实验一：线性回归模型—波士顿房价预测

任务流程：

1.简单的线性回归模型

①导入sklearn中的线性回归模型

②给出一组简单的数据，进行模型训练

③训练完成后进行预测，查看权重项和偏置项的值

Q1.1：此模型权重项和偏置项的值各为多少？

array([0.5, 0.5])

1.1102230246251565e-16

Q1.2：将训练集数据改为[[3, 3], [4, 4], [5, 5]], [7, 8, 9]后，再进行一次训练，此时权重项和偏置项的值各为多少？预测值为多少？

array([0.5, 0.5])

4.0

7

2.单特征波士顿房价预测

①导入所需库

②提取数据库，并查看详情

③只取每栋住宅的房间数作为模型唯一的特征

④训练模型

⑤绘制拟合曲线

Q2.1：附上你绘制的拟合曲线，在单个特征下模型的拟合效果是好还是不好？讨论原因。

3.多特征波士顿房价预测

①导入库

②加载波士顿数据

③将数据划分为训练集和测试集

④建立线性回归模型

⑤在测试集上进行预测，并进行性能度量

Q3.1：预测前20个结果是什么？

预测前20个结果为：

[21.12953164 19.67578799 22.01735047 24.62046819 14.45164813 23.32325459

16.6468677 14.9175848 33.58466804 17.48328609 25.50385719 36.60215179

25.95309333 28.48503161 19.34928078 20.16966217 25.9788081 18.25959831

16.52754056 17.08448854]

Q3.2：结合模型的平均绝对误差、均方误差、中值绝对误差、可解释方差值、R方值等性能度量参数，讨论多特征下模型的性能。

Boston数据线性回归模型的平均绝对误差为： 3.377642697362791

Boston数据线性回归模型的均方误差为： 31.15059667690467

Boston数据线性回归模型的中值绝对误差为： 1.777421315736218

Boston数据线性回归模型的可解释方差值为： 0.7105949626282937

Boston数据线性回归模型的R方值为： 0.7068954225782444

附：波士顿房价数据集说明

<https://blog.csdn.net/qq_41185868/article/details/87691801>

Boston波士顿房价数据集的简介

该数据集包含美国人口普查局收集的美国马萨诸塞州波士顿住房价格的有关信息, 数据集很小，只有506个案例。

数据集都有以下14个属性:

CRIM--城镇人均犯罪率 ------【城镇人均犯罪率】

ZN - 占地面积超过25,000平方英尺的住宅用地比例。 ------【住宅用地所占比例】

INDUS - 每个城镇非零售业务的比例。 ------【城镇中非商业用地占比例】

CHAS - Charles River虚拟变量（如果是河道，则为1;否则为0 ------【查尔斯河虚拟变量，用于回归分析】

NOX - 一氧化氮浓度（每千万份） ------【环保指标】

RM - 每间住宅的平均房间数 ------【每栋住宅房间数】

AGE - 1940年以前建造的自住单位比例 ------【1940年以前建造的自住单位比例 】

DIS -波士顿的五个就业中心加权距离 ------【与波士顿的五个就业中心加权距离】

RAD - 径向高速公路的可达性指数 ------【距离高速公路的便利指数】

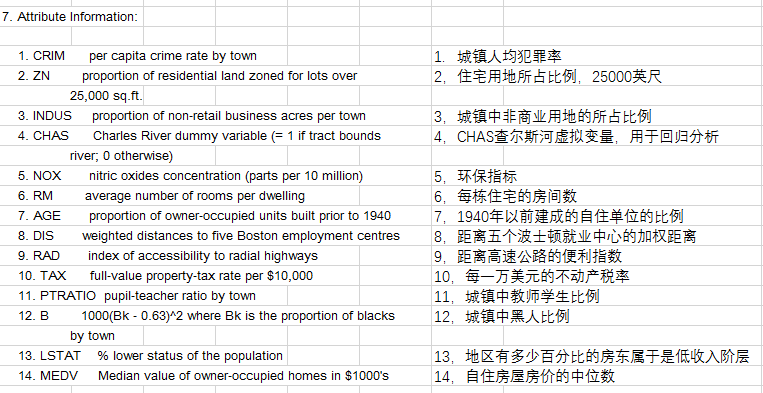
TAX - 每10,000美元的全额物业税率 ------【每一万美元的不动产税率】

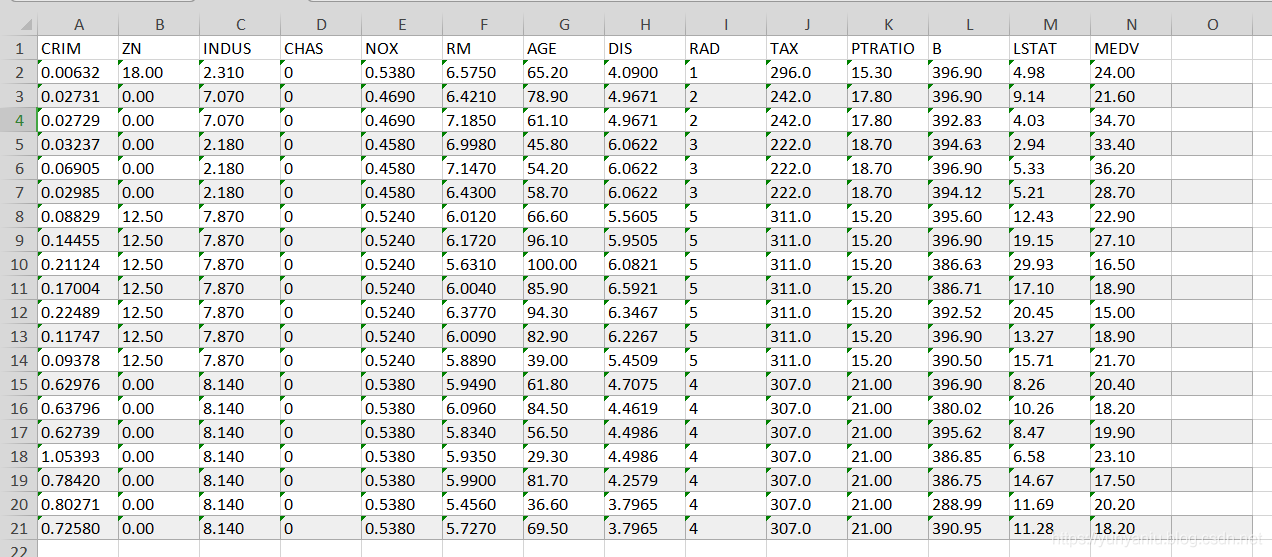
PTRATIO - 城镇的学生与教师比例 ------【城镇中教师学生比例】

B - 1000（Bk - 0.63）^ 2其中Bk是城镇黑人的比例 ------【城镇中黑人比例】

LSTAT - 人口状况下降％ ------【房东属于低等收入阶层比例】

MEDV - 自有住房的中位数报价, 单位1000美元 ------【自住房屋房价中位数】





## Boston波士顿房价数据集的下载

下载地址：<https://archive.ics.uci.edu/ml/machine-learning-databases/housing/housing.data>