姓名和学号：李胜志 2210180232

章节名称：线性回归模型的性能评估

知识目标： 使用最小二乘法和梯度下降法训练模型并求出模型的均方误差、均方根误差和预测准确率（R²分数）对模型进行评估。

能力目标：能够解释模型的参数估计，包括回归系数、截距著性。

素质目标：学习基础知识，提高选择合适方法解决不同问题的能力；养成分析问题、事先做好准本的良好习惯。

知识重点：回归任务中常使用的评估方法，均方误差、均方根误差和确定系数。

知识难点：线性回归的数学原理，包括最小二乘法、梯度下降等优化算法；均方误差、均方根误差、预测准确率（R²分数）等。

举例评估

评估模型：分别计算出使用最小二乘法和梯度下降法训练模型的均方误差、均方根误差和预测准确率

表 2-4二手房房屋销售数据（测试集）

|  |  |  |  |
| --- | --- | --- | --- |
| 面积/（m2） | 售价/（万元） | 面积/（m2） | 售价/（万元） |
| 103 | 301 | 115 | 344 |
| 90 | 275 | 89 | 276 |
| 60 | 206 | 70 | 210 |
| 50 | 160 | 45 | 124 |
| 55 | 190 | 78 | 235 |

1. ，average()函数计算平均值
2. ，sqrt()函数计算平方根
3. 使用score()函数对模型的预测准确率进行评估代码：

yy1 = model1.predict(x) #求LinearRegression()模型预测值

yy2 = model2.predict(x) #求SGDRegressor()模型预测值

#输入测试集

data\_test = {"面积":[103,90,60,50,55,115,89,70,45,78],"售价":[301,275,206,160,190,344,276,210,124,235]}

# 将数据集转换成DataFrame类型

data\_test\_frame = pd.DataFrame(data\_test)

# 划分特征和标签

x\_test = data\_test\_frame["面积"].values.reshape(-1,1)

y\_test = data\_test\_frame["售价"]

#计算最小二乘法训练模型的均方误差，均方根误差，准确率

mse1 = np.average((yy1 - y)\*\*2) #均方误差

rmse1 = np.sqrt(mse1) #均方根误差

score1 = model1.score(x\_test,y\_test) #预测准确率

#计算梯度下降法训练模型的均方误差，均方根误差，准确率

mse2 = np.average((yy2 - y)\*\*2) #均方误差

rmse2 = np.sqrt(mse2) #均方根误差

score2 = model2.score(x\_test,y\_test) #预测准确率

# 输出三个值

print("最小二乘法训练模型均方误差为：", mse1)

print("最小二乘法训练模型均方根误差为：", rmse1)

print("最小二乘法训练模型预测准确率为：", score1)

print("梯度下降法训练模型均方误差为：", mse2)

print("梯度下降法训练模型均方根误差为：", rmse2)

print("梯度下降法训练模型预测准确率为：", score2)

运行结果：

|  |  |  |  |
| --- | --- | --- | --- |
| 优化算法 | 均方误差 | 均方根误差 | 预测准确率 |
| 最小二乘法 | 1697.4442789901582 | 41.200051929459484 | 0.8148333769437266 |
| 梯度下降法 | 2368.636941163581 | 48.66864433250202 | 0.9408238259540125 |

预测准确率越接近1，说明模型对数据的拟合度越好。

利用糖尿病数据集训练模型并对模型进行评估：

1. 导入糖尿病数据集、线性回归模型及划分样本的方法
2. 将数据集划分为训练集和测试集
3. 分别使用最小二乘法和梯度下降法创建模型并训练
4. 分别计算出模型的均方误差、均方根误差和预测准确率并输出

代码：

#导入糖尿病数据集、线性回归模型及划分样本的方法

from sklearn.linear\_model import LinearRegression,SGDRegressor #导入线性回归模型

from sklearn.datasets import load\_diabetes #导入糖尿病数据集

from sklearn.model\_selection import train\_test\_split #导入划分样本的方法

import numpy as np

#将数据集划分为训练集和测试集

x,y = load\_diabetes().data, load\_diabetes().target

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size=0.25, random\_state=8)

model1 = LinearRegression() #建立基于最小二乘法的线性回归模型

model1.fit(x\_train, y\_train) #开始训练模型

model2 = SGDRegressor(loss='huber', max\_iter=5000, random\_state=42) #建立基于梯度下降法的线性回归模型

model2.fit(x\_train,y\_train) #开始训练模型

#评估模型

yy1 = model1.predict(x) #求LinearRegression()模型预测值

yy2 = model2.predict(x) #求SGDRegressor()模型预测值

#计算最小二乘法训练模型的均方误差，均方根误差，准确率

mse1 = np.average((yy1 - y)\*\*2) #均方误差

rmse1 = np.sqrt(mse1) #均方根误差

score1 = model1.score(x\_test,y\_test) #预测准确率

#计算梯度下降法训练模型的均方误差，均方根误差，准确率

mse2 = np.average((yy2 - y)\*\*2) #均方误差

rmse2 = np.sqrt(mse2) #均方根误差

score2 = model2.score(x\_test,y\_test) #预测准确率

# 输出三个值

print("最小二乘法训练模型均方误差为：", mse1)

print("最小二乘法训练模型均方根误差为：", rmse1)

print("最小二乘法训练模型预测准确率为：", score1)

print("梯度下降法训练模型均方误差为：", mse2)

print("梯度下降法训练模型均方根误差为：", rmse2)

print("梯度下降法训练模型预测准确率为：", score2)

运行结果：

|  |  |  |  |
| --- | --- | --- | --- |
| 优化算法 | 均方误差 | 均方根误差 | 预测准确率 |
| 最小二乘法 | 2886.629861264137 | 53.727366036910254 | 0.45934221748744386 |
| 梯度下降法 | 18469.661109853438 | 135.90313134675534 | -2.0990238775628765 |