PRML 第一次作业报告

22371375 谭海涛

Abstract

本研究选取二维数据集 Data4Regression 作为基准，分别使用最小二乘法、梯度下降法和牛顿法对训练数据进行线性拟合，并通过测试数据评估模型的泛化性能。实验结果表明，三种线性回归方法在训练误差和测试误差上表现一致，其中牛顿法因引入二阶导数信息，其收敛速度明显优于梯度下降法。

Introduction

在实际问题中，数据通常包含复杂的非线性规律，简单的线性模型难以充分描述这种关系。为验证这一现象，本文以二维数据集 Data4Regression 为例，首先采用最小二乘法、梯度下降法和牛顿法对训练数据进行线性拟合，并对比了各方法的训练误差和测试误差。尽管三种方法在数值上表现出较高的一致性，但线性模型仍存在拟合不足的局限性。为进一步提升拟合精度，本文引入了基于核函数的 KNN 回归模型，该模型通过局部加权平均方法，利用高斯核函数为不同距离的邻居分配不同权重，从而更有效地捕捉数据的局部非线性特征，同时梯度下降法加入早停机制（tolerance 参数），牛顿法改为单次迭代（理论最优解）。本文将系统阐述各方法的理论基础、实验结果及对比分析，并探讨模型选择与参数调优的关键因素。

Methodology

在下文中，我们将详细描述所使用的数据拟合方法。

M1：Ordinary Least Squares

最小二乘法（Ordinary Least Squares, OLS）是一种通过最小化预测值与真实值的残差平方和来估计模型参数的优化方法。其核心目标是找到一条直线（或超平面），使得所有数据点到该直线的垂直距离平方和最小。它广泛应用于线性 回归中，尤其适用于特征与目标变量近似线性相关的场景。

对于二维数据集 $\{(x_1,y_1),(x_2,y_2),...,(x_n,y_n)\}$，设拟合直线方程为

$y = \beta_0 + \beta_1 x$，目标 是找到参数 β0 和 β1 使得预测值与真实值的残差平方和$\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))^2$ 最小。

常用解法是通过矩阵形式求解，设计矩阵 X（含截距项）与观测向量 y 的参数解为：

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y}$$

其中：

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

M2：Gradient Descent Method

梯度下降法（Gradient Descent，GD）是一种通过迭代调整参数来最小化损失函数的优化算法。其核心思想是沿损失函数的负梯度方向逐步更新参数，直至收敛到局部最小值。

损失函数定义为：

$$J(\beta_0, \beta_1) = \frac{1}{2n}\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))^2$$

参数更新规则为：

$$\beta_0 = \beta_0 - \alpha\frac{\partial J}{\partial \beta_0}, \quad \beta_1 = \beta_1 - \alpha\frac{\partial J}{\partial \beta_1}$$

其中 α 为学习率（步长）。

M3：Newton’s Method

牛顿法（Newton’s Method）是一种利用二阶导数信息加速优化的迭代算法，通过构建目标函数的二次近似模型，直接求解极值点。迭代公式为：

$$\beta^{(k+1)} = \beta^{(k)} - \frac{f'(\beta^{(k)})}{f''(\beta^{(k)})}$$

其中 f' 为一阶导数，f'' 为二阶导数。

在求解损失函数的最小值时，需计算损失函数的梯度$\nabla J(\beta)$和海森矩阵（二阶导数矩阵）H。则参数更新公式为：

$$\boldsymbol{\beta} = \boldsymbol{\beta} - \mathbf{H}^{-1}\nabla J(\boldsymbol{\beta})$$

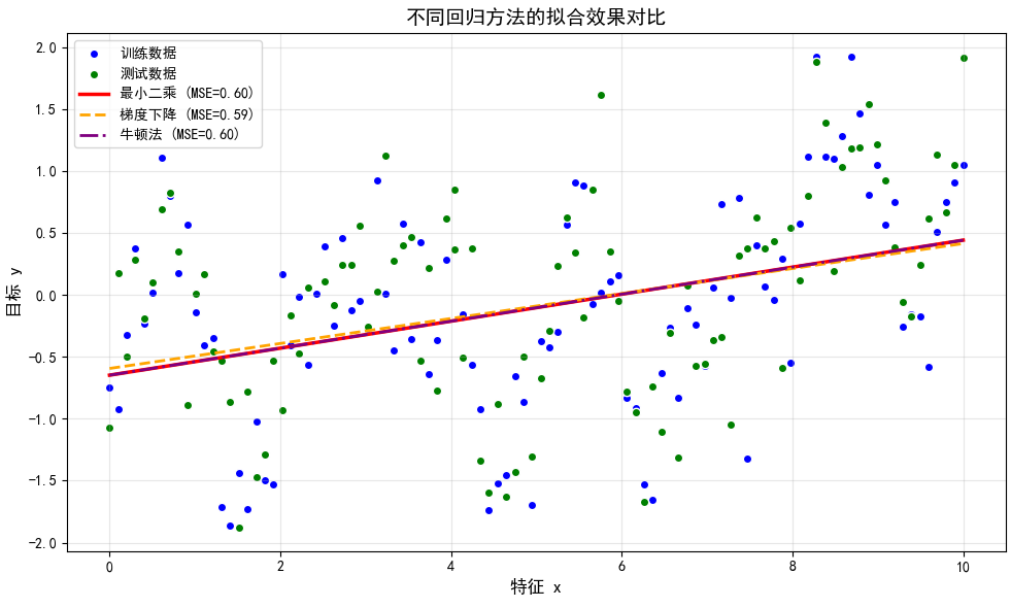
相比梯度下降法，牛顿法收敛速度更快（二阶收敛），但需计算 Hessian 矩阵及其逆矩阵，计算复杂度较高。

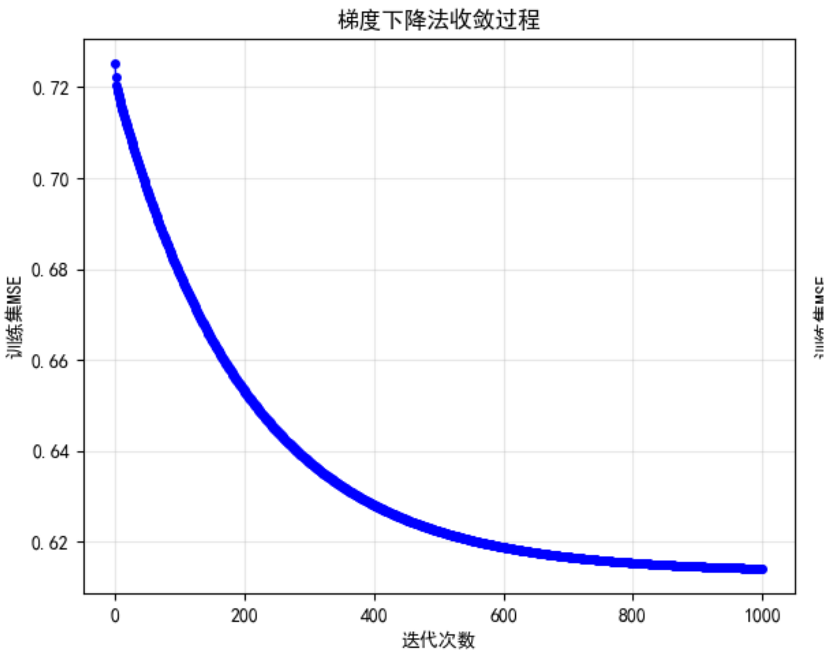
Eperimental Studies

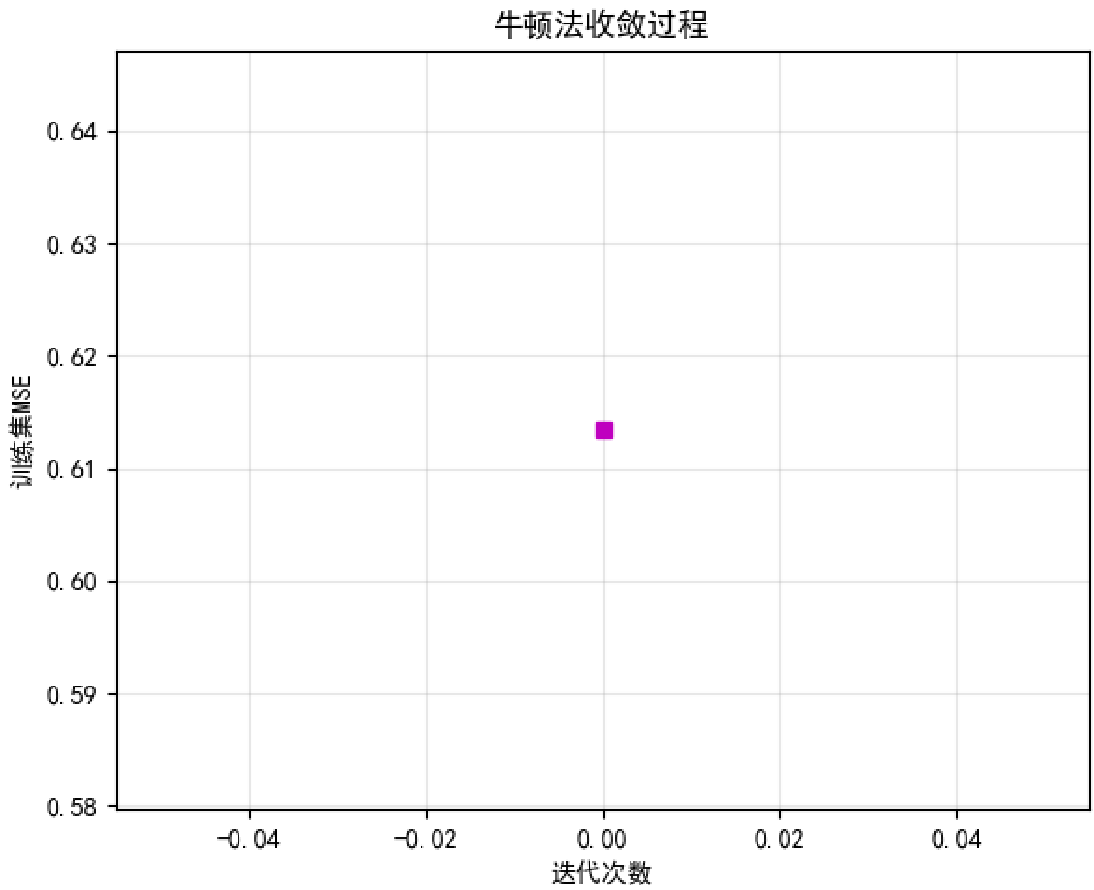
能效对比图

方法	训练集 MSE	测试集 MSE	收敛迭代次数	最终训练 MSE
最小二乘法	0.6134	0.5950		0.6134
梯度下降法	0.6141	0.5934	1000	0.6142
牛顿法	0.6134	0.5950	1	0.6134

可视化结果







Conclusions

在求解线性回归问题时，最小二乘法、梯度下降法和牛顿法均能够达到相近的训练误差和测试误差水平，表明它们在模型拟合能力上具有一致性。然而，三种方法在迭代次数和收敛速度上存在显著差异。最小二乘法通过解析解直接求得最优参数，无需迭代，计算效率最高；梯度下降法通过一阶导数信息逐步更新参数，收敛速度较慢，通常需要大量迭代才能接近最优解；而牛顿法利用二阶导数信息，能够更快地逼近最优解，显著减少了迭代次数。实验结果表明，牛顿法在收敛速度上具有明显优势，但其计算复杂度较高，尤其是在高维数据中可能面临计算瓶颈。因此，在实际应用中，需根据数据规模和计算资源权衡选择合适的方法。