



中国科学院大学

University of Chinese Academy of Sciences

模式识别与机器学习

081203M04004H

Chap 5 课程作业解答

2022 年 10 月 24 号

Professor: 黄庆明



学生: 周胤昌

学号: 202228018670052

学院: 网络安全学院

所属专业: 网络安全

方向: 安全协议理论与技术

Problem 1

自己编程实现课堂上 Polynomial Curve Fitting 的例子, 体会过拟合.

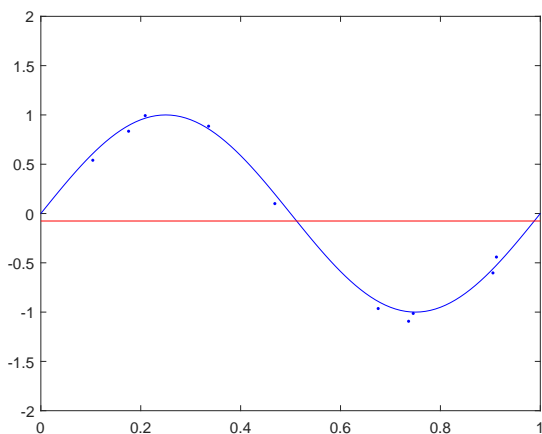
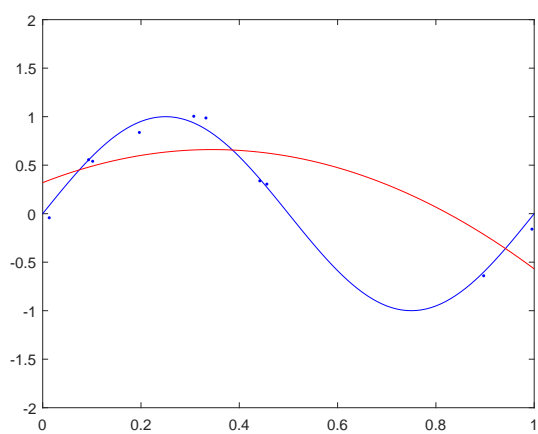
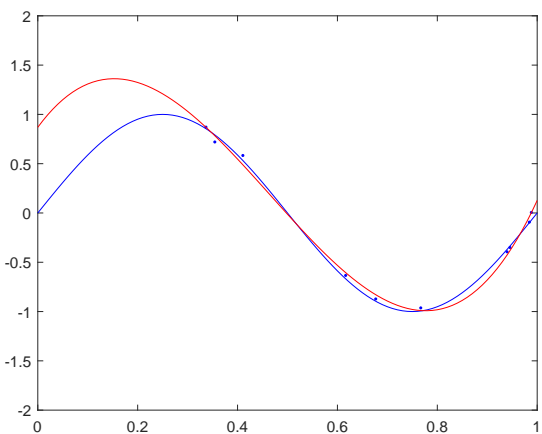
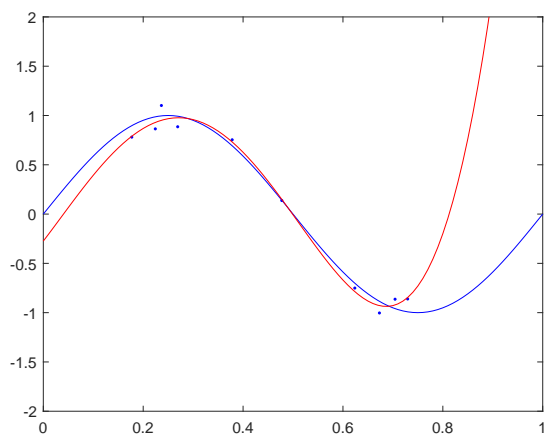
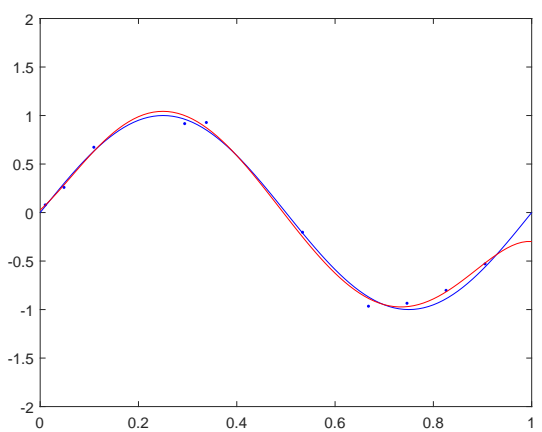
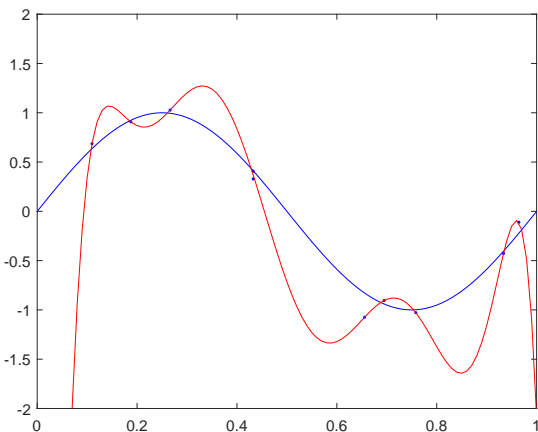
Solution: 先根据均匀分布 $U(0, 1)$ 随机产生 10 个数据点, 再利用函数 $y(x) = \sin(2\pi x)$ 并添加高斯噪声 (噪声均值为 0, 方差为 $\sigma^2 = 0.08^2$) 产生对应的 y 数据. 我们分别用 $M = 0, 2, 5, 9$ 次多项式对上述产生的数据点进行多项式拟合. 具体的 Matlab 代码如下所示:

```

1 N = 10; % 产生的数据点的个数
2 x = rand(10,1); % 产生均匀分布 U(0,1) 的 10 个点
3 noise_sigma = 0.08; % 噪声的方差为 noise_sigma^2
4 M = 9;
5 y = sin(2*pi*x) + randn(10,1)*noise_sigma; %randn 产生 N(0,1) 正态分布的数据
6 figure(1)
7 axis([0 1 -2 2])
8 plot(x, y, 'b.')
9 x_r = 0: 0.01: 1;
10 y_r = sin(2*pi*x_r);
11 hold on
12 plot(x_r, y_r, 'b'); % 数据的真实曲线 (蓝色的)
13 p_x = [];
14
15 for m = 0 : M
16     p_x = [p_x, x.^m]; % 产生 [x^0; x^1; x^2, ..., x^M]
17 end
18 p_x = p_x';
19 w = pinv(p_x*p_x')*p_x*y;
20 y_est = w'*p_x;
21 figure(1);
22 hold on
23 x_cur = 0:0.01:1;
24
25 y_cur = zeros(size(x_cur));
26 for m = 0 : M
27     y_cur = y_cur + w(m + 1)*(x_cur.^m);
28 end
29 axis([0 1 -2 2])
30 plot(x_cur, y_cur, 'r') % 画出红色的拟合曲线

```

代码的运行结果和具体的多项式曲线拟合情况见后页图1,2,3,4,5,6中所示. 可以看出的是: 当 $M = 3, 4, 5$ 时, 过拟合效应较小, 当 $M = 9$ (过大) 时, 多项式拟合出现了强烈的过拟合现象.

图 1: $M = 0$ 时的拟合情况图 2: $M = 2$ 时的拟合情况图 3: $M = 3$ 时的拟合情况图 4: $M = 4$ 时的拟合情况图 5: $M = 5$ 时的拟合情况图 6: $M = 9$ 时的拟合情况