradient of loss function w.r.t. to  $\eta$ , we get the results as follows:-

$$\begin{aligned} -\frac{\partial L_i}{\partial \eta} &= \frac{1}{f_{Y_i}(y_i|\eta)} * \frac{\partial f_{Y_i}(y_i|\eta)}{\partial \eta} \\ &= \frac{1}{f_{Y_i}(y_i|\eta)} * \frac{-f'_{Z_i}(z_i)}{y\sigma^2} \\ &= \frac{-f'_{Z_i}(z_i)}{\sigma f_{Z_i}(z_i|\eta)} \end{aligned} \quad (17)$$

### 1.3 Hessian

Hessian is the second derivative of Loss with respect to  $\eta$ . As we have already calculated the negative gradient, further we need to take one more partial derivative with respect to  $\eta$  of negative gradient with negative sign.

$$\begin{aligned} \frac{\partial^2 L_i}{\partial \eta^2} &= \frac{\partial}{\partial \eta} \frac{\partial L_i}{\partial \eta} \\ &= \frac{\partial}{\partial \eta} \frac{f'_{Z_i}(z_i)}{\sigma f_{Z_i}(z_i|\eta)} \\ &= \frac{\partial}{\partial z_i} \frac{f'_{Z_i}(z_i)}{\sigma f_{Z_i}(z_i|\eta)} \frac{\partial z_i}{\partial \eta} \\ &= \frac{\partial}{\partial z_i} \frac{f'_{Z_i}(z_i)}{\sigma f_{Z_i}(z_i|\eta)} \frac{\partial z_i}{\partial \eta} \\ &= -\frac{f_{Z_i}(z_i)f''_{Z_i}(z_i) - [f'_{Z_i}(z_i)]^2}{\sigma^2 f_{Z_i}^2(z_i)} \end{aligned} \quad (18)$$

## 2 Left Censored Data

### 2.1 Loss Function

$$L_i = -\log \text{lik}_i = -\log(F_{Y_i}(y_i|\eta)) \quad (19)$$

where, F is the cdf of  $Y_i|\eta$ .

$$L_i = -\log(F(z_i)) \quad (20)$$

## 2.2 Negative Gradient

$$-\frac{\partial L_i}{\partial \eta} = -\frac{\partial -\log(F(z))}{\partial \eta} \quad (21)$$

$$-\frac{\partial L_i}{\partial \eta} = \frac{\partial \log(F(z))}{\partial \eta} \quad (22)$$

$$= \frac{F'(z_i)}{F(z_i)} \frac{\partial z_i}{\partial \eta} \quad (23)$$

Therefore, combining results of chain rule and pdf, below is the final result for negative gradient,

$$-\frac{\partial L_i}{\partial \eta} = \frac{-f(z_i)}{\sigma F(z_i)} \quad (24)$$

## 2.3 Hessian

Hessian is the second derivative of Loss with respect to  $\eta$ . As we have already calculated the negative gradient, further we need to take one more partial derivative with respect to  $\eta$  of negative gradient with negative sign.

$$\frac{\partial^2 L_i}{\partial \eta^2} = \frac{\partial}{\partial \eta} \frac{\partial L_i}{\partial \eta} \quad (25)$$

After inputting the values of negative gradient calculated to above equation.

$$\begin{aligned} &= \frac{\partial}{\partial \eta} * \frac{f(z_i)}{\sigma F(z_i)} \\ &= \frac{\partial}{\partial z_i} \frac{f(z_i)}{\sigma F(z_i)} * \frac{\partial z_i}{\partial \eta} \\ &= \frac{F(z_i)f'(z_i) - f^2(z_i)}{\sigma F^2(z_i)} * \frac{-1}{\sigma} \\ &= -\frac{F(z_i)f'(z_i) - f^2(z_i)}{\sigma^2 F^2(z_i)} \end{aligned} \quad (26)$$

## 3 Right Censored Data

### 3.1 Loss Function

$$L_i = -\log \text{lik}_i = -\log(1 - F_{Y_i}(y_i|\eta)) \quad (27)$$

where, F is the cdf of  $Y_i|\eta$ .

$$L_i = -\log(1 - F(z_i)) \quad (28)$$

### 3.2 Negative Gradient

$$-\frac{\partial L_i}{\partial \eta} = -\frac{\partial -\log(1 - F(z_i))}{\partial \eta} \quad (29)$$

$$-\frac{\partial L_i}{\partial \eta} = \frac{\partial \log(1 - F(z_i))}{\partial \eta} \quad (30)$$

Using similar steps, we have done for left censored.

$$-\frac{\partial L_i}{\partial \eta} = -\frac{F'(z_i)}{1 - F(z_i)} \frac{\partial z_i}{\partial \eta} \quad (31)$$

which is nothing but negative of the negative gradient of left censored data. Therefore,

$$-\frac{\partial L_i}{\partial \eta} = \frac{f(z_i)}{\sigma(1 - F(z_i))} \quad (32)$$

### 3.3 Hessian

$$\frac{\partial^2 L_i}{\partial \eta^2} = \frac{(1 - F(z_i))f'(z_i) + f^2(z_i)}{\sigma^2(1 - F(z_i))^2} \quad (33)$$

## 4 Interval Censored Data

### 4.1 Loss Function

This is combination of left censored and right censored data.

$$L_i = -\log \text{lik}_i = -\log(F_{Y_i^u}(y_i^u|\eta) - F_{Y_i^l}(y_i^l|\eta)) \quad (34)$$

where, F is the cdf of  $Y_i|\eta$ ,  $y_i^u$  is the upper limit of time and  $y_i^l$  is the lower limit of the time. Above equation is written in terms of  $\phi$  notations as below.

$$L_i = -\log(F(z_i^u) - F(z_i^l)) \quad (35)$$

### 4.2 Negative Gradient

$$-\frac{\partial L_i}{\partial \eta} = -\frac{\partial -\log(F(z_i^u) - F(z_i^l))}{\partial \eta} \quad (36)$$

Using chain rule for two variables

$$\frac{\partial L_i}{\partial \eta} = \frac{\partial L_i}{\partial z_i^u} \frac{\partial z_i^u}{\partial \eta} + \frac{\partial L_i}{\partial z_i^l} \frac{\partial z_i^l}{\partial \eta} \quad (37)$$

$$\frac{\partial L_i}{\partial z_i^u} = -\frac{F'(z_i^u)}{F(z_i^u) - F(z_i^l)} \quad (38)$$

$$\frac{\partial L_i}{\partial z_i^l} = \frac{F'(z_i^l)}{F(z_i^u) - F(z_i^l)} \quad (39)$$

$$\frac{\partial L_i}{\partial \eta} = \frac{\partial L_i}{\partial z_i^u} \frac{\partial z_i^u}{\partial \eta} + \frac{\partial L_i}{\partial z_i^l} \frac{\partial z_i^l}{\partial \eta} \quad (40)$$

$$\begin{aligned} -\frac{\partial L_i}{\partial \eta} &= \frac{F'(z_i^u)}{F(z_i^u) - F(z_i^l)} \frac{\partial z_i^u}{\partial \eta} - \frac{F'(z_i^l)}{F(z_i^u) - F(z_i^l)} \frac{\partial z_i^l}{\partial \eta} \\ &= \frac{f(z_i^u)}{F(z_i^u) - F(z_i^l)} \frac{-1}{\sigma} - \frac{f(z_i^l)}{F(z_i^u) - F(z_i^l)} \frac{-1}{\sigma} \\ &= -\frac{f(z_i^u) - f(z_i^l)}{\sigma(F(z_i^u) - F(z_i^l))} \end{aligned} \quad (41)$$

### 4.3 Hessian

Hessian is the second derivative of Loss with respect to  $\eta$ . As we have already calculated the negative gradient, further we need to take one more partial derivative with respect to  $\eta$  of negative gradient with negative sign.

$$\frac{\partial^2 L_i}{\partial \eta^2} = \frac{\partial}{\partial \eta} \frac{\partial L_i}{\partial \eta} \quad (42)$$

Then we apply the Chain Rule. Since  $\partial L_i / \partial \eta$  is now a function of two variables  $z_i^u$  and  $z_i^l$ , the previous equation is broken down to two components:

$$\frac{\partial^2 L_i}{\partial \eta^2} = \frac{\partial}{\partial \eta} \frac{\partial L_i}{\partial \eta} = \frac{\partial}{\partial z_i^u} \frac{\partial L_i}{\partial \eta} \cdot \frac{\partial z_i^u}{\partial \eta} + \frac{\partial}{\partial z_i^l} \frac{\partial L_i}{\partial \eta} \cdot \frac{\partial z_i^l}{\partial \eta} \quad (43)$$

Let us simplify the first term:

$$\begin{aligned} \frac{\partial}{\partial z_i^u} \frac{\partial L_i}{\partial \eta} \cdot \frac{\partial z_i^u}{\partial \eta} &= \frac{\partial}{\partial z_i^u} \frac{f(z_i^u) - f(z_i^l)}{\sigma\{\phi(z_i^u) - \phi(z_i^l)\}} \frac{\partial z_i^u}{\partial \eta} \\ &= \frac{\{F(z_i^u) - F(z_i^l)\} f'(z_i^u) - f(z_i^u)(f(z_i^u) - f(z_i^l))}{\sigma\{F(z_i^u) - F(z_i^l)\}^2} \frac{-1}{\sigma} \\ &= \frac{-\{F(z_i^u) - F(z_i^l)\} f'(z_i^u) + f^2(z_i^u) - f(z_i^u)f(z_i^l)}{\sigma^2\{F(z_i^u) - F(z_i^l)\}^2} \end{aligned} \quad (44)$$

Similarly, we simplify the second term:

$$\begin{aligned} \frac{\partial}{\partial z_i^l} \frac{\partial L_i}{\partial \eta} \cdot \frac{\partial z_i^l}{\partial \eta} &= \frac{\partial}{\partial z_i^l} \frac{f(z_i^u) - f(z_i^l)}{\sigma\{\phi(z_i^u) - \phi(z_i^l)\}} \frac{\partial z_i^l}{\partial \eta} \\ &= \frac{-\{F(z_i^u) - F(z_i^l)\} f'(z_i^l) + f(z_i^l)(f(z_i^u) - f(z_i^l))}{\sigma\{F(z_i^u) - F(z_i^l)\}^2} \frac{-1}{\sigma} \\ &= \frac{\{F(z_i^u) - F(z_i^l)\} f'(z_i^l) + f^2(z_i^l) - f(z_i^l)f(z_i^u)}{\sigma^2\{F(z_i^u) - F(z_i^l)\}^2} \end{aligned} \quad (45)$$

Combining the terms will give the full expression for  $\partial^2 L_i / \partial \eta^2$ :

$$\frac{\partial^2 L}{\partial \eta^2} = - \frac{\{F(z_i^u) - F(z_i^l)\} \{f'(z_i^u) - f'(z_i^l)\} - (f(z_i^u) - f(z_i^l))^2}{\sigma^2 \{F(z_i^u) - F(z_i^l)\}^2} \quad (46)$$