

# Homework 01 Curve Fitting

马倩

2016 年 5 月 24 日

## 1 实验内容

使用多项式对带有噪声的  $\sin(x)$  数据进行拟合，实验内容一共有五个部分：

- (1) 10 个样本点，多项式最高阶为 3
- (2) 10 个样本点，多项式最高阶为 9
- (3) 15 个样本点，多项式最高阶为 9
- (4) 100 个样本点，多项式最高阶为 9
- (5) 10 个样本点，多项式最高阶为 9，带有正则项， $\lambda = -18=-18$

## 2 实验方法

使用最小二乘方法选取多项式参数。设拟合多项式为：

$$y = a_0 + a_1x + \dots + a_kx^k$$

简写为  $Y = AX$ ，将偏差平方和作为目标函数，通过对  $A$  求导使得目标函数最小，最后可得： $A = (X' \times X)^{-1} \times X' \times Y$ 。

单纯考虑误差最小可能会导致过拟合，为解决这个问题，可以在目标函数中加入正则项，求导后可得： $A = (X' \times X - \lambda)^{-1} \times X' \times Y$ 。

## 3 实验数据及结果

数据：使用  $\sin(x)$  函数产生样本点同时加上缩小后的均值为 0，方差为 1 的高斯分布噪音。

结果：

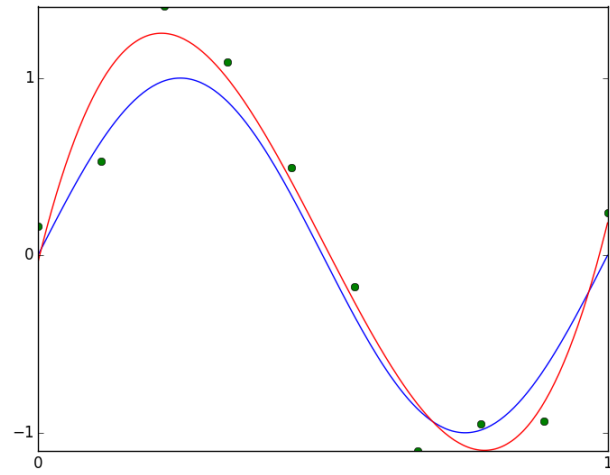


Figure 1: Sample=10 M=3

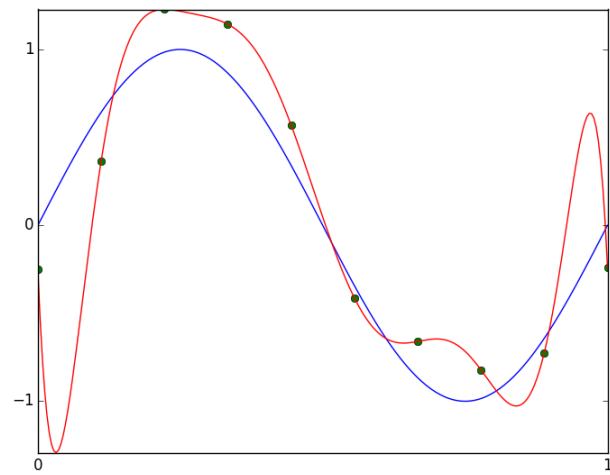


Figure 2: Sample=10 M=9

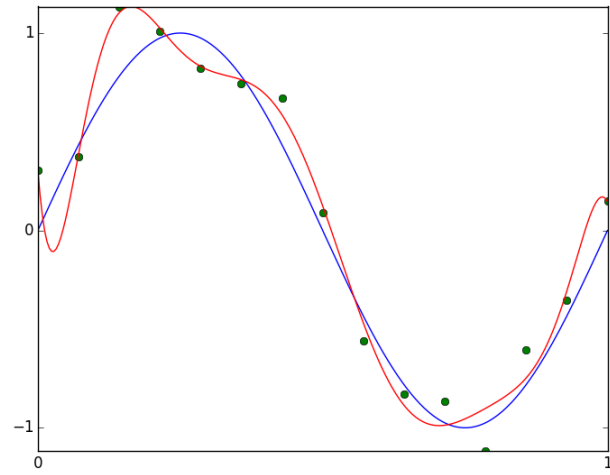


Figure 3: Sample=15  $M=9$

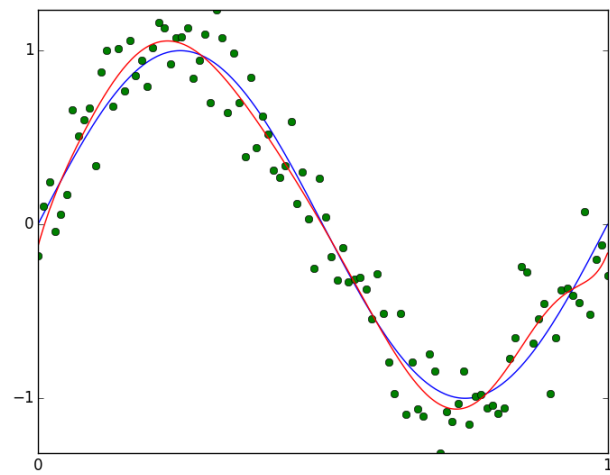


Figure 4: Sample=100  $M=9$

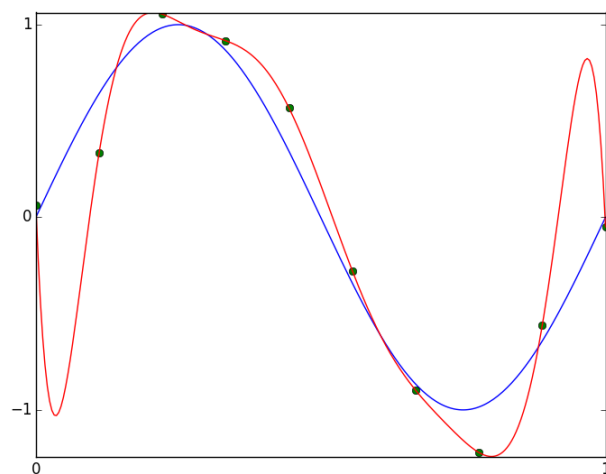


Figure 5: Sample=10 M=9  $\lambda=-18$

## 4 讨论与分析

在  $M$  较大而样本数量较小时，所产生的曲线虽然能够符合所有的数据点，但无法与  $\sin$  曲线相吻合，这样在其它  $\sin$  数据上就会产生较大的误差。这是由于有着更大  $M$  的多项式被过分地调参，使得多项式被调节成了与目标值的随机噪声相符。对应于这种情况，可以添加正则项，控制产生系数的大小，从而抑制  $M$  较大时的过拟合情况。

## References

- [1] <http://blog.csdn.net/jairuschan/article/details/7517773/>
- [2] [http://blog.csdn.net/lotus\\_\\_\\_\\_/article/details/20546259](http://blog.csdn.net/lotus____/article/details/20546259)