CSCI 5525 Machine Learning Fall 2019

Lecture 5: Logistic Regression

Feb 10 2020

Lecturer: Steven Wu *Scribe: Steven Wu*

Last lecture, we give several convex surrogate loss functions to replace the zero-one loss function, which is NP-hard to optimize. Now let us look into one of the examples, logistic loss: given parameter $\mathbf{w}$ and example $(x_i, y_i) \in \mathbb{R}^d \times \{\pm 1\}$, the logistic loss of $\mathbf{w}$ on example (x_i, y_i) is defined as

$$\ln\left(1 + \exp(-y_i \mathbf{w}^\intercal x_i)\right)$$

This loss function is used in logistic regression. We will introduce the statistical model behind logistic regression, and show that the ERM problem for logistic regression is the same as the relevant maximum likelihood estimation (MLE) problem.

1 MLE Derivation

For this derivation it is more convenient to have $\mathcal{Y} = \{0, 1\}$. Note that for any label $y_i \in \{0, 1\}$, we also have the "signed" version of the label $2y_i - 1 \in \{-1, 1\}$. Recall that in general supervised learning setting, the learner receive examples $(x_1, y_1), \ldots, (x_n, y_n)$ drawn iid from some distribution P over labeled examples. We will make the following parametric assumption on P:

$$y_i \mid x_i \sim \mathrm{Bern}(\sigma(\mathbf{w}^\intercal x_i))$$

where Bern denotes the Bernoulli distribution, and σ is the *logistic function* defined as follows

$$\sigma(z) = \frac{1}{1 + \exp(-z)} = \frac{\exp(z)}{1 + \exp(z)}$$

See Figure 1 for a visualization of the logistic function. In general, the logistic function is a useful function to convert real values into probabilities (in the range of $(0, 1)$). If $\mathbf{w}^\intercal x$ increases, then