姓名和学号：李胜志 2210180232

章节名称：举例用套索回归求解模型参数W和b

知识目标：掌握套索回归的基本原理与参数调节方法。

能力目标：能够理解并应用L1正则化技术来防止过拟合。

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：L1正则化的原理及其在回归模型中的应用。

知识难点：正则化技术：L1正则化。

使用糖尿病数据集，用套索回归训练模型，并对其进行评估：

1. 导入套索回归模型、糖尿病数据集和划分数据集方法
2. 创建数据集，划分特征和标签，划分训练集和测试集
3. 创建模型并训练
4. 对模型进行评估

代码：

# 导入套索回归模型、糖尿病数据集和划分数据集方法

from sklearn.linear\_model import Lasso

from sklearn import datasets

from sklearn.model\_selection import train\_test\_split

import numpy as np

# 创建数据集

mydata = datasets.load\_diabetes()

# 划分特征和标签

x = mydata.data

y = mydata.target

# 划分训练集和测试集

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

# 创建模型并训练

ls\_model = Lasso()

ls\_model.fit(x\_train, y\_train)

# 计算模型的斜率w 和 截距b

print("斜率w =", ls\_model.coef\_, "截距b =", ls\_model.intercept\_)

# 计算模型的准确率

score = ls\_model.score(x\_test, y\_test)

print("准确率 = ", score)

# 计算模型中w值不等于0的个数,即特征变量的使用个数

a1 = np.sum(ls\_model.coef\_ != 0)

print("套索回归使用的特征个数为:", a1)

# 创建模型并训练,指定alpha的值为0.1

ls\_model2 = Lasso(alpha=0.1, max\_iter=100000)

ls\_model2.fit(x\_train, y\_train)

# 计算模型的斜率w 和 截距b

print("斜率w =", ls\_model2.coef\_, "截距b =", ls\_model2.intercept\_)

# 计算模型的准确率

score2 = ls\_model2.score(x\_test, y\_test)

print("准确率 = ", score3)

a2 = np.sum(ls\_model2.coef\_ != 0)

print("套索回归使用的特征个数为:", a2)

运行结果：

没有设置alpha的值时，默认为1

斜率w = [ 0. 0. 344.69763764 0. 0.

-0. 0. 339.5257565 0. ]

截距b = 149.68848985932246

准确率 = 0.38584702936904414

套索回归使用的特征个数为: 2

设置alpha的值为0.1时：

斜率w = [ -2.29966573 -129.97289116 489.36105746 244.99355751 -38.8887861

-0. -224.03390797 0. 477.63949949 74.98333519] 截距b = 148.91027450952257

准确率 = 0.5750927105592569

套索回归使用的特征个数为: 8

在套索回归中，可以通过改变alpha参数的值来控制减少部分特征，alpha的值默认为1。