

Assignment 1

Abstract

本报告探讨不同线性拟合方法在二维数据集上的表现，重点比较**最小二乘法**、**梯度下降法**和**牛顿法**的拟合效果。

首先，分别用三个方法在训练集上进行线性拟合，获得参数和训练误差。

然后，利用测试集对上面所得参数进行测试，获得测试误差，对比几种方法。

最后，由于训练集用线性方法拟合结果并不理想，更换非线性拟合方式再次进行数据拟合。本报告中采用**高斯核函数**的方式进行拟合。

Introduction

假设给定二维度数据集

$$(x_i, y_i) (i = 1, 2, \dots, m)$$

考虑采用线性模型进行线性拟合

$$h(x) = \theta_0 + \theta_1 x$$

拟合目标是使得**均方误差**最小，即下面表达式取到最小值

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h(x_i) - y_i)^2$$

为了方便后面数据处理，将数据集中输入人为添加一列，便于后面进行矩阵计算，变化后为：

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

损失函数写成矩阵形式

$$J(\theta) = \frac{1}{2m} \|X\theta - Y\|^2$$

Methodology (Linear Models)

采用三种线性模型。

1.最小二乘法

目标求解：

$$\nabla_{\theta} J(\theta) = 0$$

求梯度：

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

令其为零：

$$X^T X \theta = X^T Y$$

如果左侧矩阵可逆，则有解析解：

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

2. 梯度下降法

梯度下降法采用迭代更新方式，设学习率为alpha，按以下公式更新：

$$\theta = \theta - \alpha \nabla_{\theta} J(\theta)$$

即：

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

其中，alpha需要调整，以确保收敛。

3. 牛顿法

牛顿法使用二阶导数信息，Hessian矩阵进行迭代：

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

其中Hessian矩阵为：

$$H = \nabla_{\theta}^2 J(\theta) = \frac{1}{m} X^T X$$

梯度：

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

代入更新公式：

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

Methodology (Nonlinear Model)

采用高斯核函数进行非线性模型拟合。在核方法中，通过选择合适的核函数（如高斯核），训练各个位置上的权重。

首先根据距离计算相似度

$$K_{ij} = K(x_i, x_j) = \exp \left(-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right)$$

优化的函数可以写成下面形式

$$L(\alpha) = \|K\alpha - Y\|^2 + \lambda \|\alpha\|^2$$

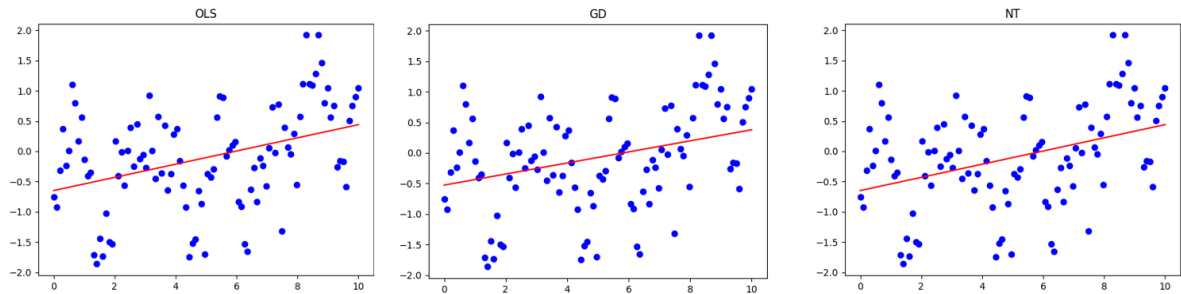
其中alpha为训练参数，lamda为预先设置的参数，避免模型出现过拟合的状况

预测结果如下

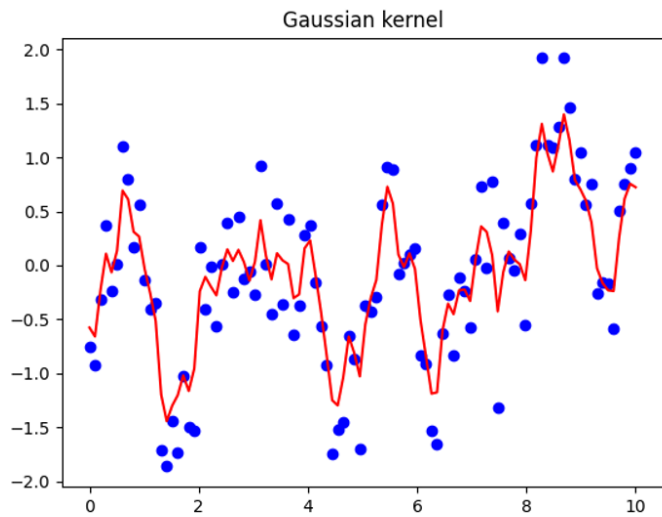
$$Y_{pred} = K\alpha$$

Experimental Studies

线性模型拟合结果



非线性拟合结果



线性拟合具体数值

	theta 0	theta 1	Training loss	Test loss
最小二乘 (OLS)	-0.6487467	0.10894739	0.3067012140725026	0.2975216930733648
梯度下降 (GD)	-0.52685403	0.09061958	0.30858674921819673	0.2967235531976619
牛顿法 (NT)	-0.6487467	0.10894739	0.30670121407250256	0.2975216930733648

非线性拟合

	Training loss	Test loss
高斯核	0.04525650660829045	0.13248094131254676

Conclusion

对于线性模型来说，最小二乘法可以在有最优解的时候一定得到最优解，但由于需要计算维度较大的矩阵的逆，因此不适用于大数据量情况。梯度下降在本案例中收敛速度偏慢，且学习率如果太大可能出现不收敛的情况，但是当轮次达到一定大小后，也基本可以趋于并最终取到最优解（大概1000轮次以上）。本案例中牛顿法收敛速度最快，仅一次即可得到最优解，但是由于需要计算海森矩阵核逆矩阵，相对计算量也较大。

对于本例中的数据分布，采用非线性模型效果明显更好。前面采用高斯核函数方式进行拟合，训练误差和测试误差都远小于线性模型线性拟合得到的结果。高斯核函数相当于把一维度数据通过计算和其余数据的相似度的方式投影到高维空间中（使得数据具有更多维度的特征），然后对于全局数据进行一定权重（可训练参数）的加权求和。这个模型中 λ ，是人为设置的超参数（只调整到了小数点后一位）。这个参数代表选择的权重 α ，不能为了拟合训练数据而变得数值过于复杂。如果过小，在小数数据集中会出现明显的过拟合；过大则会出现欠拟合。

总之对于这个数据集而言，非线性模型拟合会表现的比线性模型更好。这三个线性模型经过，基本都会收敛到特定的最优解处。