

# 对数几率回归模型的实现

## 实验报告

姓名： 孙健  
班级： 计算机 1603  
学号： 1611640302

1. 问题描述:

编程实现对数几率回归模型，并对 Iris 数据集进行分类以验证模型的效能

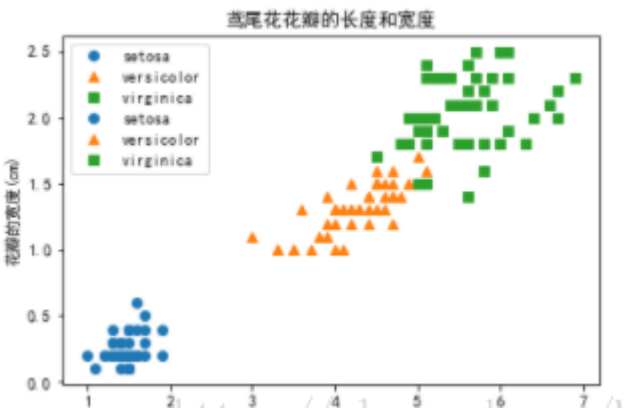
- (1).将数据集的 50%作为训练集,50%作为训练集，检验模型在测试集上的分类正确率.
- (2).将数据集的 70%作为训练集,30%作为训练集，检验模型在测试集上的分类正确率.
- (3).将数据集的 90%作为训练集,10%作为训练集，检验模型在测试集上的分类正确率.

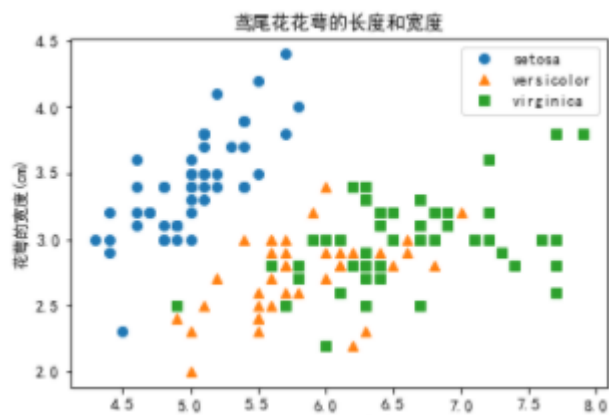
2. 数据集描述

Iris 数据集是常用的分类实验数据集，由 Fisher, 1936 收集整理. Iris 也称鸢尾花卉数据集，是一类多重变量分析的数据集. 数据集包含 150 个数据集，分为 3 类，每类 50 个数据，每个数据包含 4 个属性. 可通过花萼长度，花萼宽度，花瓣长度，花瓣宽度 4 个属性预测鸢尾花卉属于（Setosa-山鸢尾, Versicolour-杂色鸢尾, Virginica-维吉尼亚鸢尾）三个种类中的哪一类.

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

Sepal.Length（花萼长度），单位是 cm； Sepal.Width（花萼宽度），单位是 cm；  
Petal.Length（花瓣长度），单位是 cm； Petal.Width（花瓣宽度），单位是 cm；





### 3. 实验结果图:

实验结果如下表:

测试集比例	训练集准确率	测试机准确率
10%	100%	100%
30%	100%	100%
50%	100%	100%

有上表可知，二分类情况下训练集和测试集的精度都达到了 100%，表明了如果只有两类数据标签，模型可以很好地进行分类.

### 4. 实验结果分析:

该实验训练了一个对数几率回归模型，训练的数据集为 Iris 数据集，分类模型为二分类，验证了 Iris 数据集的效能.