Medium1

Avinash Barnwal

March 2019

Problem Write a vignette in LaTeX or MathJax explaining how to use the logistic loss with non-uniform weights to get the binomial loss function in xgboost.
Answer - Following is likelihood function of binomial loss with count of trials and success :-

$$L(\theta; x) = \prod_{i=1}^{n} \binom{n_i}{n_i y_i} p(\theta; x)^{(n_i y_i)} (1 - p(\theta; x))^{(n_i - n_i y_i)} \tag{1}$$

$$log(L(\theta; x)) = \sum_{i=1}^{n} n_i y log(p(\theta; x)) + (n_i - n_i y_i) log(1 - p(\theta; x)) + \sum_{i=1}^{n} log(\binom{n_i}{n_i y_i}) \tag{2}$$

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} n_i y_i log(p(\theta; x)) + (n_i - n_i y_i) log(1 - p(\theta; x)) + \sum_{i=1}^{n} log(\binom{n_i}{n_i y_i}) \tag{3}$$

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} n_i \{y_i log(p(\theta; x)) + (1 - y_i) log(1 - p(\theta; x))\} + \sum_{i=1}^{n} log(\binom{n_i}{n_i y_i}) \tag{4}$$

Above equation is nothing but logistic loss with weight n_i and y_i is proportions of success rather than 0-1 like in logistic loss.
Therefore , for xgboost y_i(label in design matrix) will change to proportion of success and weight will be passed with values equal to number of trials with logistic loss.