**Machine Learning Exercise 1.1**

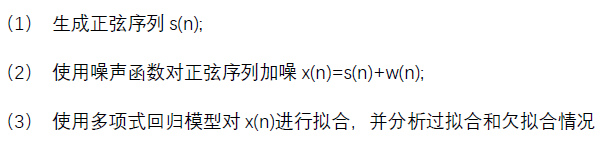
**姓名：沈金龙**

**学号：18214806**

**实验题目：**

编写程序：模拟仿真多项式回归

要求如下：



**实验过程：**

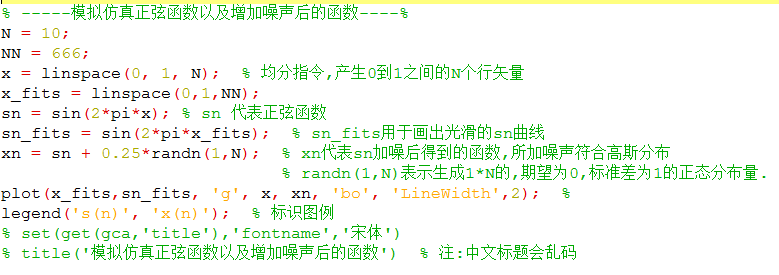
本实验采用**Octave4.4.1**完成仿真任务；

1、假设正弦序列s(n) = sin(2\*pi\*n)，利用linspace函数在[0,1]区间取10个均值点作为s(n)函数的输入；噪声函数w(n) = 0.25\*randn(1,N)，噪声符合正态分布，加噪之后的函数假设为x(n)，则x(n) = s(n) + w(n)；

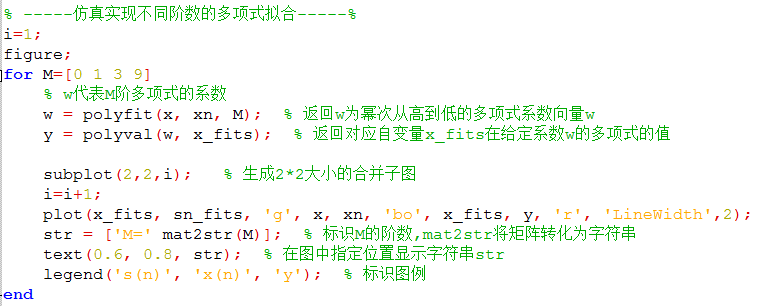
2、假设多项式函数y =w\_0 + w\_1\*x + … + w\_m\*x^M，其中M为多项式的最高阶数。本次实验利用polyfit以及polyval函数根据给定训练集（x, xn）和指定的多项式阶数M(M=0,1,3,9)来求得指定阶数多项式的参数w，以此研究不同阶数多项式过拟合以及欠拟合的情况；

**部分代码截图：**

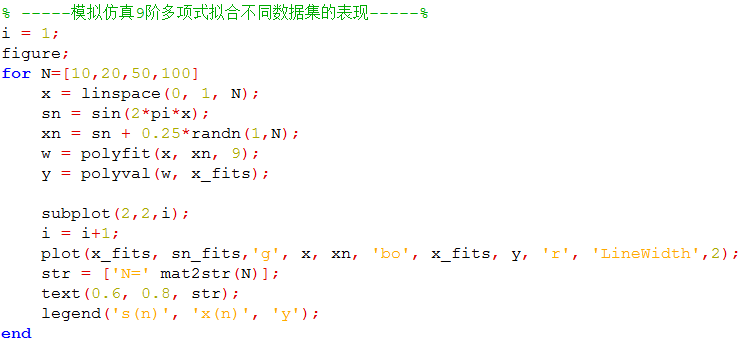
>模拟仿真正弦函数以及增加噪声后的函数：



>模拟仿真实现不同阶数的多项式拟合：

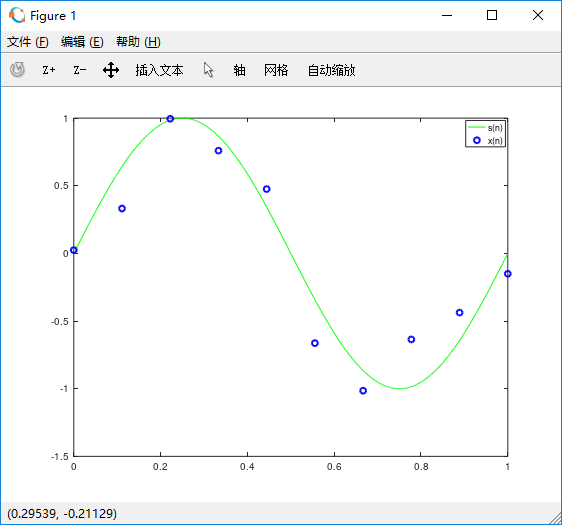


>模拟仿真9阶多项式拟合不同数据集的表现：

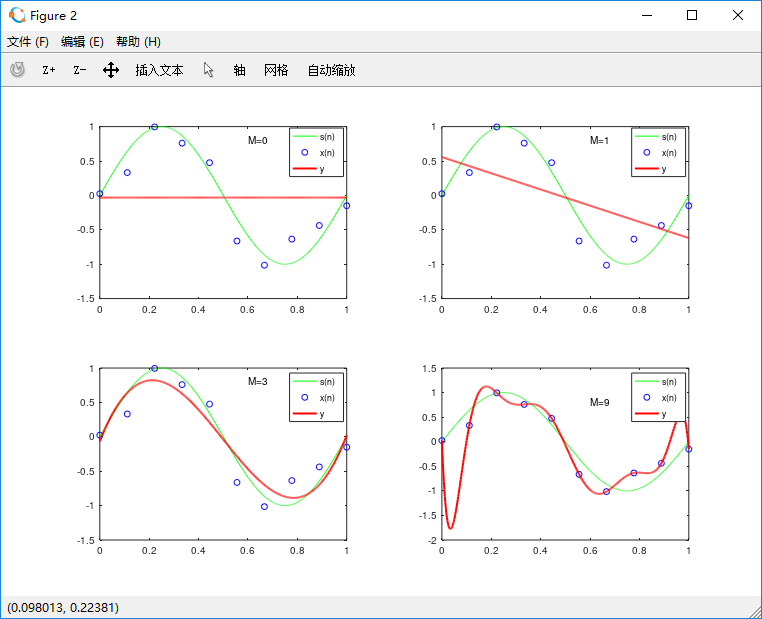


**实验结果截图：**

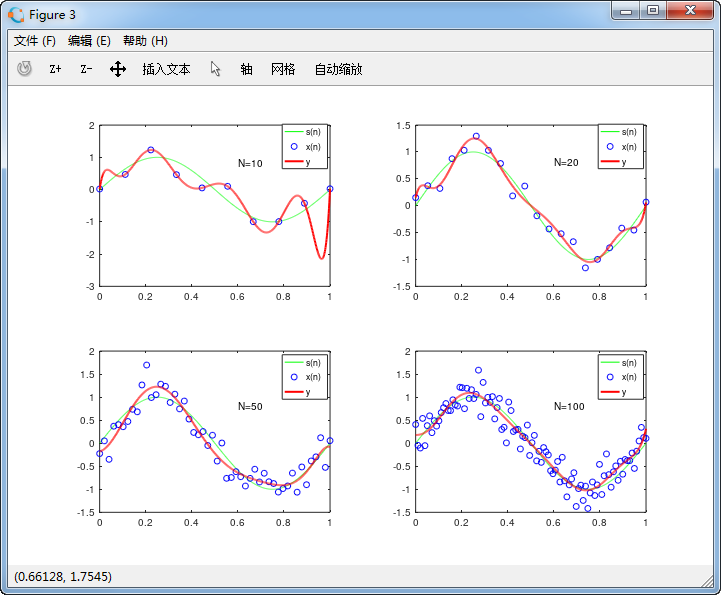
>正弦序列s(n)和增加噪声后的序列x(n)的图像如下图所示：



>不同阶数多项式y拟合x(n)图像如下图所示：



>保持M=9不变，增加数据集的大小，拟合图像如下图所示：



**实验分析：**

1. 当M比较小时，如M=0或者M=1时，会出现欠拟合现象，多项式曲线y无法很好的拟合数据集(x,xn)，主要原因是模型太简单，没有很好地捕捉到数据特征，不能够很好地拟合数据。
2. 随着M的增加，模型逐渐变得复杂，y对数据集的拟合效果也越来越好。
3. 当M=9时，出现了过拟合现象，主要原因是训练的数据量太小或者说训练数据占总数据的比例过小。
4. 当多项式的阶数固定为M=9时，随着训练集的增加，模型的过拟合问题得到了解决。