**2025年春季学期《机器学习》实验报告**

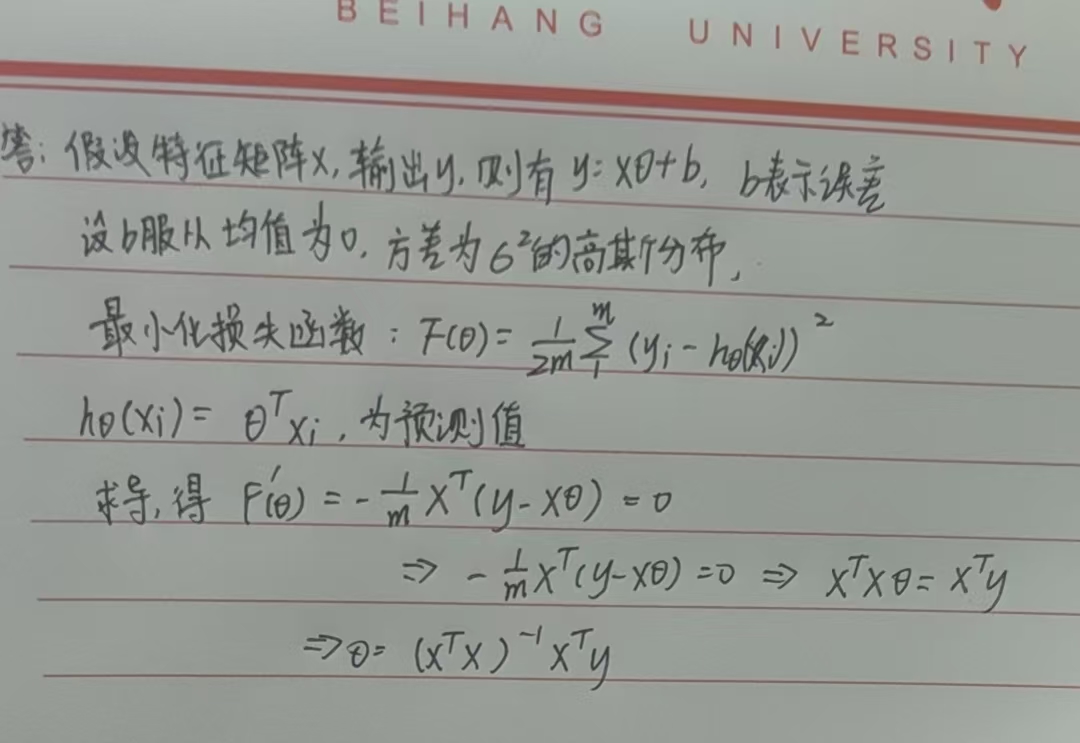
**（第一次实验）**

班级： 230615 学号： 23371540 姓名： 罗天翼

1. 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。

答：在实验中对输入数据进行了标准化处理。归一化是将数据缩放到固定范围，而标准化是将数据转换为均值为 0、标准差为 1 的分布。线性回归模型对特征尺度敏感的原因在于，梯度下降法中特征尺度差异大会导致梯度更新速度不同，增加优化难度；解析解法中尺度差异会导致矩阵条件数变差，求解过程不稳定。

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。



标准方程组法相较于梯度下降法的优势是无需选择学习率且可直接得到全局最优解，但劣势在于计算复杂度较高，样本量较大时矩阵求逆成本高。

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？

答：均方误差（MSE）衡量预测值与真实值之间差异的平方和的平均值，反映模型预测精度；均方根误差（RMSE）是 MSE 的平方根，与数据同量纲，便于理解误差大小；决定系数（R²）衡量模型对数据变异的解释能力，取值在 0 到 1 之间。某次实验若 R² 为负，可能是模型预测效果极差，不如简单使用均值预测。

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。

答：若特征与目标变量存在非线性关系，可采用以下方法提升模型性能：一是多项式回归，通过构造多项式特征拟合非线性关系；二是核方法，使用核函数将数据映射到高维空间；三是使用非线性模型，如决策树、随机森林、支持向量回归等。

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？

答：本次实验课程内容丰富，从python基础内容讲起，再逐步向常用的库扩展，感觉实验的材料和ppt都制作的很用心，可以很快弄懂。课程形式为上课与实验相结合，课程压力小，比较满意。平台使用不太熟悉，但在指导下可以很快上手，总体很不错。建议是感觉上的比较快，可以稍微放慢一点。