Lecture 2 - LOGISTIC REGRESSION

Ha Minh Duc

August 23rd, 2022

1 Proof: $\frac{\delta L}{\delta W} = \frac{1}{N} X^T(\hat{y} - y)$

$$L = -\frac{1}{N}[y \log \hat{y} + (1-y)\log(1-\hat{y})]$$

$$\rightarrow \frac{\delta L}{\delta \hat{y}} = -\frac{1}{N}(\frac{y}{\hat{y}} - \frac{1-y}{1-\hat{y}}) = \frac{1}{N}\frac{\hat{y}-y}{\hat{y}(1-\hat{y})}$$

$$X = \begin{bmatrix} 1 & {x_1}^{(1)} & {x_2}^{(1)} \\ 1 & {x_1}^{(2)} & {x_2}^{(2)} \\ ... & ... & ... \\ 1 & {x_1}^{(n)} & {x_2}^{(n)} \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{bmatrix}, W = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$

$$\hat{y} = \sigma(XW) = \frac{1}{1+e^{-XW}}$$

$$\begin{aligned} \frac{\delta \hat{y}}{\delta w} &= X^T \frac{e^{-XW}}{(1+e^{-XW})^2} \\ &= X^T \frac{1}{1+e^{-XW}}(1 - \frac{1}{1+e^{-XW}}) \\ &= X^T \hat{y}(1-\hat{y}) \end{aligned}$$

$$\frac{\delta L}{\delta W} = \frac{\delta L}{\delta \hat{y}}\frac{\delta \hat{y}}{\delta w} = \frac{1}{N}X^T(\hat{y} - y)$$

2 Proof that MSE Loss is non-convex, BCE Loss is convex with Logistic Regression

MSE: $L = (\hat{y} - y)^2$

$$\begin{aligned}
\frac{\delta L}{\delta \hat{y}} &= 2(\hat{y} - y) \\
\frac{\delta \hat{y}}{\delta W} &= -X^T \hat{y}(1 - \hat{y}) \\
\rightarrow \frac{\delta L}{\delta W} &= -2X^T(\hat{y} - y)\hat{y}(1 - \hat{y}) \\
&= -2X^T(\hat{y}^2 - y\hat{y} - \hat{y}^3 + y\hat{y}^2) \\
\rightarrow \frac{\delta^2 L}{\delta^2 W} &= 2(X^T)^2\hat{y}(1 - \hat{y})(2\hat{y} - y - 3\hat{y}^2 + y\hat{y})
\end{aligned}$$

$\hat{y} \in (0, 1) \rightarrow \hat{t}(1 - \hat{y}) > 0 \rightarrow 2(X^T)^2\hat{y}(1 - \hat{y}) > 0$
When y=0:

$$2\hat{y} - y - 3\hat{y}^2 + y\hat{y} = 2\hat{y} - 3\hat{y}^2 \ngeq 0 \; \forall \hat{y}$$

When y=1:

$$2\hat{y} - y - 3\hat{y}^2 + y\hat{y} = -3\hat{y}^2 + 3\hat{y} - 1 < 0 \; \forall \hat{y}$$

$\Rightarrow \frac{\delta^2 L}{\delta^2 W} \ngeq 0 \; \forall \hat{y} \Rightarrow$ non convex

BCE: $L = -[ylog\hat{y} + (1 - y)log(1 - \hat{y})]$

$$\begin{aligned}
\frac{\delta L}{\delta W} &= X^T(\hat{y} - y) \\
\rightarrow \frac{\delta^2 L}{\delta^2 W} &= (X^T)^2\hat{y}(1 - \hat{y}) > 0 \; \forall \hat{y} \Rightarrow convex
\end{aligned}$$

Read more at https://towardsdatascience.com/why-not-mse-as-a-loss-function-for-logistic-regression-589816b5e03c