CMA 使用不同误差计算方法对非线性曲线拟合结果的影响

1. 误差计算方式

对于函数$f(x)$, 有大小为 n 的观测样本 S={(x_1, $\hat{f}(x_1)$, x_2, $\hat{f}(x_2)$,…, x_n, $\hat{f}(x_n)$)}, 通过观测样本拟合函数$f(x)$，误差计算方式分为三种：

均方误差（Mean Squared Error, MSE）：

$$\frac{\sum(f(x_i)-\hat{f}(x_i))^2}{n}$$

百分比误差（Mean Percent Error, MPE）：

$$\frac{\sum(\frac{f(x_i)-\hat{f}(x_i)}{f(x_i)})^2}{n}$$

绝对值误差（Mean Absolute Error, MAE）：

$$\frac{\sum\left|\frac{f(x_i)-\hat{f}(x_i)}{f(x_i)}\right|}{n}$$

2. 函数

构造非线性曲线，选用一元二次多项式函数：$f_1(x)=\omega_2x^2+\omega_1x+\omega_0$ ，令$\omega_0=0$经过原点；此外选用指数函数$f_2(x)=a(e^{\frac{x}{b}}-1)$, 经过原点。

3. 数据集

为模拟对函数$f_1(x)$ 采样过程，构造噪点函数:

$$f_{11}(x)=\omega_2x^2+\omega_1x+\omega_0+\frac{1}{20\sqrt{2\pi}}e^{(-\frac{x^2}{800})}$$

为模拟对函数$f_1(x)$ 采样过程，构造噪点函数:

$$f_{21}(x)=a(e^{\frac{x}{b}}-1)+\frac{1}{20\sqrt{2\pi}}e^{(-\frac{x^2}{800})}$$

另外为模拟数据点离原点远近对其集中程度的影响，使用指数分布

$$f(x;\beta) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$$

函数f_{11}使用参数向量(2,1,0)，函数f_{12}使用参数向量(0.5,2)，最后构造 6 个数据集：S0~S5，其中 S0~S2 使用多项式函数产生噪点数据，S3~S5 使用指数函数产生噪点数据，对应如下图 1 所示。

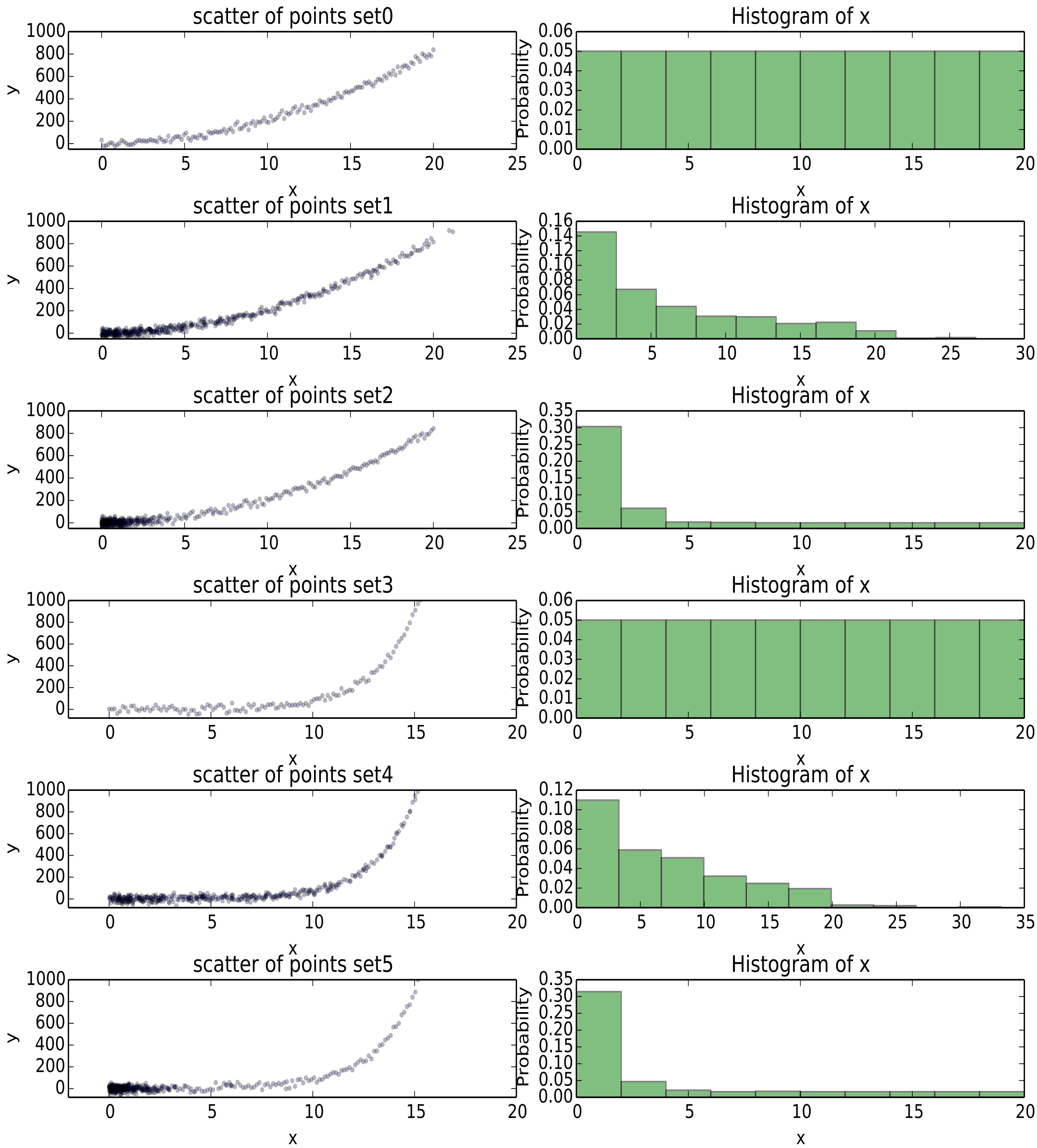


图 1　6 个数据集与其分布情况

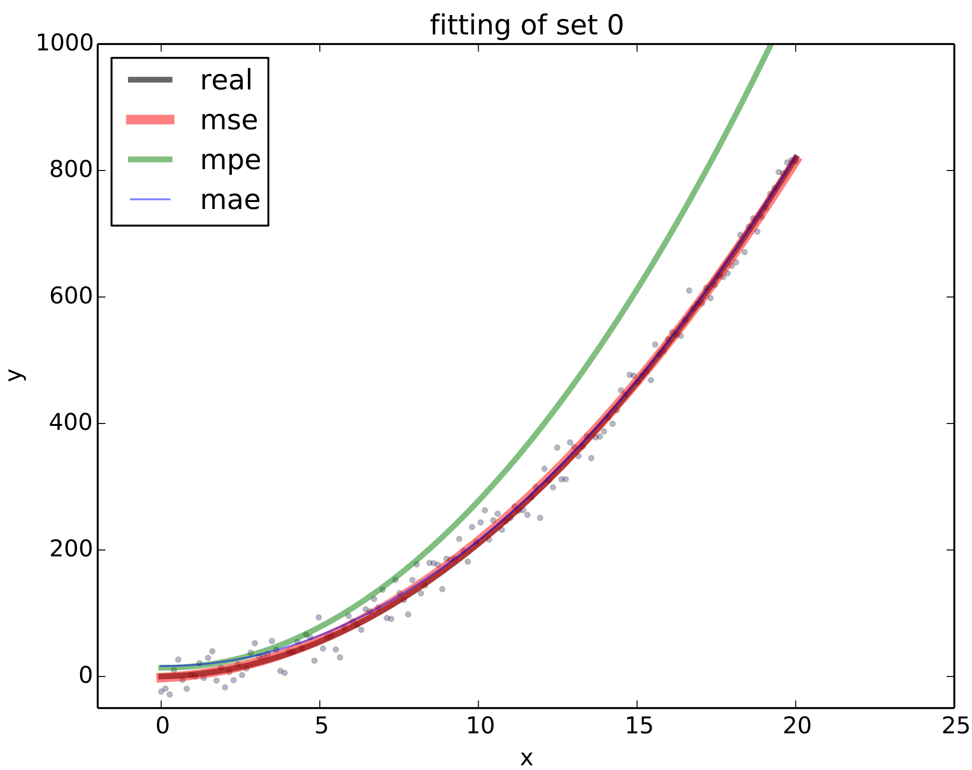
4. 测试结果

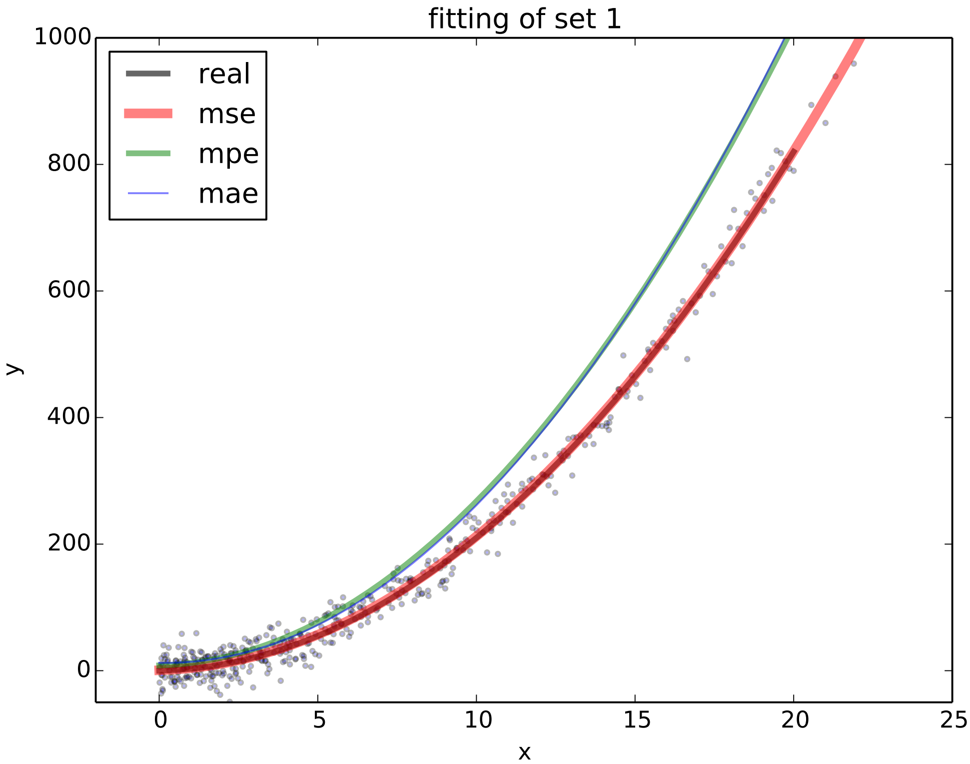
下表为使用 CMA 算法在上面数据集基础上计算残差函数到达最小值时的函数参数。

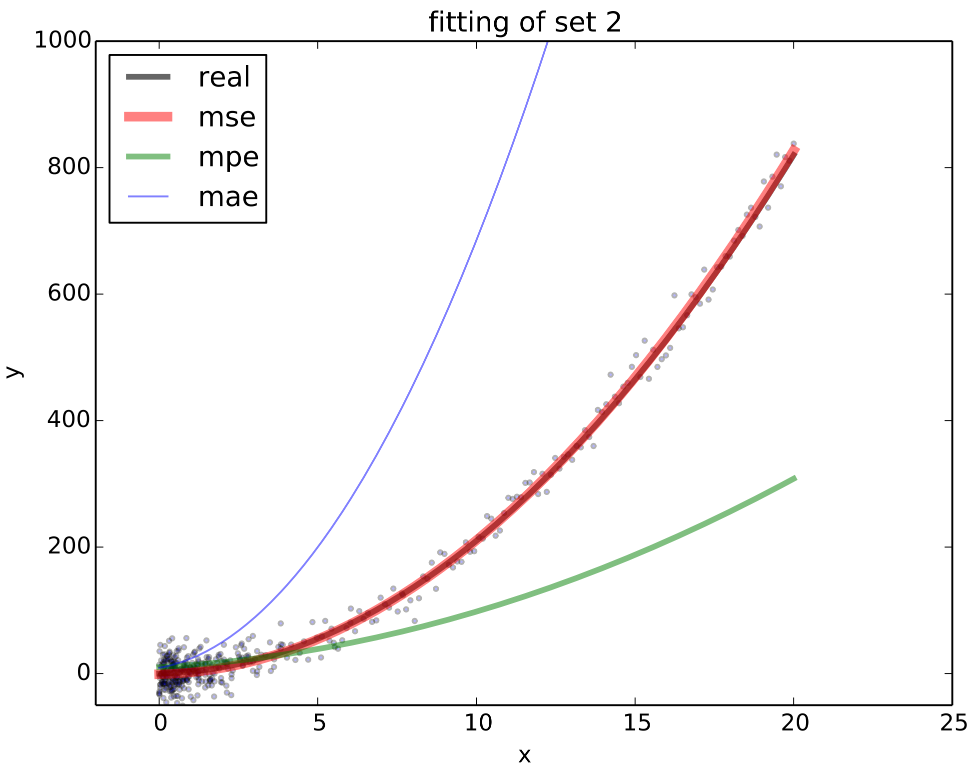
表 1　CMA 函数参数结果

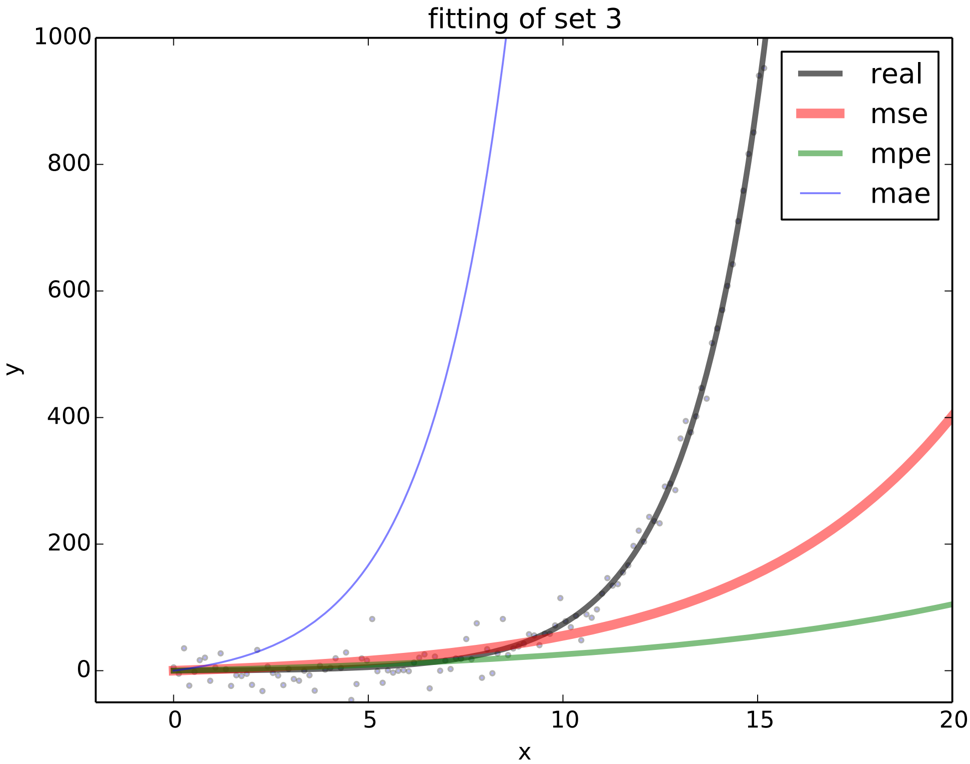
数据集 参数向量	S0	S1	S2	S3	S4	S5
pmse	1.9261, 2.3326, -1.8857	1.9858, 1.2944, 0.7171	2.0026, 0.9484, -0.9666	10.3715, 5.4289	6.8239, 3.0361	14,7554, -1.5484
pmpe	2.7009, -0.6332, 13.8089	2.4613, 1.3265, 8.3007	0.6115, 2.6682, 10.1617	12.1946, 8.8389	11.24, 1.9114	5.5104, 2.0464
pmae	2.0378, -0.5021, 16.5252	2.5931, -1.0274, 12.4453	5.8871, 8.7492, 9.0464	16.2185, 2.0636	9.805, 5.9454	12.1325, 2.6738

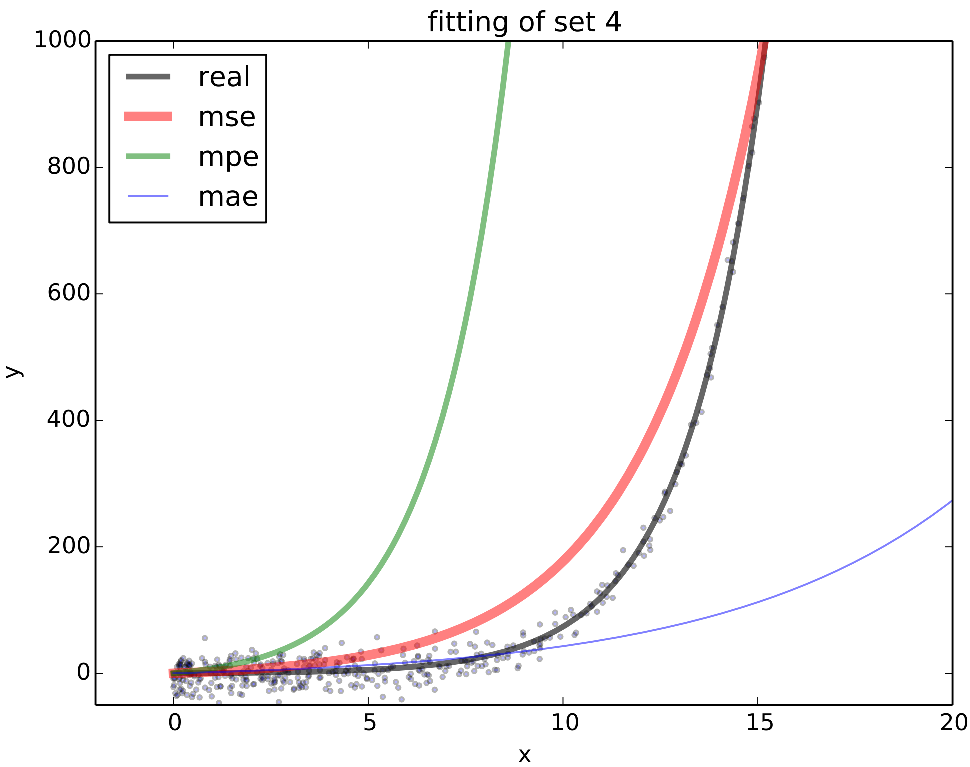
下图为 cma 拟合函数的可视化结果

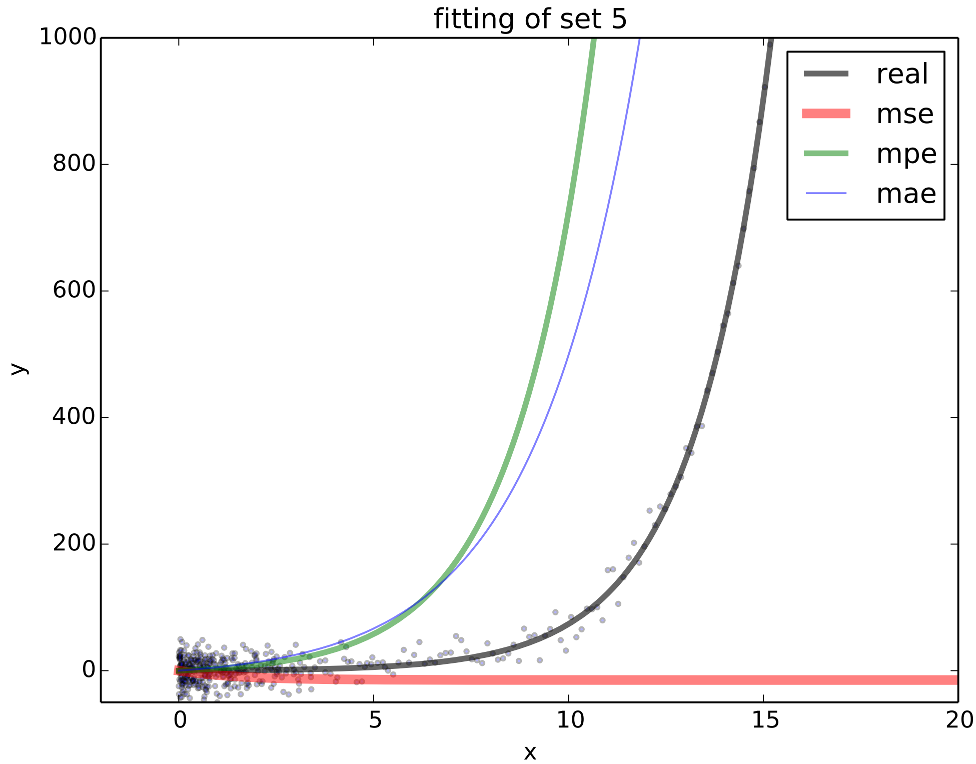



fitting of set 1
real
mse
mpe
mae
y
x
0
200
400
600
800
1000
5
10
15
20
25


fitting of set 2
real
mse
mpe
mae
y
x
0
200
400
600
800
1000
5
10
15
20
25


fitting of set 3
1000
800
600
400
200
0
y
0
5
10
15
20
x
real
mse
mpe
mae


fitting of set 4
1000
800
600
400
200
0
y
0
5
10
15
20
x
real
mse
mpe
mae


fitting of set 5
1000
800
600
400
200
0
y
0
5
10
15
20
x
real
mse
mpe
mae