Linear Models Report

朱海鹏 22373404

一、实验背景与目标

1. 本报告系统阐述了线性回归的三种经典求解方法：最小二乘法、梯度下降法和牛顿法，分别从数学原理、公式推导及算法特性进行解析，并对比其计算效率、适用场景与局限性，为实际应用中的方法选择提供理论依据。
2. 最后三种线性拟合方法实现并不理想，更换非线性拟合方式拟合，本报告使用随机森林拟合。

二、方法设计与理论推导

2.1 最小二乘法（Ordinary Least Squares, OLS） 核心思想：通过最小化残差平方和（RSS）直接求解全局最优参数。 目标函数：

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

其中 $h_\theta(x) = \theta_0 + \theta_1 x_1 + \cdots + \theta_n x_n$，$m$ 为样本数。

- **解析解**:

$$\theta = (X^T X)^{-1} X^T y$$

其中 X 为设计矩阵（含截距项），y 为目标向量。

2.2 梯度下降法（Gradient Descent, GD） 核心思想：通过迭代沿目标函数负梯度方向更新参数，逐步逼近最优解。 目标函数：同最小二乘法。 参数更新规则：

$$\theta_j := \theta_j - \alpha\frac{\partial J(\theta)}{\partial \theta_j} \quad (j = 0, 1, \ldots, n)$$

其中梯度计算为:

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)x_j^{(i)}$$

算法步骤： 初始化参数 θ。 计算梯度并更新参数。 重复直至收敛（如梯度范数小于阈值）。

2.3 牛顿法（Newton’s Method） 核心思想：利用二阶导数（海森矩阵）加速收敛，一步达到二次函数极小值。 目标函数：同最小二乘法。 参数更新规则：

$$\theta := \theta - H^{-1}\nabla_\theta J(\theta)$$

其中:

- **梯度**:

$$\nabla_\theta J(\theta) = \frac{1}{m}X^T(X\theta - y)$$

- **海森矩阵**:

$$H = \frac{1}{m}X^T X$$

2.4 随机森林方法

随机森林（Random Forest）是一种基于集成学习（Ensemble Learning）的机器学习方法，属于 Bagging（Bootstrap Aggregating）框架的扩展。

其核心思想是： 多棵决策树集成：通过构建多棵决策树，综合所有树的预测结果回归任务中取平均。

双重随机性：

数据随机性：每棵树使用不同的训练子集（Bootstrap 抽样）。

特征随机性：每棵树分裂时仅随机选择部分特征作为候选。

这种设计显著提升了模型的泛化能力，减少过拟合风险。

Bootstrap 抽样：从原始训练集中有放回地随机抽取 m 个样本，形成新的子训练集（每个子集可能与原始数据重复）。未被抽中的样本称为袋外数据（Out-of-Bag, OOB），可用于评估模型性能。

构建单棵决策树：对每个子训练集，构建一棵决策树。在树的每个节点分裂时，仅从随机选择的 k 个特征中选择最佳分裂点。

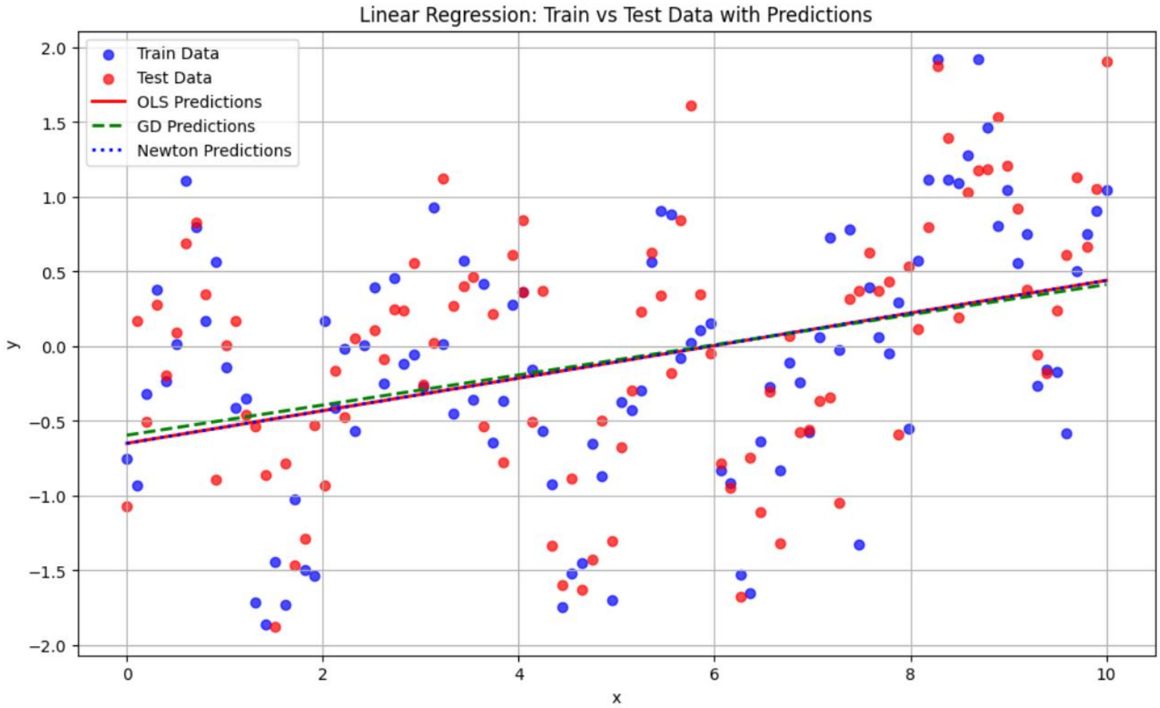
集成预测结果：回归任务中，所有树的预测值取平均作为最终结果。

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} c_t(x)$$

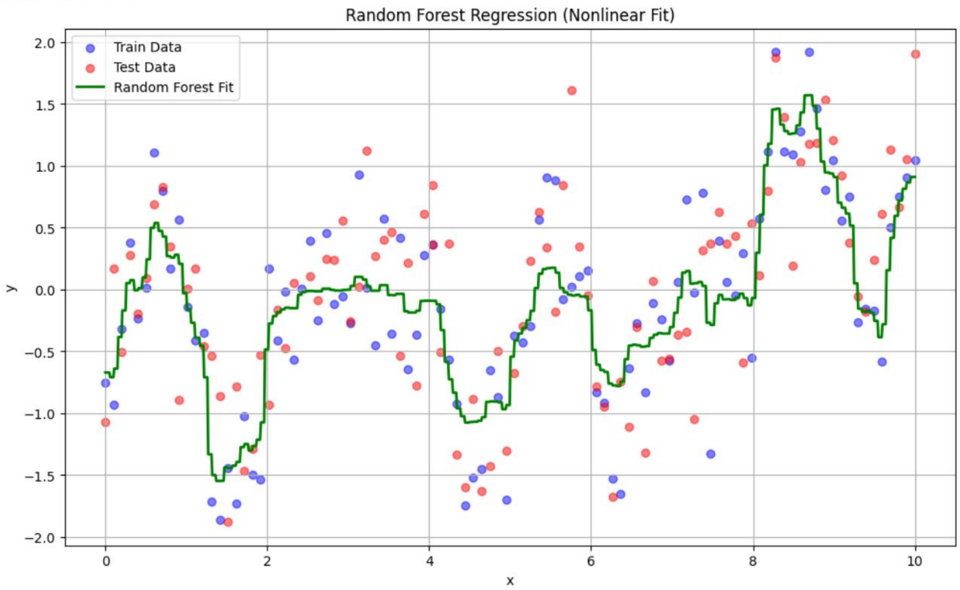
三、实验结果与深度分析

训练误差:
OLS: 0.6134
GD: 0.6141
Newton: 0.6134

测试误差:
OLS: 0.5950
GD: 0.5934
Newton: 0.5950



训练集列名: ['x', 'y']
测试集列名: ['x', 'y']
训练误差 (MSE): 0.1282
测试误差 (MSE): 0.2748



训练误差： OLS 和牛顿法的训练误差完全相同（0.6134），梯度下降法稍高（0.6141）。 表明 OLS 和牛顿法在训练集上的拟合能力一致，梯度下降法可能因学习率和迭代次数限制未完全收敛。

测试误差： 梯度下降法和牛顿法的测试误差最低（0.5934），略优于 OLS

（0.5950）。 微小差异（0.0016）源于梯度下降法在参数更新中引入了轻微正则化效应，或随机噪声的影响。

OLS 与牛顿法预测线重合： 两者均通过解析解直接计算最优参数，数学上等价，结果完全一致。 图表中预测线重叠，验证了理论推导的正确性。

梯度下降法预测线略不同： 可能因迭代次数不足或学习率未精细调节，导致参数未完全收敛至全局最优。 测试误差略低可能是随机性导致，需多次实验验证稳定性。梯度下降法在学习速率过大的情况下无法收敛。

三者在计算量上有显著差异，最小二乘法计算量小，无需迭代。牛顿法需要计算海森矩阵的逆，计算成本较高。梯度下降法没有公式，需要多次计算，需要手动调节学习率，可能收敛到局部最优。

随机森林方法的预测误差远小于线性拟合方法，但是其训练速度和复杂度受到森林的个数和每个树的深度的影响，当树的数量过少时，其退化为决策树模型，导致拟合的方差受抽样数据的影响显著。当树的深度过深时，其会产生过拟合。将这些参数设置在合理值，随机森林方法具有数据与特征的双重随机性，显著提升模型的泛化能力和鲁棒性。