

机器学习实战

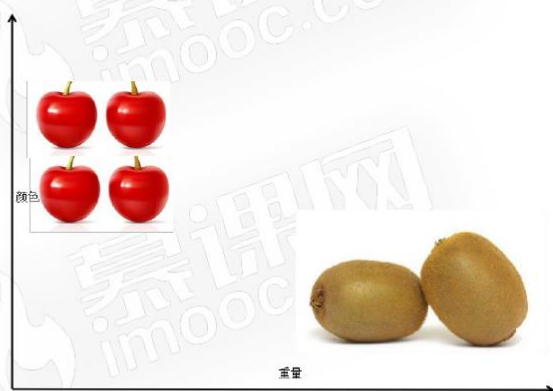
目录

- ◆ 逻辑回归模型原理
- ◆ 逻辑回归模型实战

逻辑回归模型原理

一个简单的机器学习问题

◆ 如何区分樱桃和猕猴桃？

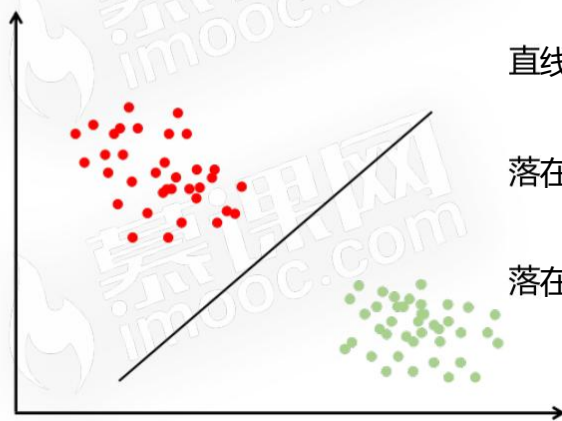


用于区分樱桃和猕猴桃的特征

樱桃的颜色亮，重量小；猕猴桃的颜色暗，重量大；选择两个有区分度的特征-颜色 y ，重量 x ，组成特征向量 (x, y) ，樱桃位于 xy 平面的左上角，猕猴桃位于右下角

如何进行建模求解

- ◆ 把樱桃和猕猴桃抽象成平面上的点，用一条直线可以将这两种类型分开，这条直线称为线性分类器



用直线将两类水果分开

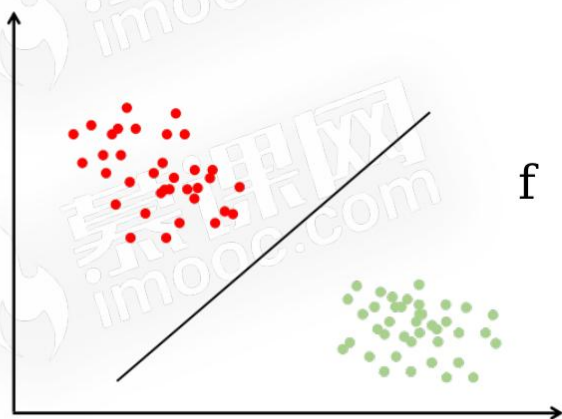
直线方程为 $ax + by + c = 0$

落在直线上方的点被判定成樱桃 $ax + by + c > 0$

落在直线下方的点被判定成猕猴桃 $ax + by + c < 0$

如何进行建模求解

- ◆ 将机器学习模型表示为 f ，输入特征向量为 (x, y) ，樱桃的类别标签为 -1 ，猕猴桃的类别标签为 $+1$ ，直线方程的参数 (a, b, c) 就是机器学习模型的参数



$f(x, y)$

$-1 \quad a \quad b \quad c \quad 0$

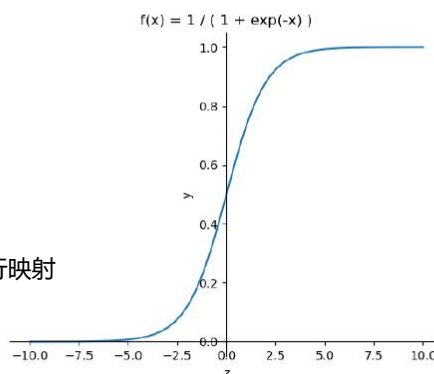
$+1 \quad a \quad b \quad c \quad 0$

逻辑回归 (logistic) 模型

- 又被称为对数回归模型，实际上是一种二分类算法，从一个样本的特征向量 x ，预测出它是正样本的概率值 $p(y=1|x)$

$$h(x) = \frac{1}{1 + \exp(-w^T x - b)}$$

首先用线性函数对特征向量进行映射，然后再用logistic函数进行映射



logistic函数是一种sigmoid函数(S函数)

如何使用逻辑回归模型进行分类

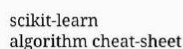
- 对数似然比，样本属于正样本和负样本概率值比的对数

$$\log \frac{p(y=1|x)}{p(y=0|x)} = \log \frac{\frac{1}{1 + \exp(-w^T x)}}{\frac{1}{1 + \exp(w^T x)}} = w^T x$$

如果属于正样本的概率大于负样本的概率则样本被判定为正样本，即 $w^T x > 0$

sklearn机器学习库

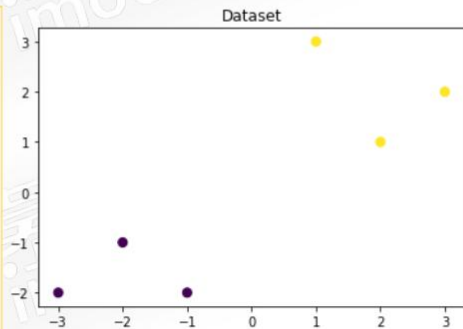
- 安装命令: `pip install scikit-learn`



使用sklearn解决逻辑回归问题

- 构建包含两个类的简单二维数据并进行可视化

```
## 构造数据集
x_features = np.array([[ -1, -2], [-2, -1], [-3, -2], [1, 3], [2, 1], [3, 2]])
y_label = np.array([0, 0, 0, 1, 1, 1])
## 可视化构造的数据样本点
plt.figure()
plt.scatter(x_features[:,0],x_features[:,1], c=y_label, s=50, cmap='viridis')
plt.title('Dataset')
plt.show()
```



使用sklearn解决逻辑回归问题

- 导入库进行训练

```
## 导入矩阵库和绘图库
import numpy as np
import matplotlib.pyplot as plt
## 导入逻辑回归模型函数
from sklearn.linear_model import LogisticRegression
## 调用逻辑回归模型
lr_clf = LogisticRegression()
## 用逻辑回归模型拟合构造的数据集
lr_clf = lr_clf.fit(x_features, y_label) #其拟合方程为  $y=w_0+w_1*x_1+w_2*x_2$ 
## 查看其对应模型的w
print('the weight of Logistic Regression:',lr_clf.coef_)
## 查看其对应模型的w0
print('the intercept(w0) of Logistic Regression:',lr_clf.intercept_)
```

the weight of Logistic Regression:
[[0.73455784 0.69539712]]

the intercept(w0) of Logistic
Regression: [-0.13139986]

$ax + by + c = 0$

weight intercept

数据集

- 鸢尾花分类



变量	描述
sepal length	花萼长度(cm)
sepal width	花萼宽度(cm)
petal length	花瓣长度(cm)
petal width	花瓣宽度(cm)
target	鸢尾的三个亚属类别,'setosa'(0), 'versicolor'(1), 'virginica'(2)

下次预告：神经网络基础