作业要求

一、数据导入

利用 sklearn.datasets.load diabetes 函数导入数据集。

二、数据标准化

将各维度的特征数据标准化为均值为 0, 方差为 1 的数据。

三、模型构建

分别利用 sklearn.linear_model 中的 LinearRegression、Ridge、Lasso 构建线性 回归模型、Ridge 回归模型、LASSO 回归模型,其中 Ridge 回归模型和 LASSO 回归模型的正则化系数设定需要使用 sklearn.model_selection 中的 GridSearchCV 进行交叉验证,从而选择出最优的超参数值。

四、模型比较

比较上一步中学习得到的三个模型在 R-squared 值和特征回归系数上的差异。

五、拓展分析

绘制 Ridge 回归模型和 LASSO 回归模型中特征回归系数随着正则化系数增大时的变化情况。