Notes on Logistic Loss Function

Liangjie Hong

October 3, 2011

1 Logistic Function & Logistic Regression

The common definition of Logistic Function is as follows:

$$P(x) = \frac{1}{1 + \exp(-x)} \tag{1}$$

where $x \in \mathbb{R}$ is the variable of the function and $P(x) \in [0, 1]$. One important property of Equation (1) is that:

$$
\begin{aligned}
P(-x) &= \frac{1}{1 + \exp(x)} \\
&= \frac{1}{1 + \frac{1}{\exp(-x)}} \\
&= \frac{\exp(-x)}{1 + \exp(-x)} \\
&= 1 - \frac{1}{1 + \exp(-x)} \\
&= 1 - P(x) \tag{2}
\end{aligned}
$$

The form of Equation (1) is widely used as the form of Logistic Regression (e.g., [1, 2, 3]):

$$
\begin{aligned}
P(y = 1 \,|\, \boldsymbol{\beta}, \mathbf{x}) &= \frac{\exp(\boldsymbol{\beta}^T \mathbf{x})}{1 + \exp(\boldsymbol{\beta}^T \mathbf{x})} \\
P(y = 0 \,|\, \boldsymbol{\beta}, \mathbf{x}) &= \frac{1}{1 + \exp(\boldsymbol{\beta}^T \mathbf{x})} \tag{3}
\end{aligned}
$$

where $\mathbf{x}$ is a feature vector and $\boldsymbol{\beta}$ is a coefficient vector. By using Equation (2), we also have:

$$P(y = 1 \,|\, \boldsymbol{\beta}, \mathbf{x}) = 1 - P(y = 0 \,|\, \boldsymbol{\beta}, \mathbf{x}) \tag{4}$$

This formalism of Logistic Regression is used in [1, 2] where labels $y \in \{0, 1\}$ and the functional form of the probability to generate different labels is different. Another formalism introduced in [3] unified the two forms into one single equation by integrating the label and the prediction together:

$$P(g = \pm 1 \,|\, \boldsymbol{\beta}, \mathbf{x}) = \frac{1}{1 + \exp(-g\boldsymbol{\beta}^T \mathbf{x})} \tag{5}$$

where $g \in \{\pm 1\}$ is the label for data item x. It is also easily to verify that $P(g = 1 \,|\, \boldsymbol{\beta}, \mathbf{x}) = 1 - P(g = -1 \,|\, \boldsymbol{\beta}, \mathbf{x})$.

2 The Equivalence of Two Forms of Logistic Regression

At first glance, the form (3) and the form (5) looks very different. However, the equivalence between these two forms can be easily established. Starting from the form (3), we can have:

$$
\begin{aligned}
P(y = 1 \,|\, \boldsymbol{\beta}, \mathbf{x}) &= \frac{\exp(\boldsymbol{\beta}^T \mathbf{x})}{1 + \exp(\boldsymbol{\beta}^T \mathbf{x})} \\
&= \frac{1}{\frac{1}{\exp(\boldsymbol{\beta}^T \mathbf{x})} + 1} \\
&= \frac{1}{\exp(-\boldsymbol{\beta}^T \mathbf{x}) + 1} \\
&= P(g = 1 \,|\, \boldsymbol{\beta}, \mathbf{x})
\end{aligned}
\tag{6}
$$

We can also establish the equivalence between $P(y = 0 \,|\, \boldsymbol{\beta}, \mathbf{x})$ and $P(g = -1 \,|\, \boldsymbol{\beta}, \mathbf{x})$ easily by using property (2). Another way to establish the equivalence is from the classification rule. For the form (3), we have the following classification rule:

$$
\begin{aligned}
\frac{\frac{\exp(\boldsymbol{\beta}^T \mathbf{x})}{1 + \exp(\boldsymbol{\beta}^T \mathbf{x})}}{\frac{1}{1 + \exp(\boldsymbol{\beta}^T \mathbf{x})}} &> 1 \;\; \rightarrow \;\; y = 1 \\
\exp(\boldsymbol{\beta}^T \mathbf{x}) &> 1 \\
\boldsymbol{\beta}^T \mathbf{x} &> 0
\end{aligned}
\tag{7}
$$

An exactly same classification rule for the form (5) can also be obtained as:

$$
\begin{aligned}
\frac{\frac{1}{1 + \exp(-\boldsymbol{\beta}^T \mathbf{x})}}{\frac{1}{1 + \exp(\boldsymbol{\beta}^T \mathbf{x})}} &> 1 \;\; \rightarrow \;\; g = 1 \\
\frac{1 + \exp(\boldsymbol{\beta}^T \mathbf{x})}{1 + \exp(-\boldsymbol{\beta}^T \mathbf{x})} &> 1 \\
\exp(\boldsymbol{\beta}^T \mathbf{x}) &> 1 \\
\boldsymbol{\beta}^T \mathbf{x} &> 0
\end{aligned}
\tag{8}
$$

Therefore, we can see that two forms essentially learn the same classification boundary.

3 Logistic Loss

Since we establish the equivalence of two forms of Logistic Regression, it is convenient to use the second form as it can be explained by a general classification framework. Here, we assume y is the label of data and $\mathbf{x}$ is a feature vector. The classification framework can be formalized as follows:

$$
\arg\min \sum_i L\Big(y_i, f(\mathbf{x}_i)\Big)
\tag{9}
$$

where f is a hypothesis function and L is loss function. For Logistic Regression, we have the following instantiation:

$$
\begin{aligned}
f(\mathbf{x}) &= \boldsymbol{\beta}^T \mathbf{x} \\
L\Big(y, f(\mathbf{x})\Big) &= \log\Big(1 + \exp(-y f(\mathbf{x}))\Big)
\end{aligned}
\tag{10}
$$

where $y \in \{\pm 1\}$.

References

[1] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning.* Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

[2] Tom M. Mitchell. *Machine learning.* McGraw Hill series in computer science. McGraw-Hill, 1997.

[3] Jason D. M. Rennie. Logistic Regression. http://people.csail.mit.edu/jrennie/writing, April 2003.