Logistic

Weike Pan

College of Computer Science and Software Engineering
Shenzhen University

Bernoulli distribution

The Bernoulli distribution of binary random variable $c \in \{0, 1\}$:

$$f(c; p) = p^c(1 - p)^{1-c}, \tag{1}$$

where $p = Pr(y = 1)$.

Logit

The log-odds of an event is called the logit of the probability of the event,

$$logit(p) = \log \frac{p}{1-p} = \log(p) - \log(1-p), \tag{2}$$

where $p = Pr(y = 1)$ denotes the probability of the event.

Let $\theta = logit(p) = \log \frac{p}{1-p}$, we have $p = \frac{1}{1+\exp(-\theta)}$.

Logistic function

The (unitary) Logistic function (i.e., sigmoid function),

$$\sigma(\theta) = \frac{\exp(\theta)}{1+\exp(\theta)} = \frac{1}{1+\exp(-\theta)} \in (0,1), \tag{3}$$

and we have,

$$\begin{aligned}
\sigma(-\theta) &= 1-\sigma(\theta), \\
\frac{\partial \sigma(\theta)}{\partial \theta} &= \sigma(\theta)(1-\sigma(\theta)) = \sigma(\theta)\sigma(-\theta), \\
\frac{\partial \log \sigma(\theta)}{\partial \theta} &= \frac{1}{\sigma(\theta)}[\sigma(\theta)(1-\sigma(\theta))] = 1-\sigma(\theta) = \sigma(-\theta).
\end{aligned}$$

Logistic loss function (1/3)

The Logistic loss function ($\{0,1\}$ case) between y and $\sigma(\boldsymbol{w}'\boldsymbol{x})$,

$$\begin{aligned} \ell_{01}(y, \sigma(\boldsymbol{w}'\boldsymbol{x})) &= -[y \log \sigma(\boldsymbol{w}'\boldsymbol{x}) + (1-y)\log(1-\sigma(\boldsymbol{w}'\boldsymbol{x}))] \\ &= \boldsymbol{w}'\boldsymbol{x} - y\boldsymbol{w}'\boldsymbol{x} + \log(1+\exp(-\boldsymbol{w}'\boldsymbol{x})), \end{aligned} \tag{4}$$

where $y \in \{0,1\}$ is the true label, and $\sigma(\boldsymbol{w}'\boldsymbol{x}) \in (0,1)$ is the prediction, i.e., the probability that instance $\boldsymbol{x}$ belongs to category "1".

- when $y = 1$: $-\log \sigma(\boldsymbol{w}'\boldsymbol{x}) = \log(1+\exp(-\boldsymbol{w}'\boldsymbol{x}))$
- when $y = 0$: $-\log(1-\sigma(\boldsymbol{w}'\boldsymbol{x})) = \boldsymbol{w}'\boldsymbol{x} + \log(1+\exp(-\boldsymbol{w}'\boldsymbol{x}))$

Logistic loss function (2/3)

The Logistic loss function ($\{-1,1\}$ case) between y and $\boldsymbol{w}'\boldsymbol{x}$,

$$\ell_{\pm 1}(y,\boldsymbol{w}'\boldsymbol{x}) = \log(1+\exp(-y\boldsymbol{w}'x)) = -\log\sigma(y\boldsymbol{w}'\boldsymbol{x}), \quad (5)$$

where $y \in \{-1,1\}$ is the true label, and $\boldsymbol{w}'\boldsymbol{x}$ is the prediction.

- when $y = 1$: $-\log\sigma(\boldsymbol{w}'\boldsymbol{x})$
- when $y = -1$/: $-\log\sigma(-\boldsymbol{w}'\boldsymbol{x}) = -\log(1-\sigma(\boldsymbol{w}'\boldsymbol{x}))$

Notes:

- The Logistic loss function in Eq.(4) and the Logistic loss function in Eq.(5) is equivalent.

Logistic loss function (3/3)

The cross entropy loss function between y and $\boldsymbol{w}'\boldsymbol{x}$ [Wu et al., 2016],

$$\begin{aligned} \ell_{ce}(y, \boldsymbol{w}'\boldsymbol{x}) &= -[y \log \sigma(\boldsymbol{w}'\boldsymbol{x}) + (1-y)\log(1-\sigma(\boldsymbol{w}'\boldsymbol{x}))] \\ &= \boldsymbol{w}'\boldsymbol{x} - y\boldsymbol{w}'\boldsymbol{x} + \log(1+\exp(-\boldsymbol{w}'\boldsymbol{x})), \end{aligned} \tag{6}$$

where $y \in \{0, 1\}$ is the true label, and $\boldsymbol{w}'\boldsymbol{x}$ is the prediction.

- when $y = 1$: $-\log \sigma(\boldsymbol{w}'\boldsymbol{x}) = \log(1+\exp(-\boldsymbol{w}'\boldsymbol{x}))$
- when $y = 0$: $-\log(1-\sigma(\boldsymbol{w}'\boldsymbol{x})) = \boldsymbol{w}'\boldsymbol{x} + \log(1+\exp(-\boldsymbol{w}'\boldsymbol{x}))$

Notes:

- The cross entropy loss function in Eq.(6), the Logistic loss function in Eq.(4) and the Logistic loss function in Eq.(5) are equivalent.

Objective function (1/2)

The objective function,

$$\min_{\boldsymbol{w}} \frac{1}{2}\|\boldsymbol{w}\|_F^2 + C\sum_{i=1}^{N} \ell_{01}(y_i, \sigma(\boldsymbol{w}'\boldsymbol{x}_i)), \tag{7}$$

where $y_i \in \{0, 1\}$ is the true label of instance $\boldsymbol{x}_i$.

Objective function (2/2)

The objective function,

$$\min_{\boldsymbol{w}} \frac{1}{2}||\boldsymbol{w}||_F^2 + C\frac{1}{N}\sum_{i=1}^{N} \ell_{\pm 1}(y_i, \boldsymbol{w}'\boldsymbol{x}_i), \quad (8)$$

where $y_i \in \{-1, 1\}$ is the true label of instance $\boldsymbol{x}_i$.

Notes:

- The objective function in Eq.(7) and the objective function in Eq.(8) is equivalent.
- In the liblinear toolbox, the loss function is $\ell_{\pm 1}(y_i, \boldsymbol{w}'\boldsymbol{x}_i)$.

Derivations

For a randomly generated sample $(\boldsymbol{x}_i, y_i)$, we have the tentative objective function of the Logistic regression (LR) model as follows,

$$f(\boldsymbol{w}; \boldsymbol{x}_i, y_i) = \log(1 + \exp(-y_i \boldsymbol{w}' x_i)) + \frac{\lambda}{2} ||\boldsymbol{w}||_F^2, \tag{9}$$

where $y_i \in \{-1, 1\}$ is the true label of instance $\boldsymbol{x}_i$.

We then have the gradient,

- $\nabla \boldsymbol{w} = \frac{\partial f(\boldsymbol{w}; \boldsymbol{x}_i, y_i)}{\partial \boldsymbol{w}} = -\sigma(-y_i \boldsymbol{w}' \boldsymbol{x}_i) y_i \boldsymbol{x}_i + \lambda \boldsymbol{w}$

Finally, we have the update rule,

- $\boldsymbol{w} = \boldsymbol{w} - \gamma \nabla \boldsymbol{w}$

where γ is the learning rate.

Wu, Y., DuBois, C., Zheng, A. X., and Ester, M. (2016).
Collaborative denoising auto-encoders for top-n recommender systems.
In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, San Francisco, CA, USA, February 22-25, 2016*, pages 153–162.