Multinomial Logistic Regression

The Multinomial Logistic Regression is a generalization of the logistic regression to multivariate classification problems. The theoretical background of multinomial logistic regression algorithm is as follows:

Assume the training dataset consist of m (x_i, y_i) pairs and k is the number of all possible classes. The probability of given x to be classified as y is:

$$\boldsymbol{p}(\boldsymbol{y}^{(i)} = \boldsymbol{j} \,|\, \boldsymbol{x}^{(i)};\theta) = \frac{\exp(\theta_{\boldsymbol{j}}^{\boldsymbol{T}} \boldsymbol{x}^{(i)})}{\sum_{\boldsymbol{i}=1}^{\boldsymbol{k}} \exp(\theta_{\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{x}^{(i)})}$$

Thus the hypothesis function will return a k dimensional vector with the estimated probabilities, as follows:

$$\boldsymbol{h}_\theta(\boldsymbol{x}^{(i)}) = \left[\boldsymbol{p}_1, \boldsymbol{p}_{2,\dots}\boldsymbol{p}_{\boldsymbol{k}}\right]^{\boldsymbol{T}} = \frac{1}{\sum_{\boldsymbol{i}=1}^{\boldsymbol{k}} \exp(\theta_{\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{x}^{(i)})} \left[\exp(\theta_1^{\boldsymbol{T}} \boldsymbol{x}^{(i)}), \exp(\theta_2^{\boldsymbol{T}} \boldsymbol{x}^{(i)}), \dots \exp(\theta_{\boldsymbol{k}}^{\boldsymbol{T}} \boldsymbol{x}^{(i)})\right]^{\boldsymbol{T}}$$

The category with the highest probability will be chosen as the classification for the entry x. The parameter θ is calculated by minimizing the cost function J using gradient descent. Thus, the Multinomial Logistic Regression algorithm requires significantly more time to be trained comparing to Naive Bayes,

$$\boldsymbol{J}(\theta) = -\frac{1}{\boldsymbol{m}} \left[\sum_{\boldsymbol{i}=1}^{\boldsymbol{m}} \sum_{\boldsymbol{j}=1}^{\boldsymbol{k}} \{\boldsymbol{y}^{(i)} = \boldsymbol{j}\} \log \frac{\exp(\theta_{\boldsymbol{j}}^{\boldsymbol{T}} \boldsymbol{x}^{(i)})}{\sum_{\boldsymbol{i}=1}^{\boldsymbol{k}} \exp(\theta_{\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{x}^{(i)})} \right]$$

Here $\{\boldsymbol{y}^{(\boldsymbol{i})} = \boldsymbol{j}\}$ is the proportion of y(i) over sum of all y(i) values.

Reference:

http://blog.datumbox.com/machine-learning-tutorial-the-multinomial-logistic-regression-softmax-regression/