  通过中国国家统计局和快易理财网的数据，共收集到九组数据，分别是中国每年出生人口、总人口统计(万/人)、人均GPA(按照美元算)、性别比例(按照女生等于100人计算)、人口自然增长率、城镇人口和乡村人口(两者加起来等于100)、美元兑换人民币汇率、总就业人口(万/人)。

我采用了BP 神经网络模型来预测中国每年的出生人口数据，表格中(详见“人口.xlsx”)每一行，分别利用总人口、人均GPA、性别比例、自然增长率、城镇人口、乡村人口、美元兑换人民币汇率、就业人口共八列数据作为特征值，来影响中国每年出生人口的数量。

首先对八个特征值近10年-20年的数据进行分析，得到两种拟合的方向:

1、初步分析呈线性的数据:

包括“人口”表中的总人口、人均GPA、性别比例、城镇人口、乡村人口共5列的数据。我通过灰色模型GM(1,1)来对这五项数据进行预测，目的是得到未知的2020年的5项人口特征值。图表在GM(1,1)模型求解中展示

通过以上的分析，可以得到近似地拟合得到2020年的总人口数为140908万人，人均GDP为11158美元，男女比例为104.58:100，城乡村人口比例为61.65:38.35。之所以采用近十年来预测，是因为近十年我国在科技、军事、文化等方面飞速发展，综合国力和文化软实力显著提升，采用近十年的数据能较好地反应出中国在这一段时间内数据的总体趋势。

x = xlsread('人口.xlsx',1,'A2:A54')';

y2 = xlsread('人口.xlsx',1,'C2:C54');

y3 = xlsread('人口.xlsx',1,'D2:D54');

y4 = xlsread('人口.xlsx',1,'E2:E54');

y5 = xlsread('人口.xlsx',1,'F2:F54');

y6 = xlsread('人口.xlsx',1,'G2:G54');

y7 = xlsread('人口.xlsx',1,'H2:H54');

y8 = xlsread('人口.xlsx',1,'I2:I54');

y9 = xlsread('人口.xlsx',1,'J2:J54');

y=[y2 y3 y4 y5 y6 y7 y8 y9];%采用8个因素来衡量

y1 = xlsread('人口.xlsx',1,'B2:B54'); %输出的结果，人口数

temp = randperm(size(y,1)-1);

temp(53)=53;

% %训练集-43个样本，要有足够的代表性，训练模型才能好，这里我采用随机产生的方法

P\_train = y(temp(1:43),:)';%P代表输入，T代表输出

T\_train = y1(temp(1:43),:)';

%测试集--10个样本

Year=x(temp(44:end));

P\_test = y(temp(44:end),:)';

T\_test = y1(temp(44:end),:)';

N = size(P\_test,2);%10

%固定训练

P\_train = y(1:42,:)';

T\_train = y1(1:42,:)';

Year=x(43:end);

P\_test = y(43:end,:)';

T\_test = y1(43:end,:)';

N = size(P\_test,2);

temp = randperm(size(y,1));

P\_train = y(temp(1:42),:)';

T\_train = y1(temp(1:42),:)';

Year=x(temp(43:end));

P\_test = y(temp(43:end),:)';

T\_test = y1(temp(43:end),:)';

N = size(P\_test,2);

[p\_train, ps\_input] = mapminmax(P\_train,0,1);

p\_test = mapminmax('apply',P\_test,ps\_input);

[t\_train, ps\_output] = mapminmax(T\_train,0,1);

%% IV.BP（Back Propagation）神经网络建立、训练及仿真测试

%1.创建网络

net = newff(p\_train,t\_train,9);%信号前向传播，误差反向传播，连接权值和阈值随机生成的

%单隐含层神经元

%%

%2.设置训练参数

net.trainParam.epochs = 1000;

net.trainParam.goal = 1e-3;%训练要求精度

net.trainParam.lr = 0.001; %学习速率

%%

%3.训练网络

net = train(net,p\_train,t\_train);

%%

%4.仿真测试

t\_sim = sim(net,p\_test);

%%

%5.数据反归一化

T\_sim = mapminmax('reverse',t\_sim,ps\_output);

%% V.性能评价

%%

% 1.相对误差error

error = abs(T\_sim - T\_test)./T\_test;

% 2.决定系数R^2

R2 = (N \* sum(T\_sim .\* T\_test) - sum(T\_sim) \* sum(T\_test))^2 / ((N \* sum((T\_sim).^2) - (sum(T\_sim))^2) \* (N \* sum((T\_test).^2) - (sum(T\_test))^2));

% 3.结果对比

result = [T\_test' T\_sim' error'];

%% VI.绘图

figure

plot(1:N,T\_test,'b:\*',1:N,T\_sim,'r-o');%2020年的真实值是用matlab拟合函数工具箱的拟合值，2020年的预测值是神经网络模型训练的预测值

legend('真实值','预测值'); %右上角标注

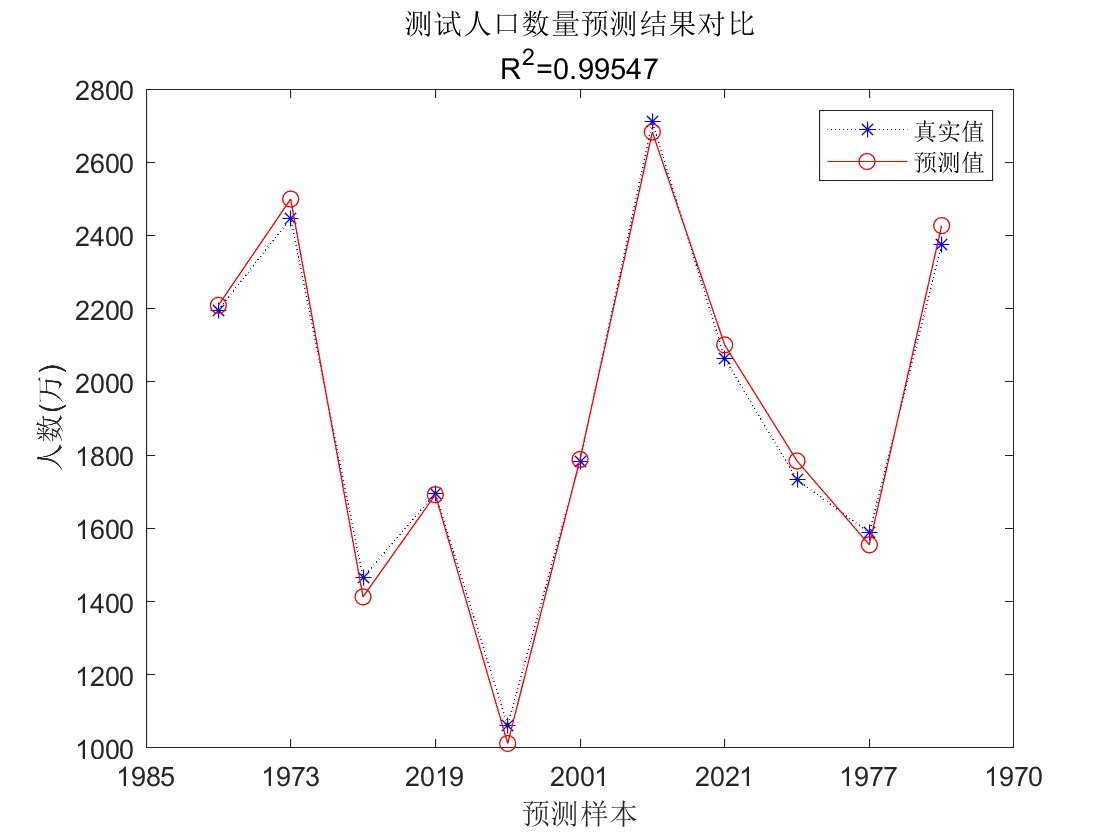
xlabel('预测样本') %x轴坐标描述

ylabel('人数(万)') %y轴坐标描述

set(gca, 'XTickLabels', Year);

string = {'测试人口数量预测结果对比';['R^2=' num2str(R2)]};

title(string);

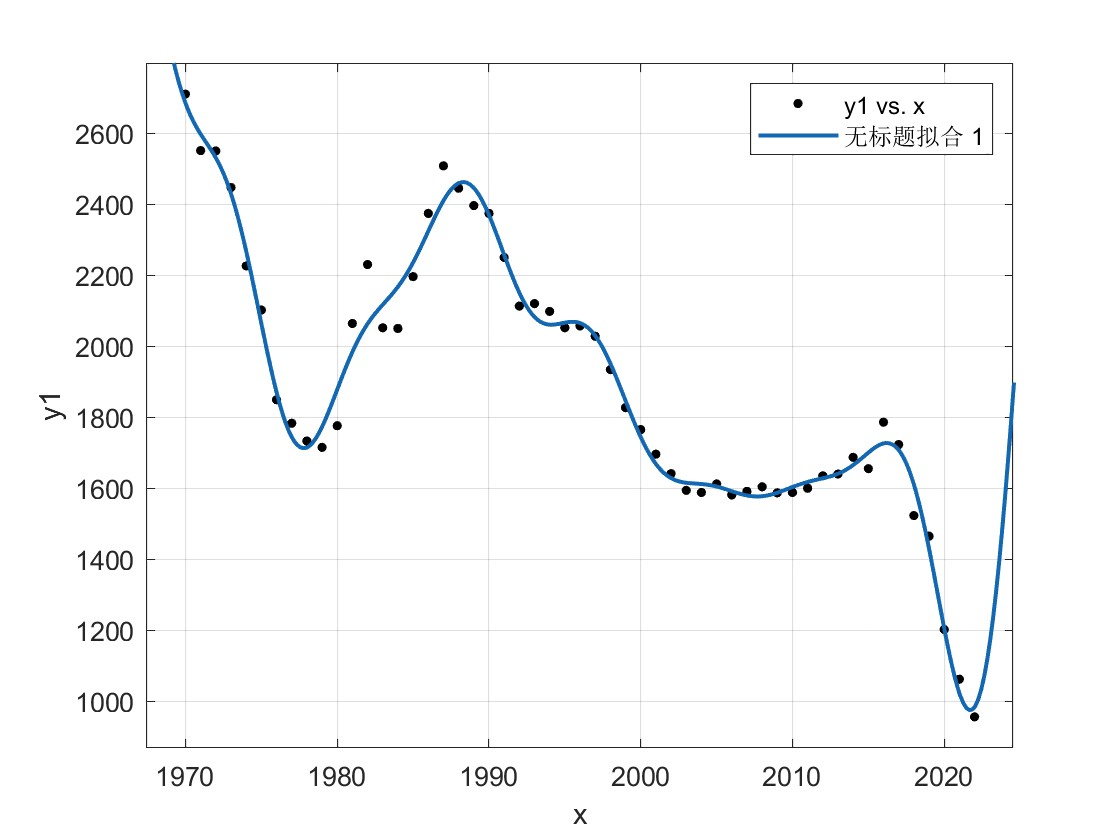


在随机产生训练集下，测试集的拟合效果很好

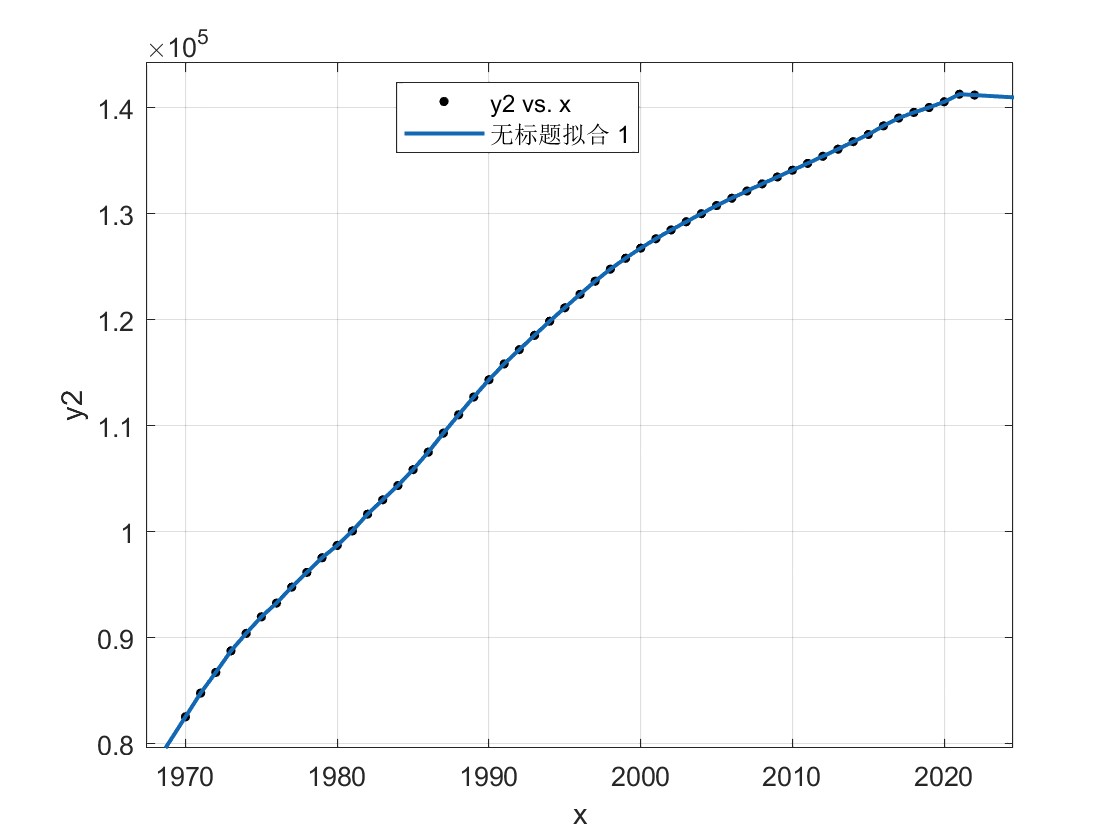
下面预测未来五年的具体结果

首先是八个因素的拟合图

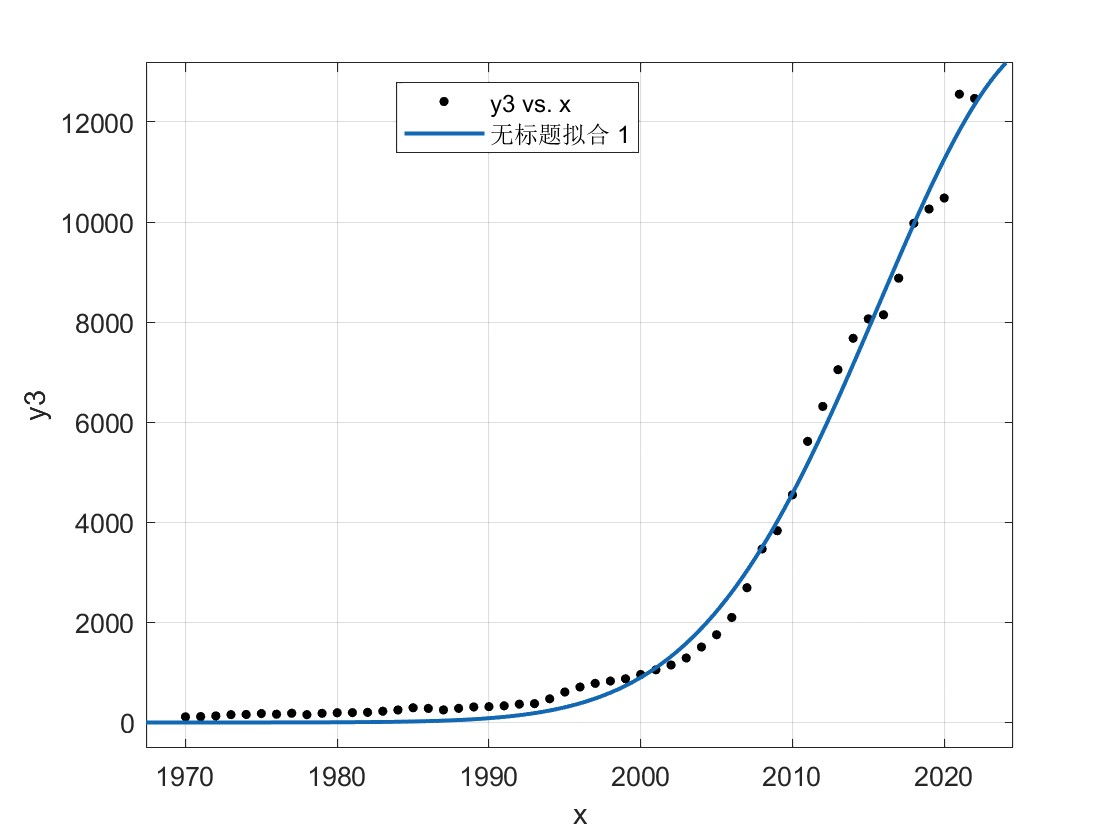
1.出生人口拟合



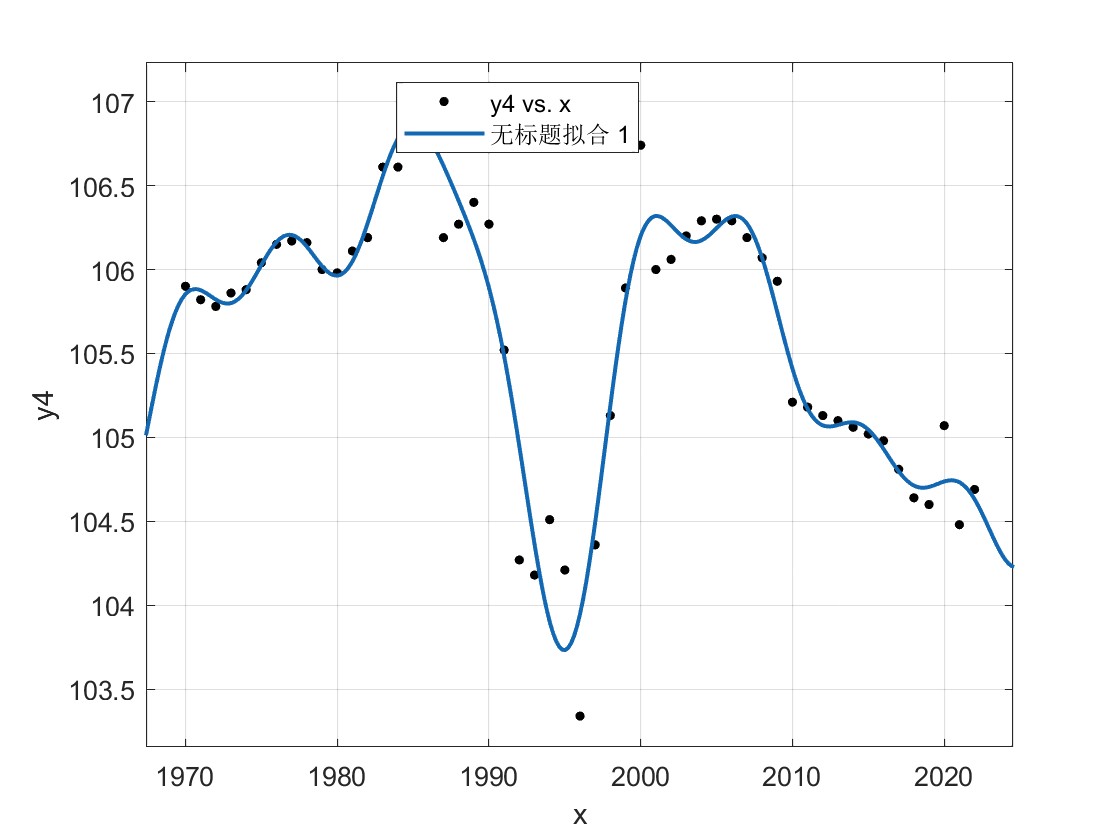
1. 总人口(万人)



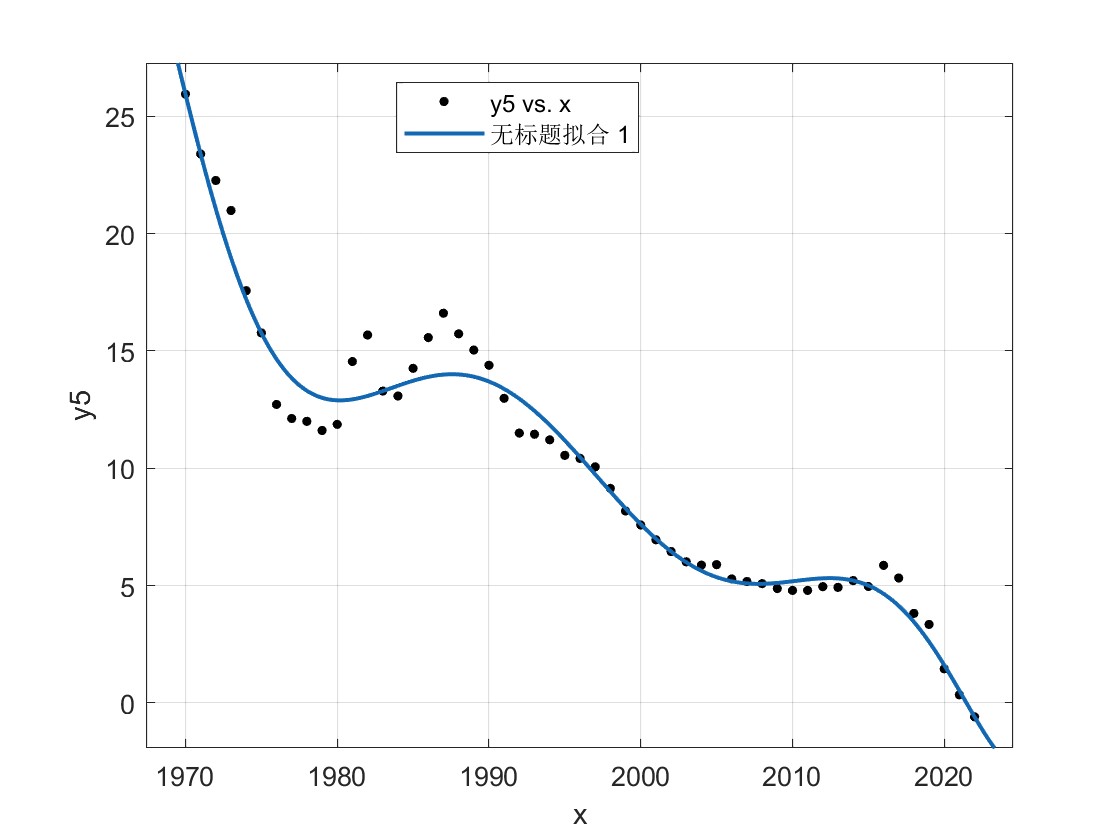
1. 中国人均GPA(美元计)



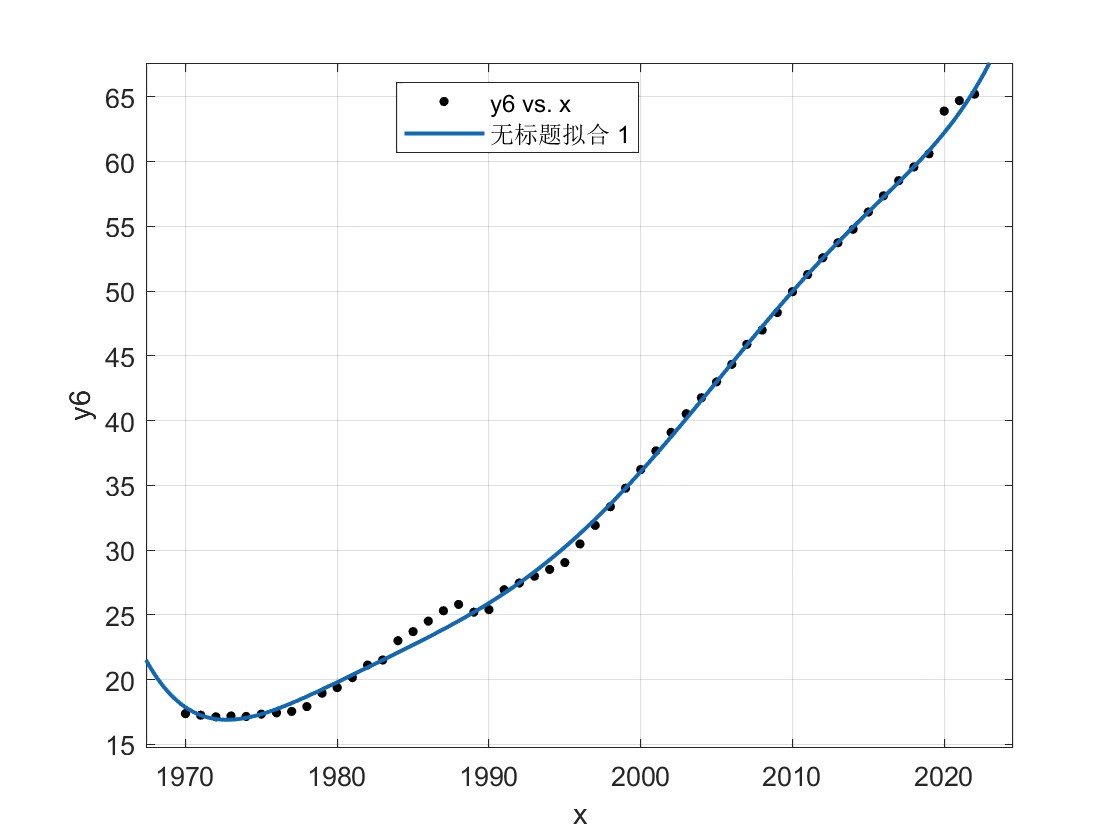
1. 中国性别比例(按照女生=100)



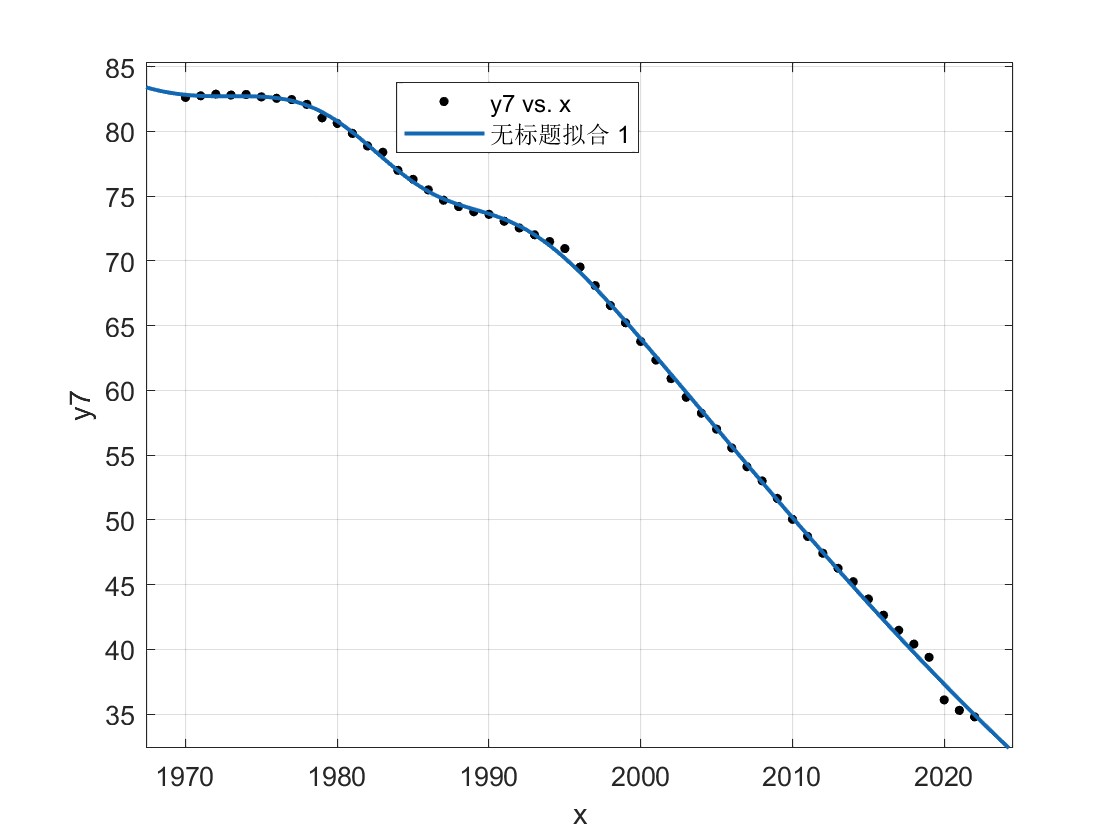
1. 自然增长率(%)



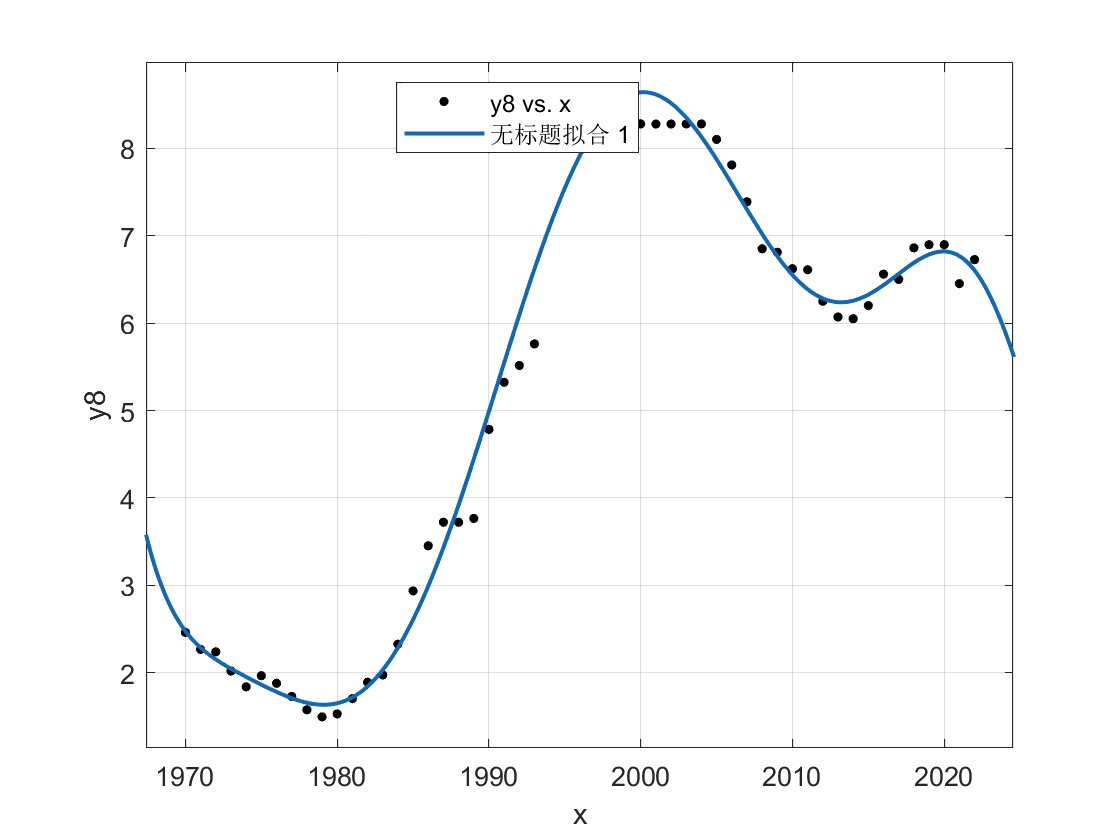
1. 城镇人口(城镇+乡村=100)



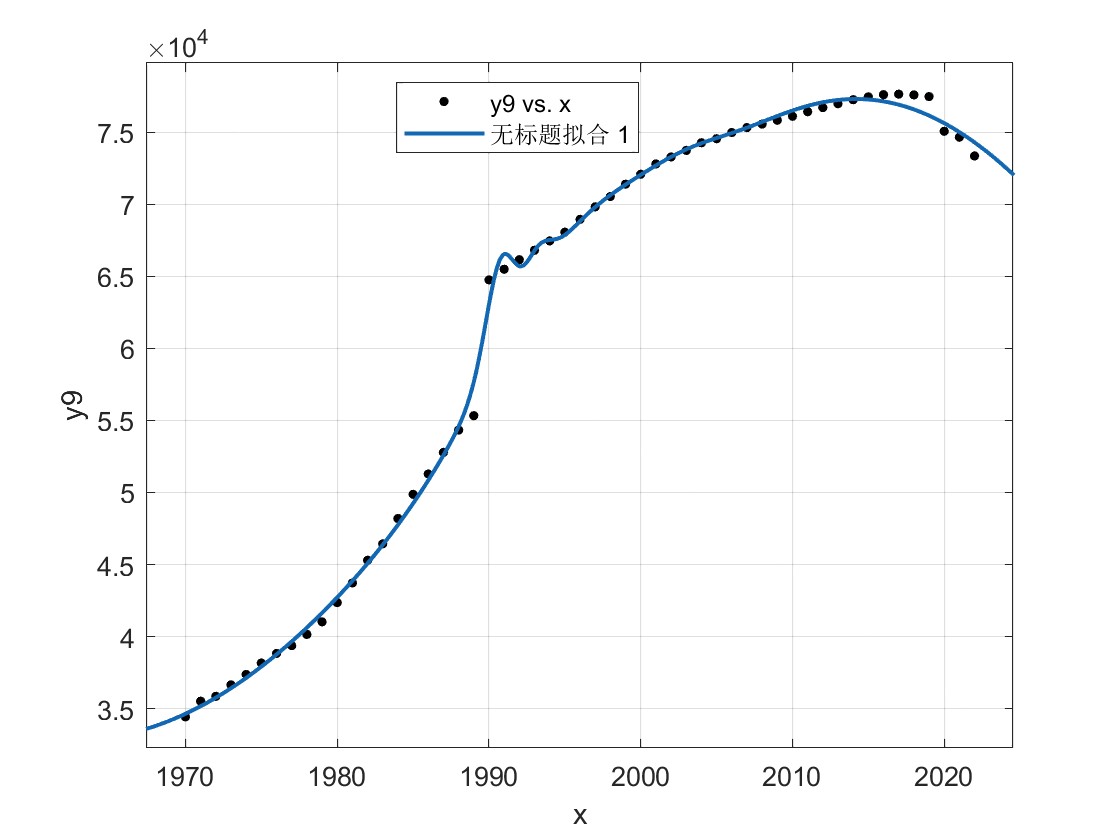
1. 乡村人口



1. 美元兑换人民币汇率



1. 美元兑换人民币汇率



%读取新数据

new\_data = xlsread('人口.xlsx',1,'C55:C64');

% 归一化处理

p\_new = mapminmax('apply',new\_data',ps\_input);

% 进行预测

t\_pred = sim(net,p\_new);

% 反归一化处理

t\_unscaled = mapminmax('reverse',t\_pred,ps\_output);

结果为

2023 141082

2024 141011

2025 140747

2026 140563

2027 140021