

**2022 – 2023 学年第 1 学期**

**机器学习课程实验报告**

|  |  |
| --- | --- |
| **姓 名：** | 张建才 |
| **学 号：** | 209074458 |
| **专 业：** | 智能202 |

|  |
| --- |
| **安徽工业大学** |
| **计算机科学与技术学院** |

目 录

[实验1 3](#_Toc115034217)

[实验2 4](#_Toc115034218)

[实验3 6](#_Toc115034219)

[实验4 7](#_Toc115034220)

[实验5 8](#_Toc115034221)

# 实验1

**实验内容：**

**实验步骤：**

**实验体会：**

LSM：最小二乘法需要求解最优参数图片包含 形状

描述已自动生成，­­­­­­­­­­­­­­­



最小二乘法将向量表达形式转为矩阵表达形式可以解出最优参数的值：

文本

描述已自动生成

图形用户界面, 文本, 应用程序

描述已自动生成



图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

深刻理解了数据处理时预处理的重要性，数据类型的转换关系。

同时，对岭回归的参数也有了深刻认识。

对数据训练过程也有进一步的认识。

# 实验2

**实验内容：**

**实验步骤：**

**实验体会：**

感知机基本原理：

假定训练数据集是基本可分，通过定义损失函数，并使之最小，就达到了二分类的目的。

假设输入空间是图标

低可信度描述已自动生成输出空间是图示, 图标

描述已自动生成输入 图标

描述已自动生成表示实例的特征向量。输出图标

中度可信度描述已自动生成表示实例的类别。由输入空间到输出空间感知机是如下函数：

文本, 信件

描述已自动生成

文本

描述已自动生成叫做权值, 图片包含 图标

描述已自动生成叫做偏置。画里面的卡通动物

中度可信度描述已自动生成是符号函数

图示

描述已自动生成

参数调整之前：

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

参数调整之后：

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

额外：

图表, 日程表

描述已自动生成

从上述热力图可以看出一些明显的特征，如糖尿病的标签Outcome和葡萄糖测试值Glucose正相关系数比较大，说明葡萄糖测试值高的话，有可能患有糖尿病。同理，年龄Age和怀孕次数Pregnancies之间的相关性也比较强。

# 实验3

**实验内容：**

**实验步骤：**

**实验体会：**

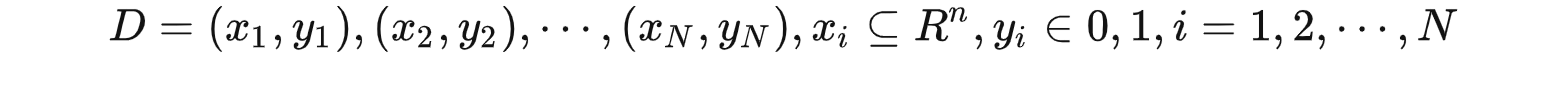
假设数据服从这个分布，然后使用极大似然估计做参数的估计。

假设函数，损失函数，

图示, 示意图

描述已自动生成

分类概率 P(Y=1)，给定数据集D：



使用文本

低可信度描述已自动生成拟合条件概率图片包含 文本

描述已自动生成

当文本

低可信度描述已自动生成的值越接近正无穷，图片包含 文本

描述已自动生成概率值也就越接近 1。

因此逻辑回归的思路是，先拟合决策边界(不局限于线性，还可以是多项式)，再建立这个边界与分类的概率联系，从而得到了二分类情况下的概率。

实验内容：

图形用户界面, 文本, 应用程序

描述已自动生成

使用时不需要创建模型比较方便。

因为循环次数少，却可以达到一个比较高的精度，可能这个数据集之间的关系比较明确。

# 实验4

**实验内容：**

**实验步骤：**

**实验体会：**

支持向量机是一类按监督学习方式对数据进行二元分类的广义线性分类器其决策边界是对学习样本求解的最大边距超平面。

为了把两组数据分开，在空心点的类别找到一个或多个点离实心点最近，在实心点中找到一个或多个点与空心点最近，分类实心和空心点取决于这些边界上的点而与离边界较远的点，即在分割两类别点的时候，只需要考虑支持向量，通过支持向量确定分割直线。假设此直线有宽度，左边贴合一个边界点，右边贴和另一组，实现把点分开并且宽度最大。

文本

描述已自动生成

内容：

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

sklearnSVM的运算使用CPU，一次计算不方便。

尝试使用thundersvm库进行计算。

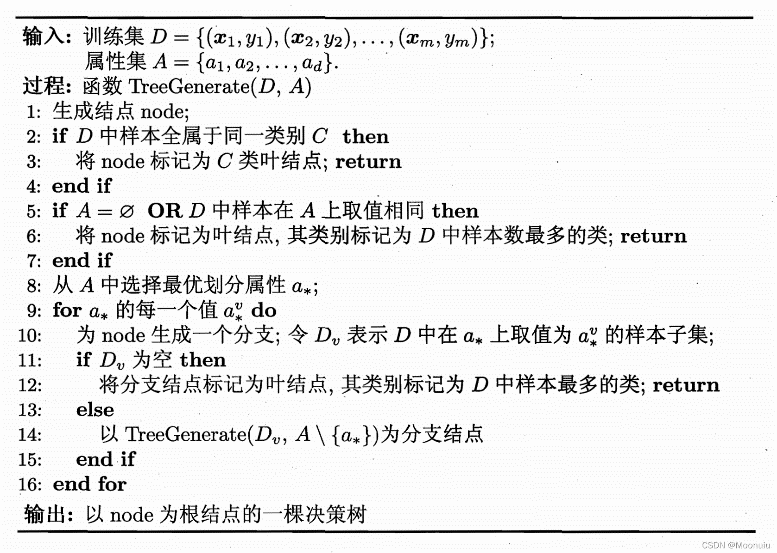
# 实验5

**实验内容：**

**实验步骤：**

**实验体会：**

**决策树思想：**

[](https://blog.csdn.net/weixin_43182102/article/details/122106265)[[1]](#footnote-1)

**信息熵**是度量样本集合纯度最常用的一种指标。假定当前样本集合D中第类样本所占的比例来为 ，则有：

文本

中度可信度描述已自动生成

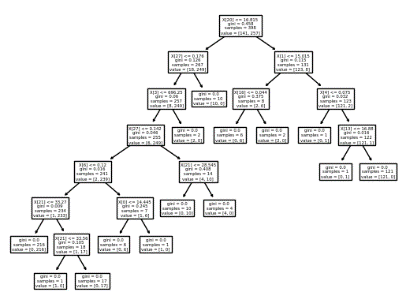
实验内容：

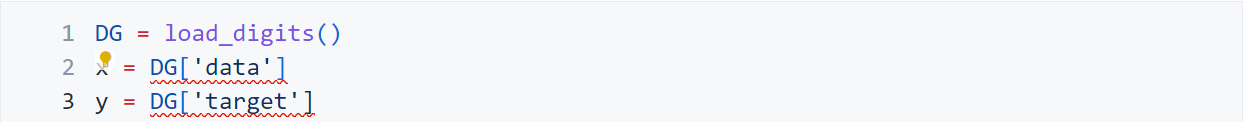
形状

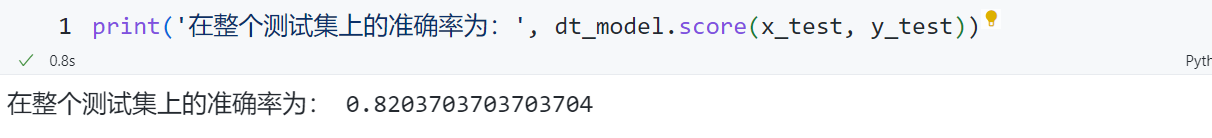
低可信度描述已自动生成

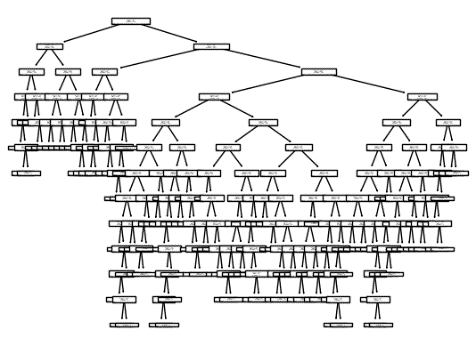
图形用户界面, 文本, 应用程序

描述已自动生成









决策树是基于树结构来进行决策的，一棵决策树包含一个根结点、若干个内部节点和若干个叶结点，最终目的是将样本越分越纯。

1. https://blog.csdn.net/weixin\_43182102/article/details/122106265 [↑](#footnote-ref-1)