Linear Models Report

潘旻坤 22371303

1. 实验背景与目标

给定一组二维数据，包含训练数据和测试数据。数据分布呈现显著的非线性特征。本实验目标如下：

1) 线性模型分析：使用最小二乘法、梯度下降法（GD）和牛顿法进行线性回归，对比训练与测试误差。
2) 非线性模型改进：通过多项式回归优化模型，验证非线性模型的性能提升。

2. 方法描述

2.1 线性回归模型

1) 最小二乘法（OLS）

OLS 的目标是通过最小化观测值和模型预测值之间的垂直距离（残差）的平方和，找到最适合一组数据点的线（或超平面）来对数据分布进行拟合。找到一条能使得观测值和模型预测值之间的垂直距离和最小的直线，即

$$\min\left\|X\theta-Y\right\|_2^2$$

其结果为：

$$\theta=(X^TX)^{-1}X^TY$$

2) 梯度下降法（GD）

对于一个线性模型来说可定义损失函数

$$J(\theta_0,\theta_1)=\frac{1}{2N}\sum_{i=1}^{N}(\hat{y}^{(i)}-y^{(i)})^2=\frac{1}{2N}\sum_{i=1}^{N}(f(x^{(i)})-y^{(i)})^2=\frac{1}{2N}\sum_{i=1}^{N}(\theta_0+\theta_1x^{(i)}-y^{(i)})^2$$

其中

$$\theta_0=\theta_0-\alpha\frac{\partial J}{\partial\theta_0},\theta_1=\theta_1-\alpha\frac{\partial J}{\partial\theta_1}$$

$$\frac{\partial J}{\partial\theta_0}=\frac{1}{N}\sum_{i=1}^{N}(\hat{y}^{(i)}-y^{(i)})^2,\frac{\partial J}{\partial\theta_1}=\frac{1}{N}\sum_{i=1}^{N}(\hat{y}^{(i)}-y^{(i)})^2x_i$$

3) 牛顿法

利用损失函数的二阶导数（Hessian 矩阵）加速收敛，通过二阶泰勒展开逼近最优解损失函数与最小二乘法相同。参数更新公式：

$$\theta^{(k+1)}=\theta^{(k)}-\alpha H^{-1}\nabla J(\theta^{(k)})$$

2.2 非线性模型（多项式回归）

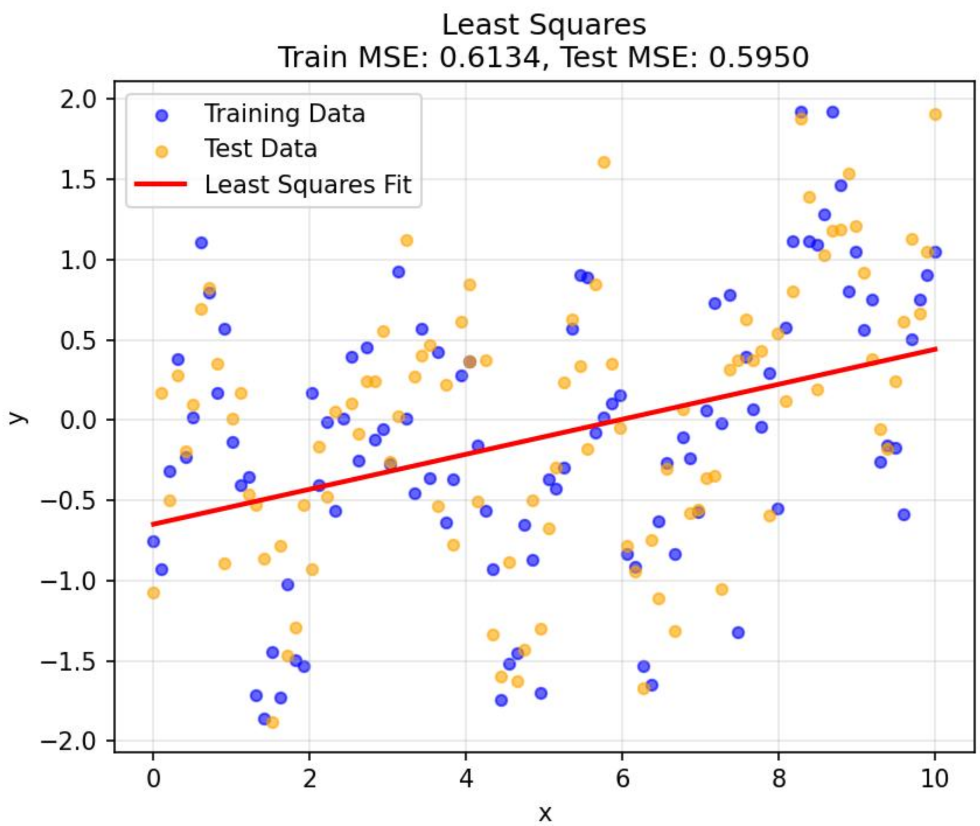
多项式数据拟合是一种常用的拟合方法，用于通过多项式函数对一组数据进行逼近或拟合。可将表达式写为：

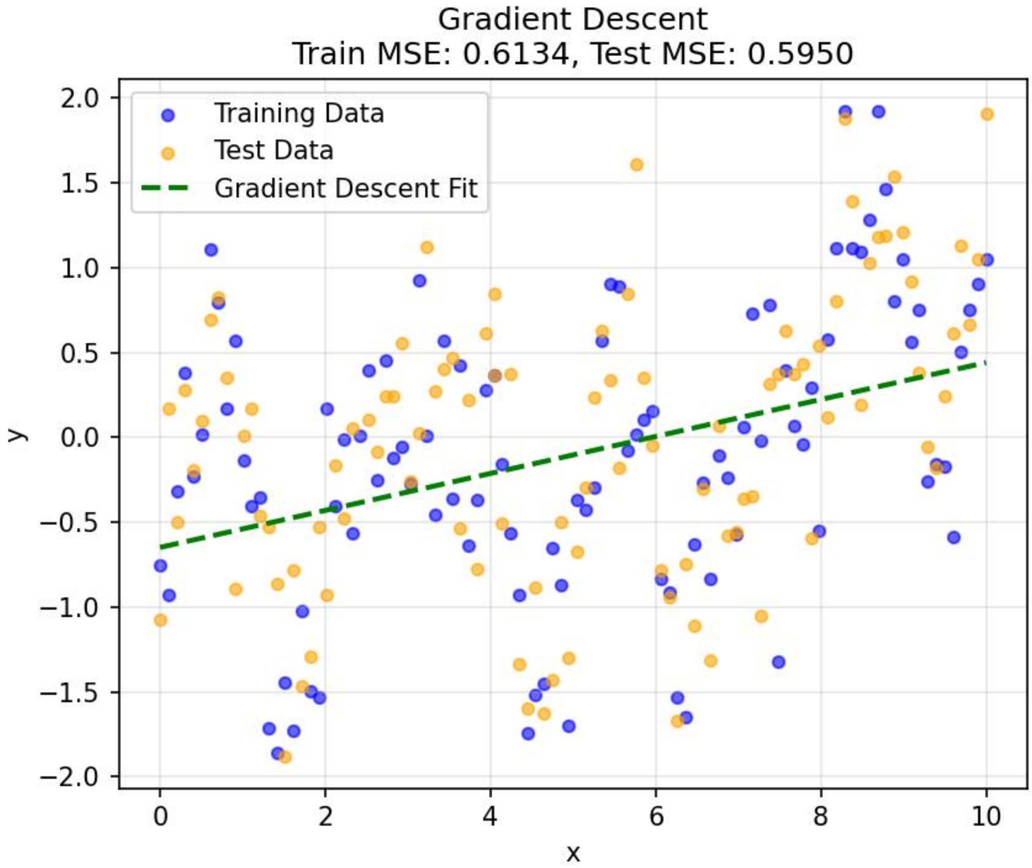
$$f(x)=\sum_{k=0}^{n}a_k x^k$$

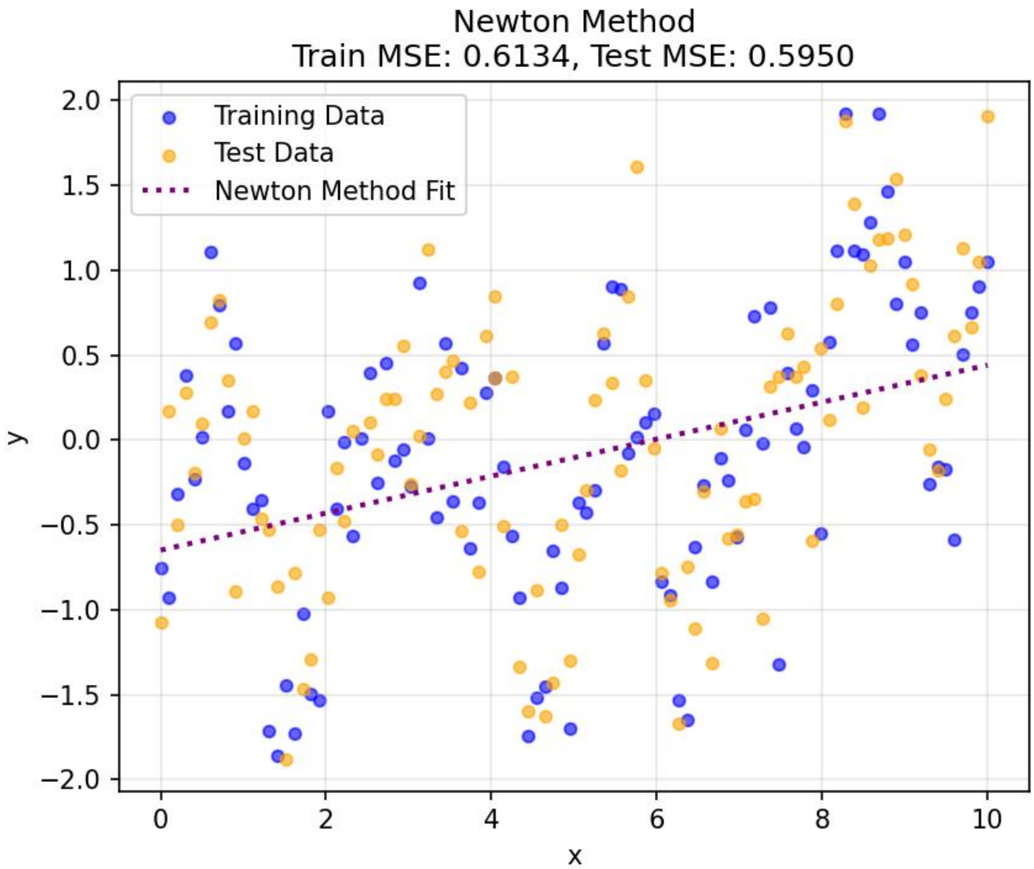
3. 实验结果与分析

3.1 线性模型性能对比

最小二乘法、梯度下降法和牛顿法的训练集和测试集拟合结果如下图所示：




Gradient Descent
Train MSE: 0.6134, Test MSE: 0.5950
Training Data
Test Data
Gradient Descent Fit
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
y
0
2
4
6
8
10
x

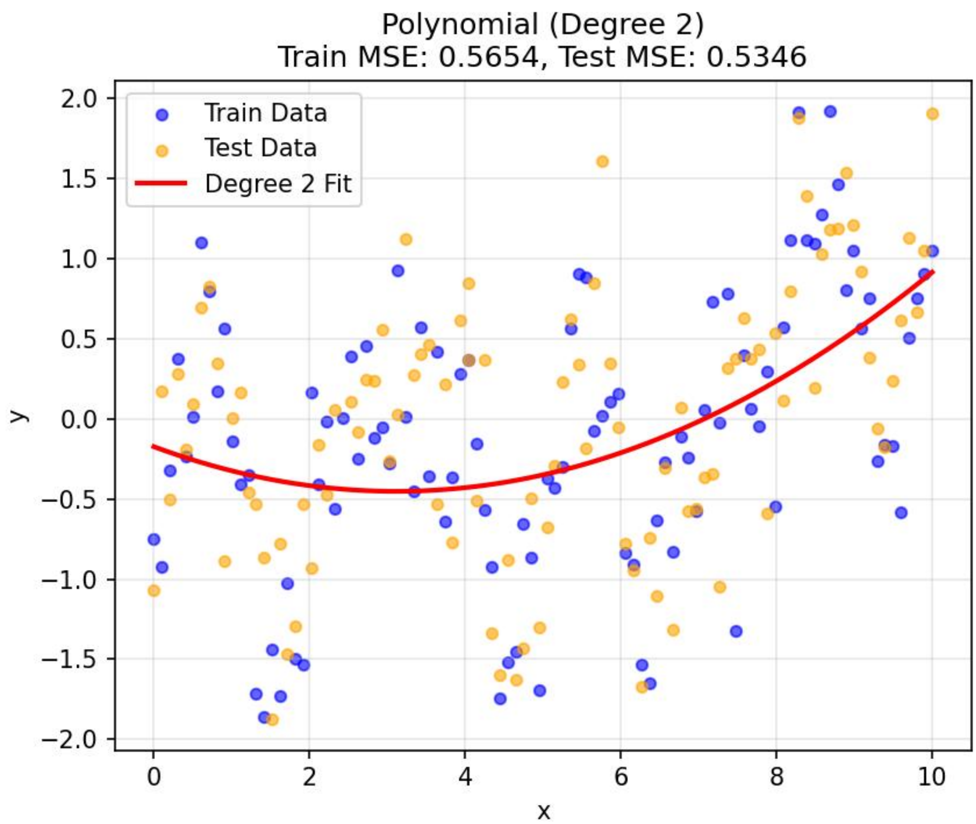

Newton Method
Train MSE: 0.6134, Test MSE: 0.5950
Training Data
Test Data
Newton Method Fit
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
y
0
2
4
6
8
10
x

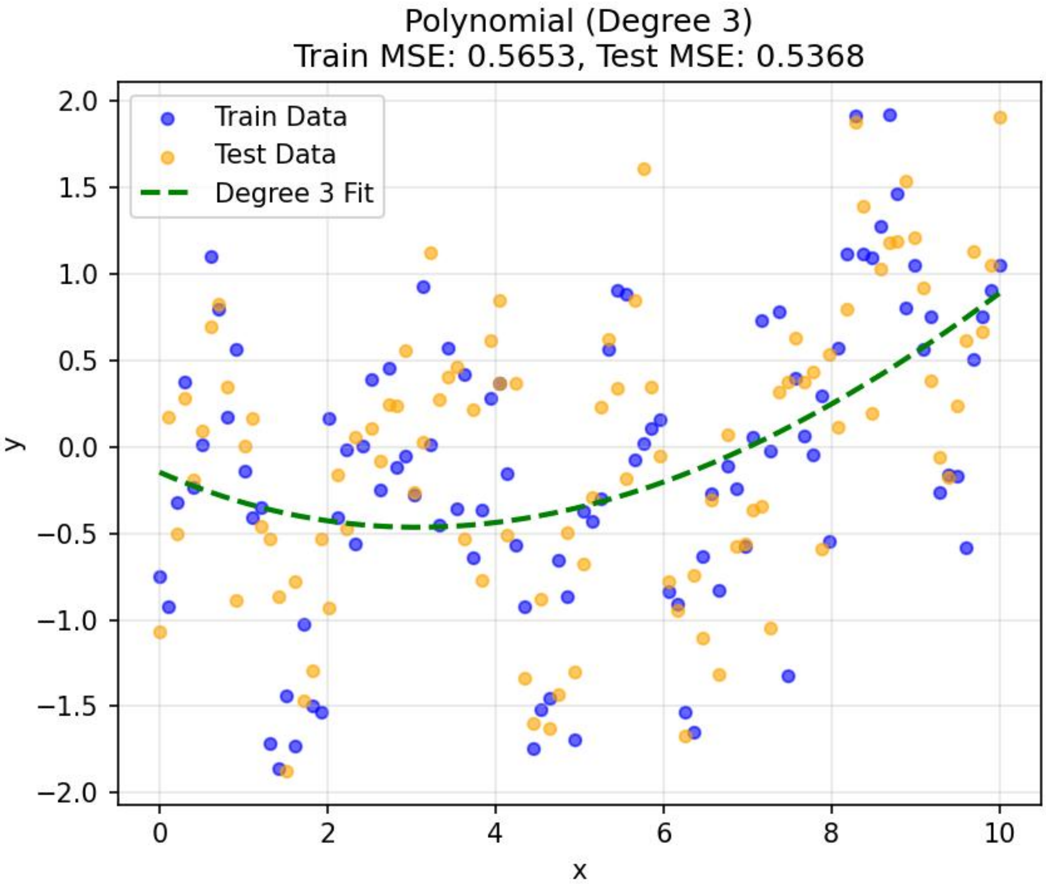
三种方法计算得训练误差和测试误差如下：

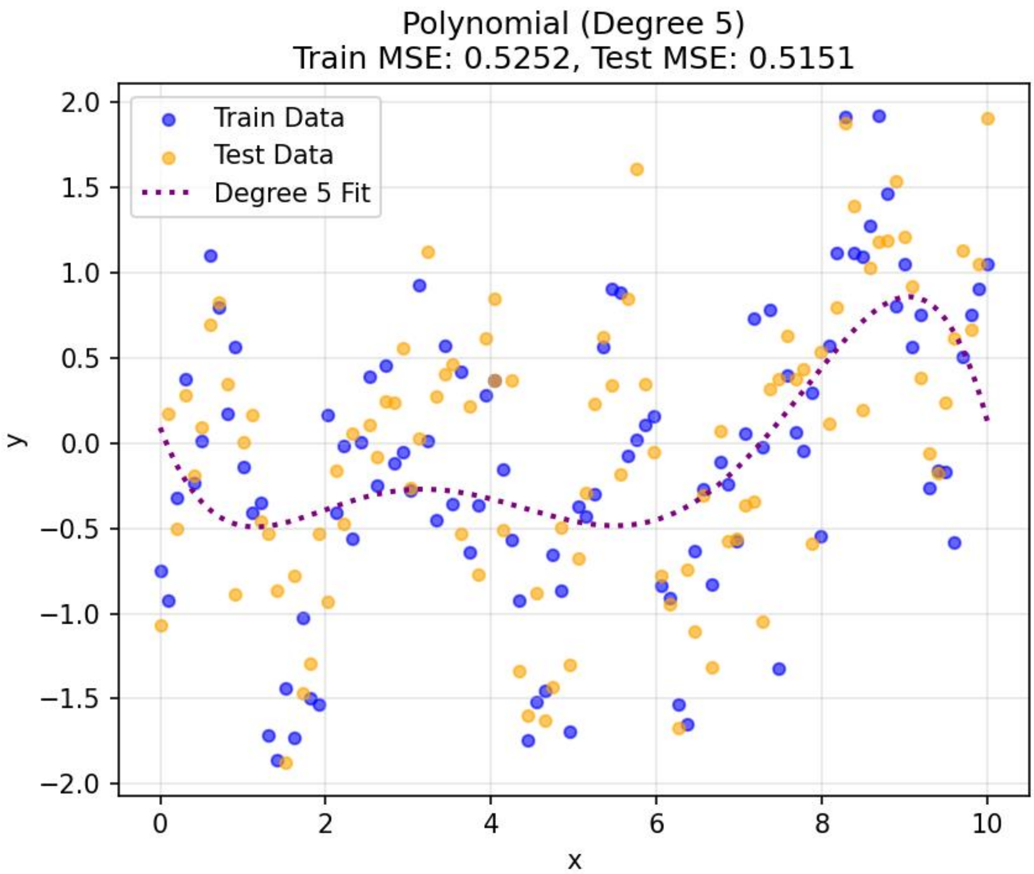
```
最小二乘法:
训练误差(MSE): 0.613402, 测试误差(MSE): 0.595043

梯度下降法:
训练误差(MSE): 0.613402, 测试误差(MSE): 0.595043

牛顿法:
训练误差(MSE): 0.613402, 测试误差(MSE): 0.595043
```

三种方法结果一致，验证了算法的正确性。但误差较高，表明线性模型无法捕捉数据非线性特征。

3.2 非线性模型性能对比




Polynomial (Degree 3)
Train MSE: 0.5653, Test MSE: 0.5368
Train Data
Test Data
Degree 3 Fit
y
x
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
2
4
6
8
10


Polynomial (Degree 5)
Train MSE: 0.5252, Test MSE: 0.5151
Train Data
Test Data
Degree 5 Fit
y
x
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
2
4
6
8
10

二次和三次时基本捕捉数据趋势，但部分区域拟合不足；五次时达到最佳MSE，训练误差最低且测试误差同步下降，未出现明显过拟合现象。

4. 总结

在本次线性模型实验中，通过使用 python 编写三种线性模型的算法，我对于这三种线性模型有了更深的理解和认识。当选择合适的学习率且迭代次数足够大时，梯度下降的结果会无限接近于最小二乘法的结果。但由于要拟合的数据本身具有非线性，所以三种线性模型在上面的表现均不佳。在多项式回归过程中没有出现明显的过拟合现象，但是随着次数增加，过拟合现象可能会发生。