Content

- Pseudo Residual for log loss

Pseudo Residual for log loss

To show that the negative of the derivative of the loss function (log-loss) is psuedo residual [$-(\frac{d(L^{log})}{d(F_k(x_i))})$] which is similar to the residual (difference between actual and predicted values)

Let us assume our loss function be Log-loss and our problem is a binary classification problem

- The output of the k_{th} model $F_k(x_i) = p_i = p((y_i = 1)|x_i)$ that is probability of y_i being 1, given x_i
    - where y_i is the actual class label

We know the **loss function is log loss**

- $L = logloss(y_i, p_i) = y_i log p_i + (1 - y_i) log(1 - p_i)$

Therefore,

- $-(\frac{\partial L(y_i, p_i)}{\partial (p_i)}) = -\frac{\partial ([y_i log p_i + (1-y_i) log(1-p_i))}{\partial p_i}$

Using product rule,

$$\frac{\partial L(y_i, p_i)}{\partial (p_i)} = y_i . \frac{\partial (log p_i)}{\partial p_i} + log p_i . \frac{\partial (y_i)}{\partial p_i} + (1 - y_i) \frac{\partial (log(1-p_i))}{\partial p_i} + log(1 - p_i) . \frac{\partial (1-y_i)}{\partial p_i}$$

y_i is constant when we take derivative w.r.t p_i

The derivative comes out to be

$$-\frac{\partial L(y_i, p_i)}{\partial (p_i)} = -[\frac{y_i}{p_i} - \frac{1-y_i}{1-p_i}] = \frac{p_i(i-y_i) - y_1(1-p_i)}{p_i(1-p_i)}$$

so, $-(\frac{dL}{d(p_i)}) = \frac{p_i - y_i}{p_i(1-p_i)}$

- Here the numerator is capturing the difference which is the $(\hat{y_i} - y_i)$ which is key a
- Consider denomination $p_i(1 - p_i)$ as a normalizing factor
- Hence proved that the when we use log-loss the psuedo residual behaves like a residual.