

python第四次实验相关资料

相关介绍

1. Scikit-Learn

Scikit-Learn（也称为sklearn）是针对Python编程语言的免费软件机器学习库，它具有各种分类，回归和聚类算法，包括支持向量机，随机森林，梯度提升，k均值和DBSCAN，并且旨在与Python数值科学库NumPy和SciPy联合使用。

1. 打开命令行界面，分别安装scikit-learn工具库

```
pip install scikit-learn
如果因网速问题无法下载上述两个工具库，可以使用以下命令
pip install -i https://pypi.tuna.tsinghua.edu.cn/simple scikit-learn
```

2. 糖尿病数据集介绍

属性名	属性值
样本数据个数	442
特征个数（数据维度）	10
各特征含义	['age','sex','bmi','bp','s1','s2','s3','s4','s5','s6']
特征取值范围	(-0.2,0.2)
标签含义	基于病情进展一年后的定量测量
标签取值范围	[25,346]

['age','sex','bmi','bp','s1','s2','s3','s4','s5','s6']，对应年龄、性别、体质指数、平均血压、S1~S6一年后疾病级数指标。Targets为一年后患疾病的定量指标，值在25到346之间。

3. 本次实验所用到的一些方法

```
from sklearn.datasets import load_diabetes
# 加载糖尿病数据集
data_diabetes = load_diabetes()

# 导入scikit-learn中的线性回归函数
from sklearn.linear_model import LinearRegression
# 加载线性回归函数
regr=LinearRegression()
regr.coef_ # 训练线性回归函数后的模型权值
regr.intercept_ # 训练线性回归函数后的模型偏置
```

```
# 导入scikit-learn中的均方误差损失函数
from sklearn.metrics import mean_squared_error
# 使用均方误差损失函数计算误差
error = mean_squared_error(y, y_hat) # y表示观测到的数据, y_hat表示预测的数据

# 导入scikit-learn中的数据集划分方法
from sklearn.model_selection import train_test_split
```