**信息科学技术学院**

**上机报告**

**课 程 名 称：** 机器学习

**专 业：**

**学 号：**

**学 生 姓 名：**

**日 期：**

# 实验1 房价回归

## 1. 实验内容

该实验是依据房屋的属性信息，包括城镇人均犯罪率、住宅用地所占比例、城镇中非商业用地占比例、环保指标、每栋住宅房间数等13条属性，对房屋的价格进行预测。

1. 数据分析与处理
2. 线性回归算法应用

## 2. 实验步骤

1. 数据分析与处理
2. 导入数据

导入datasets模块下的“波士顿房价” 数据集load\_boston。

1. 数据可视化分析

导入matplotlib模块下的pyplot，画出房价关于各个特征的散点图。

1. 数据划分

导入model\_selection模块下的KFold，定义数据划分方法。

1. 数据归一化

根据数据可视化分析结果，判断是否需要进行归一化操作。

1. 模型训练

导入linear\_model模块下的线性回归模型，建立线性回归模型。

1. 模型评估

导入model\_selection模块下的cross\_val\_score对模型进行评估。

## 3. 实验要求

1. 尝试不同的数据集划分比例（如5折和10折），分析对预测结果的影响。
2. 尝试不同的归一化方法，min-max和z-score（preprocessing模块MinMaxScaler和StandardScaler），分析对预测结果的影响。
3. 尝试不同的线性回归模型求解方法，最小二乘法和梯度下降法（linear\_model模块LinearRegression和SGDRegressor）。
4. 尝试不同的性能度量准则，MSE和MAE（metrics模块mean\_squared\_error, mean\_absolute\_error）。

## 4. 实验代码

## 5. 实验结果

# 实验2 鸢尾花分类

## 1. 实验内容

该实验是依据鸢尾花的属性信息，包括花萼长度、花萼宽度、花瓣长度、花瓣宽度4条属性，对鸢尾花的品种（山鸢尾、变色鸢尾还是维吉尼亚鸢尾）进行预测。

1. 数据分析与处理
2. logistic回归算法应用

## 2. 实验步骤

1. 数据分析与处理
2. 导入数据

导入datasets模块下的数据集 “鸢尾花” load\_iris。

1. 数据可视化分析

导入matplotlib模块下的pyplot，画出样本关于各个特征的散点图。

1. 数据划分

导入model\_selection模块下的train\_test\_split，按比例划分训练集和测试集。

1. 数据归一化

根据数据可视化分析结果，判断是否需要进行归一化操作。

1. 模型训练

导入linear\_model模块下的LogisticRegression，利用训练集对模型参数进行估计。

* 训练fit(X, y)

1. 模型评估

利用估计的模型参数对测试集进行预测，导入metrics模块下的性能度量准则，评估模型预测性能。

* 预测predict(X)
* 准确率accuracy\_score

## 3. 实验要求

1. 鸢尾花数据集包含3种类别，依次选取两种类别的组合进行二分类器的训练。
2. 尝试不同的数据集划分比例（如9:1,8:2,7:3,…,1:9），分析对预测结果的影响，画出性能关于不同比例的折线图。
3. 尝试不同的归一化方法，min-max和z-score（preprocessing模块MinMaxScaler和StandardScaler），分析对预测结果的影响。

## 4. 实验代码

## 5. 实验结果