Implementing Binomial Loss in XGBoost

Aditya Samantaray

March 22, 2019

The binomial distribution is a finite discrete distribution and arises in situations where one is observing a sequence of Bernoulli trials. The density function of a binomial distribution is defined as

$$f(y,N,p) = {}^{N}C_y p^y (1-p)^{N-y} \tag{1}$$

where, N represents the number of trials and p represents the
success in each trial. $f(y,N,p)$ represents the probability of exactly y successes out of N trials

Using the maximum likelihood estimation, the likelihood for (1) becomes

$$L(p,y,N) = {}^{N}C_y p^y.(1-p)^{N-y} \tag{2}$$

In maximum likelihood estimation, we try to maximize ${}^{N}C_y p^y (1-p)^{N-y}$, but maximizing it is the same as maximizing $p^y(1-p)^{N-y}$. Hence our likelihood function can be rewritten as

$$L(p,y,N) = p^y.(1-p)^{N-y} \tag{3}$$

Now,
For our loss function, we take the negative logarithm of the likelihood obtained in (3)

$$loss(y_i) = -\log_e(p^{y_i}.(1-p)^{1-y_i}) \tag{4}$$

$$\therefore loss(y_i) = -(y_i \log_e p + (1-y_i)\log_e(1-p)) \tag{5}$$

We see that the loss function obtained is similar to the logistic loss function. Hence we can implement binomial loss by setting the objective parameter to *binary.logistic* within the parameter list