姓名和学号：李胜志 2210180232

章节名称：岭回归

知识目标：掌握岭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L2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：L2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：L2正则化。

岭回归（Ridge Regression）原理

岭回归是一种用于解决多元线性回归问题的正则化方法，特别适用于处理数据的特征维度很高，且特征之间存在着多重共线性（即特征之间相互关联）的情况。岭回归通过保持模型所有的特征变量而减小特征变量的系数值，来减小特征变量对预测结果的影响。

岭回归的目标是最小化一个包含平方误差项和正则化项的损失函数。对于线性回归模型，标准的损失函数是均方误差（MSE）：

岭回归回归在损失函数中加入L2正则化减小特征变量的系数来减小特征变量对预测结果的影响，从而实现特征选择。即：

其中，λ 是一个非负的正则化参数，用于控制正则化的强度。λ 的值越大，正则化作用越强，模型的复杂度越低，但同时可能会导致模型欠拟合。

在岭回归中，可以通过改变alpha参数的值来控制减小特征变量系数的程度，alpha的值默认为1。

岭回归的特点：

岭回归是一种改良的最小二乘估计法，通过放弃最小二乘法的无偏性，以损失部分信息、降低精度为代价获得回归系数，它是更为符合实际、更可靠的回归方法，对存在离群点的数据的拟合要强于最小二乘法。

岭回归的优缺点：

优点：

1. 处理多重共线性：岭回归能够有效缓解因多重共线性导致的标准最小二乘估计不稳定的问题，使得模型在面对高度相关的特征时依然保持较好的预测能力。
2. 防止过拟合：通过L2正则化，岭回归能够在一定程度上避免模型过于复杂，减少过拟合现象的发生。
3. 解的存在唯一性：即使数据矩阵不可逆，岭回归也能保证解的存在性和唯一性。

缺点：

1. 参数选择：正则化参数λ的选择对于模型性能至关重要，但往往需要通过经验或特定的交叉验证过程才能找到合适的取值。
2. 模型解释性：相比简单线性回归，正则化后的模型可能会牺牲部分变量的解释性，特别是当某些系数因正则化而趋近于零时。
3. 对于离群点敏感：尽管岭回归在一定程度上增强了模型稳健性，但它仍属于线性模型范畴，对异常值较为敏感。