上海理工大学光电信息与计算机工程学院

**《智能仿真实验》报告**

****

**专　　业 智能科学与技术**

**姓 名 高浩琦**

**学　 号 2035060413**

**年　　级 2020 级**

**指导教师 陈玮**

**成 绩：**

**教师签字：**

目录

[实验一 多项式曲线拟合及最小二乘曲线拟合 2](#_Toc105167492)

[一．实验目的 2](#_Toc105167493)

[二．实验原理 2](#_Toc105167494)

[2.1多项式拟合原理 3](#_Toc105167495)

[2.2最小二乘拟合原理 3](#_Toc105167496)

[2.2.3协方差矩阵 3](#_Toc105167497)

[2.3 MATLAB中多项式拟合方法 3](#_Toc105167498)

[2.3.1 polyfit函数 3](#_Toc105167499)

[2.3.2 polyval函数 3](#_Toc105167500)

[2.3.3 拟合优度分析方法 3](#_Toc105167501)

[1. SSE(残差平方和): the Sum of Square due to Error 3](#_Toc105167502)

[2. RMSE(均方根误差、标准差)：Root mean squared error 3](#_Toc105167503)

[3. MSE(均方差、方差)：Mean squared error 3](#_Toc105167504)

[4. R-square(R^2)(决定系数):Coefficient of determination 3](#_Toc105167505)

[5. (校正决定系数): adjust R-square 3](#_Toc105167506)

[2.4 MATLAB最小二乘拟合方法 3](#_Toc105167507)

[三．实验内容 3](#_Toc105167508)

[四、分析过程与部分代码 3](#_Toc105167509)

[1. 散点图绘制 3](#_Toc105167510)

[2. 拟合曲线和残差图绘制 4](#_Toc105167511)

实验一 多项式曲线拟合及最小二乘曲线拟合

一．实验目的

1.1掌握多项式曲线拟合和最小二乘拟合的原理和方法；

1.2掌握中多项式拟合及最小二乘曲线拟合的方法。

二．实验原理

2.1多项式拟合原理

已知变量x,y之间的函数关系为



现希望通过实验获得一组测量数据，确定出系数，这类问题就是多项式拟合。

2.2最小二乘拟合原理

最小二乘法多项式曲线拟合，根据给定的m个点,并不要求这条曲线精确地经过这些点，而是曲线y=f(x)的近似曲线y= φ(x)。使偏差平方和最小，即

按偏差平方和最小的原则选取拟合曲线，并且采取二项式方程为拟合曲线的方法,称为最小二乘法。

推导过程：

I．设拟合多项式为：

II．各点到这条曲线的距离之和，即偏差平方和如下：



III．为了求得符合条件的ai值，对等式右边求ai偏导数，得到了：

IV．将等式左边进行一下化简，然后应该可以得到下面的表达式：

，

…



V．把这些等式表示成矩阵的形式，就可以得到下面的矩阵：



VI．将这个范德蒙得矩阵化简后可得到:



VII．也就是说，便得到了系数矩阵，同时，我们也就得到了拟合曲线。

2.2.3协方差矩阵

]

数字越接近于1 说明相关性越强

2.3 MATLAB中多项式拟合方法

2.3.1 polyfit函数

常用函数用法：，，返回次数为n的多项式p(x)的系数，该阶数是y中数据的最佳拟合（在最小二乘方式中）p中的系数按降幂排列，p的长度为n+1，函数模型为：



返回的结构体S，可用作的输入来获取误差估计值。 返回的mu，是一个二元素向量，包含中心化值和缩放值。mu(1)是mean(x)，mu(2)是std(x)。使用这些值时，将x的中心置于零值处并缩放为具有单位标准差: 这种中心化和缩放变换可同时改善多项式和拟合算法的数值属性

2.3.2 polyval函数

常用函数用法：或，

计算多项式p在x的每个点处的值。参数p是长度为n+1的向量，其元素是n次多项式的系数，并且系数是降幂排序的

使用等函数都可以计算p中的多项式系数，但也可以事先为系数指定任何向量。如果要以矩阵方式计算多项式，可以用函数。

使用生成的可选输出结构体S来生成误差估计值。delta是使用p(x)预测x处的未来观测值时的标准误差估计值。

或使用生成的可选输出mu来中心化和缩放数据。mu(1)为mean(x)，mu(2)为std(x)。使用这些值时，将x的中心置于零值处并缩放为具有单位标准差:  这种中心化和缩放变换同样可改善多项式的数值属性。

2.3.3 拟合优度分析方法

1. SSE(残差平方和): the Sum of Square due to Error

SSres =

该参数越接近于0,表示曲线拟合越成功

2. RMSE(均方根误差、标准差)：Root mean squared error

RMSE =

每个误差对 RMSE 的影响与平方误差的大小成正比；因此较大的误差对 RMSE 的影响非常大,所以RMSE 对异常值很敏感。

3. MSE(均方差、方差)：Mean squared error

数理统计中均方误差是指参数估计值与参数值之差平方的期望值，记为MSE。MSE是衡量“平均误差”的一种较方便的方法，MSE可以评价数据的变化程度，MSE的值越小，说明预测模型描述实验数据具有更好的精确度。预测数据和原始数据对应点误差平方和的均值.

MSE =

4. R-square(R^2)(决定系数):Coefficient of determination

​ R^2即判定系数，也称为拟合优度，拟合优度越大，自变量对因变量的解释程度越高，自变量引起的变动占总变动的百分比就越高。观察点在回归直线附近越密集。取值范围：0-1.

回归平方和：**SSR(Sum of Squares for regression)** = **ESS(explained sum of squares)**

残差平方和：**SSE(Sum of Squares for Error）** = **RSS(residual sum of squares)**

总离差平方和：**SST(Sum of Squares for total)** = **TSS(total sum of squares)**

5. (校正决定系数): adjust R-square

R^2评价拟合模型的好坏具有一定的局限性，即使向模型中增加的变量没有统计学意义，R^2值仍会增大。因此需对其进行**校正**，从而形成了校正的决定系数**(Adj R^2)** 。与R^2不同的是，当模型中增加的变量没有统计学意义时，Adj R^2会减小，因此AdjR^2是衡量所建模型好坏的重要指标之一，AdjR^2越大，模型拟合得越好。

校正决定系数（AdjR^2）引入了样本数量和特征数量，公式如下

2.4 MATLAB最小二乘拟合方法

已知样本点：，已知原型函数：

则定义拟合误差的最小二乘：

求解函数：

三．实验内容

1.以中美两国1980年至2016年的GDP历史数据为基础，用多项式拟合进行曲线拟合，确定其数学模型，并给出拟合过程及分析。

2.用最佳的多项式拟合模型进行最小二乘拟合，并观察模型误差。

3.用拟合模型预测2017年至2020年中、美GDP，并计算模型的均方误差MSE和拟合优度R2。

数据：

(1) 中国1980年至2016年GDP，单位：万亿美元。CH=[305350,290724,286729,307683,316666,...              %1980~1984

312616,303340,330303,411923,461066,...

398623,415604,495671,623054,566471,...              %1990~1994

736870,867224,965320,1032576,1097133,...

1214912,1344097,1477483,1671072,1966223,...         %2000~2004

2308786,2774308,3571451,4604285,5121681,...

6066351,7522103,8570348,9635025,10534526,...        %2010~2014

   11226186,11232108].'\*1e-6;    %2015~2016

(2) 美国1980年至2016年GDP，单位：万亿美元。US=[2862475,3210950,3345000,3638125,4040700,...     %1980~1984

4346750,4590125,4870225,5252625,5657700,...

5979575,6174050,6539300,6878700,7308775,...         %1990~1994

7664050,8100175,8608525,9089150,9660625,...

10284750,10621825,10977525,11510675,12274925,...    %2000~2004

13093700,13855900,14477625,14718575,14418725,...

14964400,15517925,16155250,16691500,17393100,...    %2010~2014

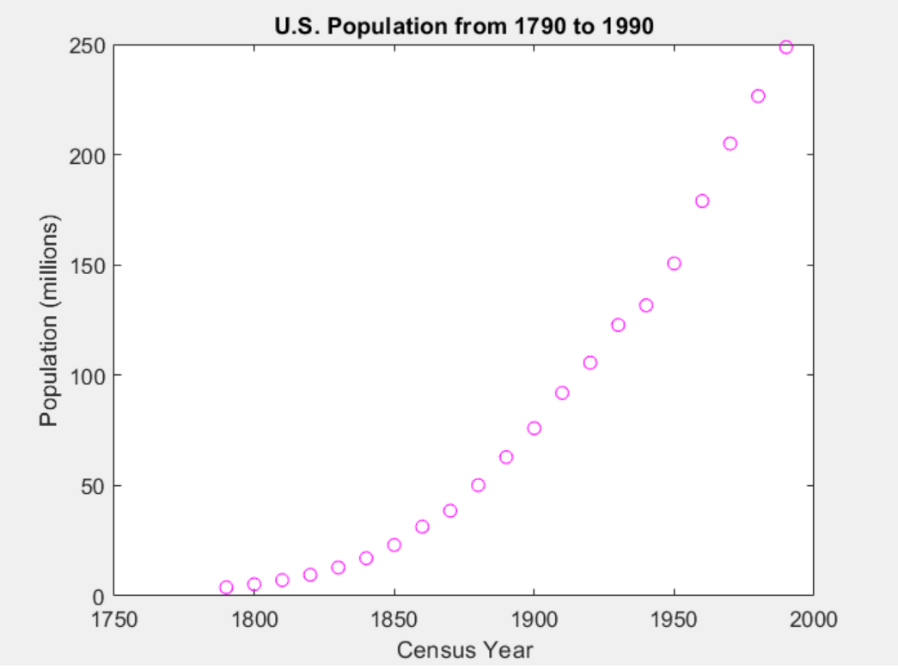
18120700,18624450].'\*1e-6;

四、分析过程与部分代码

1. 散点图绘制

1. plot(cdate,pop,'om')
2. title('U.S. Population from 1790 to 1990');
3. xlabel('Census Year');
4. ylabel('Population (millions)');

计算协方差矩阵

1. >> corrcoef(cdate,pop)
2. ans =
4. 1.0000    0.9597
5.       0.9597    1.0000

图一

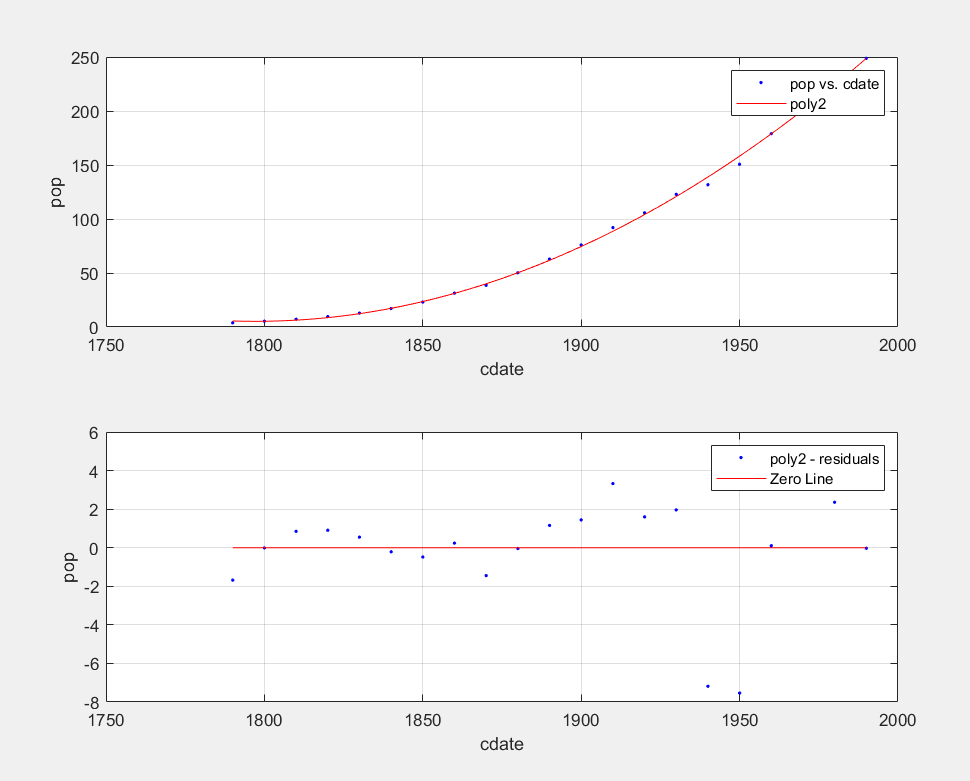
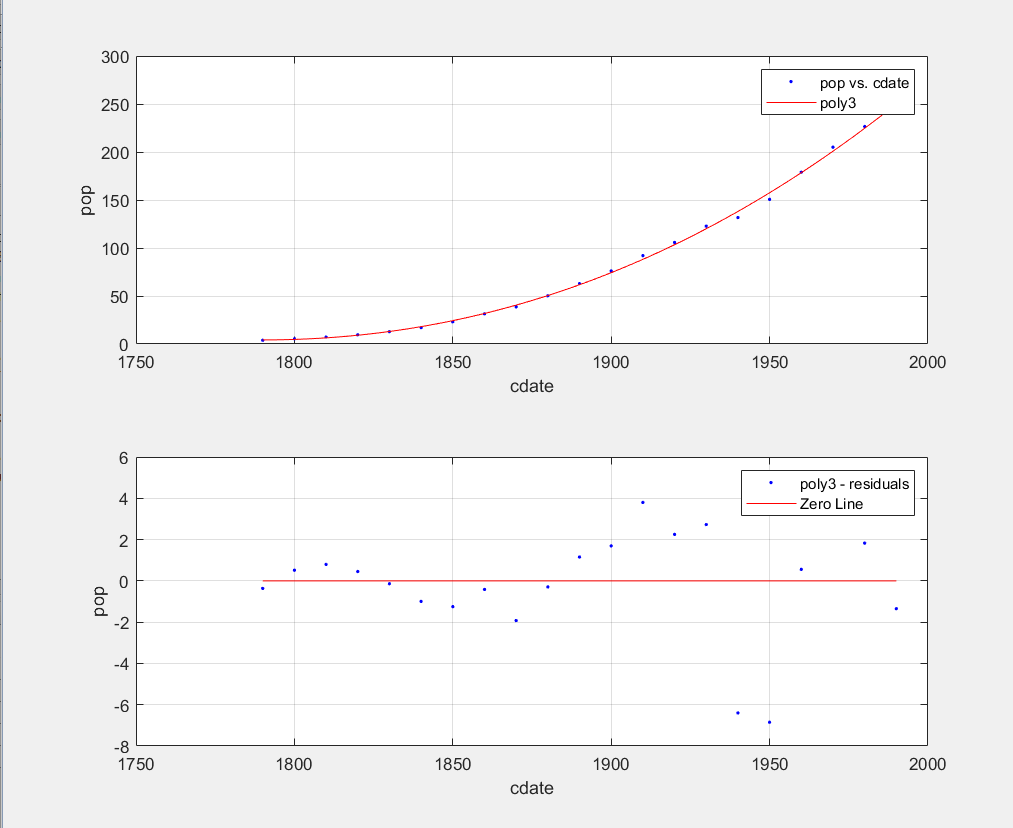
结论：

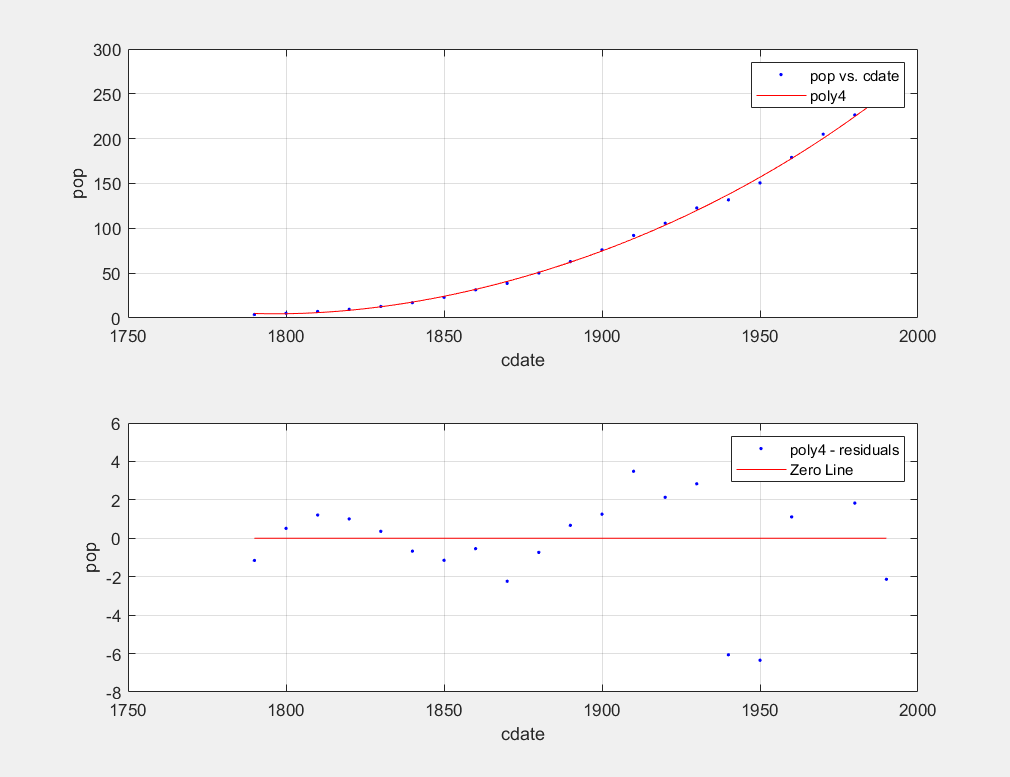
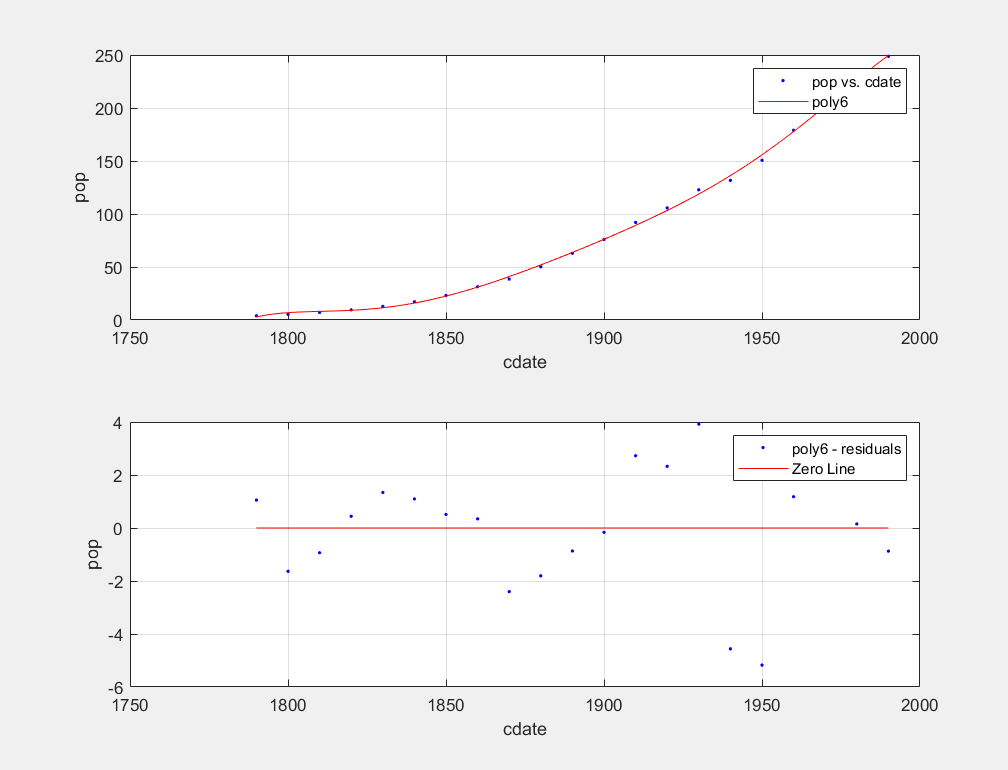
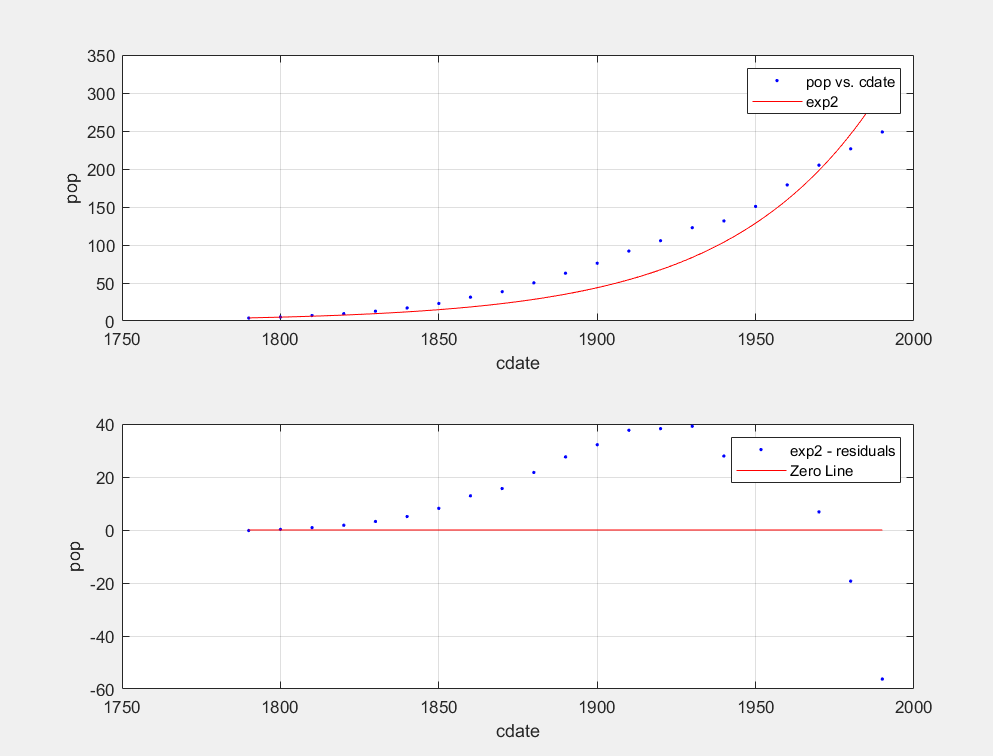
1. 从图一上可以直观地看出原始样本数据具有较强的相关性

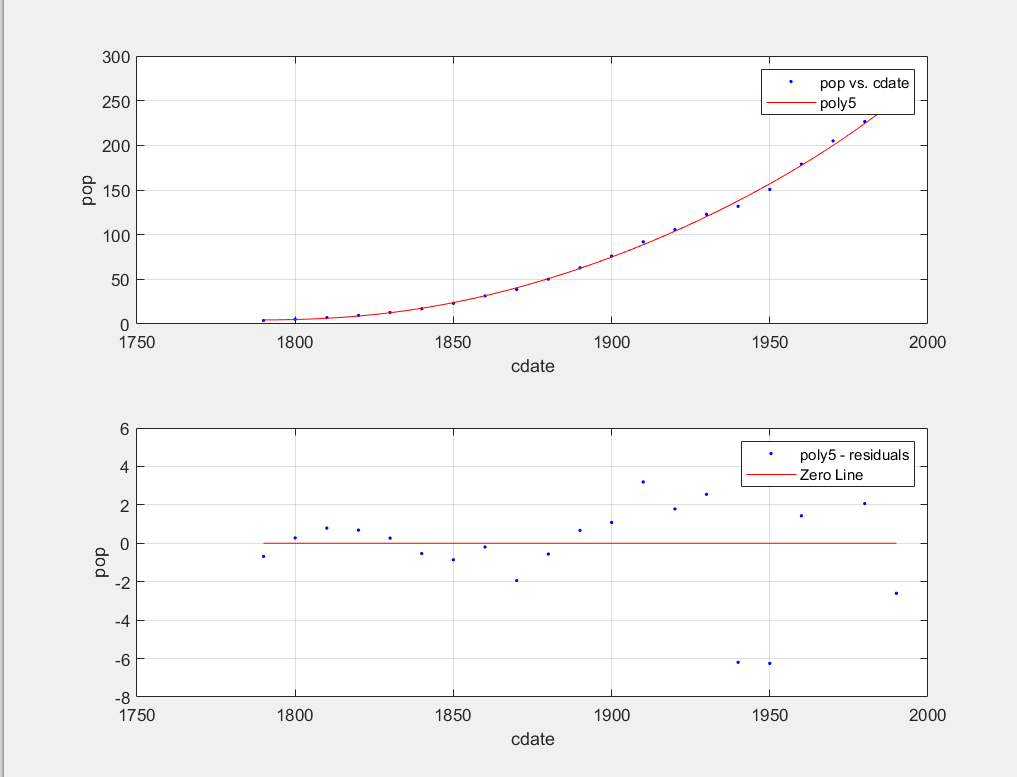
2. 协方差矩阵系数都接近于1,进一步证明数据的强相关性

2. 拟合曲线和残差图绘制

1. %% 初始化数据
2. %% Initialization.
4. % Initialize arrays to store fits and goodness-of-fit.
5. fitresult = cell( 2, 1 );
6. gof = **struct**( 'sse', cell( 2, 1 ), ...
7. 'rsquare', [], 'dfe', [], 'adjrsquare', [], 'rmse', [] );
8. %% 拟合曲线(注:这里只贴出poly2的源码)
9. %% Fit: 'poly2'.
10. [xData, yData] = prepareCurveData( cdate, pop );
12. % Set up fittype and options.
13. ft = fittype( 'poly2' );
15. % Fit model to data.
16. [fitresult{1}, gof(1)] = fit( xData, yData, ft, 'Normalize', 'on' );
18. % Create a figure **for** the plots.
19. figure( 'Name', 'poly2' );
21. % Plot fit with data.
22. subplot( 2, 1, 1 );
23. h = plot( fitresult{1}, xData, yData );
24. legend( h, 'pop vs. cdate', 'poly2', 'Location', 'NorthEast' );
25. % Label axes
26. xlabel cdate
27. ylabel pop
28. grid on
30. % Plot residuals.
31. subplot( 2, 1, 2 );
32. h = plot( fitresult{1}, xData, yData, 'residuals' );
33. legend( h, 'poly2 - residuals', 'Zero Line', 'Location', 'NorthEast' );
35. % Label axes
36. xlabel cdate
37. ylabel pop
38. grid on

poly2~poly6,exp2的拟合曲线及残差图





3. 数据分析