PRML第一次作业报告

Ziheng Wang
wangziheng@buaa.edu.cn

Abstract

本报告探讨了线性回归模型在二维数据集Data4Regression上的适用性，并评估了最小二乘法、梯度下降法（GD）和牛顿法在训练与测试阶段的性能。实验结果表明，这三种方法的拟合效果相近，但可视化分析显示，它们均未能有效捕捉数据的特征。为进一步挖掘数据的真实模式，本研究突破线性拟合的限制，引入多层感知机（MLP），以更精确地拟合该非线性数据集。

Introduction

数据拟合是统计学和机器学习中的核心问题，广泛应用于预测建模、参数估计及科学数据分析等领域。线性回归作为一种常见方法，假设数据可用线性方程描述。然而，许多实际数据集呈现非线性特征，使得线性模型的拟合能力受限。

本研究探讨不同拟合方法对二维数据集的适用性，采用最小二乘法（Least Squares）、梯度下降法（Gradient Descent, GD）和牛顿法（Newton’s Method）进行实验，并以均方误差（MSE）评估模型的拟合效果。通过实验分析这些方法在训练数据和测试数据上的表现，比较它们在不同情况中的拟合效果。

在初步线性回归分析后，本报告进一步探讨当线性模型无法有效拟合数据时，如何通过更复杂的模型提升拟合效果。鉴于数据的非线性特征，实验选择目前较为流行且适应性较强的多层感知机（MLP）进行拟合，并比较其与线性模型的表现。通过实验分析，本研

究旨在探讨模型选择对数据拟合结果的影响，为未来的研究提供参考。随着数据科学和机器学习的快速发展，合理选择模型对于提高拟合精度至关重要，本研究希望能为实际数据分析提供有益的见解。

Methodology

本部分将详细阐述数据拟合过程中所用回归模型的原理，包括具体的实现步骤和过程。

M1: Least Squares

最小二乘法是一种统计方法，通过最小化观察值与模型预测值之间的垂直距离（残差）来估计线性回归模型的参数。其线性模型可用如下形式表示：

$$h_\theta(\boldsymbol{x}^{(i)}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

因此，我们需要找到一组最合适的参数$\boldsymbol{\theta} = [\theta_0, \theta_1, \theta_2, \cdots, \theta_n]^T$来满足下述方程：

$$\begin{cases} \theta_0 + \theta_1 {x_1}^{(1)} + \theta_2 {x_2}^{(1)} + \cdots + \theta_n {x_n}^{(1)} = y^{(1)} \\ \theta_0 + \theta_1 {x_1}^{(2)} + \theta_2 {x_2}^{(2)} + \cdots + \theta_n {x_n}^{(2)} = y^{(2)} \\ \vdots \\ \theta_0 + \theta_1 {x_1}^{(N)} + \theta_2 {x_2}^{(N)} + \cdots + \theta_n {x_n}^{(N)} = y^{(N)} \end{cases}$$

通常，我们无法找到这组方程的解，所以我们需要找到最合适的一组参数，来使得计算所得输出和真实输出尽可能的接近。即我们需要解决如下问题：

$$\min ||\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{Y}||_2^2$$

式中，$\boldsymbol{X}$为$m \times (n+1)$矩阵，表示有m组输入，每组输入维度为$(n+1)$，这里指加上偏置之后的维度。最小二乘法的优化方法也很简单，直接通过矩阵运算，对损失函数$\boldsymbol{J}(\boldsymbol{\theta})$求导，并令导数为零：

$$\frac{\partial \boldsymbol{J}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{m} \boldsymbol{X}^T (\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{Y}) = 0$$

式中，m为样本数量，解得：

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

实际实现时，直接导入输入和相应输出到上述表达式即可得到相应的最优参数。不过需要注意，在实现过程中要先对输入向量加一列偏置。

M2: Gradient Descent

梯度下降法分析时，同样假设线性模型可用如下形式表示：

$$h_\theta(\boldsymbol{x}^{(i)}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

采用均方误差（MSE）作为损失函数：

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

之后计算损失函数对参数$\boldsymbol{\theta}$的梯度，沿着梯度的负方向更新参数，逐步找到最优解，具体计算公式如下：

$$\frac{\partial \boldsymbol{J}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} = \frac{1}{m}\sum_{i=1}^{m}\left(h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) - y^{(i)}\right)\boldsymbol{x}_{i,j}$$

用矩阵表示：

$$\nabla \boldsymbol{J}(\boldsymbol{\theta}) = \frac{1}{m}\boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{Y})$$

梯度下降更新公式：

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \alpha \cdot \frac{1}{m}\boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{Y})$$

其中，α为自定义的学习率。

具体实现时，需自行定义学习率以及迭代次数（本实验定义的学习率为0.01，迭代次数为1000次），之后将输入和输出导入公式进行计算即可。

M3: Newton’s Method

牛顿法是一种二阶优化方法，前面的模型假设和损失函数与前面均相同，只是参数更新公式不同。其利用梯度和Hessian矩阵（二阶导数矩阵）来更新参数，更新公式为：

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

其中，梯度计算公式：

$$\nabla \boldsymbol{J}(\boldsymbol{\theta}) = \frac{1}{m} \boldsymbol{X}^{\boldsymbol{T}} (\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{Y})$$

Hessian矩阵计算公式：

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

所以，牛顿法的更新规则变为：

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

牛顿法相比于梯度下降法，收敛速度更快，理论上仅需少量迭代便可找到最优解。本实验中，迭代次数设置为10次。

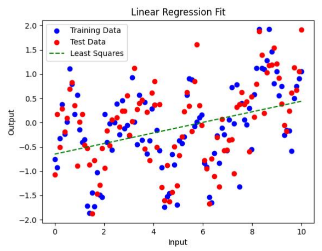
M4: Multilayer Perception

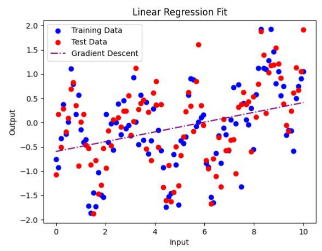
多层感知机（MLP）是一种前馈神经网络，由多个神经元层组成，可以学习非线性的映射关系。本实验中，定义BPNet，神经网络由8层（7个隐藏层+1个输出层）组成，使用Xavier正态分布初始化权重，激活函数选取ReLU函数，Dropout设置为靠近输入和输出层的置0，剩余层依次增加，以防止过拟合。实验过程中设置学习率为0.0001，循环次数分别设置为5000次和10000次。

Experimental Studies

Table 1 : Linear Regression ($y = \beta_0 + \beta_1 x$)

New Table	β_0	β_1	MSE(train)
M1	-0.6487467	0.10894739	0.613402
M2	-0.59455395	0.10079896	0.614148
M3	-0.6487467	0.10894739	0.613402





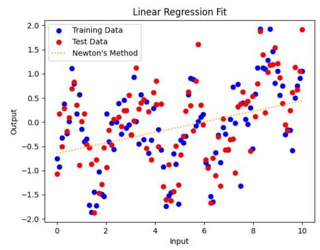


Figure 1 : Visual linear regression analysis

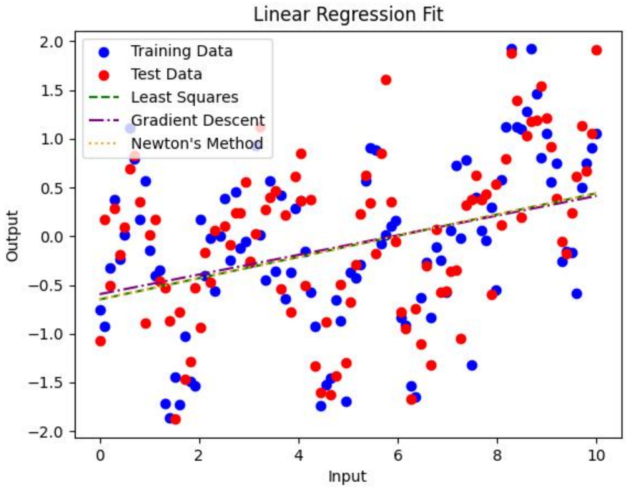


Figure 2 : Linear Regression

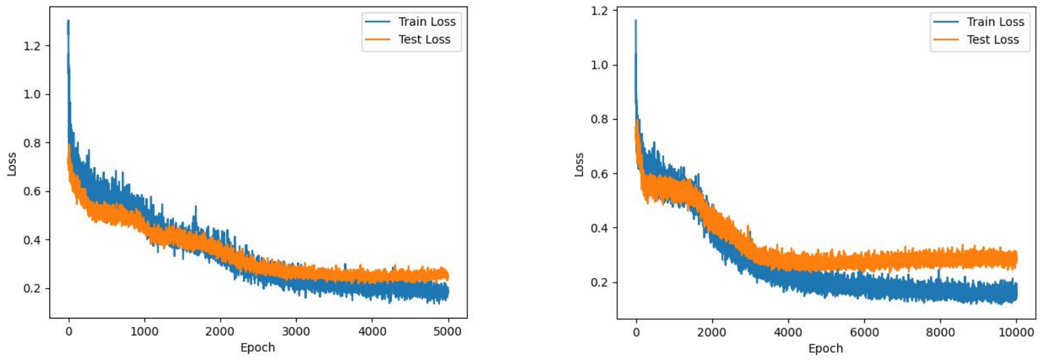


Figure 3： Loss-value curve

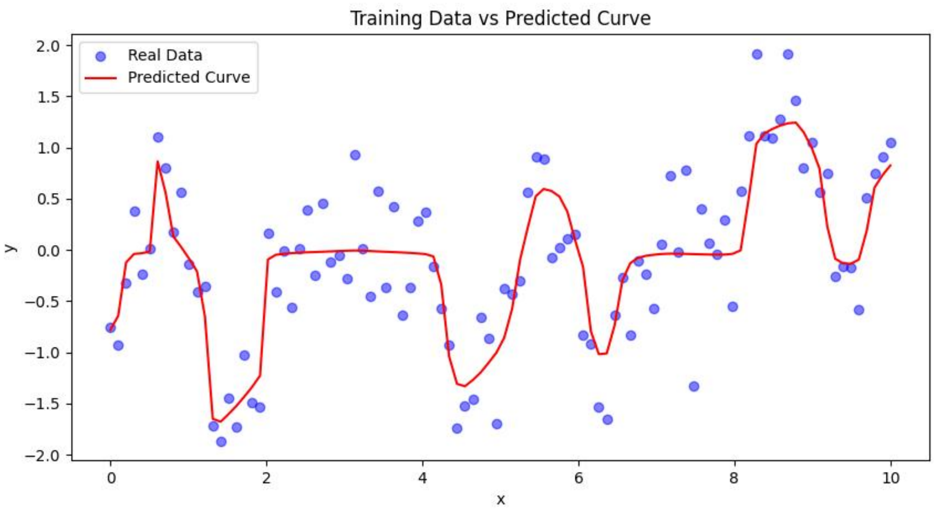


Figure 4 : MLP fitting curve (Training set)

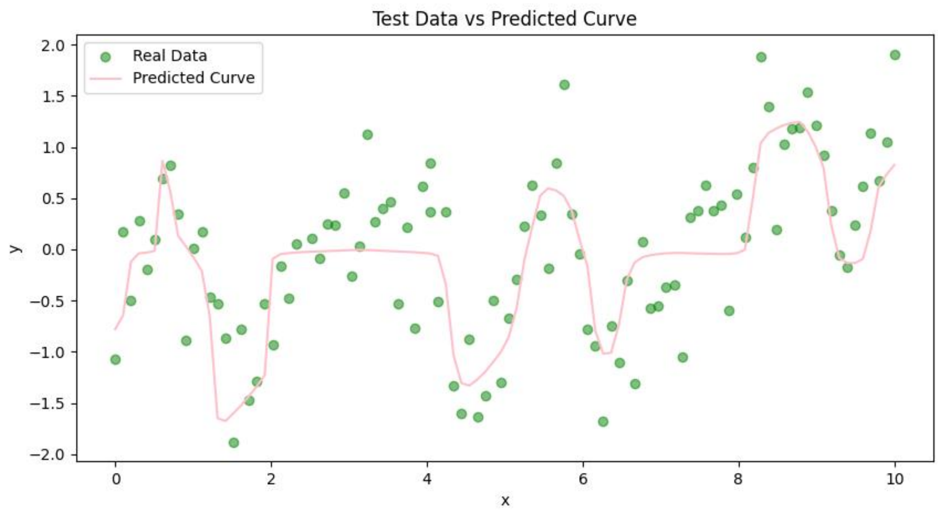


Figure 5 : MLP fitting curve (Test set)

最终，训练集的损失函数稳定在0.17左右，测试集的损失函数稳定在0.27左右。

Conclusions

本研究评估了线性回归模型在二维数据集上的适用性，并对比了最小二乘法、梯度下降法和牛顿法的拟合效果。实验结果表明，这三种方法的均方误差相近，但均未能有效捕捉数据的非线性特征，导致拟合能力受限。为提升拟合效果，我们引入了多层感知机，其非线性映射能力使模型在训练集上的损失降至 0.17，在测试集上降至 0.27，显著优于线性方法。尽管 MLP 能够更准确地拟合数据，但测试集损失略高，可能存在一定程度的过拟合。为提升模型的泛化能力，实验采用 Dropout 技术，并调整训练轮数，以优化拟合效果。整体来看，MLP 在非线性数据拟合方面优于传统线性模型，未来研究可尝试更复杂的神经网络架构或正则化方法，以进一步提升模型的稳健性和泛化能力。