线性拟合与非线性拟合报告

邓子凡 22231042

摘要

本报告对比分析四种核心优化方法：最小二乘法、梯度下降法、牛顿法和多项式拟合，通过实验验证其性能差异。

引言

在数据驱动建模日益重要的背景下，理解优化算法原理对自然语言处理和深度学习应用至关重要。尽管现代工具已实现自动化，数学基础仍是掌握本质的核心。

方法论

方法 1：最小二乘法

最小二乘法是一种通过最小化误差平方和来寻找数据最佳函数匹配的优化技术。其核心目标函数定义为：

$$J(\theta) = \sum_{i=1}^{n}(y_i - f(x_i;\theta))^2$$

其中，$f(x_i;\theta)$ 是模型预测值，θ 为待求参数向量。通过求解 $\nabla J(\theta) = 0$，可得闭式解：

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

式中 X 为设计矩阵，y 为观测值向量

方法 2：梯度下降法

梯度下降法通过沿目标函数负梯度方向迭代更新参数：

$$\theta_{t+1} = \theta_t - \eta \nabla J(\theta_t)$$

其中 η 为学习率，$\nabla J(\theta_t)$ 是损失函数的梯度。对于线性回归，梯度计算为：

$$\nabla J(\theta) = \frac{2}{m} X^T (X\theta - y)$$

该方法的收敛性依赖于学习率的选择，需通过线性搜索或自适应方法（如 Adam）调整

方法 3：牛顿法

牛顿法利用二阶导数信息加速收敛，迭代公式为：

$$\theta_{t+1} = \theta_t - H^{-1}(\theta_t)\nabla J(\theta_t)$$

其中 $H(\theta_t)$ 是目标函数的 Hessian 矩阵。对于二次函数，牛顿法可在单步内收敛。

方法 4：多项式拟合法

多项式拟合将非线性关系转化为高维线性模型：

$$f(x) = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M$$

通过构造多项式特征矩阵 $X_{\text{poly}} = [x, x^2, \ldots, x^M]$，转化为线性回归问题：

$$\min_w \| y - X_{\text{poly}} w \|^2$$

其解仍可通过最小二乘法或梯度下降求解。

实验分析

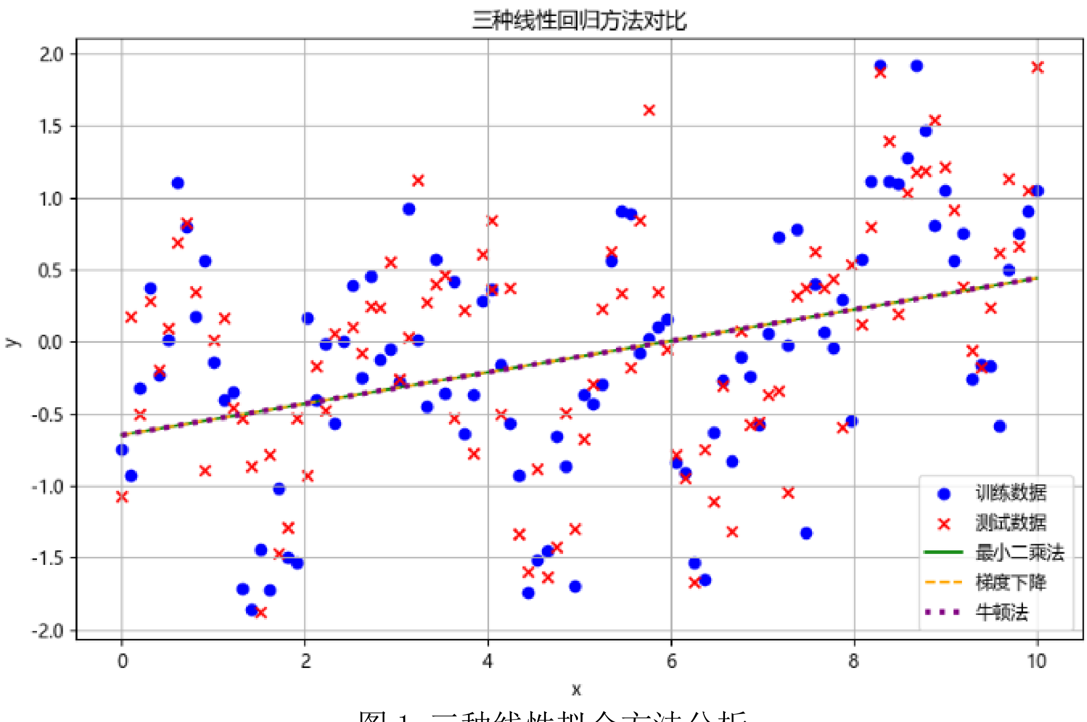

图 1. 三种线性拟合方法分析

```
【训练误差】
最小二乘法: 0.6134
梯度下降法: 0.6134
牛顿法:     0.6134

【测试误差】
最小二乘法: 0.5950
梯度下降法: 0.5950
牛顿法:     0.5950
```

图 2. 三种线性拟合训练误差和测试误差

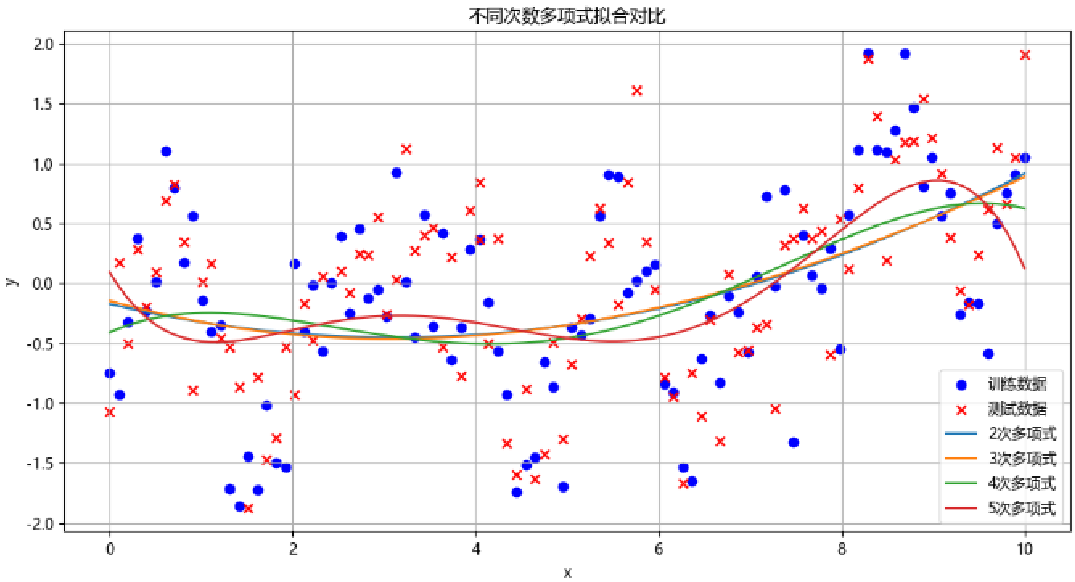

图 3. 多项式拟合方法分析（2-5 次多项式）

```
不同次数模型表现:
次数 | 训练误差 | 测试误差
2次 | 0.5654 | 0.5346
3次 | 0.5653 | 0.5368
4次 | 0.5559 | 0.5418
5次 | 0.5252 | 0.5151
```

图 4. 多项式拟合训练误差和测试误差（2-5 次多项式）

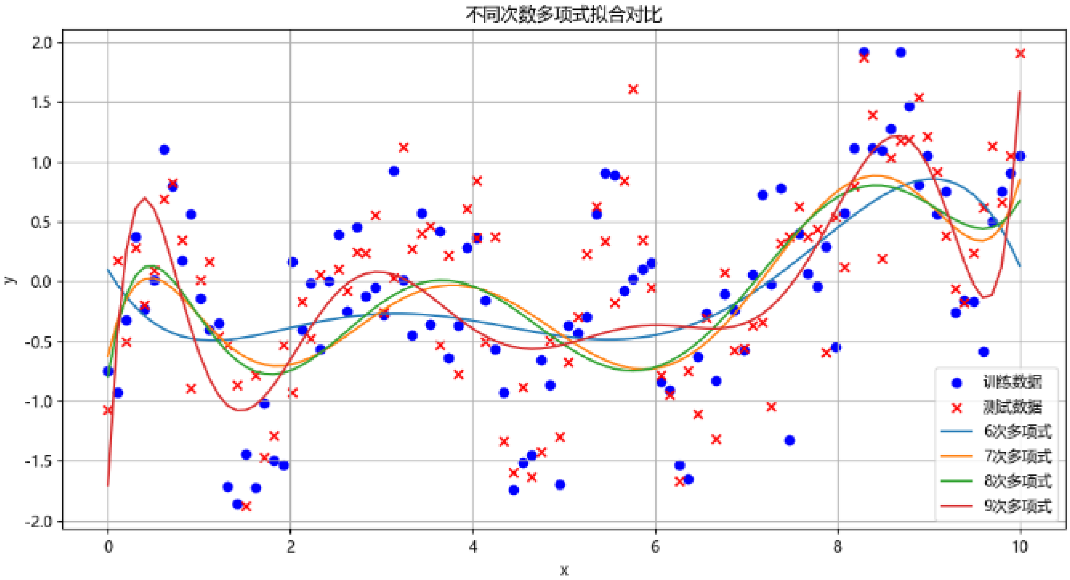

图 5. 多项式拟合方法分析（6-9 次多项式）

不同次数模型表现：

次数	训练误差	测试误差
6次	0.5252	0.5151
7次	0.4651	0.4631
8次	0.4614	0.4620
9次	0.3541	0.3875

图 6. 多项式拟合训练误差和测试误差（6-9 次多项式）

实验结论

线性回归模型下，最小二乘法，梯度下降法以及牛顿法得到的拟合直线一致，最小二乘法适合小规模线性问题，计算量小；梯度下降法适合大规模的高维数据；牛顿法利用二阶导数信息加快收敛。

多项式拟合过程中，采用最小二乘法来拟合，随着多项式次数的增加，拟合的效果越来越好，误差不断减小，。