**2025年春季学期《机器学习》实验报告**

**（第一次实验）**

班级： 学号： 姓名：

一、实验过程中，是否对输入数据进行了归一化或标准化处理？试说明这两种方法的区别，并分析为什么线性回归模型可能对特征的尺度敏感。

本次实验对输入数据进行了归一化处理。归一化是将数据缩放到[0,1]范围内，而标准化是将数据变换为均值为 0，标准差为 1 的分布。

在线性回归模型中，如果使用梯度下降方法来估计模型参数，特征值会直接参与计算，尺度大的特征会导致其对应的参数过快或过慢。

二、对于线性回归目标函数，推导给出参数的解析解形式，并思考对于实验所使用的数据集而言，采用标准方程组法求解参数相较于梯度下降法有何优势或劣势。

方程组法不需要迭代调参，能够快速拟合出精确解，适合于类似本次实验的小规模数据集

三、实验中使用的评估指标（如均方误差MSE、均方根误差RMSE、决定系数）分别反映了模型的哪些性能？如果某次实验的值为负，可能是什么原因导致的？

MSE（均方误差）：反映预测值与真实值的平均平方误差。

RMSE（均方根误差）**：**MSE的平方根，与目标变量单位一致。

决定系数：模型解释的目标变量方差比例。

决定系数为负的原因是模型预测效果差于直接用均值预测。常见于严重过拟合或特征与目标变量无关

四、在实验中，如果原始数据中存在非线性关系（如特征与目标变量呈二次函数关系），直接使用线性回归会导致模型性能不佳，思考通过何种方式能够更好的拟合特征与目标变量之间的关系。

我们可能可以尝试多项式回归模型，或采用特征变换，对特征取对数指数等，也可以尝试决策树等非线性模型。

五、你对本次实验课程内容、课程形式、实践平台使用等方面有哪些意见及改进建议？

通过本次实验可以发现，实验平台无法承受短时间内大量的访问请求。希望课程组能够更换实验平台或与平台协商，优化平台使用体验。