GDP增长趋势问题

摘 要

本文主要研究了多项式拟合及其应用，文中给出了多项式拟合原理和方法，通过对J省GDP等数据的拟合以及对未来的预测，建立了最小二乘法拟合高阶多项式的模型以及时间序列模型。

本文分为四个部分：第一，预备知识，给出了最小二乘法相关定义、定理，并证明了有关定理；第二，多项式拟合的基本原理.介绍了多项式的方法和原理，对高阶多项式拟合的G矩阵非奇异进行了讨论，证明了法方程有唯一解；第三，多项式拟合与程序实现，对J省的GDP增长趋势多项式拟合，建立数学模型，而后使用MATLAB实现模型，对得到的多项式进行检验，通过验证R方的值判断拟合效果，而后使用时间序列模型分析并对未来做出预测；第四，多项式拟合的应用，利用其对给出点数据拟合的优越性，分析J省第一产业近年来的增长趋势，直观显现并分析其未来增长趋势。

关键词： 最小二乘法；多项式拟合； 时间序列模型；MATLAB；

**GDP GROWTH TEWND ISSUE**

**ABSTRACT**

This article mainly studies polynomial fitting and its application. The principle and method of polynomial fitting are given in the article. Through the fitting of the GDP and other data of Province J and the prediction of the future, a model of the least squares method to fit high-order polynomials is established. And time series model.

This article is divided into four parts: first, preliminary knowledge, given the relevant definitions and theorems of the least squares method, and proved the relevant theorems; second, the basic principles of polynomial fitting. Introduces the methods and principles of polynomials. The non-singularity of the G matrix fitted by a polynomial of order is discussed, which proves that the normal equation has a unique solution; third, the polynomial fitting and program realization, the polynomial fitting of the GDP growth trend of J province, the establishment of a mathematical model, and then the realization of MATLAB Model, test the obtained polynomial, judge the fitting effect by verifying the value of R-square, and then use the time series model to analyze and predict the future; fourth, the application of polynomial fitting, use it to simulate the given point data Analyze the growth trend of the primary industry in Province J in recent years, visualize and analyze its future growth trend.

**Key words:** Least squares method; polynomial fitting; time series model; MATLAB;

目 录

1 预备知识........................................................1

1.1 相关定义....................................................1

1.2 相关定理....................................................1

2 多项式拟合的基本原理............................................2

3 多项式拟合与程序实现............................................3

3.1 最小二乘法建立模型..........................................3

3.1.1 GDP数据及处理...........................................3

3.1.2 基于数据多项式拟合.......................................3

3.2 MATLAB实现模型..............................................6

3.3 实验结果分析.................................................9

4 多项式拟合的应用.................................................9

5 课程设计的总结与体会............................................11

参考文献......................................................... 12

附录.............................................................13

1. 预备知识
   1. 相关定义

在函数的最佳平方逼近中，如果只在一组离散点集上给出，这就是科学实验中经常见到的实验数据的曲线拟合，这里，要求一个函数与所给的数据拟合，若记误差，，设是上线性无关函数族，在中找一函数，使误差平方和

 (1.1)

这里

 (1.2)

以上即为曲线拟合的最小二乘法

* 1. 相关定理

证明 的确是所求最小二乘解，可以通过哈尔条件。即设的任意线性组合在点集上至多只有n个不同的零点，则称在点集上满足哈尔条件。

1. 多项式拟合的基本原理

使用最小二乘法求拟合曲线时，首先要确定的形式，通常要从问题的运动规律或给定数据描图，确定的形式，并通过实际计算选出较好的结果。的一般表达式都为(1.1)式表达的线性形式，若是k次多项式，就是n次多项式，为使问题的提法更有一般性，通常在最小二乘法中都考虑为加权平方和

 (2.1)

这里是上的权函数，它表示不同点处的数据比重不同。用最小二乘法多项式拟合，就是在形如(1.2)式的中求一函数，使(2.1)式取得最小，转化为求多元函数的极小点问题，则根据求多元函数极值的必要条件有；

 (2.2)

若记

 (2.3)

(2.2)式可改写为

 (2.4)

线性方程组(2.4)称为法方程，可将其写成矩阵形式，其中

 (2.5)

要使法方程(2.4)有唯一解，则需要矩阵G非奇异，此处可引用上文哈尔条件证明

1. 多项式拟合与程序实现
   1. 最小二乘法建立模型
      1. GDP数据及处理

J省2000-2019年GDP数据如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 年份 | 地区生产总值（亿元） | 第一产业增加值 | 第二产业增加值 | 第三产业增加值 |
| 2000 | 1820 | 400 | 800 | 620 |
| 2001 | 2032.48 | 409.6 | 880.84 | 742.04 |
| 2002 | 2242.96 | 456.1 | 964.55 | 822.31 |
| 2003 | 2521.8 | 488.8 | 1140.8 | 892.2 |
| 2004 | 2958.21 | 560.96 | 1379.31 | 1017.94 |
| 2005 | 3614.92 | 607 | 1605.13 | 1402.79 |
| 2006 | 4249.23 | 686 | 1886.59 | 1676.64 |
| 2007 | 5226.08 | 813.48 | 2389.87 | 2022.73 |
| 2008 | 6424.06 | 916.7 | 3064.63 | 2442.73 |
| 2009 | 7203.18 | 980.5 | 3491.96 | 2730.72 |
| 2010 | 8577.06 | 1050.15 | 4417.39 | 3109.52 |
| 2011 | 10530.71 | 1277.4 | 5601.2 | 3652.11 |
| 2012 | 11937.82 | 1412.11 | 6374.45 | 4151.26 |
| 2013 | 12981.46 | 1509.34 | 6858.23 | 4613.81 |
| 2014 | 13803.81 | 1524.56 | 7287.26 | 4991.99 |
| 2015 | 14274.11 | 1596.28 | 7337.06 | 5340.77 |
| 2016 | 14886.23 | 1498.52 | 7147.18 | 6240.53 |
| 2017 | 15288.94 | 1429.21 | 7012.85 | 6846.88 |
| 2018 | 15074.62 | 1160.75 | 6410.85 | 7503.02 |
| 2019 | 11726.8 | 1287.32 | 4134.82 | 6304.68 |

表1 J省2000-2019年GDP数据表

由于需要在拟合后进行预测，而且年份数据起始值相较于变化范围过大，所以这里取可以显示出明显变化的(1,2,···,19)。

* + 1. 基于数据多项式拟合

要求考虑GDP近年来的发展趋势，则提取表1中GDP数据以及年份数据作为给定数据，通过最小二乘法建立模型。问题即转化为：

给定数据为，求一个n次多项式

 (3.1)

即求一组参数，使

 (3.2)

取最小值 。

将 利用极值必要条件，有

 (3.3)

 (3.4)

令，则上述线性方程组变为

 (3.5)

可以证明，当互异时，该方程组有唯一解，并是最小值问题的解。法方程组可写成以下形式：

 (3.6)

令

 (3.7)

则法方程系数矩阵为，常数项为

* 1. MATLAB实现模型

首先，根据原数据画出散点图，判断是否有明显关系。

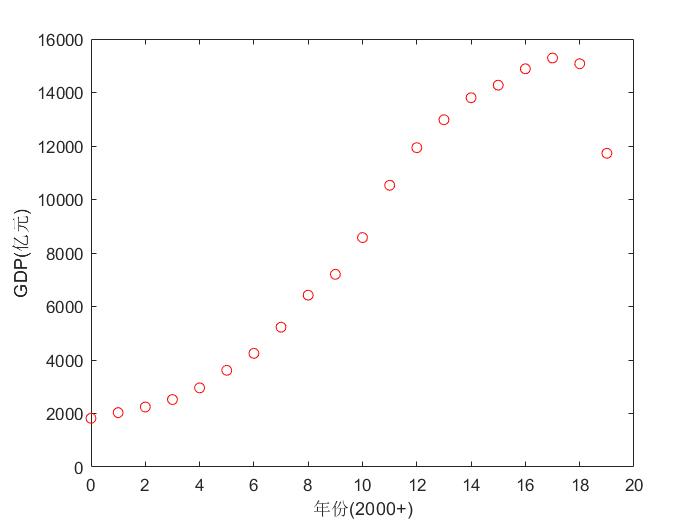


图1 GDP数据散点图

可见，数据并不是简单的线性，所以从二阶开始使用多项式拟合。根据以上模型在MATLAB中实现，同时应用MATLAB内置函数进行对比检验，结果如下：

二阶多项式拟合结果图如下（左为模型拟合，右为内置函数拟合，下同）：

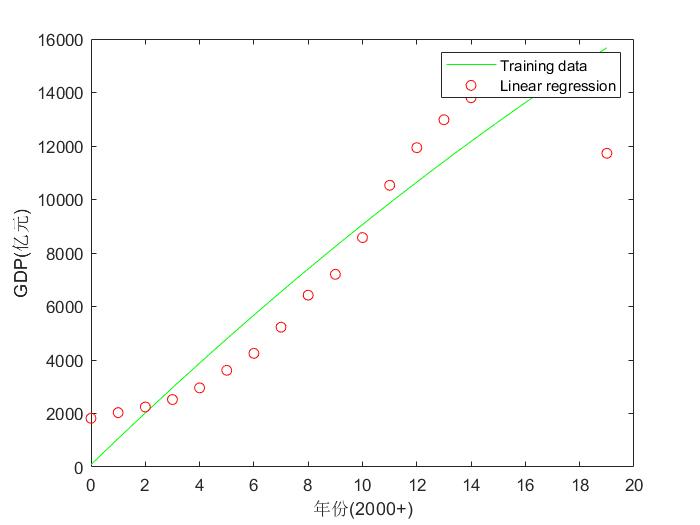
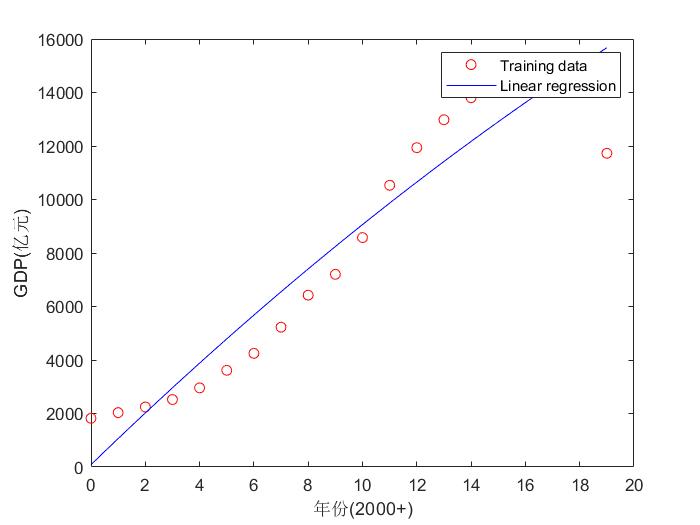


图2 二阶模型拟合 图3 二阶内置拟合

拟合多项式为：

三阶多项式拟合结果图如下：

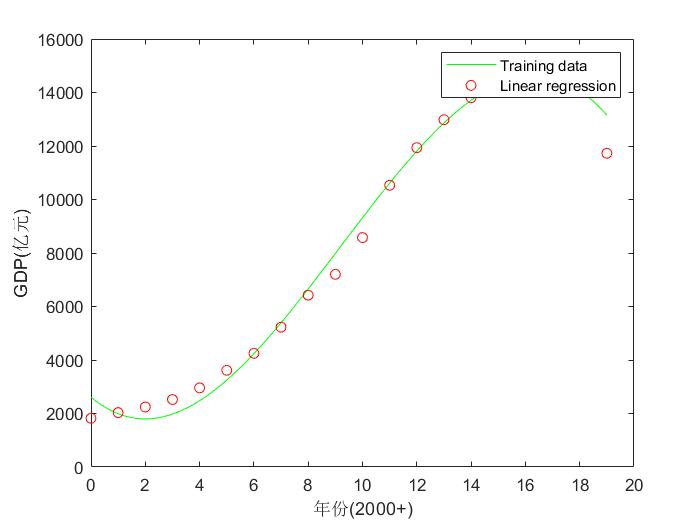
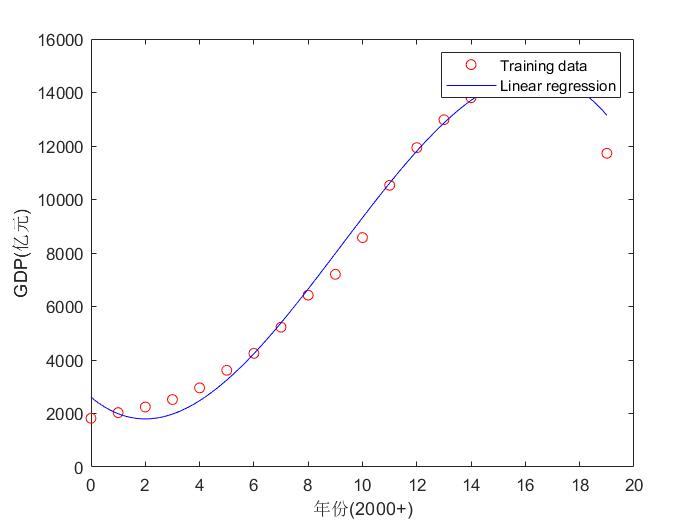


图3 三阶模型拟合 图4 三阶内置拟合

拟合多项式为：

四阶多项式拟合结果图如下：

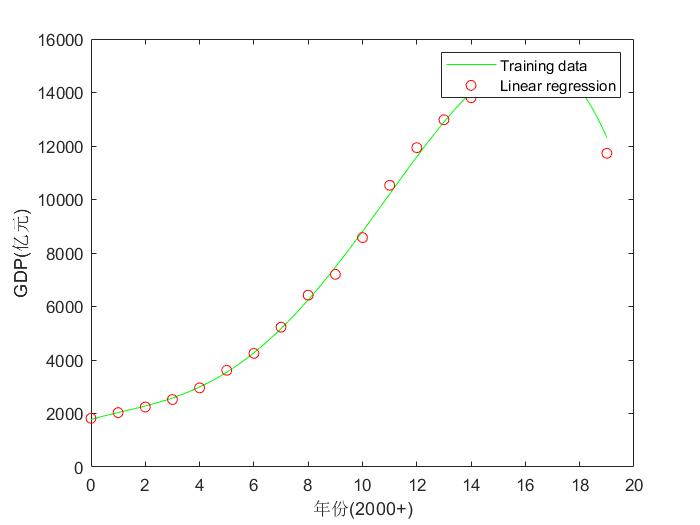
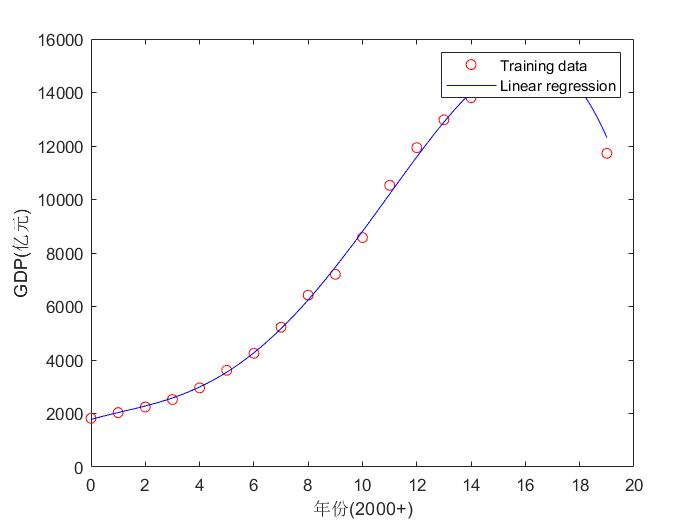


图4 四阶模型拟合 图5 四阶内置拟合

拟合多项式为：

五阶多项式拟合结果图如下：

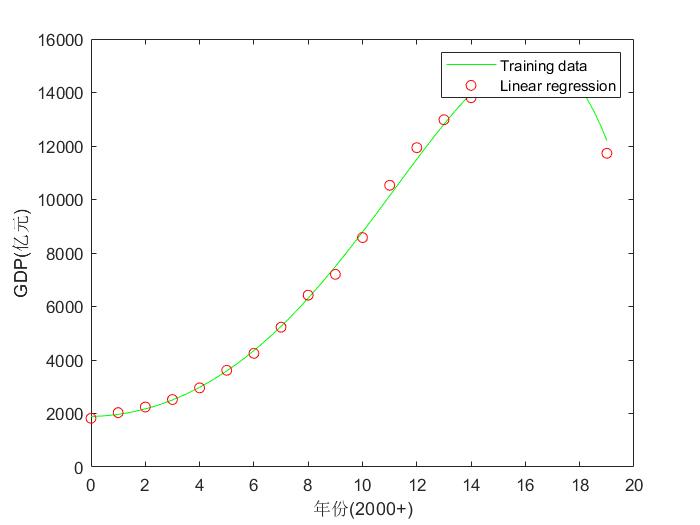
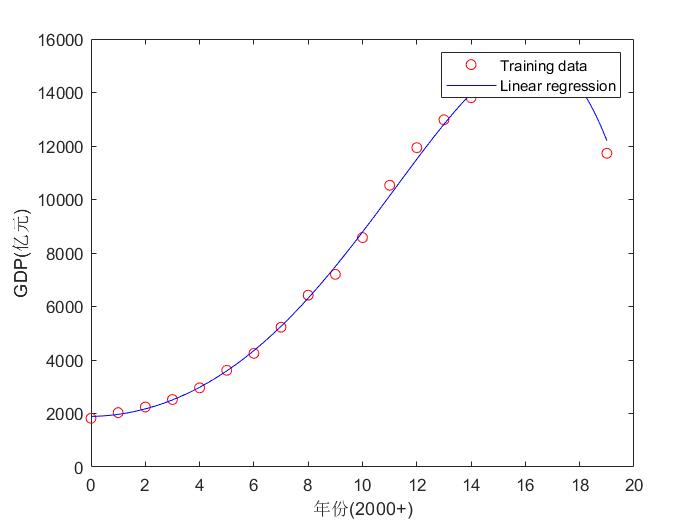


图6 五阶模型拟合 图7 五阶内置拟合

拟合多项式为：

此时，五次项的值已无限接近于0，故此不再继续。

对接下几年的J省GDP做预测，由于突发的现实原因，2019年GDP数据并不计算在预测内，建立时间序列模型预测，结果如下：

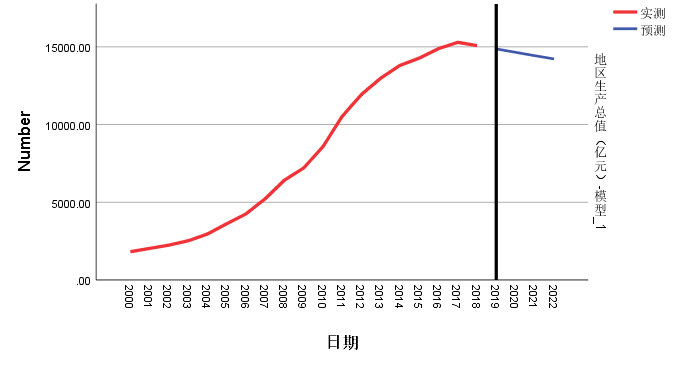


图7 时间预测模型

* 1. 实验结果分析

对得到的结果进行分析检验。如下：

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | |
| 因变量: 地区生产总值（亿元） | | | | | | | | | |
| 方程 | 模型摘要 | | | | | 参数估算值 | | | |
| R 方 | F | 自由度 1 | 自由度 2 | 显著性 | 常量 | b1 | b2 | b3 |
| 二次 | .919 | 96.165 | 2 | 17 | .000 | -903.961 | 1000.276 | -8.573 |  |
| 三次 | .987 | 400.110 | 3 | 16 | .000 | 3706.987 | -1352.497 | 264.803 | -8.679 |

表2 模型摘要和参数估算值

当判定系数R方越大时，拟合效果越好。

1. 多项式拟合的应用

由上可知，J省GDP由于现实原因影响2019年整体下滑，并不是一个很好的分析对象，所以可以选择第一产业这个受影响较小的局部数据来分析。数据如下：

|  |  |
| --- | --- |
| 年份 | 第一产业增加值 |
| 2000 | 400 |
| 2001 | 409.6 |
| 2002 | 456.1 |
| 2003 | 488.8 |
| 2004 | 560.96 |
| 2005 | 607 |
| 2006 | 686 |
| 2007 | 813.48 |
| 2008 | 916.7 |
| 2009 | 980.5 |
| 2010 | 1050.15 |
| 2011 | 1277.4 |
| 2012 | 1412.11 |
| 2013 | 1509.34 |
| 2014 | 1524.56 |
| 2015 | 1596.28 |
| 2016 | 1498.52 |
| 2017 | 1429.21 |
| 2018 | 1160.75 |
| 2019 | 1287.32 |

表3 第一产业数据表

模型构建采用上文中J省GDP趋势模型，MATLAB拟合结果如下：

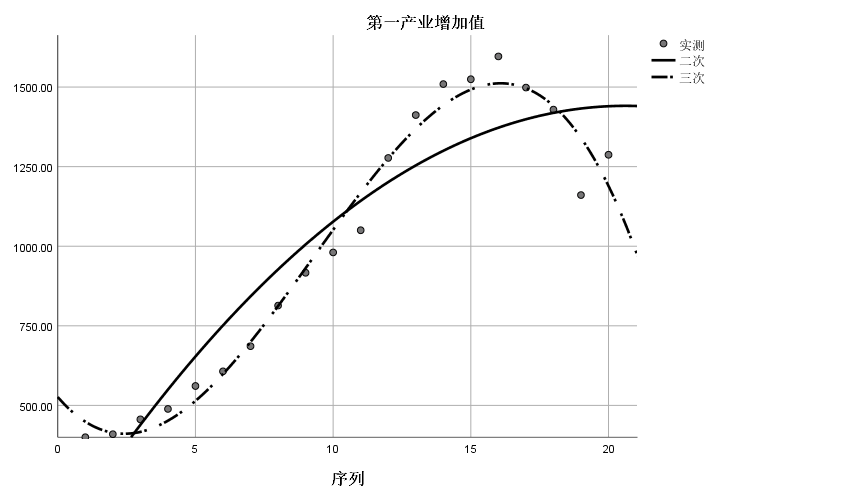


图8 J省第一产业趋势图

二阶多项式拟合为：

三阶多项式拟合为：

四阶为：

检验如下：

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | |
| 因变量: 第一产业增加值 | | | | | | | | | |
| 方程 | 模型摘要 | | | | | 参数估算值 | | | |
| R 方 | F | 自由度 1 | 自由度 2 | 显著性 | 常量 | b1 | b2 | b3 |
| 二次 | .880 | 62.471 | 2 | 17 | .000 | 68.299 | 133.208 | -3.232 |  |
| 三次 | .976 | 214.135 | 3 | 16 | .000 | 526.237 | -100.459 | 23.919 | -.862 |

表3 模型摘要和参数估算值

由R方可见拟合效果良好

1. 课程设计的总结与体会

通过此次课程设计，更加了解并掌握了最小二乘法。最小二乘法是指使因变量估计值与实测值间的相对误差平方和为最小。在研究两个变量之间的关系时，可以用回归分析的方法进行分析。当确定了描述两个变量之间的回归模型后，就可以使用最小二乘法估计模型中的参数，进而建立数学模型，然后通过MATLAB求解模型。通过本文实例模型的求解，我们学会了怎样对一组数据进行多种多项式拟合，也对MATLAB内置函数ployfit有了更深刻的理解，对今后的学习生活有很大帮助

参 考 文 献

[1] 李庆扬，王能超，易大义编．数值分析（第5版）[M]：华中科技大学出版社，2018：13-24.

[2] 姜启源，数学模型（第三版），高等教育出版社，2003.

[3] 刘宏，邱建雄，谢中科．C++程序设计教程．武汉大学出版社2005年7月.

[4] 何坚勇编著. 运筹学基础（第二版）[M]. 北京：清华大学出版社，2008.

[5] 蒋加伏．大学计算机基础．第3版．北京邮电大学出版社，2006年.

[6] 张志涌　等编著,MATLAB教程R2012a，北京航空航天大学出版社，2010年.

[7] 卜金友. 吉林省2019年国民经济和社会发展统计公报[EB/OL]. http://www.jl.gov.cn/sj/sjcx/ndbg/tjgb/202004/t20200403\_7024695.html, 2020-04-03/2020-12-28.

附 录

main.m

clc;clear

% format rat

[num,txt] = xlsread('.\GDP\_data.xlsx');

year = num(:,1);

GDP = num(:,2); % 国内生产总值

industry\_1 = num(:,3); % 第一产业

industry\_2 = num(:,4); % 第二产业

industry\_3 = num(:,5); % 第三产业

per\_capita = num(:,6); % 人均收入

[p,p\_re] = ploy\_fit(year,industry\_1,4);

p

p\_re

ploy\_fit.m

function [p,p\_re] = ploy\_fit(x,y,n)

% 多项式拟合函数，x为自变量，y为因变量，n为阶数

x = x-2000; % 数据处理，此处年份太大影响拟合效果

dx = min(x):0.01:max(x); % 步长

% 内置函数对比检验

p = polyfit(x,y,n);

dy\_1 = polyval(p,dx);

plot(dx,dy\_1,'g-');

hold on;

plot(x,y,'ro')

legend('Training data','Linear regression');

xlabel('年份(2000+)');

ylabel('GDP(亿元)');

% 最小二乘法多项式拟合

F = zeros(n+1,length(x));

F(1,:) = 1;

for i = 2:n+1

for j = 1:length(x)

F(i,j) = x(j)^(i-1);

end

end

F=F\*F';

[m,~] = size(F);

Y = zeros(m,1);

Y(1) = sum(y);

for i = 2:m

for j = 1:length(y)

Y(i) = Y(i)+y(j)\*x(j)^(i-1);

end

end

p\_re = F\Y;

p\_re = p\_re(end:-1:1)'; %数组反序

figure;

plot(x,y,'ro')

hold on;

dy\_2 = polyval(p\_re,dx);

plot(dx,dy\_2,'b-')

legend('Training data','Linear regression');

xlabel('年份(2000+)');

ylabel('GDP(亿元)');