作业一报告

钟经佑
22373168@buaa.edu.cn

Abstract

本研究针对二维回归数据，探讨了不同回归方法在数据拟合中的表现。首先，采用最小二乘法、梯度下降法和牛顿法进行线性回归拟合，并在训练集和测试集上分别评估其均方误差。实验结果表明，线性模型难以有效捕捉数据的非线性特征，导致拟合效果较差，尤其在测试集上的泛化能力较低。针对该问题，进一步选择了非线性模型

$$a\sin(bx + c) + dx + e$$

进行拟合，并利用梯度下降优化均方误差。实验结果显示，该非线性模型相较于线性模型在训练误差和测试误差上均有显著降低，并在测试集上保持较低的误差，证明了其在该问题上的有效性。

Methodology

1. 最小二乘法

最小二乘法是通过最小化误差平方和来找到最适合的数据拟合方法。假设我们有一组数据点 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$，目标是拟合一个函数 $f(x) = ax + b$（线性拟合）。

目标：
最小化以下损失函数（误差平方和）：

$$L(a, b) = \sum_{i=1}^{n} (y_i - (ax_i + b))^2$$

推导：

计算 $L(a, b)$ 对 a 和 b 的偏导数。
对偏导数分别求解，使其为 0，得到最小值。
对于 a 和 b 的偏导数：

$$\frac{\partial L}{\partial a} = -2\sum_{i=1}^{n} x_i(y_i - (ax_i + b))$$

$$\frac{\partial L}{\partial b} = -2\sum_{i=1}^{n} (y_i - (ax_i + b))$$

令偏导数为 0：

$$\sum_{i=1}^{n} x_i(y_i - (ax_i + b)) = 0$$

$$\sum_{i=1}^{n} (y_i - (ax_i + b)) = 0$$

解以上方程组即可得到 a 和 b。

2. 梯度下降法

梯度下降法是一种通过迭代更新参数以最小化损失函数的方法。我们使用最小二乘法的损失函数来演示梯度下降。

目标：
最小化损失函数：

$$L(a, b) = \sum_{i=1}^{n} (y_i - (ax_i + b))^2$$

推导：

1. 初始化参数 a_0, b_0。
2. 计算损失函数的梯度：

$$\frac{\partial L}{\partial a} = -2\sum_{i=1}^{n} x_i(y_i - (ax_i + b))$$

$$\frac{\partial L}{\partial b} = -2\sum_{i=1}^{n} (y_i - (ax_i + b))$$

3. 使用梯度下降更新参数：

$$a_{t+1} = a_t - \eta\frac{\partial L}{\partial a}$$

$$b_{t+1} = b_t - \eta \frac{\partial L}{\partial b}$$

其中，η 是学习率，t 是当前迭代步数，直到损失函数收敛。

3. 牛顿法

牛顿法是一种通过迭代计算并使用二阶导数来加速优化过程的方法。我们通过损失函数的二阶导数来实现。

目标：
最小化损失函数：

$$L(a, b) = \sum_{i=1}^{n} (y_i - (ax_i + b))^2$$

推导：

1. 计算损失函数的二阶导数。

一阶导数：

$$\frac{\partial L}{\partial a} = -2 \sum_{i=1}^{n} x_i (y_i - (ax_i + b))$$

$$\frac{\partial L}{\partial b} = -2 \sum_{i=1}^{n} (y_i - (ax_i + b))$$

二阶导数：

$$\frac{\partial^2 L}{\partial a^2} = 2 \sum_{i=1}^{n} x_i^2$$

$$\frac{\partial^2 L}{\partial b^2} = 2n$$

$$\frac{\partial^2 L}{\partial a \partial b} = 2 \sum_{i=1}^{n} x_i$$

2. 使用牛顿法更新参数：

$$\begin{pmatrix} a_{t+1} \\ b_{t+1} \end{pmatrix} = \begin{pmatrix} a_t \\ b_t \end{pmatrix} - \mathbf{H}^{-1} \nabla L(a_t, b_t)$$

其中，**H** 为海森矩阵（Hessian matrix），

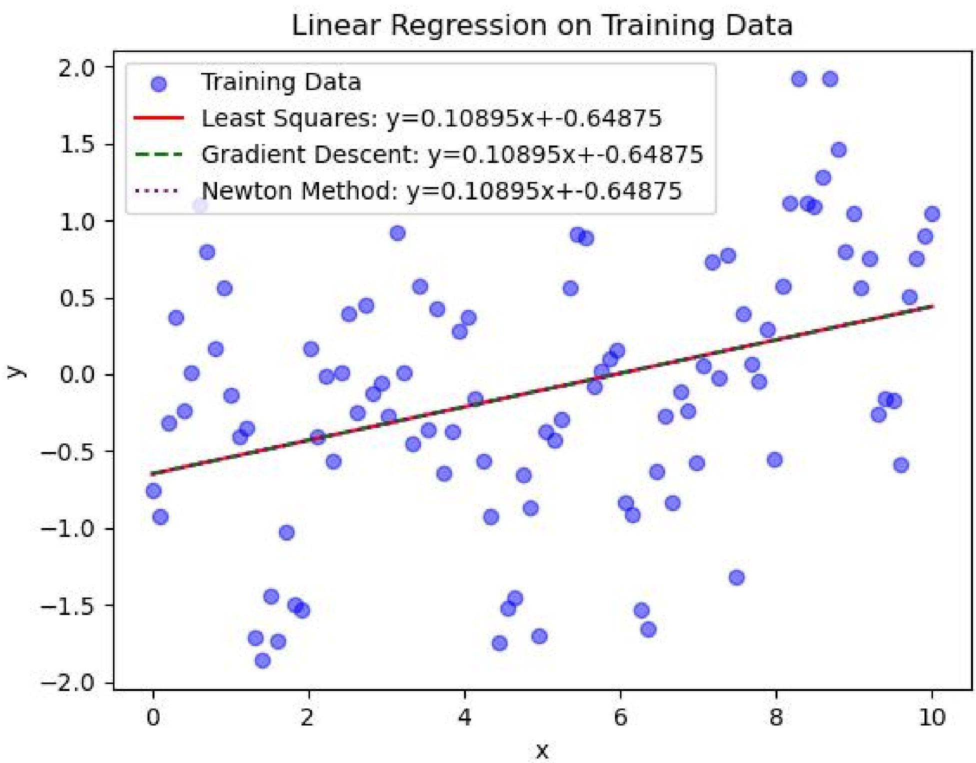
$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 L}{\partial a^2} & \frac{\partial^2 L}{\partial a \partial b} \\ \frac{\partial^2 L}{\partial a \partial b} & \frac{\partial^2 L}{\partial b^2} \end{pmatrix}$$

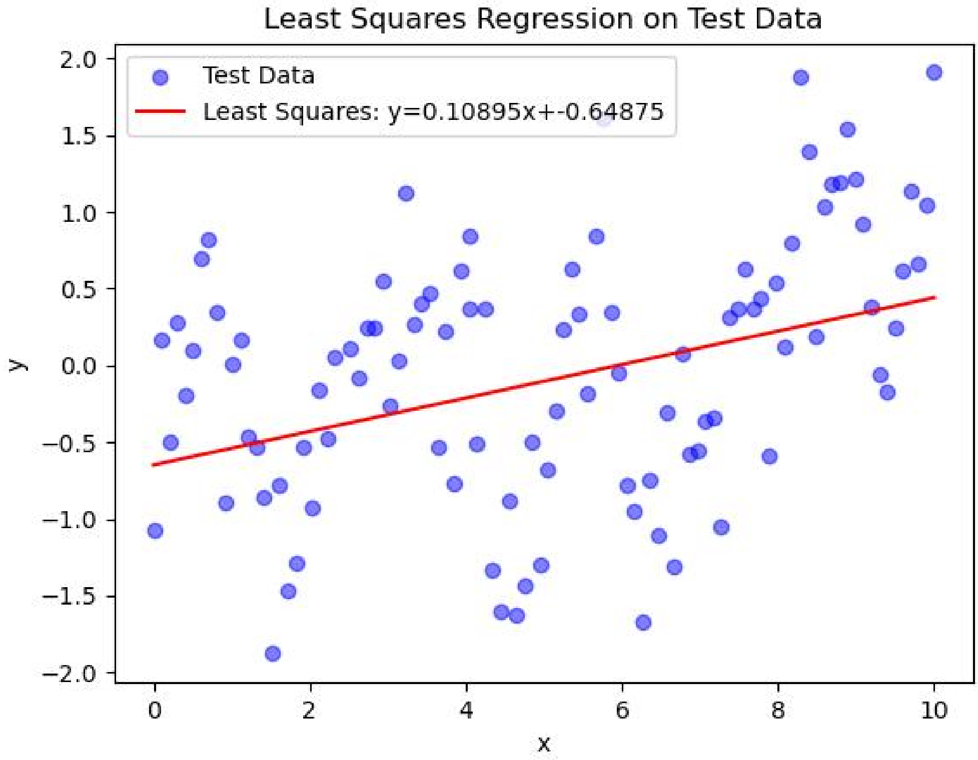
$\nabla L(a_t, b_t)$ 为梯度向量：

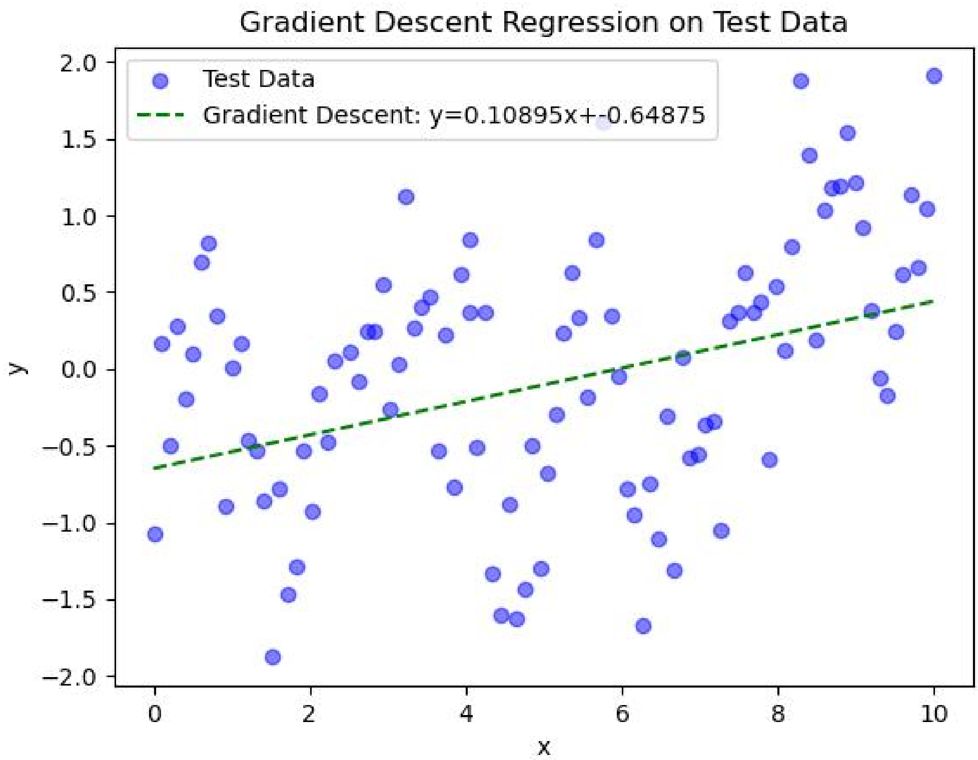
$$\nabla L(a_t, b_t) = \begin{pmatrix} \frac{\partial L}{\partial a} \\ \frac{\partial L}{\partial b} \end{pmatrix}$$

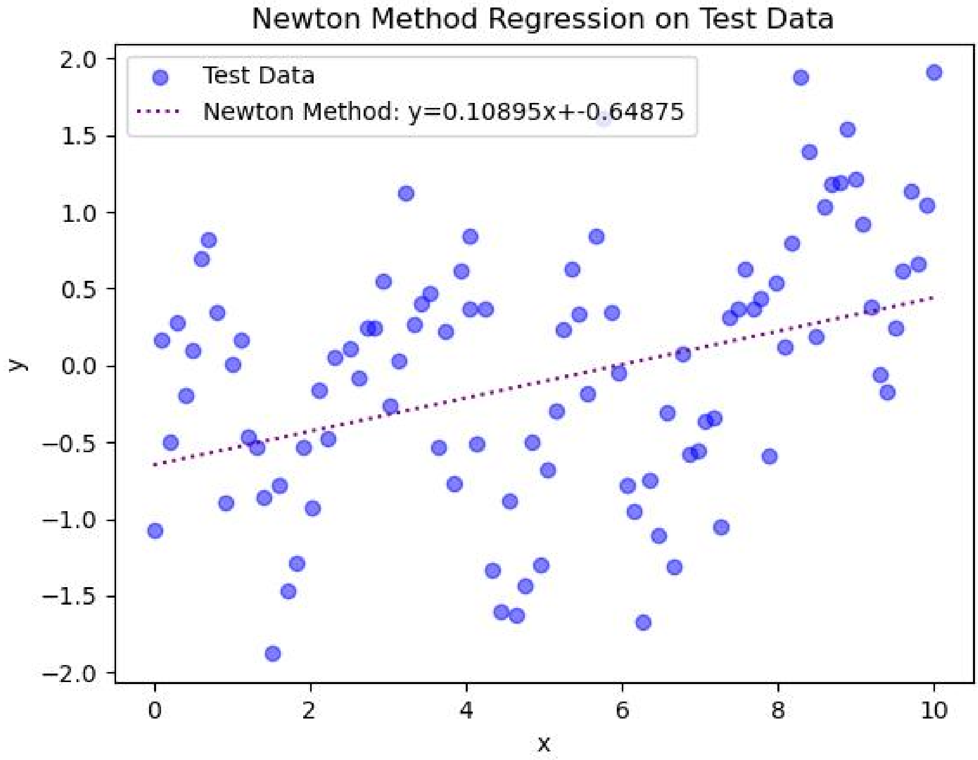
3. 迭代更新，直到收敛。

Experimental Studies




Least Squares Regression on Test Data
Test Data
Least Squares: y=0.10895x+-0.64875
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


Gradient Descent Regression on Test Data
Test Data
Gradient Descent: y=0.10895x+-0.64875
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



Training MSE - Least Squares: 0.61340

Training MSE - Gradient Descent: 0.61340

Training MSE - Newton Method: 0.61340

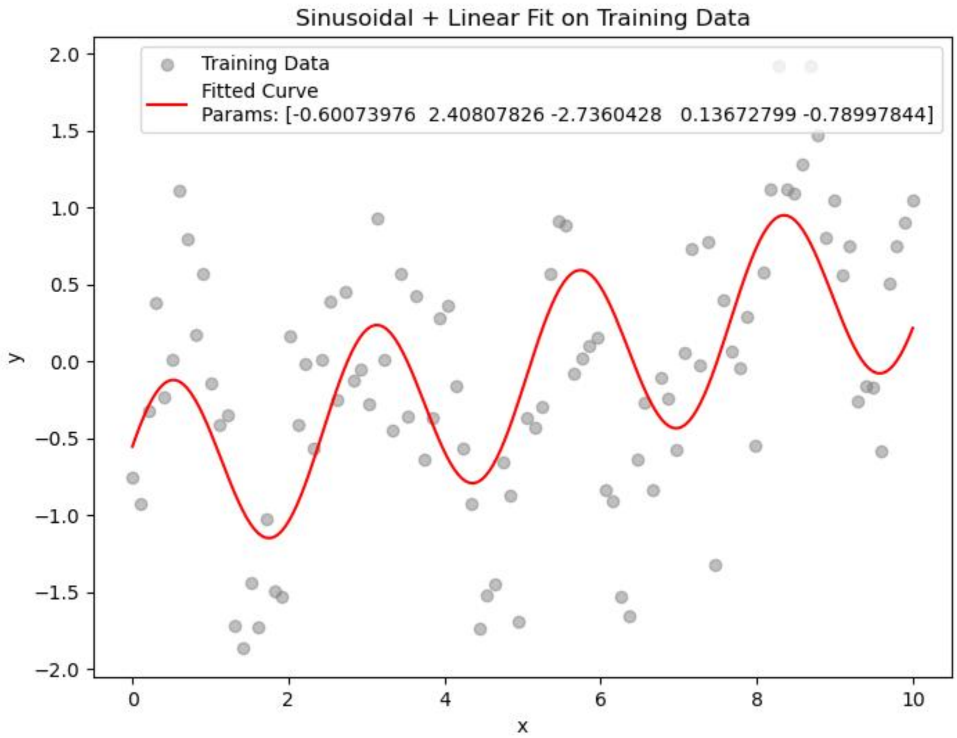
Test MSE - Least Squares: 0.59504

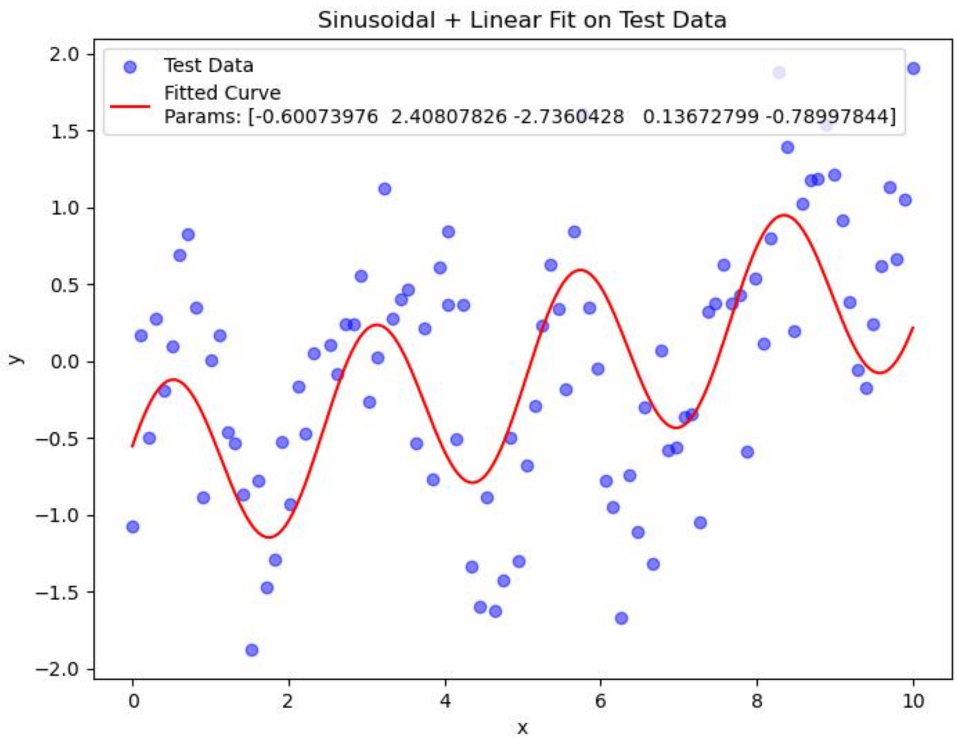
Test MSE - Gradient Descent: 0.59504

Test MSE - Newton Method: 0.59504

使用非线性模型

观察可得，数据分布与$a\sin(bx+c)+dx+e$相似，于是采用此非线性模型，使用梯度下降优化 MSE





Training MSE: 0.43425

Test MSE: 0.44529

Conclusion

三种线性拟合效果相同，说明均收敛到最优解。使用符合数据分布特征的非线性模型比线性模型效果更好