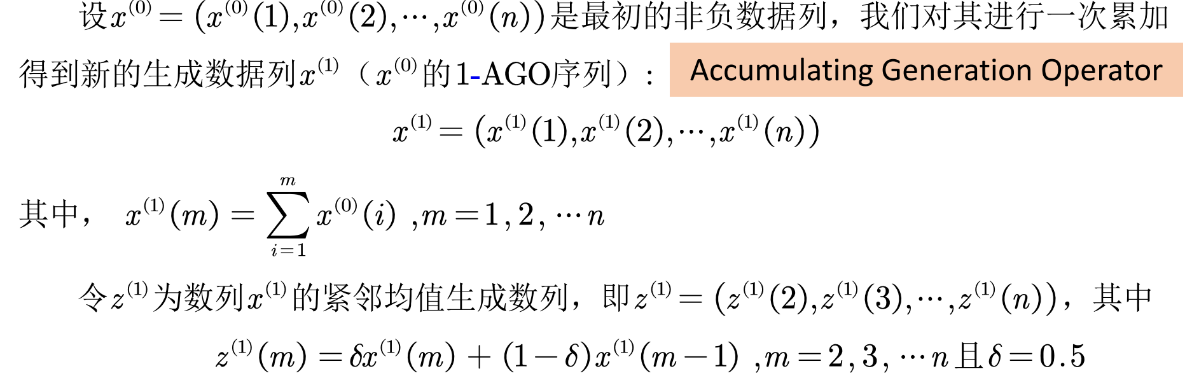
## 问题求解—数学理论

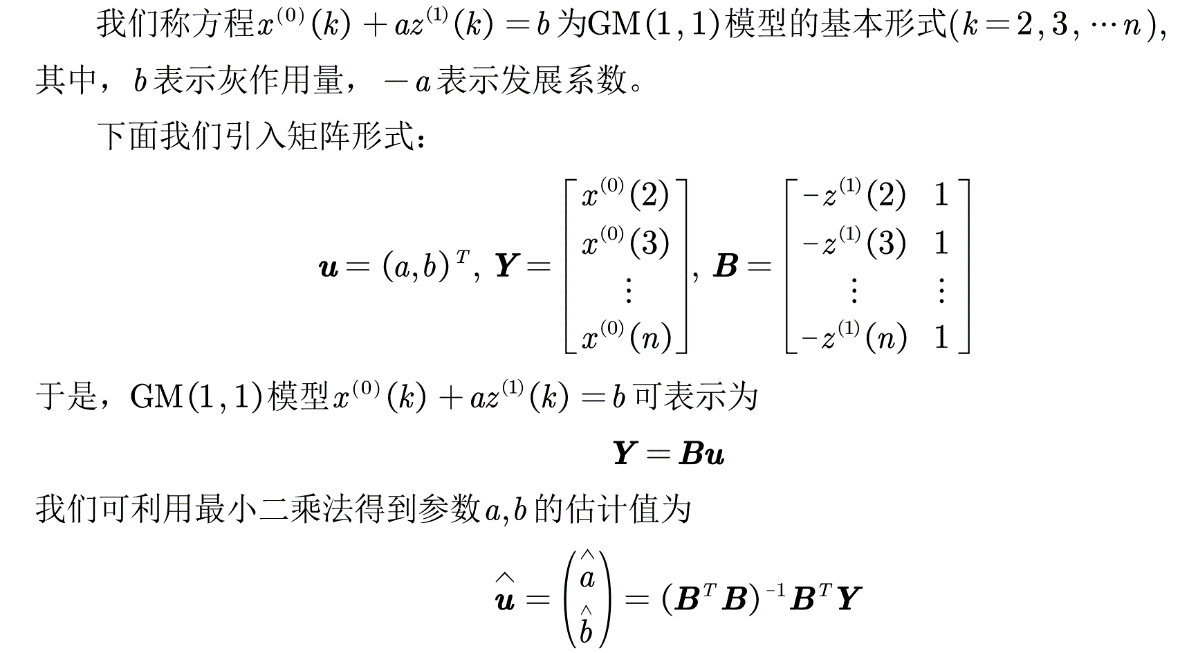
观察数据，根据其特性，我们可以利用灰色预测进行预测。

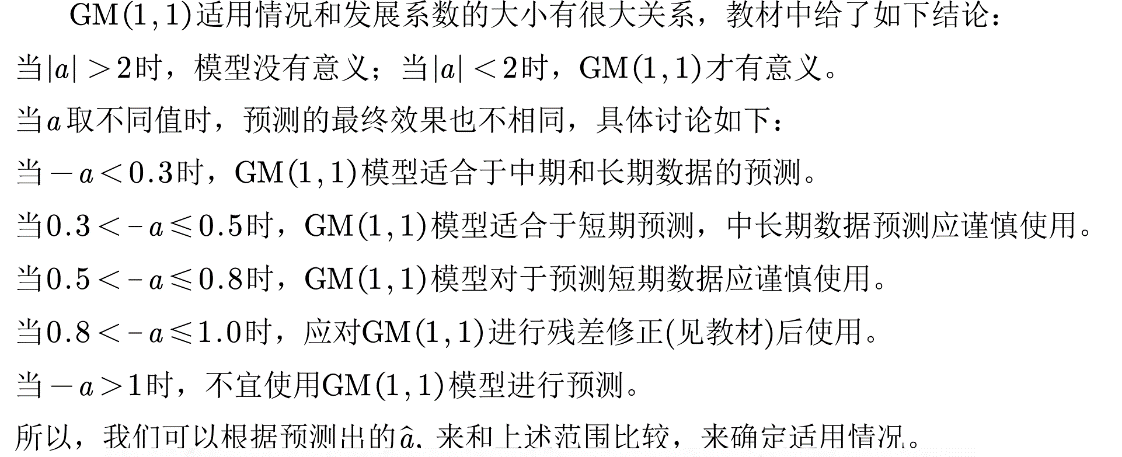
数学理论：

灰色预测是对既含有已知信息又含有不确定信息的系统进行预测，就是对在一定范围内变化的、与时间有关的灰色过程进行预测。灰色预测对原始数据进行生成处理来寻找系统变动的规律，并生成有较强规律性的数据序列，然后建立相应的微分方程模型，从而预测事物未来发展趋势的状况。

GM(1,1)是使用原始的离散非负数据列，通过一次累加生成削弱随机性的较有规律的新的离散数据列，然后通过建立微分方程模型，得到在离散点处的解经过累减生成的原始数据的近似估计值，从而预测原始数据的后续发展。







## 问题求解—程序分析与设计

1. 开发环境说明

Matlab

1. 系统分析

创建了main函数，gm11函数，metabolism\_gm11函数，new\_gm11函数。

（1）main函数的功能是加载数据以及调用gm11，metabolism\_gm11，new\_gm11函数。无参数与返回值，属于执行函数。

（2）gm11函数作用：使用传统的GM(1,1)模型对数据进行预测

% x0：要预测的原始数据

% predict\_num： 向后预测的期数

% 输出变量 （注意，实际调用时该函数时不一定输出全部结果，就像corrcoef函数一样~，可以只输出相关系数矩阵，也可以附带输出p值矩阵）

% result：预测值

% x0\_hat：对原始数据的拟合值

% relative\_residuals： 对模型进行评价时计算得到的相对残差

% eta： 对模型进行评价时计算得到的级比偏差

（3）metabolism\_gm11函数作用：使用新陈代谢的GM(1,1)模型对数据进行预测

% 输入变量

% x0：要预测的原始数据

% predict\_num： 向后预测的期数

% 输出变量

% result：预测值

（4）new\_gm11函数作用：使用新信息的GM(1,1)模型对数据进行预测

% 输入变量

% x0：要预测的原始数据

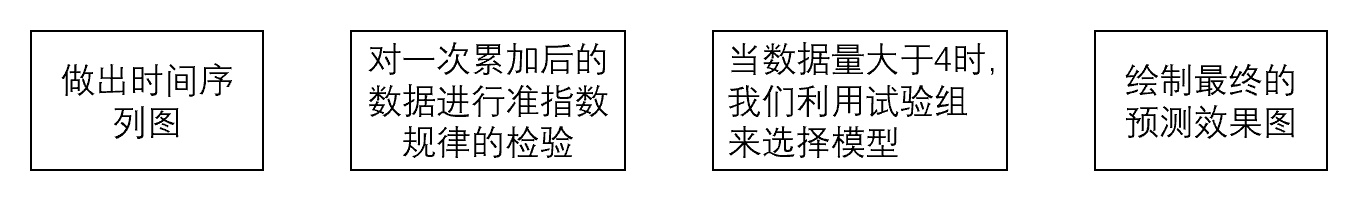
% predict\_num： 向后预测的期数

% 输出变量

% result：预测值

1. 函数设计

main流程：



1. %%  输入原始数据并做出时间序列图
2. clear;clc
3. year =[2010:2020]';  % 横坐标表示term，写成列向量的形式（加'就表示转置）
4. x0 = [412119.3  487940.2    538580  592963.2    643563.1    688858.2    746395.1    832035.9    919281.1    986515.2    1015986.2]';  %原始数据序列，写成列向量的形式（加'就表示转置）
5. % year = [2009:2015]; % 其实本程序写成了行向量也可以，因为我怕你们真的这么写了，所以在后面会有判断。
6. % x0 = [730, 679, 632, 599, 589, 532, 511];
7. % year = [2010:2017]';   % 该数据很特殊，可以通过准指数规律检验，但是预测效果却很差
8. % x0 = [1.321,0.387,0.651,0.985,1.235,0.987,0.854,1.021]';
9. % year = [2014:2017]';
10. % x0 = [2.874,3.278,3.337,3.390]';
12. % 画出原始数据的时间序列图
13. figure(1); % 因为我们的图形不止一个，因此要设置编号
14. plot(year,x0,'o-'); grid on;  % 原式数据的时间序列图
15. set(gca,'xtick',year(1:1:end))  % 设置x轴横坐标的间隔为1
16. xlabel('term');  ylabel('gpa');  % 给坐标轴加上标签
17. %% 对一次累加后的数据进行准指数规律的检验(注意，这个检验有时候即使能通过，也不一定能保证预测结果非常好，例如上面的第三组数据)
18. **if** ERROR == 0   % 如果上述错误均没有发生时，才能执行下面的操作步骤
19. disp('------------------------------------------------------------')
20. disp('准指数规律检验')
21. x1 = cumsum(x0);   % 生成1-AGO序列，cumsum是累加函数哦~    注意：1.0e+03 \*0.1740的意思是科学计数法,即10^3\*0.1740 = 174
22. rho = x0(2:end) ./ x1(1:end-1) ;   % 计算光滑度rho(k) = x0(k)/x1(k-1)
24. % 画出光滑度的图形，并画上0.5的直线，表示临界值
25. figure(2)
26. plot(year(2:end),rho,'o-',[year(2),year(end)],[0.5,0.5],'-'); grid on;
27. text(year(end-1)+0.2,0.55,'临界线')   % 在坐标(year(end-1)+0.2,0.55)上添加文本
28. set(gca,'xtick',year(2:1:end))  % 设置x轴横坐标的间隔为1
29. xlabel('term');  ylabel('原始数据的光滑度');  % 给坐标轴加上标签

32. disp(strcat('指标1：光滑比小于0.5的数据占比为',num2str(100\*sum(rho<0.5)/(n-1)),'%'))
33. disp(strcat('指标2：除去前两个时期外，光滑比小于0.5的数据占比为',num2str(100\*sum(rho(3:end)<0.5)/(n-3)),'%'))
34. disp('参考标准：指标1一般要大于60%, 指标2要大于90%，你认为本例数据可以通过检验吗？')
36. Judge = input('你认为可以通过准指数规律的检验吗？可以通过请输入1，不能请输入0：');
37. **if** Judge == 0
38. disp('亲，灰色预测模型不适合你的数据哦~ 请考虑其他方法吧 例如ARIMA，指数平滑等')
39. ERROR = 1;
40. end
41. disp('------------------------------------------------------------')
42. end
43. %% 当数据量大于4时，我们利用试验组来选择使用传统的GM(1,1)模型、新信息GM(1,1)模型还是新陈代谢GM(1,1)模型； 如果数据量等于4，那么我们直接对三种方法求一个平均来进行预测
44. **if** ERROR == 0   % 如果上述错误均没有发生时，才能执行下面的操作步骤
45. **if**  n > 4  % 数据量大于4时，将数据分为训练组和试验组(根据原数据量大小n来取，n为5-7个则取最后两年为试验组，n大于7则取最后三年为试验组)
46. disp('因为原数据的期数大于4，所以我们可以将数据组分为训练组和试验组')   % 注意，如果试验组的个数只有1个，那么三种模型的结果完全相同，因此至少要取2个试验组
47. **if** n > 7
48. test\_num = 3;
49. **else**
50. test\_num = 2;
51. end
52. train\_x0 = x0(1:end-test\_num);  % 训练数据
53. disp('训练数据是: ')
54. disp(mat2str(train\_x0'))  % mat2str可以将矩阵或者向量转换为字符串显示, 这里加一撇表示转置，把列向量变成行向量方便观看
55. test\_x0 =  x0(end-test\_num+1:end); % 试验数据
56. disp('试验数据是: ')
57. disp(mat2str(test\_x0'))  % mat2str可以将矩阵或者向量转换为字符串显示
58. disp('------------------------------------------------------------')
60. % 使用三种模型对训练数据进行训练，返回的result就是往后预测test\_num期的数据
61. disp(' ')
62. disp('\*\*\*下面是传统的GM(1,1)模型预测的详细过程\*\*\*')
63. result1 = gm11(train\_x0, test\_num); %使用传统的GM(1,1)模型对训练数据，并预测后test\_num期的结果
64. disp(' ')
65. disp('\*\*\*下面是进行新信息的GM(1,1)模型预测的详细过程\*\*\*')
66. result2 = new\_gm11(train\_x0, test\_num); %使用新信息GM(1,1)模型对训练数据，并预测后test\_num期的结果
67. disp(' ')
68. disp('\*\*\*下面是进行新陈代谢的GM(1,1)模型预测的详细过程\*\*\*')
69. result3 = metabolism\_gm11(train\_x0, test\_num); %使用新陈代谢GM(1,1)模型对训练数据，并预测后test\_num期的结果
71. % 现在比较三种模型对于试验数据的预测结果
72. disp(' ')
73. disp('------------------------------------------------------------')
74. % 绘制对试验数据进行预测的图形（对于部分数据，可能三条直线预测的结果非常接近）
75. test\_year = year(end-test\_num+1:end);  % 试验组对应的term
76. figure(3)
77. plot(test\_year,test\_x0,'o-',test\_year,result1,'\*-',test\_year,result2,'+-',test\_year,result3,'x-'); grid on;
78. set(gca,'xtick',year(end-test\_num+1): 1 :year(end))  % 设置x轴横坐标的间隔为1
79. legend('real','traditional GM(1,1)','**new** inf GM(1,1)','metabolism GM(1,1)')
80. % 注意：如果lengend挡着了图形中的直线，那么lengend的位置可以自己手动拖动
81. %'试验组的真实数据','传统GM(1,1)预测结果','新信息GM(1,1)预测结果','新陈代谢GM(1,1)预测结果'
82. xlabel('term');  ylabel('gpa');  % 给坐标轴加上标签
83. % 计算误差平方和SSE
84. SSE1 = sum((test\_x0-result1).^2);
85. SSE2 = sum((test\_x0-result2).^2);
86. SSE3 = sum((test\_x0-result3).^2);
87. disp(strcat('传统GM(1,1)对于试验组预测的误差平方和为',num2str(SSE1)))
88. disp(strcat('新信息GM(1,1)对于试验组预测的误差平方和为',num2str(SSE2)))
89. disp(strcat('新陈代谢GM(1,1)对于试验组预测的误差平方和为',num2str(SSE3)))
90. **if** SSE1<SSE2
91. **if** SSE1<SSE3
92. choose = 1;  % SSE1最小，选择传统GM(1,1)模型
93. **else**
94. choose = 3;  % SSE3最小，选择新陈代谢GM(1,1)模型
95. end
96. elseif SSE2<SSE3
97. choose = 2;  % SSE2最小，选择新信息GM(1,1)模型
98. **else**
99. choose = 3;  % SSE3最小，选择新陈代谢GM(1,1)模型
100. end
101. Model = {'传统GM(1,1)模型','新信息GM(1,1)模型','新陈代谢GM(1,1)模型'};
102. disp(strcat('因为',Model(choose),'的误差平方和最小，所以我们应该选择其进行预测'))
103. disp('------------------------------------------------------------')
104. %% 绘制最终的预测效果图
105. figure(5)  % 下面绘图中的符号m:洋红色 b:蓝色
106. plot(year,x0,'-o',  year,x0\_hat,'-\*m',  year(end)+1:year(end)+predict\_num,result,'-\*b' );   grid on;
107. hold on;
108. plot([year(end),year(end)+1],[x0(end),result(1)],'-\*b')
109. legend('basic data','fitted data','predicted data')  % 注意：如果lengend挡着了图形中的直线，那么lengend的位置可以自己手动拖动
110. %原始数据，拟合数据，预测数据
111. set(gca,'xtick',[year(1):1:year(end)+predict\_num])  % 设置x轴横坐标的间隔为1
112. xlabel('term');  ylabel('gpa');  % 给坐标轴加上标签

gm\_11流程

1. function [result, x0\_hat, relative\_residuals, eta] = gm11(x0, predict\_num)
2. % 函数作用：使用传统的GM(1,1)模型对数据进行预测
3. %     x0：要预测的原始数据
4. %     predict\_num： 向后预测的期数
5. % 输出变量 （注意，实际调用时该函数时不一定输出全部结果，就像corrcoef函数一样~，可以只输出相关系数矩阵，也可以附带输出p值矩阵）
6. %     result：预测值
7. %     x0\_hat：对原始数据的拟合值
8. %     relative\_residuals： 对模型进行评价时计算得到的相对残差
9. %     eta： 对模型进行评价时计算得到的级比偏差
11. n = length(x0); % 数据的长度
12. x1=cumsum(x0); % 计算一次累加值
13. z1 = (x1(1:end-1) + x1(2:end)) / 2;  % 计算紧邻均值生成数列（长度为n-1）
14. % 将从第二项开始的x0当成y，z1当成x，来进行一元回归  y = kx +b
15. y = x0(2:end); x = z1;
16. % 下面的表达式就是第四讲拟合里面的哦~ 当是要注意，此时的样本数应该是n-1，少了一项哦
17. k = ((n-1)\*sum(x.\*y)-sum(x)\*sum(y))/((n-1)\*sum(x.\*x)-sum(x)\*sum(x));
18. b = (sum(x.\*x)\*sum(y)-sum(x)\*sum(x.\*y))/((n-1)\*sum(x.\*x)-sum(x)\*sum(x));
19. a = -k;  %注意：k = -a哦
20. % 注意： -a就是发展系数,  b就是灰作用量
22. disp('现在进行GM(1,1)预测的原始数据是: ')
23. disp(mat2str(x0'))  % mat2str可以将矩阵或者向量转换为字符串显示
24. disp(strcat('最小二乘法拟合得到的发展系数为',num2str(-a),'，灰作用量是',num2str(b)))
25. disp('\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*分割线\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*')
26. x0\_hat=zeros(n,1);  x0\_hat(1)=x0(1);   % x0\_hat向量用来存储对x0序列的拟合值，这里先进行初始化
27. **for** m = 1: n-1
28. x0\_hat(m+1) = (1-exp(a))\*(x0(1)-b/a)\*exp(-a\*m);
29. end
30. result = zeros(predict\_num,1);  % 初始化用来保存预测值的向量
31. **for** i = 1: predict\_num
32. result(i) = (1-exp(a))\*(x0(1)-b/a)\*exp(-a\*(n+i-1)); % 带入公式直接计算
33. end
35. % 计算绝对残差和相对残差
36. absolute\_residuals = x0(2:end) - x0\_hat(2:end);   % 从第二项开始计算绝对残差，因为第一项是相同的
37. relative\_residuals = abs(absolute\_residuals) ./ x0(2:end);  % 计算相对残差，注意分子要加绝对值，而且要使用点除
38. % 计算级比和级比偏差
39. class\_ratio = x0(2:end) ./ x0(1:end-1) ;  % 计算级比 sigma(k) = x0(k)/x0(k-1)
40. eta = abs(1-(1-0.5\*a)/(1+0.5\*a)\*(1./class\_ratio));  % 计算级比偏差
41. end

metabolism\_gm11:

1. function [result] = metabolism\_gm11(x0, predict\_num)
2. % 函数作用：使用新陈代谢的GM(1,1)模型对数据进行预测
3. % 输入变量
4. %     x0：要预测的原始数据
5. %     predict\_num： 向后预测的期数
6. % 输出变量
7. %     result：预测值
8. result = zeros(predict\_num,1);  % 初始化用来保存预测值的向量
9. **for** i = 1 : predict\_num
10. result(i) = gm11(x0, 1);  % 将预测一期的结果保存到result中
11. x0 = [x0(2:end); result(i)];  % 更新x0向量，此时x0多了新的预测信息，并且删除了最开始的那个向量
12. end
13. end

new\_gm11:

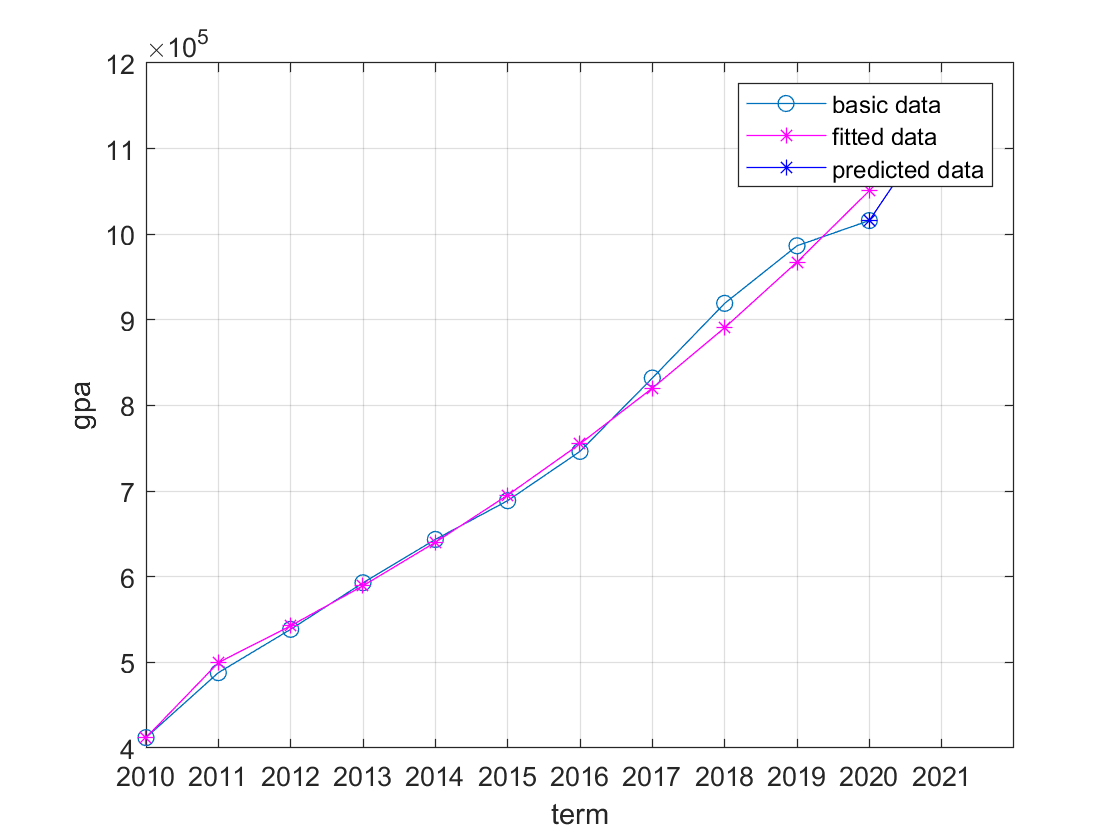
1. function [result] = new\_gm11(x0, predict\_num)
2. % 函数作用：使用新信息的GM(1,1)模型对数据进行预测
3. % 输入变量
4. %     x0：要预测的原始数据
5. %     predict\_num： 向后预测的期数
6. % 输出变量
7. %     result：预测值
8. result = zeros(predict\_num,1);  % 初始化用来保存预测值的向量
9. **for** i = 1 : predict\_num
10. result(i) = gm11(x0, 1);  % 将预测一期的结果保存到result中
11. x0 = [x0; result(i)];  % 更新x0向量，此时x0多了新的预测信息
12. end
13. end

## 问题求解结果

1. 程序测试（即程序操作流程）

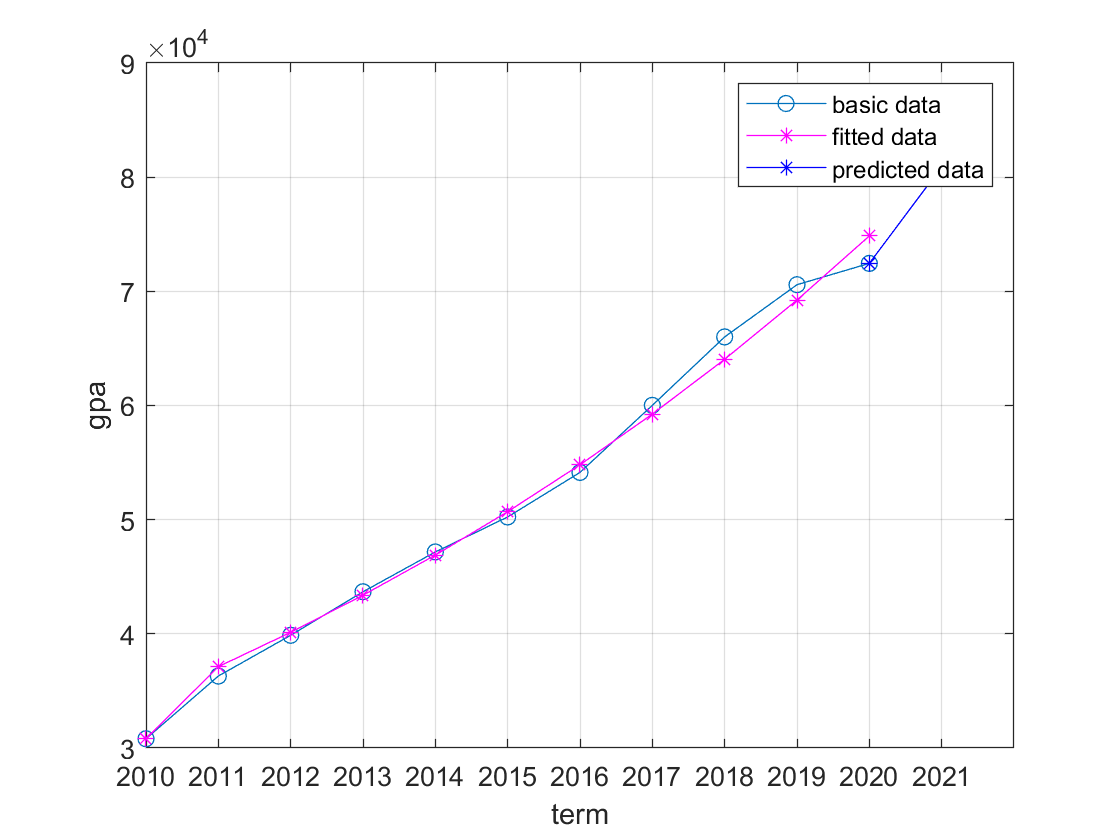
运行main即可，注意x0数据的替换，预测国内生产总值和人均国内生产总值时要修改x0的数据，year数据不用改变。

1. 结果分析（即问题求解结果分析）



往后预测1期的得到的结果：

2021 国内生产总值预测值：1141311.4748亿元。



往后预测1期的得到的结果：

2021人均国内生产总值预测值为 ：80915.0621元

1. 结论

（1）利用年份作为自变量去拟合人均国内生产总值的能力较强，从拟合平面图可知预测效果不错。

（2）利用年份作为自变量拟合国内生产总值的能力较强，从拟合平面图可知，预测曲线与原样本点相近。