Handling Imbalanced Classes In Logistic Regression

20 Dec 2017

Like many other learning algorithms in scikit-learn, `LogisticRegression` comes with a built-in method of handling imbalanced classes. If we have highly imbalanced classes and have no addressed it during preprocessing, we have the option of using the `class_weight` parameter to weight the classes to make certain we have a balanced mix of each class. Specifically, the `balanced` argument will automatically weigh classes inversely proportional to their frequency:

$$w_j = \frac{n}{kn_j}$$

where w_j is the weight to class j, n is the number of observations, n_j is the number of observations in class j, and k is the total number of classes.

Preliminaries

```
# Load libraries
from sklearn.linear_model import LogisticRegression
from sklearn import datasets
from sklearn.preprocessing import StandardScaler
import numpy as np
```

Load Iris Flower Dataset

```
# Load data
iris = datasets.load_iris()
X = iris.data
y = iris.target
```