Multi-class Classification/Multinomial Logistic Regression

In multi-class or multinomial logistic regression, the labels of the dataset contain more than 2 classes. The multinomial logistic regression setup (i.e., the cost function and optimization procedure) is structurally similar to logistic regression; the only difference is that the output of logistic regression is 2 classes, while multinomial has greater than 2 classes (see Figure 20-6).

In Figure 20-6, the multi-class logistic regression builds a one-vs.-rest classifier to construct decision boundaries for the different class memberships.

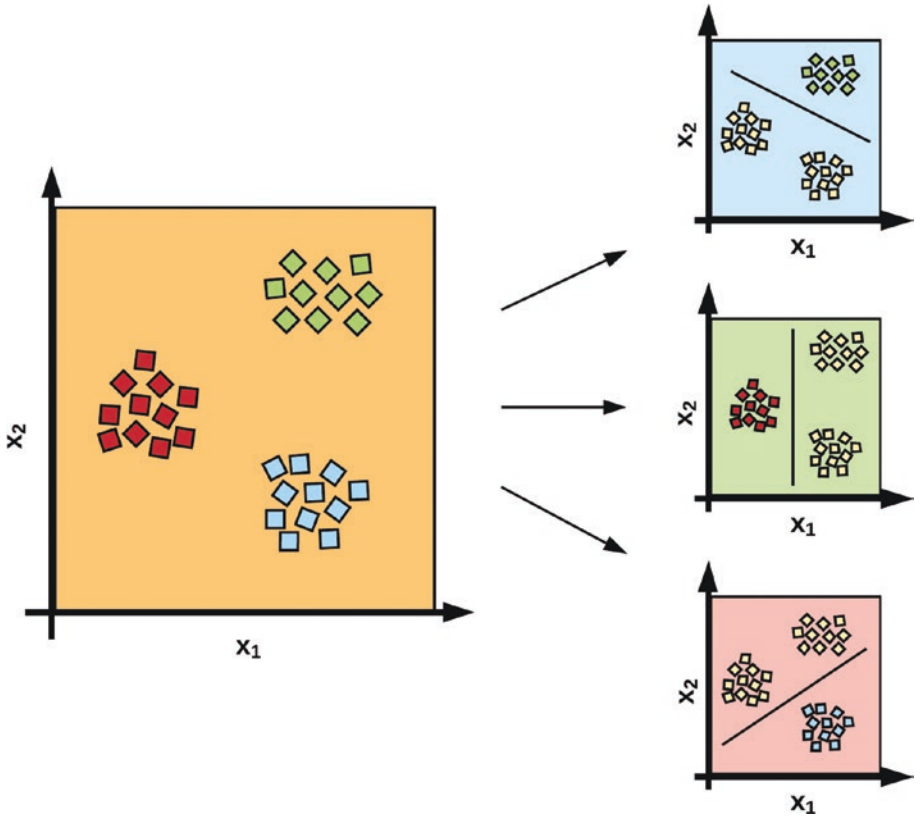

Figure 20-6. *An illustration of multinomial regression*

At this point, we introduce a critical function in machine learning called the softmax function. The softmax function is used to compute the probability that an instance belongs to one of the K classes when $K > 2$. We will see the softmax function show up again when we discuss (artificial) neural networks.

In order to build a classification model with k classes, the multinomial logistic model is formally defined as

$$\hat{y}(k) = \theta_0^k + \theta_1^k x_1 + \theta_2^k x_2 + \ldots + \theta_n^k x_n$$

The preceding model takes into consideration the parameters for the k different classes. The softmax function is formally written as

$$p(k) = \sigma\left(\hat{y}(k)\right)_i = \frac{e^{\hat{y}(k)_i}}{\sum_{j=1}^{K} e^{\hat{y}(k)_j \hat{} (k_j}}$$

where

- $i = \{1, \ldots, K\}$ classes.
- $\sigma\left(\hat{y}(k)\right)_i$ outputs the probability estimates that an example in the training dataset belongs to one of the K classes.

The cost function for learning the class labels in a multinomial logistic regression model is called the ***cross-entropy*** cost function. Gradient descent is used to find the optimal values of the parameter θ that will minimize the cost function to ***predict the class with the highest probability estimate accurately***.

Logistic Regression with Scikit-learn

In this example, we will implement a multi-class logistic regression model with Scikit-learn. The model will predict the three species of flowers from the Iris dataset. The dataset contains 150 observations and 4 features. For this example, we use the accuracy metric and confusion matrix to access the model's performance.

```
# import packages
from sklearn.linear_model import LogisticRegression
from sklearn import datasets
```