**使用 Python 实现对数**

**几率回归模型**

**课题报告**

姓名：鲁海

班级：计算机1604

学号：1611640422

时间：2018年10月11日

**目录**

1. **问题描述**
2. **数据集描述**
3. **实验结果图**
4. **实验结果分析**
5. **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%作为测试集，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%作为测试集，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%作为测试集，检验模型在测试集上的分类正确率

1. **数据集描述：**

数据集名称：Iris

实例数：150

属性数量：4

属性信息：

1. Sepal.Length     
2. Sepal.Width    
3. Petal.Length   
4. Petal.Width

类：   
1. Setosa   
2. Versicolour   
3. Virginica

摘要统计：

Min Max Mean SD Class Correlation

sepal length: 4.3 7.9 5.84 0.83 0.7826

sepal width: 2.0 4.4 3.05 0.43 -0.4194

petal length: 1.0 6.9 3.76 1.76 0.9490 (high!)

petal width: 0.1 2.5 1.20 0.76 0.9565 (high!)

类别分布：三类各占33.33%。

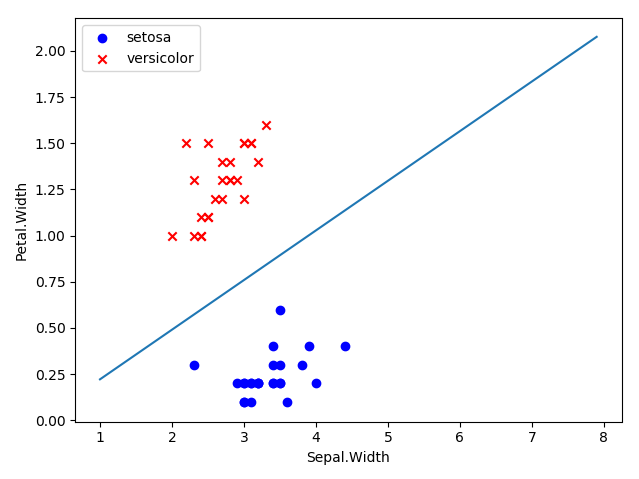
PS：这里只采用了Setosa和Versicolour两个类的100个实例，每个实例取Sepal.Width和 Petal.Width 两个属性。

为了方便编程，在读取数据时自动将Setosa置为1，Versicolour置为0。

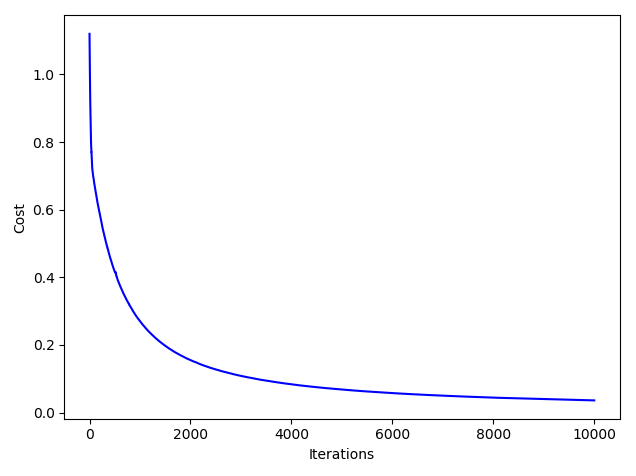
1. **实验结果图：**

1）训练集50%,测试集50%,训练次数10000,学习率0.001,小批量梯度下降法(batchsize=10)。

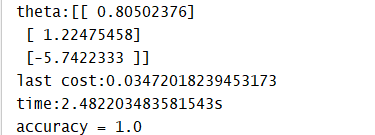
散点图及决策边界：



损失函数变化曲线：

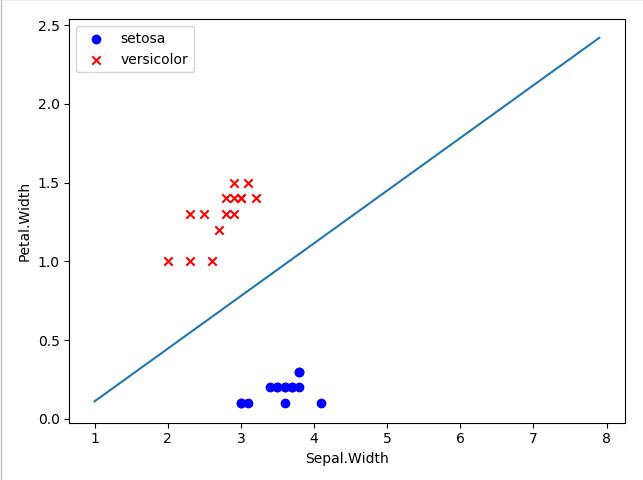


结果数据：

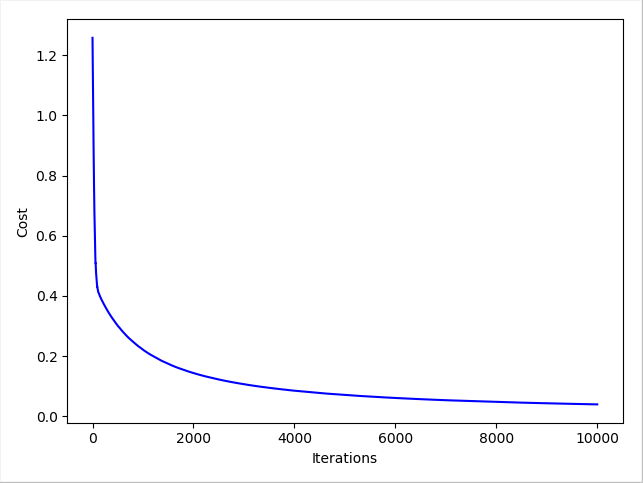


2）训练集70%,测试集30%,训练次数10000,学习率0.001,小批量梯度下降法(batchsize=10)。

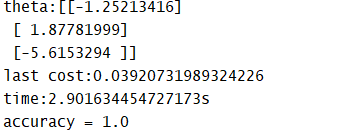
散点图及决策边界：



损失函数变化曲线：

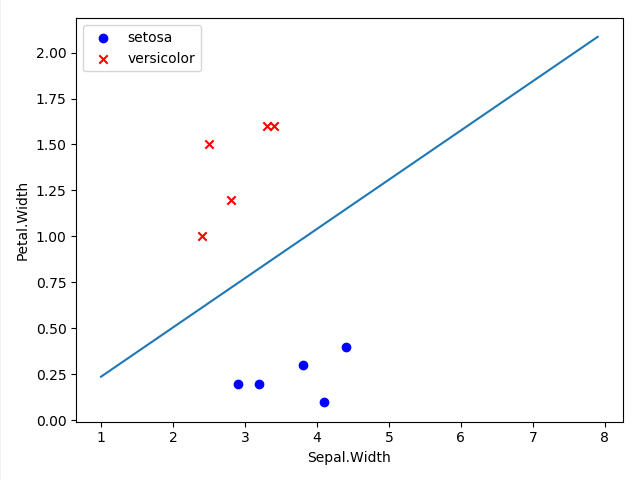


结果数据：

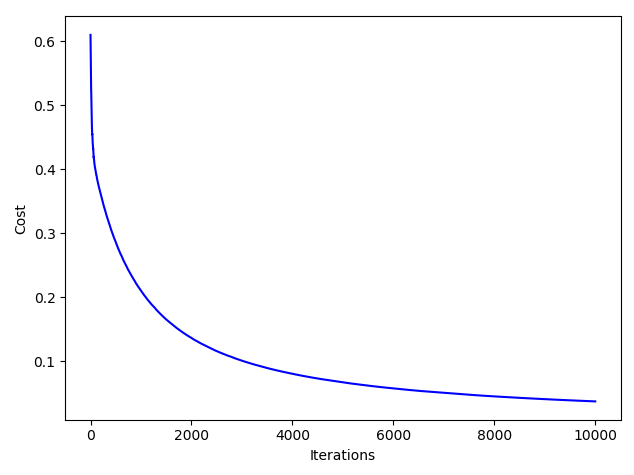


3）训练集90%,测试集10%,训练次数10000,学习率0.001,小批量梯度下降法(batchsize=10)。

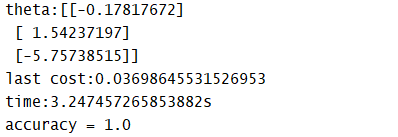
散点图及决策边界：



损失函数变化曲线：

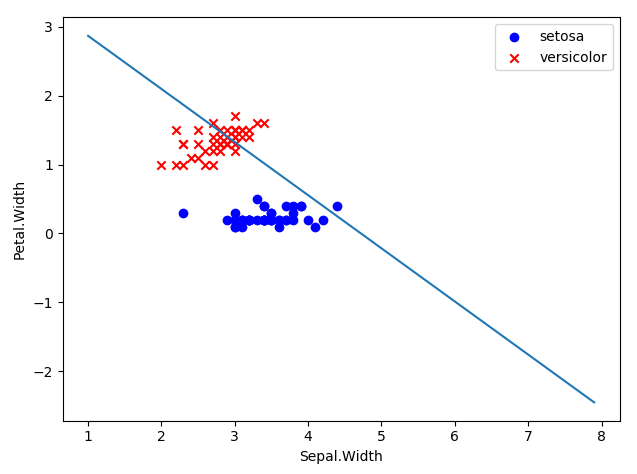


结果数据：

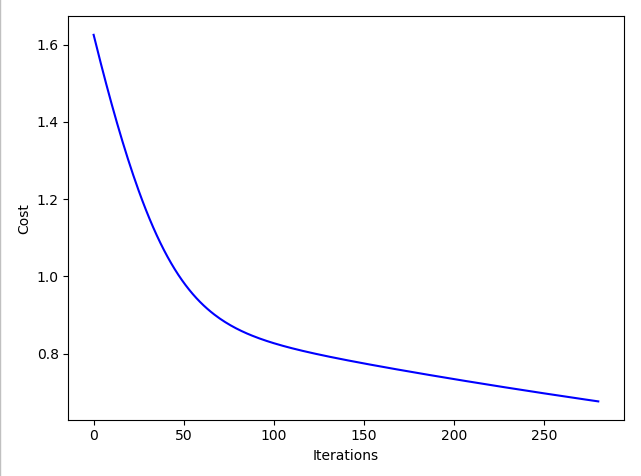


4）训练集10%,测试集90%,训练次数280,学习率0.001,小批量梯度下降法(batchsize=10)(对比用)。

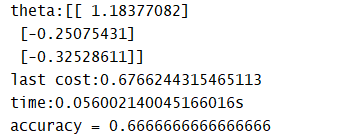
散点图及决策边界：



损失函数变化曲线：



结果数据：



1. **实验结果分析：**

因为数据集的样本数太少，所以很少的训练次数就能将准确率训练到100%。

由图可以看出,随着迭代次数的增加，损失函数不断下降，最后变化越来越小趋于一个定值，所以准确率越来越高，拟合出来的决策边界也成功的将数据集分成了两类。

当其他参数不变，只改变学习率时，学习率越高，损失函数变化的越快，但是如果过高则会导致其最后在一个区间内来回震荡，所以学习率应当设置一个合适的值。