

38. 何为最优拟合模型公式？—1stOpt 案例剖析

38.1 问题的提出

已知自变量数据 $x = [4, 25, 100]$ ，因变量数据 $y = [100, 50, 0]$ ，需要确定一个最佳拟合公式，因为只有三个数据点，对拟合公式有如下三条要求：一是模型公式的参数数要小于数据量数，也即参数数只能为1个或2个，由于1个参数不论从理论上还是实际中都无法满足要求，因此参数数实际最终只能为2个；二是拟合公式需保证无论参数取任何值，计算输出的因变量 y 值必须不为负数；三是拟合度尽可能达到百分之百的完美吻合。

该问题看似简单，但要同时满足上述三个条件亦非易事。

38.2 分析与解决

如何选择确定同时满足前述三条件的模型公式？虽然1stOpt有一个根据数据自动计算选择最好公式的工具箱，但该工具箱没有可以限定满足第一和第二条的功能，因此只能通过手动实现。首先第一条只能有两个参数，这一硬性条件明确清晰，易于掌控；第二条计算值任何情况下都不能为负，这就限定了模型公式只能是整体单一的指数函数，如： $\text{Exp}(\dots)$ ，或幂函数，如： $a^{(\dots)}$ ， a 为大于0的实数，也即只能在这两种模式下选择确定公式形式；第三条则只能在满足前两条的前提下进行不断地调试，选择最终满足拟合精度的公式。最终选定了如下11个模型公式，其中 p_1 和 p_2 为参数。

表 38-1 拟合模型公式

序号	模型公式
1	$y = \exp(p_1 x - x^{p_2})$
2	$y = \exp(p_1 + p_2 x^{p_1})$
3	$y = x^{p_1 x - x \exp(p_2 x)}$
4	$y = \exp(p_1 + p_2 \exp(p_1 x))$
5	$y = \exp\left(p_1 + p_2 \exp\left(\frac{p_2}{x}\right)\right)$
6	$y = p_1^{-p_2 + p_1 \exp(p_2 x)}$
7	$y = \exp(p_1 - p_2 \exp(p_2 x - p_1))$
8	$y = \exp(p_2 x^{p_1} - \exp(-p_2 - p_1 x))$
9	$y = p_2^{p_1 \ln\left(\frac{p_2}{x}\right) + p_1 \exp(p_2 x)}$
10	$y = \exp(p_1 - \exp(p_2 x - p_1)) \exp(p_1 x)$

11	$y = \frac{\exp\left(\exp\left(-\frac{p_1}{x}\right)p_2\right)}{\exp(p_2^x - p_1)}$
----	---

运行代码

```
Function y=exp(p1*x-x^p2);
y=exp(p1+p2*x^p1);
y=x^(p1*x-x*exp(p2*x));
y=exp(p1+p2*exp(p1*x));
y=exp(p1+p2*exp(p2/x));
y=p1^(-p2+p1*exp(p2*x));
y=exp(p1-p2*exp(p2*x-p1));
y=exp(p2*x^p1-exp(-p2-p1*x));
y=p2^(p1*ln(p2/x)+p1*exp(p2*x));
y=exp(p1-exp(p2*x-p1)*exp(p1*x));
y=exp(exp(-p1/x)*p2)/exp(p2^x-p1);

Data;
x=[4,25,100];
y=[100,50,0];
```

上述代码一次运行，可同时得到对应11个模型公式的拟合计算结果。

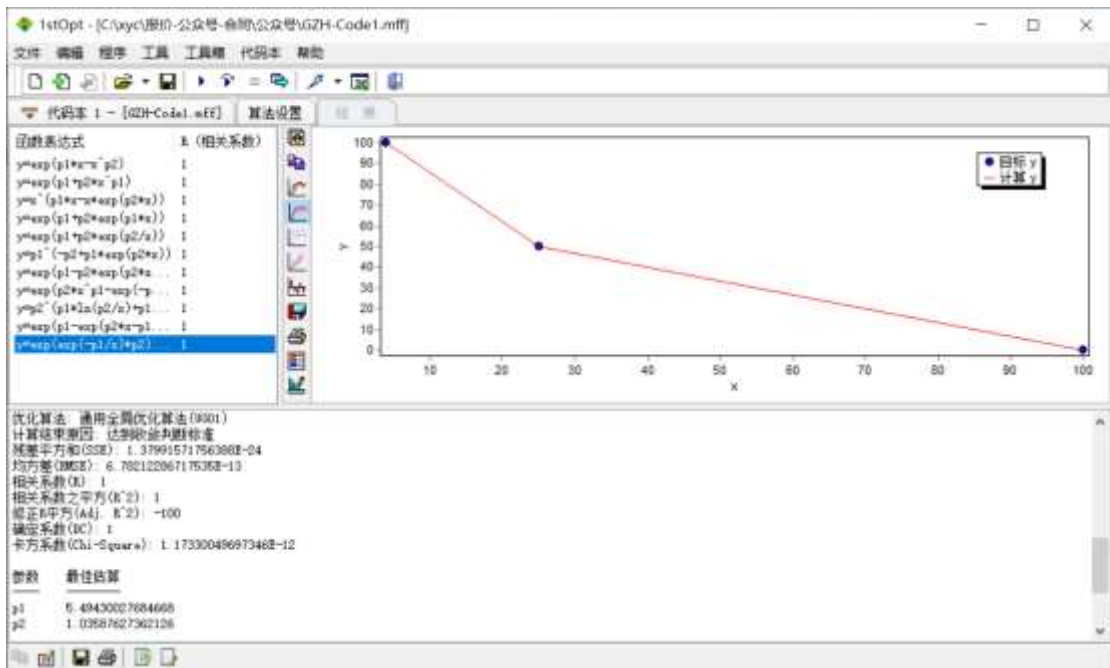


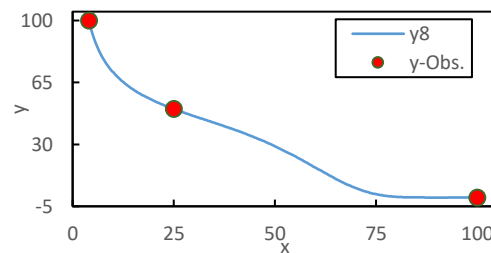
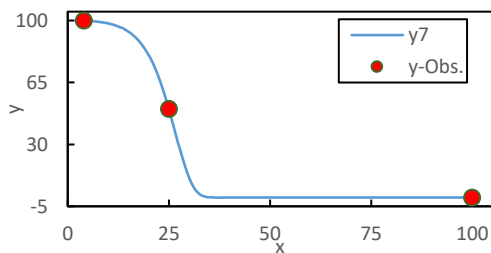
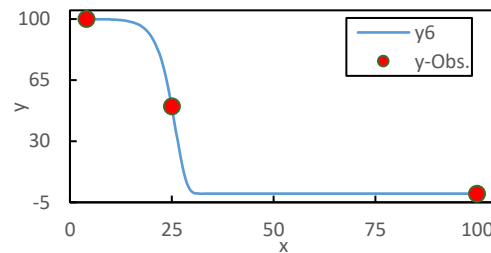
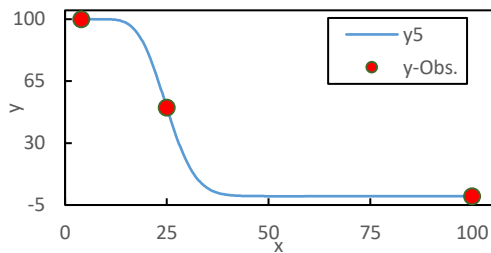
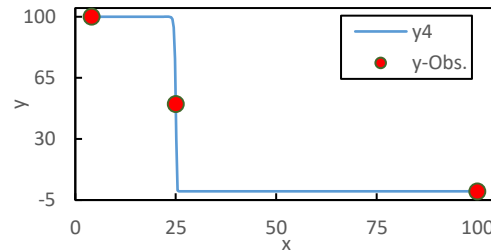
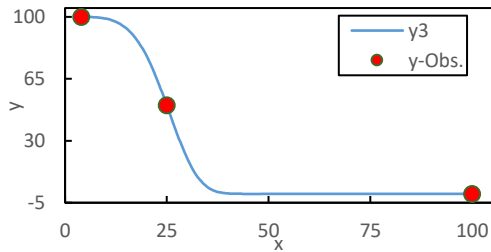
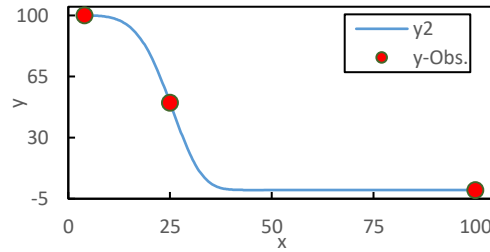
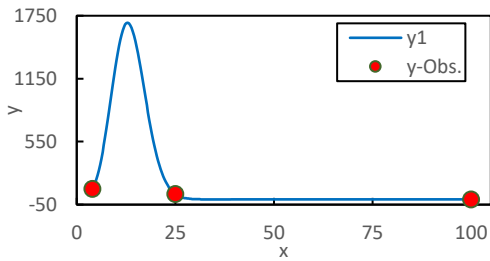
图 38-1 拟合计算示意图

计算结果

模型公式	拟合结果
1	Sum Squared Error (SSE): 5.19007366763008E-26 Root of Mean Square Error (RMSE): 1.31530397850713E-13 Correlation Coef. (R): 1 R-Square: 1 Parameter Best Estimate ----- p1 2.6277523429571

	p2	1.28107103643528
2	Sum Squared Error (SSE): 3.56337937219194E-24 Root of Mean Square Error (RMSE): 1.089859222131E-12 Correlation Coef. (R): 1 R-Square: 1 Parameter Best Estimate ----- p1 4.60532002922673 p2 -2.52905465117711E-7	
3	Sum Squared Error (SSE): 3.68555814919257E-25 Root of Mean Square Error (RMSE): 3.5050240841743E-13 Correlation Coef. (R): 1 R-Square: 1 Parameter Best Estimate ----- p1 1.93760763191348 p2 0.0254417786226765	
4	Sum Squared Error (SSE): 8.40190138046118E-19 Root of Mean Square Error (RMSE): 5.29210146679659E-10 Correlation Coef. (R): 1 R-Square: 1 Parameter Best Estimate ----- p1 4.6051701859879 p2 -6.93147180544769E-51	
5	Sum Squared Error (SSE): 9.91431457934047E-26 Root of Mean Square Error (RMSE): 1.81790305749422E-13 Correlation Coef. (R): 1 R-Square: 1 Parameter Best Estimate ----- p1 4.60517018598886 p2 -131.05283756663	
6	Sum Squared Error (SSE): 5.04870979341448E-29 Root of Mean Square Error (RMSE): 4.10232039761827E-15 Correlation Coef. (R): 1 R-Square: 1 Parameter Best Estimate ----- p1 4.72724677599644E-6 p2 0.375580462304542	
7	Sum Squared Error (SSE): 2.62532909257553E-27 Root of Mean Square Error (RMSE): 2.9582253083989E-14 Correlation Coef. (R): 1 R-Square: 1 Parameter Best Estimate ----- p1 4.61086356580846 p2 0.229053264507106	
8	Sum Squared Error (SSE): 1.29751443403376E-18 Root of Mean Square Error (RMSE): 6.57650979885675E-10 Correlation Coef. (R): 1 R-Square: 1 Parameter Best Estimate ----- p1 -0.0836301853075508 p2 5.18008616223959	
9	Sum Squared Error (SSE): 1.69092620286227E-21 Root of Mean Square Error (RMSE): 2.37411471420561E-11 Correlation Coef. (R): 1 R-Square: 1	

	Parameter Best Estimate ----- p1 0.489593324952461 p2 0.0504545387606551
10	Sum Squared Error (SSE): 3.63507105125842E-27 Root of Mean Square Error (RMSE): 3.48093428610693E-14 Correlation Coef. (R): 1 R-Square: 1 Parameter Best Estimate ----- p1 4.62463966569759 p2 -4.45320654493092
11	Sum Squared Error (SSE): 1.88594790015978E-24 Root of Mean Square Error (RMSE): 7.92874075785847E-13 Correlation Coef. (R): 1 R-Square: 1 Parameter Best Estimate ----- p1 5.49430027684667 p2 1.03587627362126



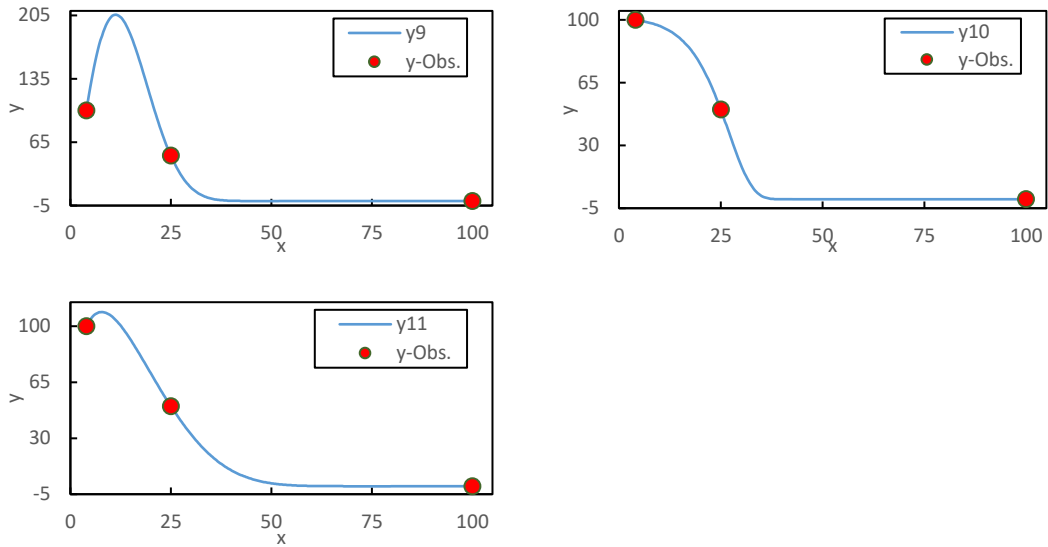


图38-2 模型公式1至11对应拟合计算对比图

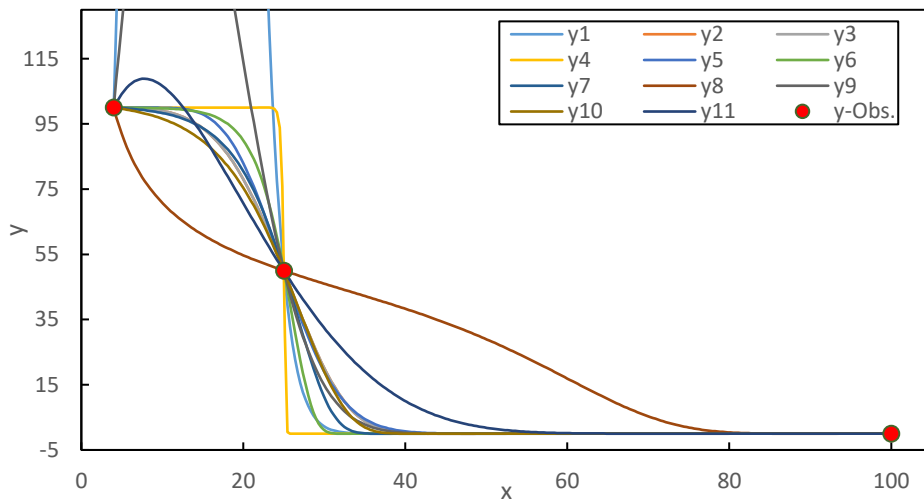


图38-3 模型公式1至11综合拟合计算对比图

从上述计算结果及对比示意图可以看出，选出的11个模型公式均满足三个限制条件，即2个参数、计算因变量值不为负以及完美拟合效果。

虽然这11个模型函数公式都满足要求，但各自对应的图形效果却相差极大，在这种多选的状况下该如何确定最终的模型公式呢？这就需要该问题更多的背景条件去选择确定，比如第1、第9和第10公式，从图形看类似于峰图形；第4公式的图形为垂直“之”字形；第8公式图形为平缓递减形，其余则为相似的到“S”形。

另外如果再添加一条限制条件：模型公式的一阶导数均为负，那么有哪些公式符合？这就需要对前述的11个模型公式分别求解各自的一阶导数解析式，如果再加以可视化图像，则会更容易判定。最简单的实现方式是在计算完成后，点击“二维-三维分析/预测”按钮，如下图38-4和图38-5所示，选择“一阶导数”或弧微分、曲率、曲线长等即可获得对应得到图像，

选择“1-5阶微分式复制”则可得相应的微分解析式。注意，当拟合方程不可微时，该项功能无效。

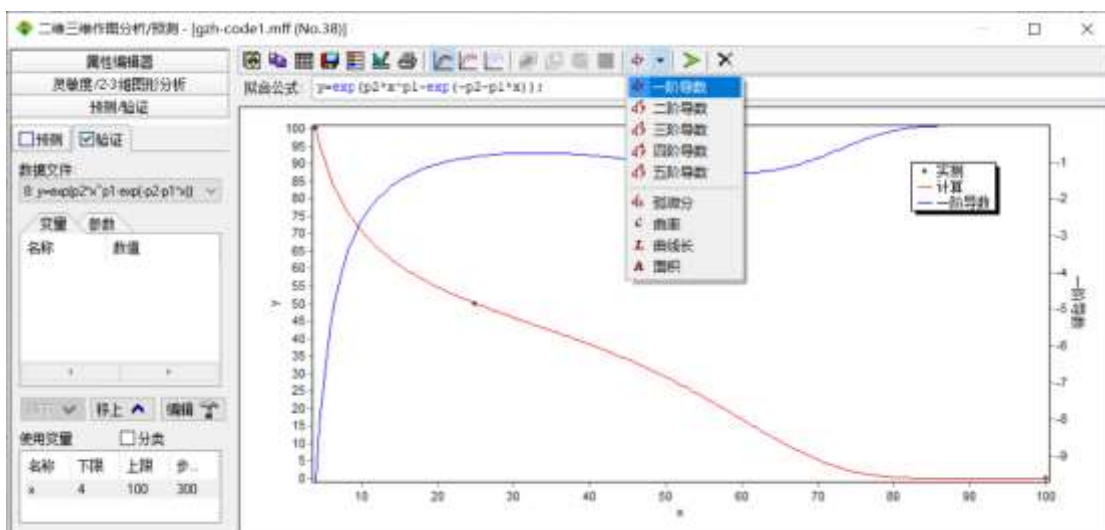


图38-4 微分方程图像

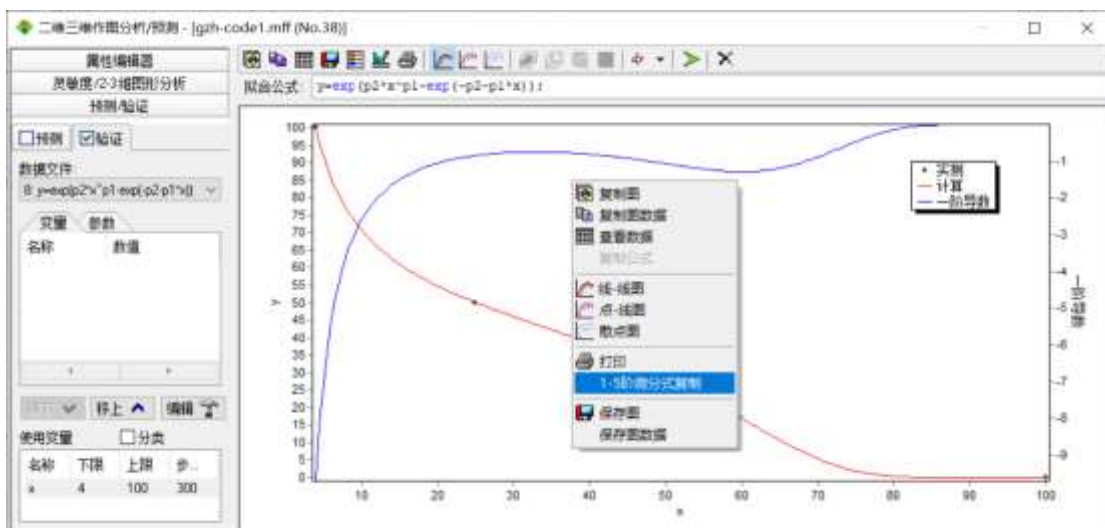


图38-5 微分方程解析式复制

此外也可以用其它软件如Matlab或Mathematical等单独求解，或者直接在1stOpt中用“Diff()”命令求解，比如对第一个公式“ $\exp(p_1 \cdot x - x^{p_2})$ ”，在1stOpt中输入“diff(exp(p1*x-x^p2),x)”，可得到其一阶导数解析式“ $\exp(p_1 \cdot x - x^{p_2}) \cdot (p_1 - p_2 \cdot x^{p_2-1})$ ”，如果输入“diff(exp(p1*x-x^p2),x,2)”，可得其二阶导数解析式“ $(\exp(p_1 \cdot x - x^{p_2}) \cdot (p_1 - p_2 \cdot x^{p_2-1})) \cdot (p_1 - p_2 \cdot x^{p_2-1}) + \exp(p_1 \cdot x - x^{p_2}) \cdot ((-p_2 \cdot ((p_2-1) \cdot x^{p_2-2}))$ ”，最高可以5阶。对11个模型公式所得一阶导数结果看，仅有公式5和公式7满足一阶导数均为负的要求。

公式5一阶导数表达式：

$$\frac{dy}{dx} = -\frac{p_2^2 \exp\left(\frac{p_2}{x}\right) \exp\left(p_1 + p_2 \exp\left(\frac{p_2}{x}\right)\right)}{x^2}$$

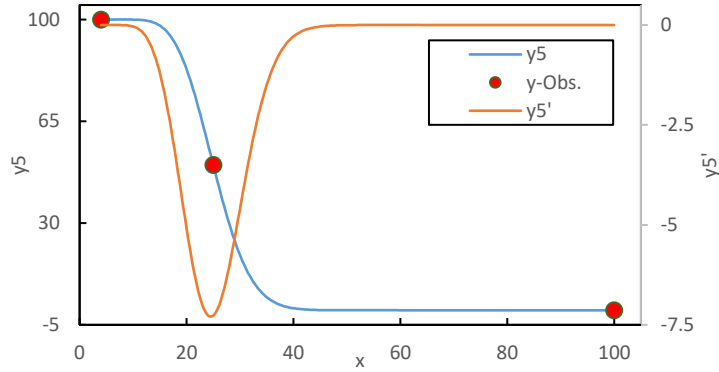


图38-6 模型公式5原函数及其一阶导数图

公式7一阶导数表达式:

$$\frac{dy}{dx} = -p_2^2 \exp(p_2 x - p_1) \exp(p_1 - p_2 \exp(p_2 x - p_1))$$

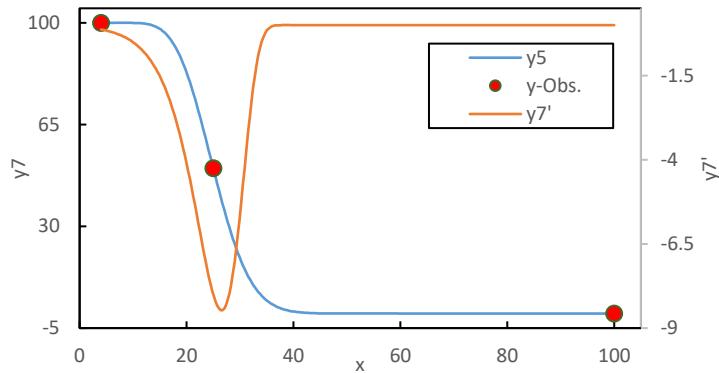


图38-7 模型公式7原函数及其一阶导数图

不论从图形还是一阶导数解析式看，不论参数 p_1 和 p_2 以及自变量 x 取任何值，模型公式5和7都满足一阶导数值均为负的要求。

反之，如果要求一阶导数值均大于0，哪个模型公式符合要求？经上述类似方法计算验证，目前所提出的11个公式都无法满足要求。是否有满足要求但目前还没有发现的？整体到“S”形曲线，应该不存在，有兴趣的可以自己动手试试。

38.3 小结

基于1stOpt强大的优化求解能力，对于过三点确定最优模型公式的问题，在满足限制条件的前提下，可以轻松尝试验证各种不同的模型拟合公式，再根据实际问题的背景，最终得到满意的结果。同时还介绍了拟合方程1-5阶导数及对应图像的求解获取方式。