

《机器学习》第一次作业

题目:使用 Python 实现对数几率回归模型

班级:计算机1602

姓名:常朝阳

学号:1611640222

一、题目

使用python实现对数几率回归模型

二、问题描述

使用python实现对数几率回归模型，对鸢尾花数据集（Iris）进行分类

三、数据集描述

该鸢尾花数据集（Iris）包含4个特征变量，1个类别变量。iris每个样本都包含了4个特征：sepal length, sepal width, petal_length, petal_width, 以及1个类别变量（label）。需要建立一个分类器，通过这4个特征来预测鸢尾花种类是属于setosa或者versicolor, 还是virginica。其中有一个类别是线性可分的，其余两个类别线性不可分，这在最后的分类结果绘制图中可观察到。

Attribute Information:

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:
 - Iris Setosa
 - Iris Versicolour
 - Iris Virginica

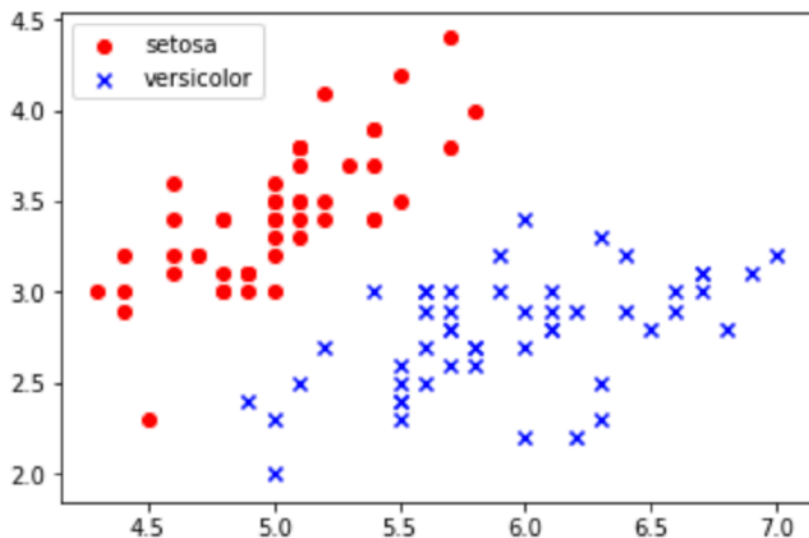
139	6.9	3.1	5.4	2.1	Iris-virginica
140	6.7	3.1	5.6	2.4	Iris-virginica
141	6.9	3.1	5.1	2.3	Iris-virginica
142	5.8	2.7	5.1	1.9	Iris-virginica
143	6.8	3.2	5.9	2.3	Iris-virginica
144	6.7	3.3	5.7	2.5	Iris-virginica
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

150 rows x 5 columns

四、实验过程

1. 数据取前两类(即setosa与versicolor), 前两列(即Sepal length与Sepal width)选为属性, 第五列选为标签

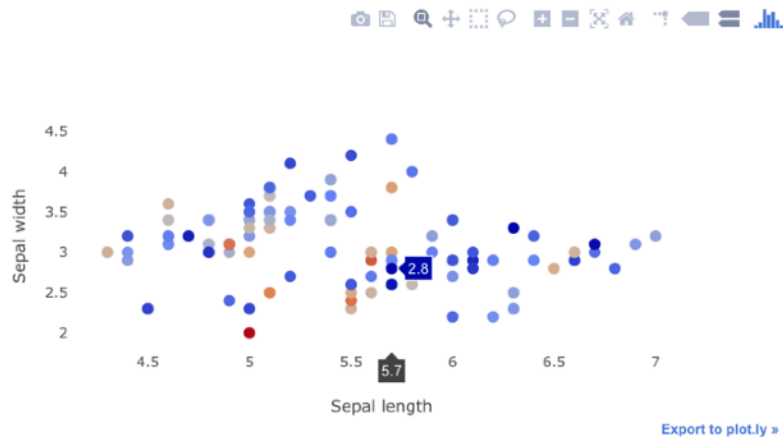
2. 绘制数据的二维可视化散点图



3. 训练集数据点

```
In [10]: #画出训练集数据点
trace = go.Scatter(x = X[:,0], y = X[:,1], mode = 'markers', marker = dict(color = np.random.randn(100), size = 10, show
scale=False))
layout = go.Layout(xaxis=dict(title='Sepal length', showgrid=False),
                    yaxis=dict(title='Sepal width', showgrid=False),
                    width = 700, height = 380)
fig = go.Figure(data=[trace], layout=layout)
```

```
In [11]: iplot(fig)
```



4. 实验结果：

```
In [13]: y_hat = lr.predict(x_test)
accuracy = metrics.accuracy_score(y_test, y_hat)
print("对数几率回归模型正确率: %.2f" % accuracy)
```

对数几率回归模型正确率: 1.00