5 Logistic Regression

黄一凡

1 Generative and Discriminative Classifiers

The most important difference between naive Bayes and logistic regression is that logistic regression is a discriminative classifier while naive Bayes is a generative classifier.

Imagine we're trying to distinguish dog images from cat images:

Generative Classifier A generative model would have the goal of understanding what dogs look like and what cats look like. Given a test image, the system then asks whether it's the cat model or the dog model that better fits the image, and choose that as its label.

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \tag{1}$$

Discriminative Classfier A discriminative model, by contrast, is only trying to learn to distinguish the classes(perhaps without learning much about them).

By contrast a discriminative model in this text categorization scenario attempts to directly compute $p(y|x)$. Perhaps it will learn to assign a high weight to document features that directly improve its ability to discriminative between possible classes, even if it couldn't generate an example of one of the classes.

2 Sigmoid Classifier

Logistic regression solves classification task by learning, from a training set, a vector of weights and a bias term.

$$z = \sum_{i=1}^{n} w_i x_i + b = w \cdot x + b \tag{2}$$

But note that z is not a legal probability, that is, to lie between 0 and 1. In fact, since weights are real -valued, the output might even be negative, z ranges from $-\infty to \infty$.

To create a probability, we'll pass z through the sigmoid function, $\sigma(z)$.

$$y = \sigma(z) = \frac{1}{1+e^{-z}} \tag{3}$$

In order to avoid the extensive human effort of feature design, recent research in NLP has focused on representation learning: ways to learn features automatically in an unsupervised way from the input.

Logistic regression is much more robust to correlated features. Naive Bayes can work extremely well on very small datasets. Furthermore, naive Bayes is easy to implement and very fast to train(there's no optimization step).

3 Learning in Logistic Regression

We use cross-entropy loss and stochastic gradient desent algorithm to learn weights.

4 Multinomial Logistic Regression

Sometimes we need more than two classes. In such cases we use multinomial logistic regerssion, also called softmax regression(or, historically, the maxnet classifier).

The multinomial logistic classifier uses a generalization of the sigmoid, called the softmax function.

$$p(c|x) = \frac{e^{w_c \cdot x + b_c}}{\sum_{j=1}^{k} e^{w_j \cdot x + b_j}} \tag{4}$$