Medium Test 3 Solutions(GSOC with R(XGBoost))

Aditya Samantaray

March 22, 2019

1. Explain why the first expression is equivalent to the second expression.

Solution: y_i can take on two discrete values i.e. 0 and 1, Hence $P(y_i|x_i)$ is defined as,

$$P(y_i|x_i) = \begin{cases} \frac{1}{1+e^{-\hat{y_i}}} & y_i = 1 \\ 1 - \frac{1}{1+e^{-\hat{y_i}}} & y_i = 0 \end{cases} \tag{1}$$

Where $\hat{y_i}$ is a prediction score and ranges from $-\infty$ to ∞ for every x_i
$\therefore$ We see that,

$$P(y_i = 1|x_i) = 1/1 + e^{-\hat{y_i}} \tag{2}$$

$$P(y_i = 0|x_i) = 1 - 1/1 + e^{-\hat{y_i}} \tag{3}$$

The above two equations can be combined to give,

$$P(y_i|x_i) = (\frac{1}{1+e^{-\hat{y_i}}})^{y_i}.(1 - \frac{1}{1+e^{-\hat{y_i}})^{1-y_i}}) \tag{4}$$

Hence, we see that equation (1) is equivalent to (4)

2. Explain how minimizing the loss function $loss(y_i, yhat_i)$ is equivalent to maximizing the probability $P(y_i|x_i)$.

Solution : Given,

$$loss(y_i, \hat{y}_i = -\log(P(y_i|x_i)) \tag{5}$$

$P(y_i|x_i)$ is constrained between 0 and 1. As the value of $P(y_i|x_i)$ increases from 0 to 1, the value of $\log(P(y_i|x_i))$ also increases. Consequently, $-\log(P(y_i|x_i))$ decreases.
Since our loss function is defined as $-\log(P(y_i|x_i))$, we can say that minimizing the loss function $loss(y_i, yhat_i)$ is equivalent to maximizing the probability $P(y_i|x_i)$.

3. Simplify the expression for $loss(y_i, yhat_i)$. Show your steps (i.e. don't just write the answer, show how you got it).

Solution :

$$\begin{aligned}
loss(y_i, \hat{y}_i) &= -\log\left(\left(\frac{1}{1+e^{\hat{y}_i}}\right)^{y_i} \times \left(1 - \frac{1}{1+e^{\hat{y}_i}}\right)^{1-y_i}\right) \\
&= -\log\left(\left(\frac{1}{1+e^{\hat{y}_i}}\right)^{y_i} \times \left(\frac{e^{-\hat{y}_i}}{1+e^{\hat{y}_i}}\right)^{1-y_i}\right) \\
&= y_i \log(1+e^{-\hat{y}_i}) - (1-y_i)\log e^{-\hat{y}_i} + (1-y_i)\log(1+e^{-\hat{y}_i}) \\
&= (y_i + (1-y_i))\log(1+e^{\hat{y}_i}) - (1-y_i)(-\hat{y}_i)
\end{aligned}$$

$$\therefore loss(y_i, \hat{y}_i) = \hat{y}_i - y_i\hat{y}_i + \log(1+e^{-\hat{y}_i}) \tag{6}$$

4. Now compute the first and second partial derivatives of $loss(y_i, yhat_i)$ with respect to the second variable $yhat_i$. Then express the two derivatives in terms of $sigmoid(yhat_i)$. Notice how simple the expressions become. Again, show your steps (i.e. don't just write the answer, show how you got it).

Solution : From (6), the loss funtion can be given as

$$loss(y_i, \hat{y}_i) = \hat{y}_i - y_i\hat{y}_i + \log(1+e^{-\hat{y}_i})$$

Now, calculating the first partial derivative w.r.t. $\hat{y}_i$

$$\begin{aligned}
\frac{\partial(loss(y_i, \hat{y}_i))}{\partial \hat{y}_i} &= -\frac{e^{-\hat{y}_i}}{1+e^{-\hat{y}_i}} - y_i + 1 \\
&= -\left(1 - \frac{1}{1+e^{-\hat{y}_i}}\right) - y_i + 1 \\
&= -(1 - sigmoid(\hat{y}_i) - y_i + 1
\end{aligned}$$

$$\therefore \frac{\partial(loss(y_i, \hat{y}_i))}{\partial \hat{y}_i} = sigmoid(\hat{y}_i) - y_i \tag{7}$$

Calculating the second partial derivative w.r.t. $\hat{y}_i$

$$\begin{aligned}
\frac{\partial(loss(y_i, \hat{y}_i))}{\partial \hat{y}_i} &= \frac{1}{1+e^{-\hat{y}_i}} - y_i \\
\frac{\partial^2(loss(y_i, \hat{y}_i))}{\partial \hat{y}_i^{\,2}} &= \left(\frac{e^{-\hat{y}_i}}{(1+e^{-\hat{y}_i})^2}\right)\left(\frac{1}{1+e^{-\hat{y}_i}}\right)
\end{aligned}$$

$$\therefore \frac{\partial^2(loss(y_i, \hat{y}_i))}{\partial \hat{y}_i^{\,2}} = (1 - sigmoid(\hat{y}_i)) \times (sigmoid(\hat{y}_i)) \tag{8}$$

Hence we see in (7) and (8) that the partial derivatives do reduce to simple expressions.

5. In the source code $src/objective/regression_loss.h$, locate the structure that implements this loss function.

Solution : (6) is implemented in the following code present at the 72nd line of code.

```
// logistic loss for binary classification task
struct LogisticClassification : public LogisticRegression {
  static const char* DefaultEvalMetric() { return "error"; }
};
```