Binomial Loss

Avinash Barnwal

June 2019

1 Binomial loss

We use the notation of the wikipedia article on "Binomial distribution"

- $n \in \mathbb{Z}_+ = \{0, 1, \dots\}$ be the total number of trials,
- $k \in \{0, 1, \dots, n\}$ be the number of successful trials,
- $p \in [0, 1]$ be the probability of a successful trial

Our model is $k \sim \text{Binomial}(n, p)$. Maximizing the log-likelihood is equivalent to minimizing a loss function $\ell : (0, 1) \to \mathbb{R}_+$ from predicted probability values to loss values:

$$\ell(p) = (k - n) \log(1 - p) - k \log p \tag{1}$$

Let $u = \log p - \log(1 - p) = \log(\frac{p}{1-p}) \in \mathbb{R}$ be a real-valued prediction variable (better for numerical stability to store predicted values as u than p values), so $p = (1 + e^{-u})^{-1}$. Thus the loss can be written as a real-valued function $g : \mathbb{R} \to \mathbb{R}_+$ from predicted values on the real line ($p = 1/2$ means $u = 0$) to loss values:

$$\begin{aligned}
\ell(p) &= (k - n) \log(1 - p) - k \log p \\
&= (k - n) \log(1 - \frac{1}{1 + e^{-u}}) - k \log(\frac{1}{1 + e^{-u}}) \\
g(u) &= (n - k) \log(1 + e^{u}) + k \log(1 + e^{-u})
\end{aligned}$$

In this last equation above we see that the overall binomial loss function g is just a weighted sum of logistic losses.

Please use this g function as the binomial loss to implement, and derive gradient/hessian for this.

In my formulas above the real-valued u variable is the equivalent of the η parameter in the sections below.

2 Binomial Loss with count upper bound

This loss function is different than logistic regression based on two ways.
Different count for each data sample. Response variable is proportion.
Important Formulas:-

$$\begin{aligned} P(y=1) &= \frac{\exp^{x\beta}}{1+\exp^{x\beta}} + \epsilon_i \\ \log \frac{P(y=1)}{1-P(y=1)} &= x\hat{\beta} \\ \eta &= x\hat{\beta} \end{aligned} \tag{2}$$

$$lik = \prod_{i=1}^{n} \binom{n_i}{y_i} p(\theta;x)^{y_i} (1-p(\theta;x))^{\{n_i - y_i\}} \tag{3}$$

$$-\,log - lik = -\sum_{i=1}^{n} \{y_i \log(p(\theta;x)) + (n_i - y_i) log(1-p(\theta;x))\} \tag{4}$$

$$L_i = -log - lik_i = -y_i \log\left(\frac{1}{1+\exp^{-\eta_i}}\right) + (-n_i + y_i) log\left(1 - \frac{1}{1+\exp^{-\eta_i}}\right) \tag{5}$$

$$L_i = -log - lik_i = y_i \log(1+\exp^{-\eta_i}) + (n_i - y_i) log(1+\exp^{\eta_i}) \tag{6}$$

3 Negative Gradient

$$\begin{aligned} -\frac{\partial L_i}{\partial \eta_i} &= -\frac{\partial y_i \log(1+\exp^{-\eta_i}) + (n_i - y_i) log(1+\exp^{\eta_i})}{\partial \eta_i} \\ &= -\{-y_i \frac{\exp^{-\eta_i}}{1+\exp^{-\eta_i}} + (n_i - y_i)\frac{\exp^{\eta_i}}{1+\exp^{\eta_i}}\} \\ &= y_i - n_i \frac{\exp^{\eta_i}}{1+\exp^{\eta_i}} \\ &= y_i - n_i p_i \end{aligned} \tag{7}$$

4 Hessian

$$\begin{aligned} \frac{\partial^2 L_i}{\partial \eta_i^2} &= \frac{\partial - y_i + n_i \frac{\exp^{\eta_i}}{1+\exp^{\eta_i}}}{\partial \eta_i} \\ = n_i \frac{(1+\exp^{\eta_i})\exp^{\eta_i} - (\exp^{\eta_i})\exp^{\eta_i}}{(1+\exp^{\eta_i})^2} \\ &= n_i \frac{\exp^{\eta_i}}{(1+\exp^{\eta_i})^2} \end{aligned} \tag{8}$$