

# 逻辑回归(Logistic Regression)

---

作者：刘伟杰 日期：2015-11-30

参考：

[1] 《统计学习方法》 李航 2012年3月第一版

[2] 《机器学习实战》 Peter Harrington

## 1. 原理

---

### 1. 概述：

这是一个分类模型。当输入一个新样本的时候，把样本的特征值输入到 $f(X)$ 函数中，再将 $f(X)$ 的结果输入到sigmoid函数中，输出即为分类结果。训练过程就是通过训练集合寻找 $f(X)$ 的系数，使得分类误差最小。 $f(X)$ 常用线性函数，训练的算法有梯度法、牛顿法等。

### 2. sigmoid函数：

logistic分布的分布函数就是sigmoid函数。

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

### 3. logistic模型：

$$Y = \text{logistic}(X) = \sigma(f(X)) = \frac{1}{1 + e^{-f(X)}}$$

其中： $f(X)$ 可以是任意函数，最简单的就是多项式线性函数 $f(X)=k_1x_1 + k_2x_2 + \dots + k_nx_n$

### 4. 训练算法：

训练过程就是搜索 $f(X)$ 的最优系数，使得风险函数（一般为0-1分类误差函数作为损失函数）最小。训练算法有很多种，梯度算法、牛顿算法等。

## 2. 实现：

---

### 1. 我的实现（训练算法为梯度算法）：

<https://github.com/autoliuweijie/MachineLearning/tree/master/regression>  
(<https://github.com/autoliuweijie/MachineLearning/tree/master/regression>)

### 2. scikit-learn:

示例：

```
#Import Library
from sklearn.linear_model import LogisticRegression
#Assumed you have, X (predictor) and Y (target) for training data set and x_test(p
# Create logistic regression object
model = LogisticRegression()
# Train the model using the training sets and check score
model.fit(X, y)
model.score(X, y)
#Equation coefficient and Intercept
print('Coefficient: n', model.coef_)
print('Intercept: n', model.intercept_)
#Predict Output
predicted= model.predict(x_test)
```