Curve fitting: perspective from machine learning

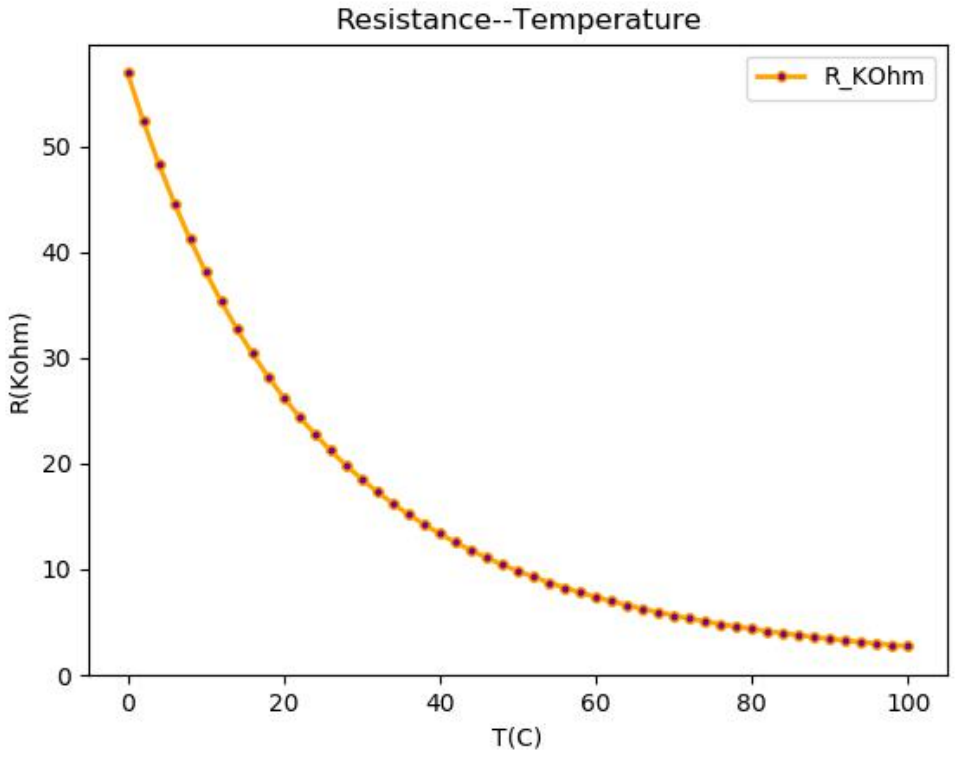
描述热敏电阻阻值与温度关系的模型可以表示为：

$$R_T = R_{T_0} e^{\beta\left(\frac{1}{T}-\frac{1}{T_0}\right)} \tag{1}$$

其中，T 为温度（单位为 K），R_T 为温度为 T 时热敏电阻阻值（单位为 kΩ），R_{T_0} 为温度为 T_0 时热敏电阻阻值（单位为 kΩ）。已知某种热敏电阻在 25℃ 时的阻值为 22kΩ，$\beta = 3100$（K），试完成如下研究工作：

1）以 2℃ 作为间隔（步长），画出该种热敏电阻在温度范围为 0℃~100℃ 间阻值随温度变化的特性曲线

答：



2）假设我们事先并不知道（1）式所描述的热敏电阻阻值—温度模型，现通过测量热敏电阻在不同温度下的阻值的实验方法对其特性加以研究，实验温度范围为 20℃~80℃。现采用如下多项式模型描述热敏电阻阻值与温度关系

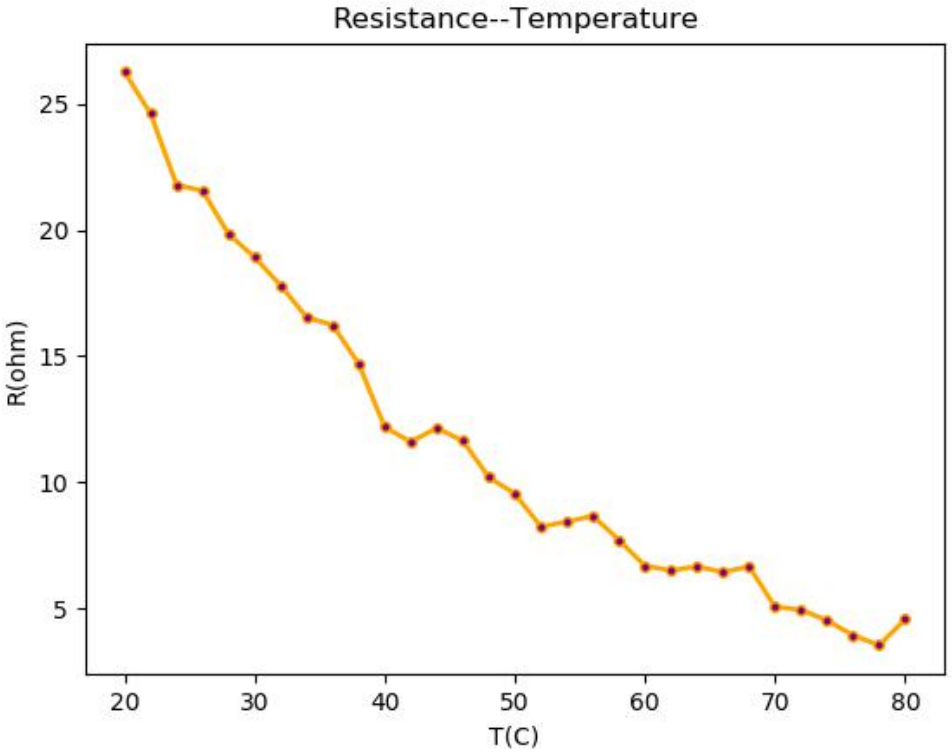
$$R_t = a_n t^n + a_{n-1} t^{n-1} + \cdots + a_1 t + a_0 \qquad (2)$$

其中，t 为温度（单位为℃），R_t 为温度为 t ℃ 时热敏电阻阻值（单位为 kΩ），n 为模型阶次，a_n 为不同阶次项系数。

在 1）中获得的 20℃~80℃ 范围的数据上添加适当噪声（以零均值、标准偏差取 500Ω的高斯噪声为例），用添加噪声后的数据模拟实验数据（添加噪声模拟实际测量过程）。针对（2）式描述的多项式模型，用模拟的实验数据作为训练数据集，采用曲线拟合最小二乘法分别获得模型阶次 $n = 1,2,3,4,5,6$ 时传感器特性曲线对应的多项式模型；分别计算不同阶次模型在温度范围 20℃~80℃（训练集）上和温度范围 0℃~100℃ 刨除 20℃~80℃ 温度范围后（测试集）上的误差（均方误差，mean squared error），观察训练集和测试集上误差随模型阶次的变化规律并加以讨论

解：添加均值为 0（单位：K 欧），方差为 0.5*0.5=0.25 的噪声。

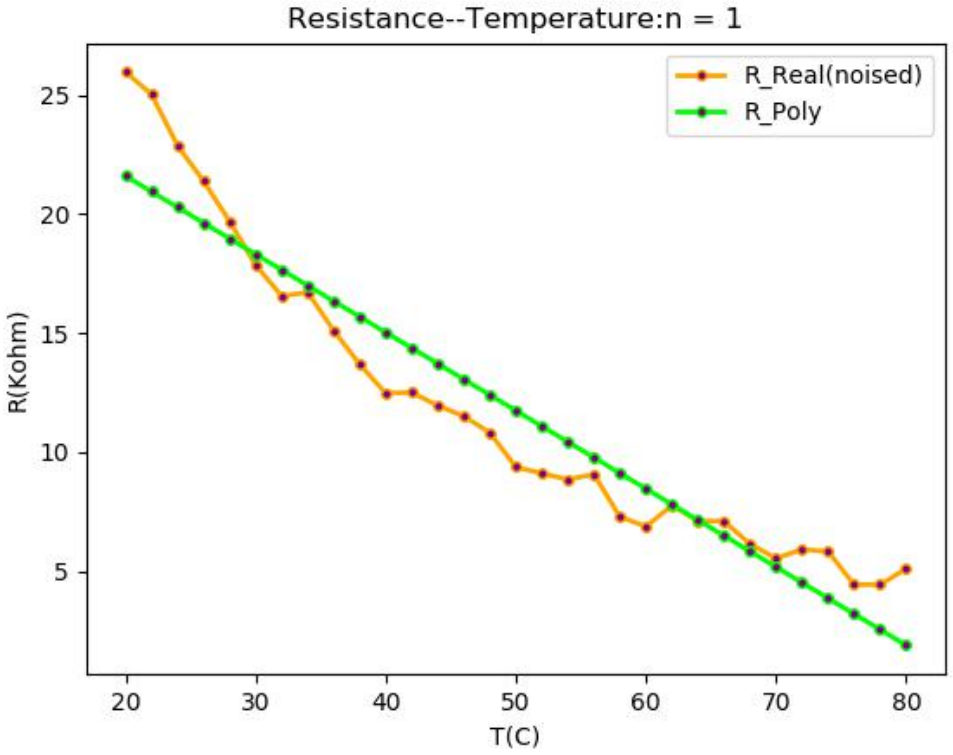
某次添加噪声后的训练数据分布：



下面进行多项式拟合:

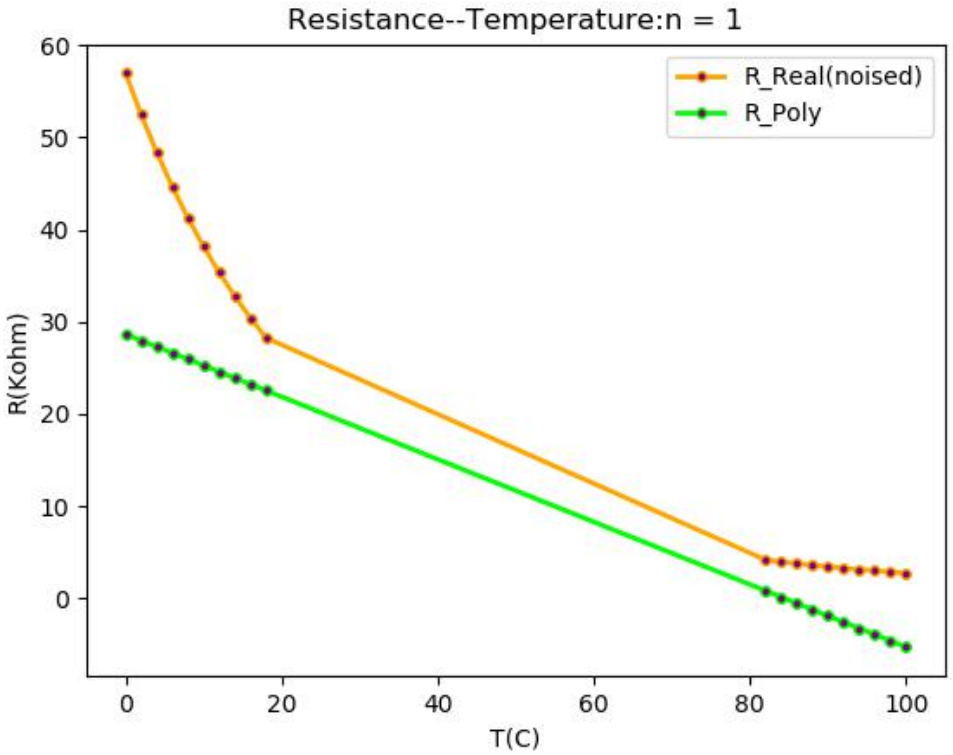
【n=1】

训练:



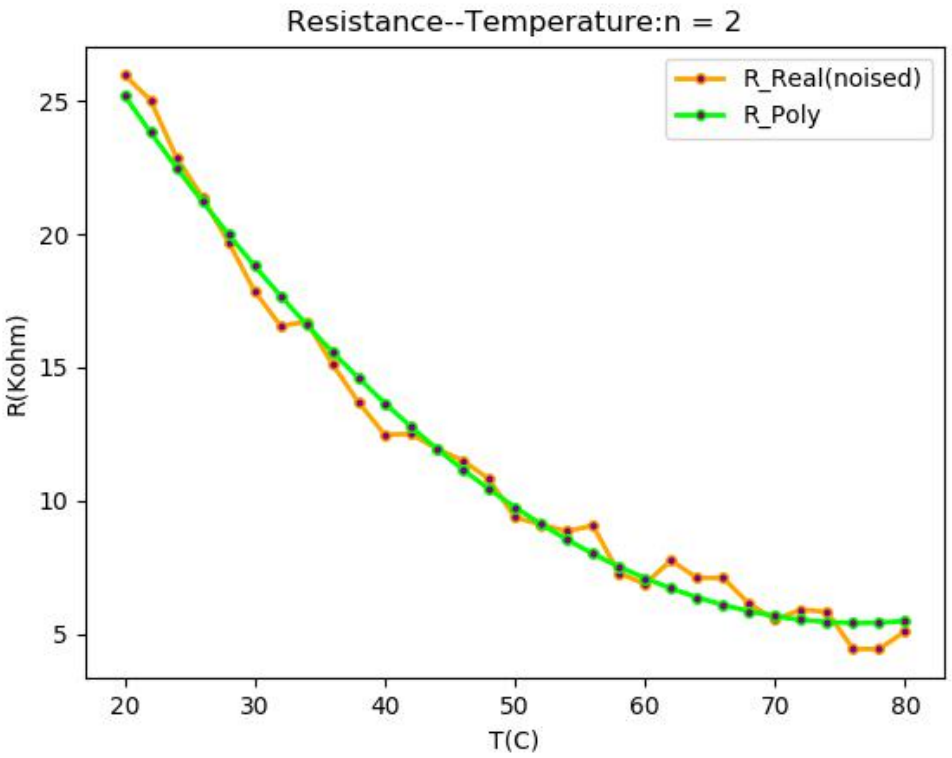
R(t) = -0.3393310492066832t**1 + 28.68631083460063

预测：



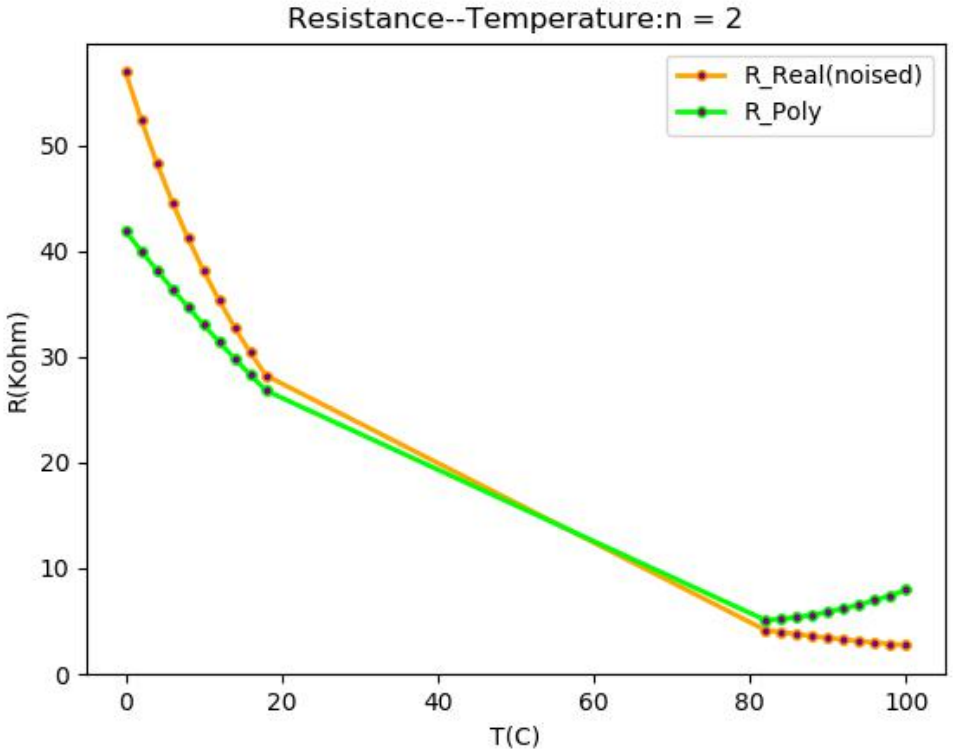
【n=2】

训练：



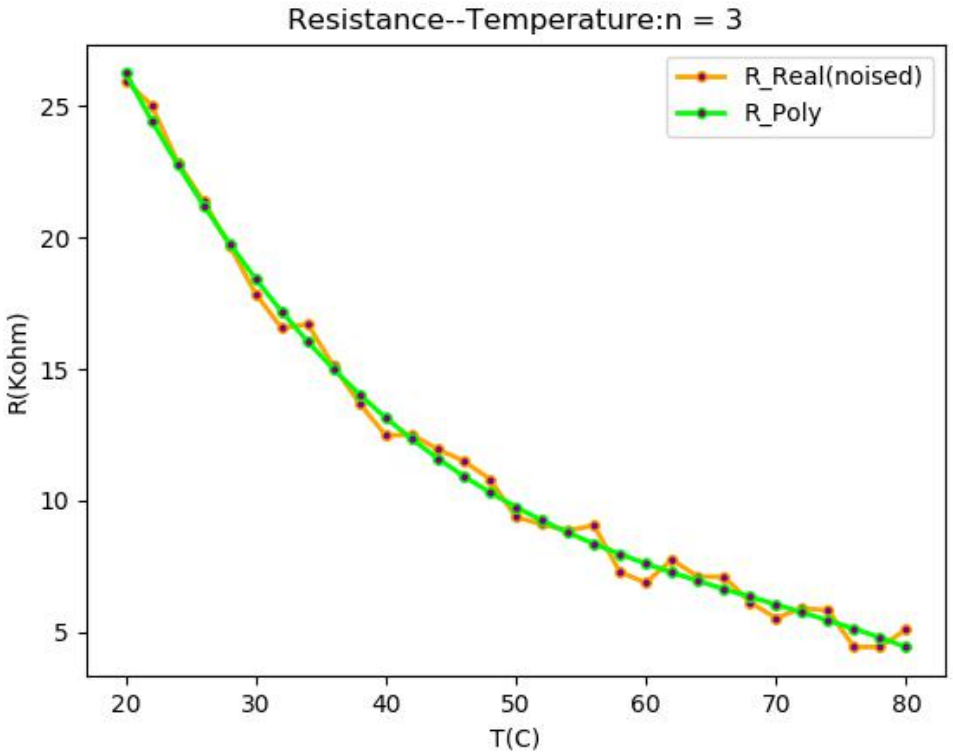
R(t) = 0.006210876071455018t**2 - 0.9604186563521848t**1 + 42.226020670372584

预测：


Resistance--Temperature:n = 2
R_Real(noised)
R_Poly
R(Kohm)
T(C)
0
10
20
30
40
50
0
20
40
60
80
100

【n=3】

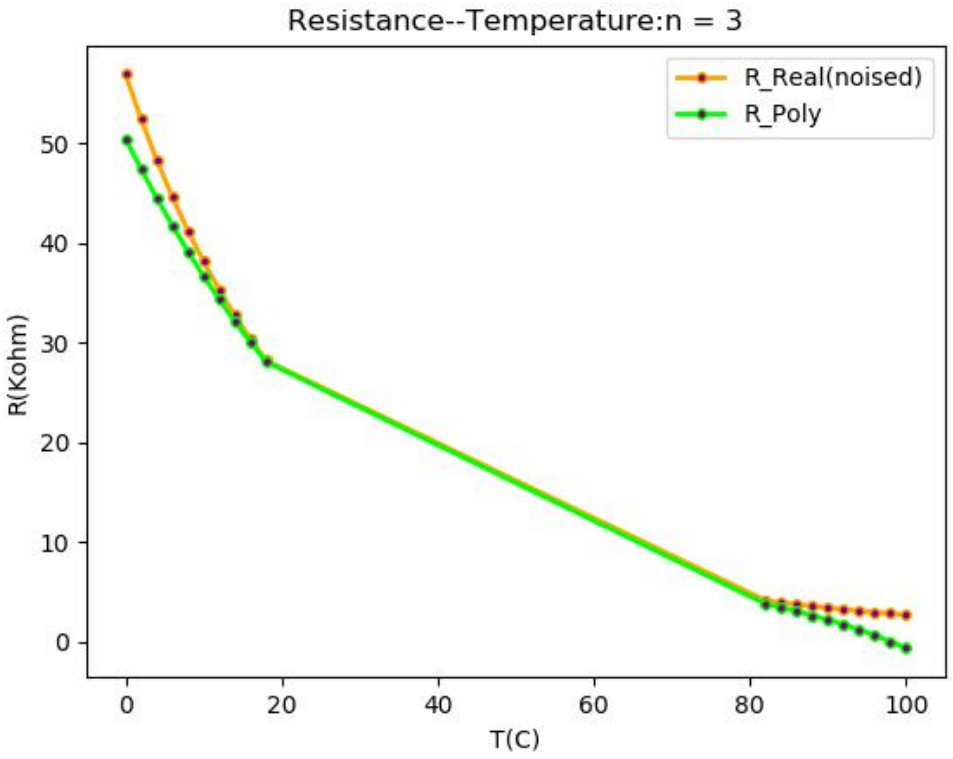
训练：


Resistance--Temperature:n = 3
R_Real(noised)
R_Poly
R(Kohm)
T(C)
5
10
15
20
25
20
30
40
50
60
70
80

R(t)=-7.499237276113137e-05t**3+0.01745973198562471t**2 -

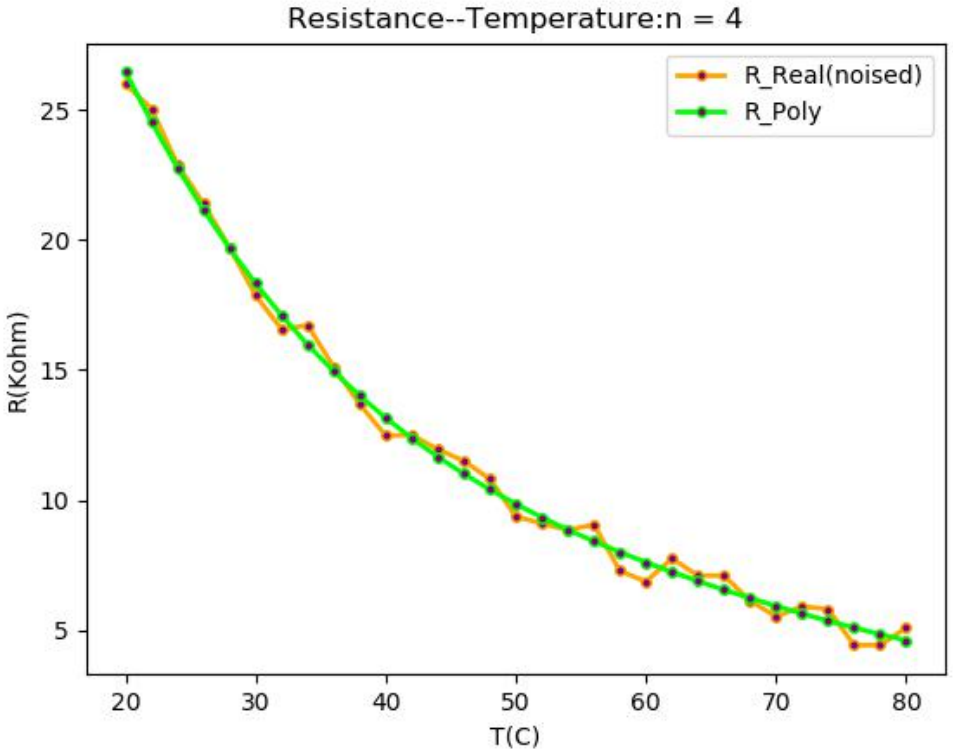
1.4797258392484666t**1 + 49.44328662490382

预测：



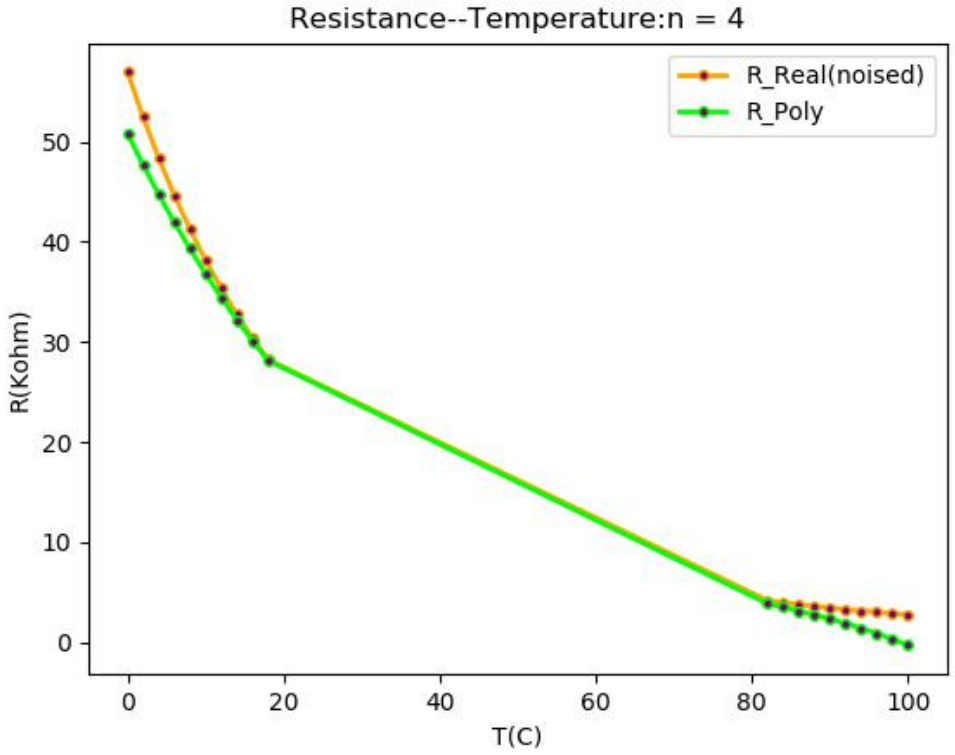
【n=4】

训练：



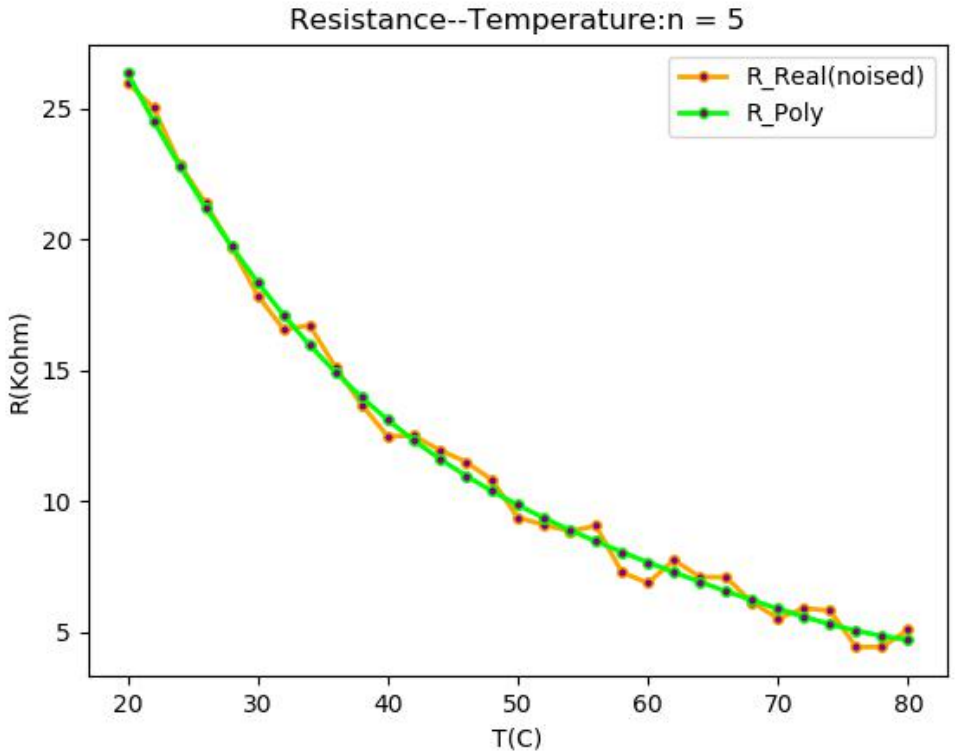
R(t) = 8.750792495621692e-07t**4 - 0.0002500082226735679t**3 + 0.029868355744416732t**2 - 1.845508965565477t**1 + 53.187169681159034

预测：


Resistance--Temperature:n = 4
R_Real(noised)
R_Poly
R(Kohm)
T(C)
0
10
20
30
40
50
0
20
40
60
80
100

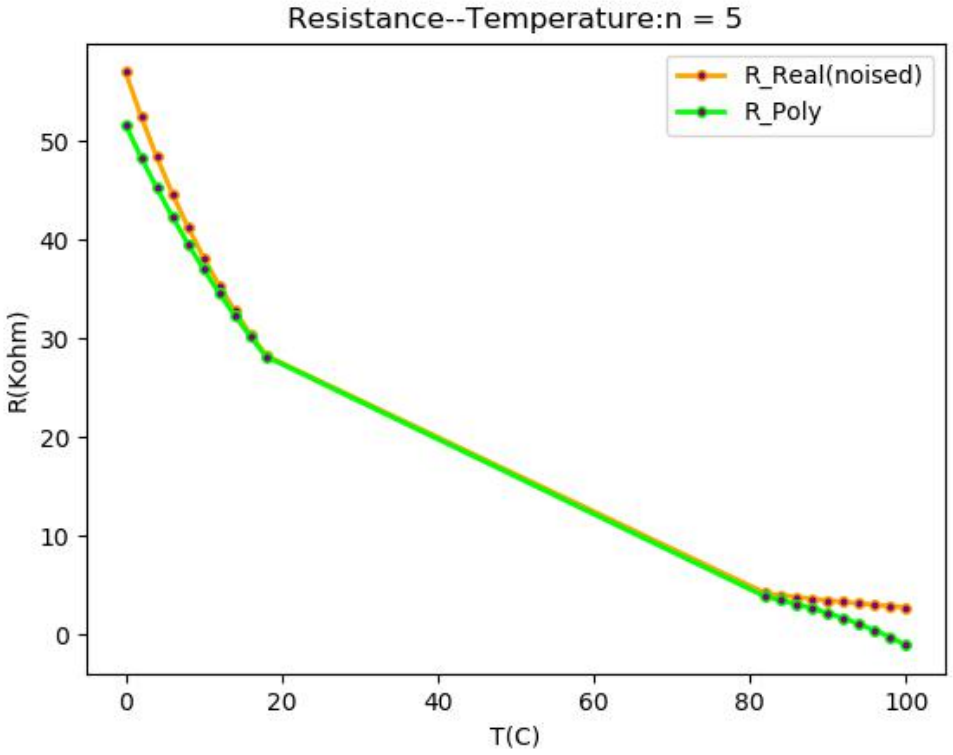
【n=5】

训练：


Resistance--Temperature:n = 5
R_Real(noised)
R_Poly
R(Kohm)
T(C)
5
10
15
20
25
20
30
40
50
60
70
80

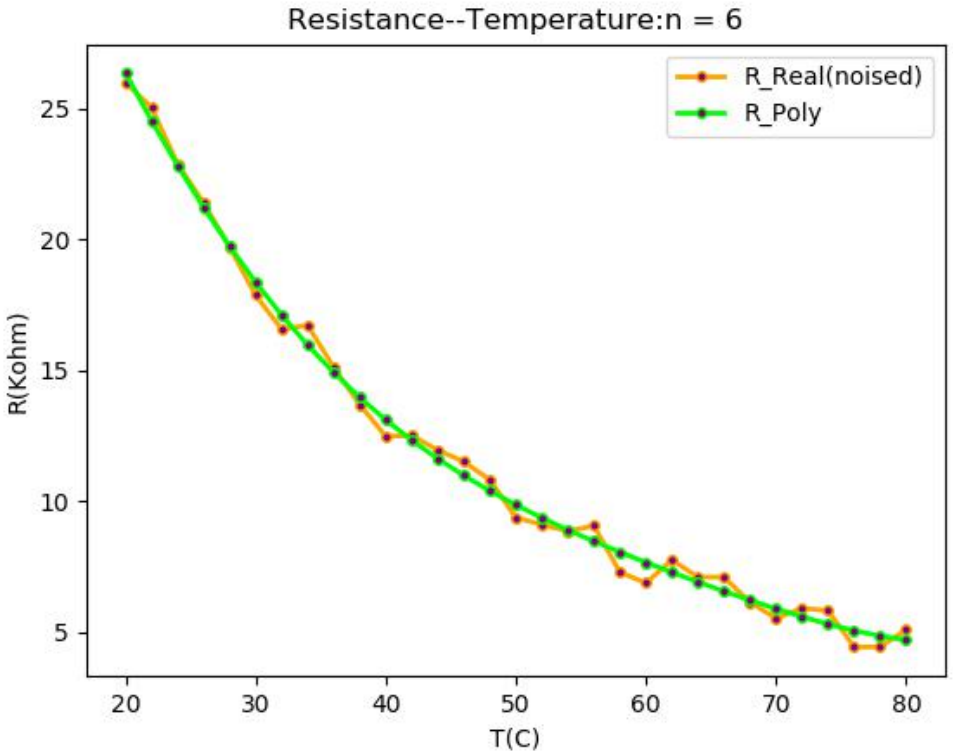
R(t) = -5.501003523197688e-08t**5 + 1.462758805755666e-05t**4 - 0.0015669484661271487t**3 + 0.08988430418250751t**2 - 3.1391460779343054t**1 + 63.684140596098324

预测：


Resistance--Temperature:n = 5
R_Real(noised)
R_Poly
R(Kohm)
T(C)
0
10
20
30
40
50
0
20
40
60
80
100

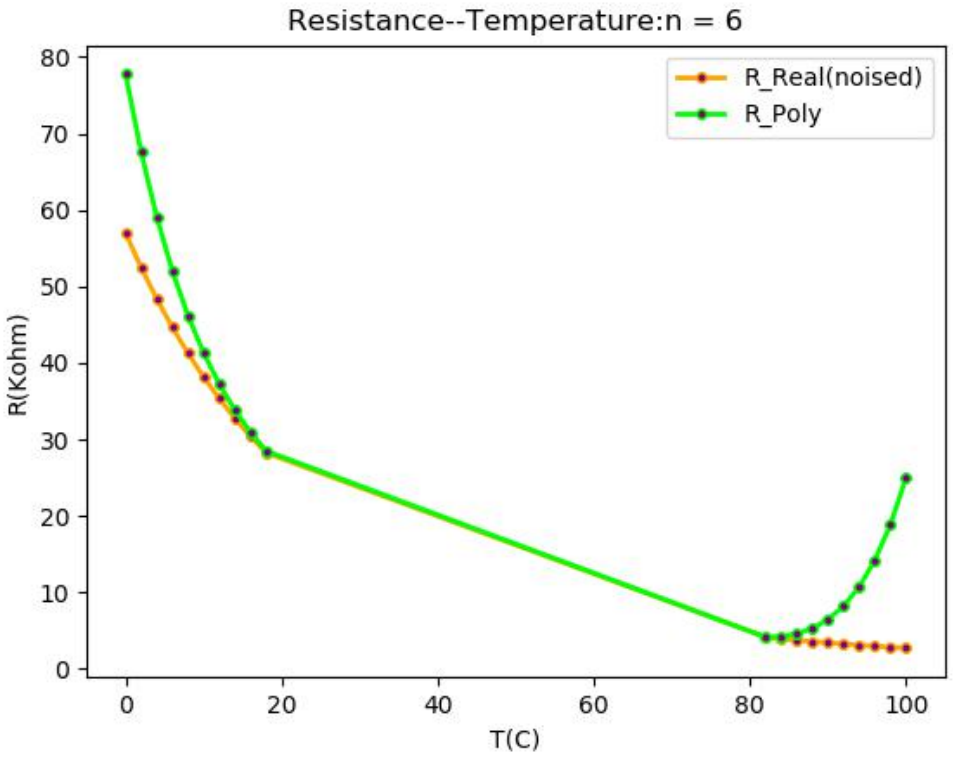
【n=6】

训练：


Resistance--Temperature:n = 6
R_Real(noised)
R_Poly
R(Kohm)
T(C)
5
10
15
20
25
20
30
40
50
60
70
80

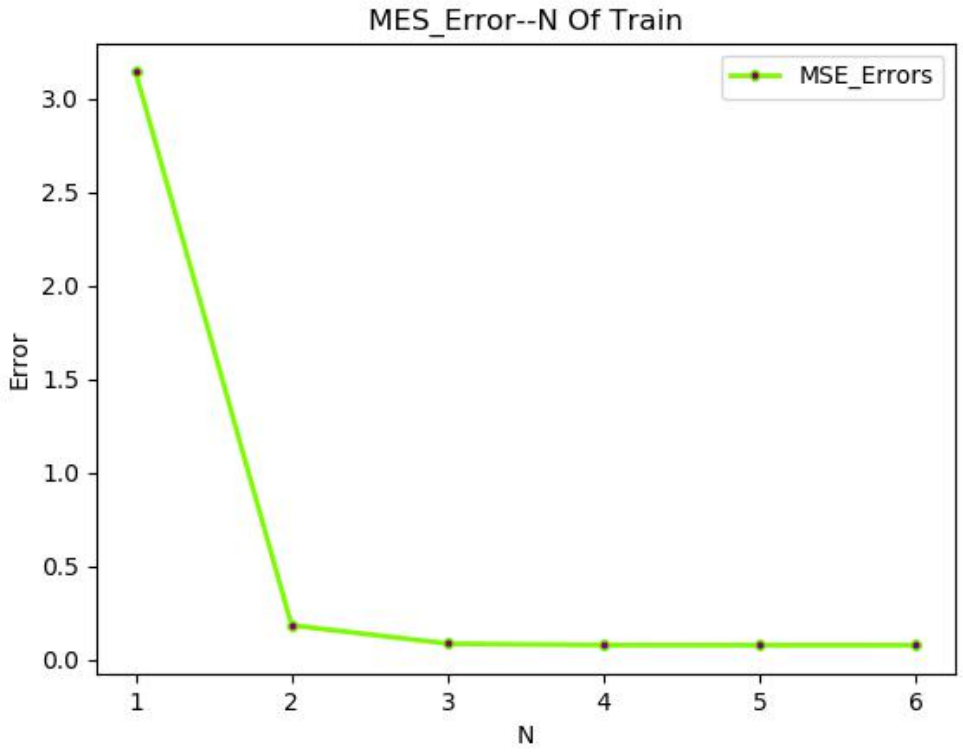
R(t) = -1.6139464165615917e-09t**6 + 4.291738897364972e-07t**5 - 4.3803141118031825e-05t**4 + 0.0020494652861827085t**3 - 0.03070129502949754t**2 - 1.0929228354753093t**1 + 49.91763449769889

预测：



在训练数据集的误差：

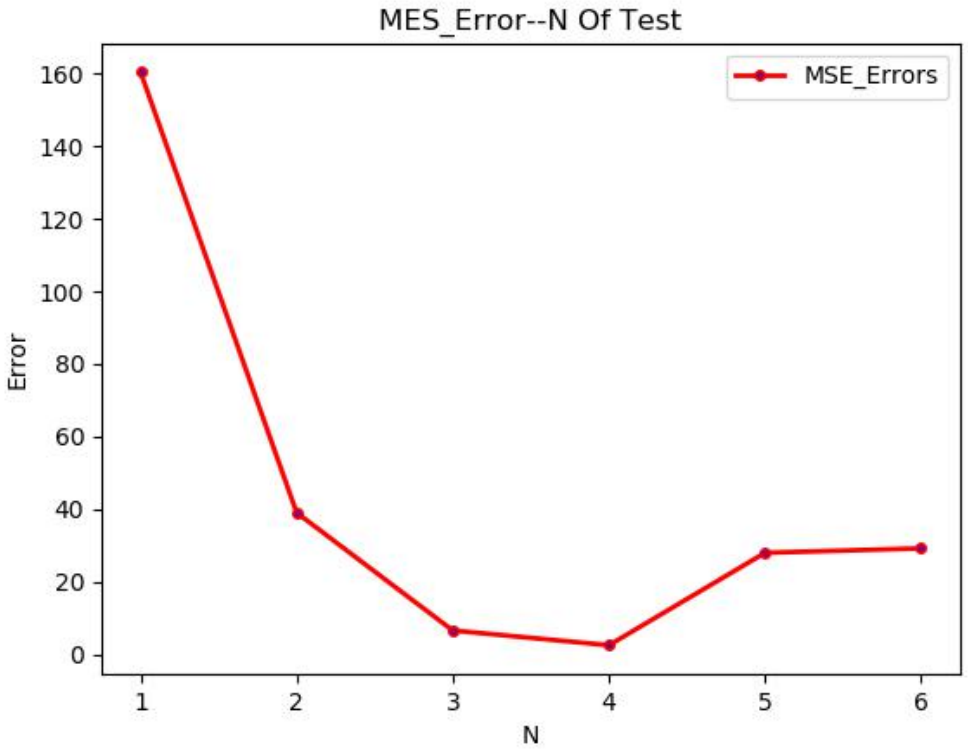
[3.122511941703357, 0.1570685180049203, 0.057325471057479656, 0.05469256990869753, 0.05457570797714998, 0.05097125119964778]（n 从 1 到 6，如下图所示）



(*横轴表示拟合次数)

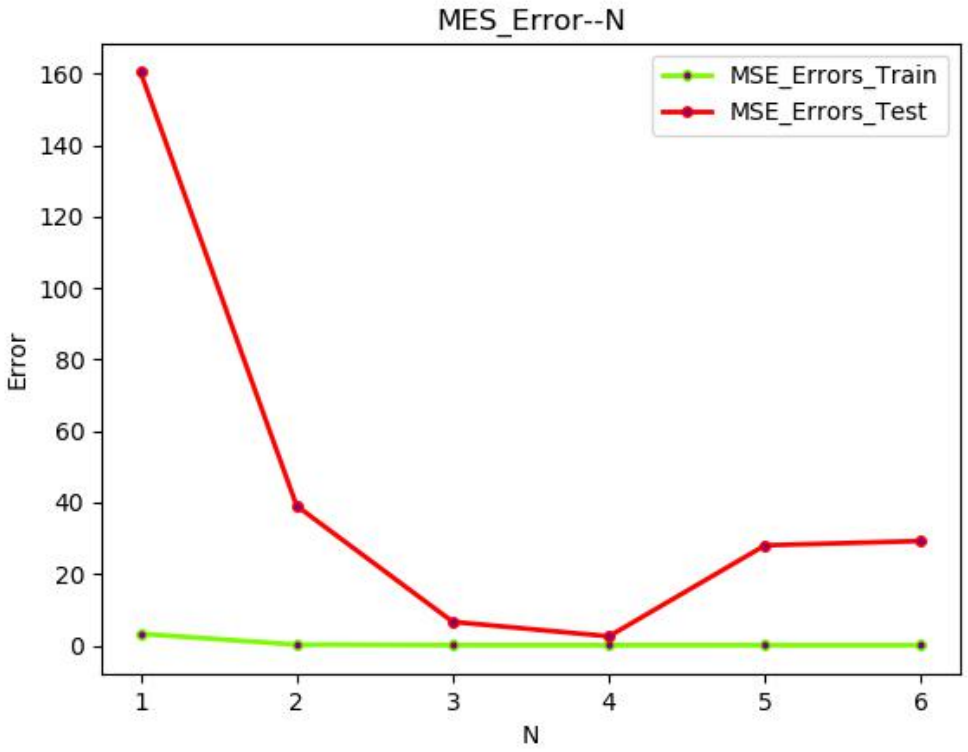
在测试数据集上的误差:

[160.5469174771094, 39.04607664742966, 9.62208772899972, 3.9249176820725027, 1.5024628695709512, 181.24754027371978]（n 从 1 到 6，如下图所示）



(*横轴表示拟合次数)

两个误差对比:



（*横轴表示拟合次数）

观察误差可以看出。对于训练集，拟合多项式阶次越高，误差越小。而对于测试集，拟合误差随着阶次的升高，先降低，后升高。升高的原因便是出现了过拟合。导致模型的普适性降低，误差反而增大。

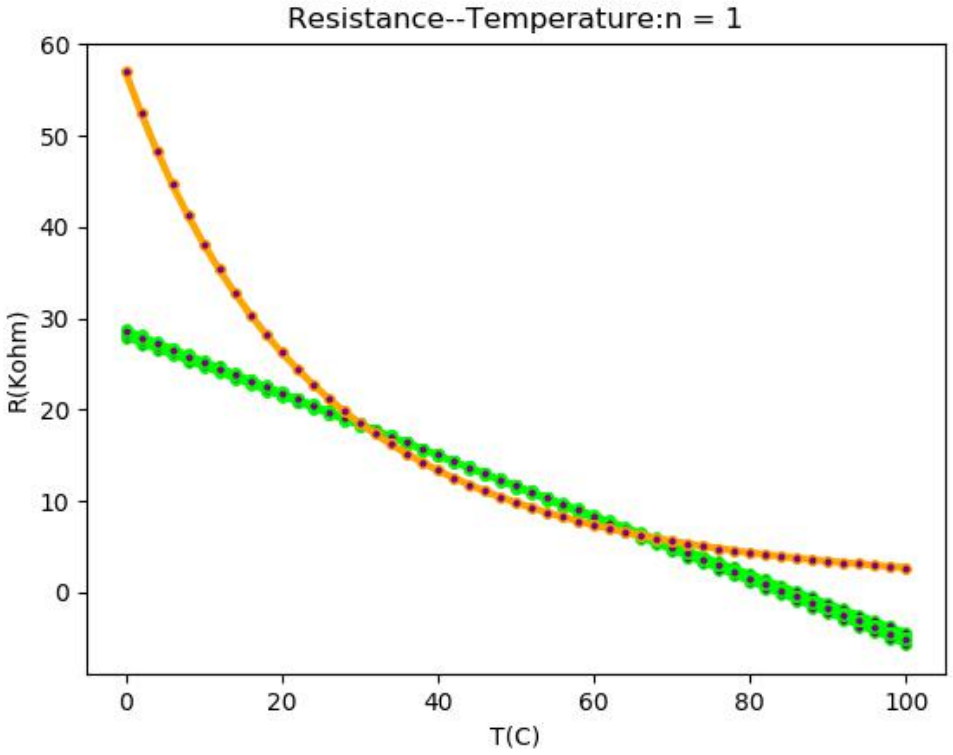
同时有两条误差曲线可以看出，当多项式阶次为 4 时，训练和拟合效果都比较好。

3）重复 2）相应内容 10 次（每次重新添加噪声模拟不同批次实验数据），观察并讨论由于采用不同训练数据给拟合（学习）结果带来的影响；

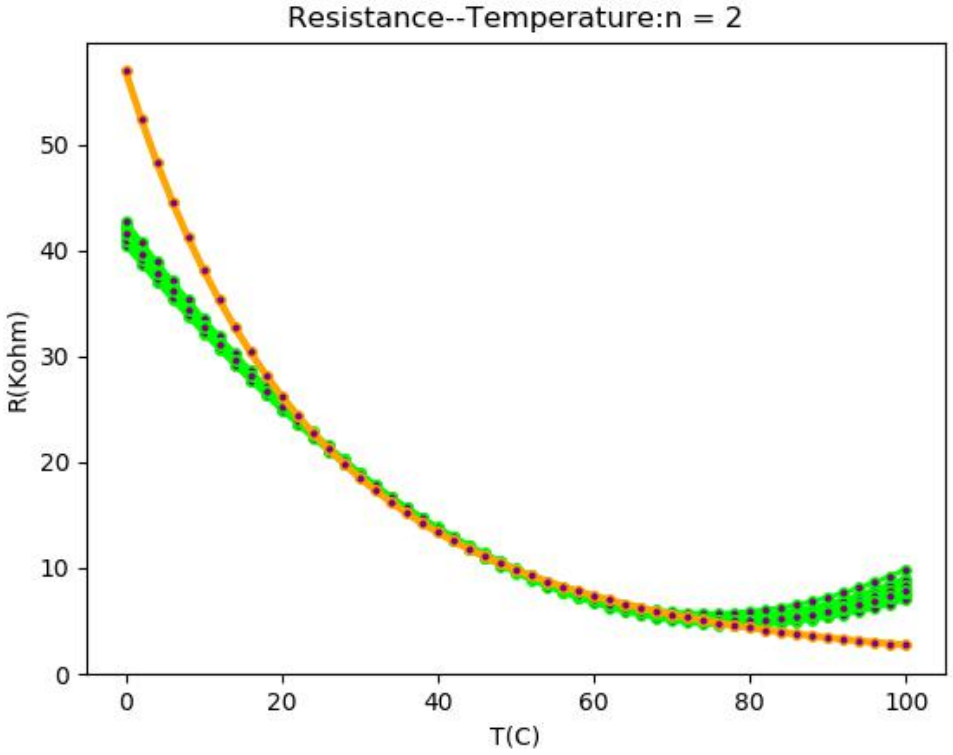
解：重复了 10 次实验

分别绘制 10 次实验的训练和测试图如下：

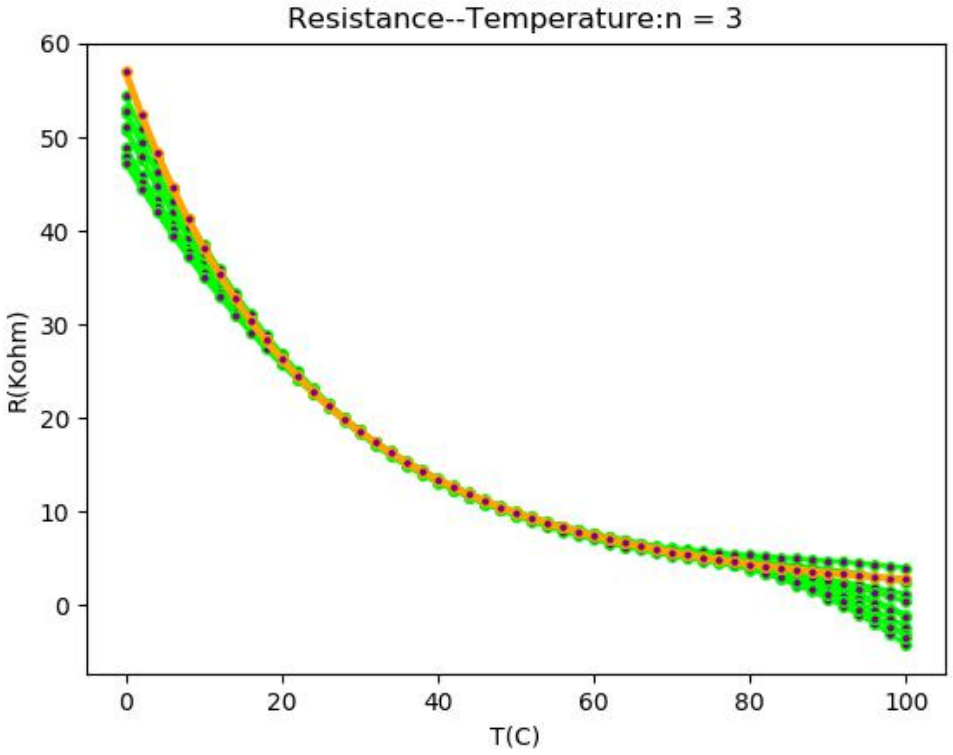
N=1


Resistance--Temperature:n = 1
R(Kohm)
T(C)
0
10
20
30
40
50
60
0
20
40
60
80
100

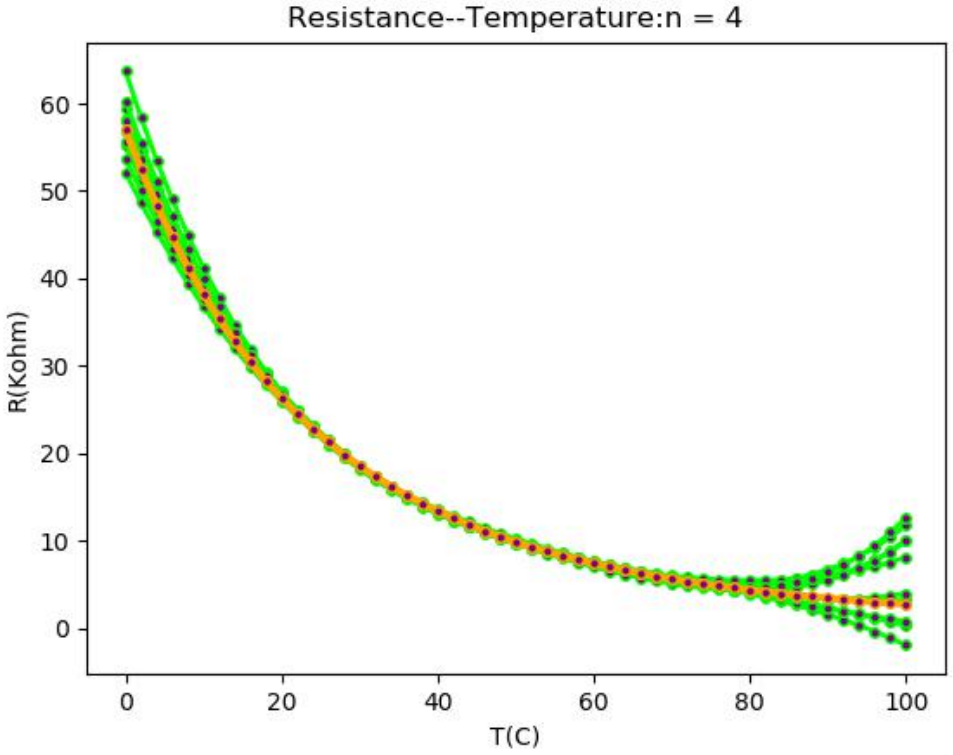
N=2


Resistance--Temperature:n = 2
R(Kohm)
T(C)
0
10
20
30
40
50
0
20
40
60
80
100

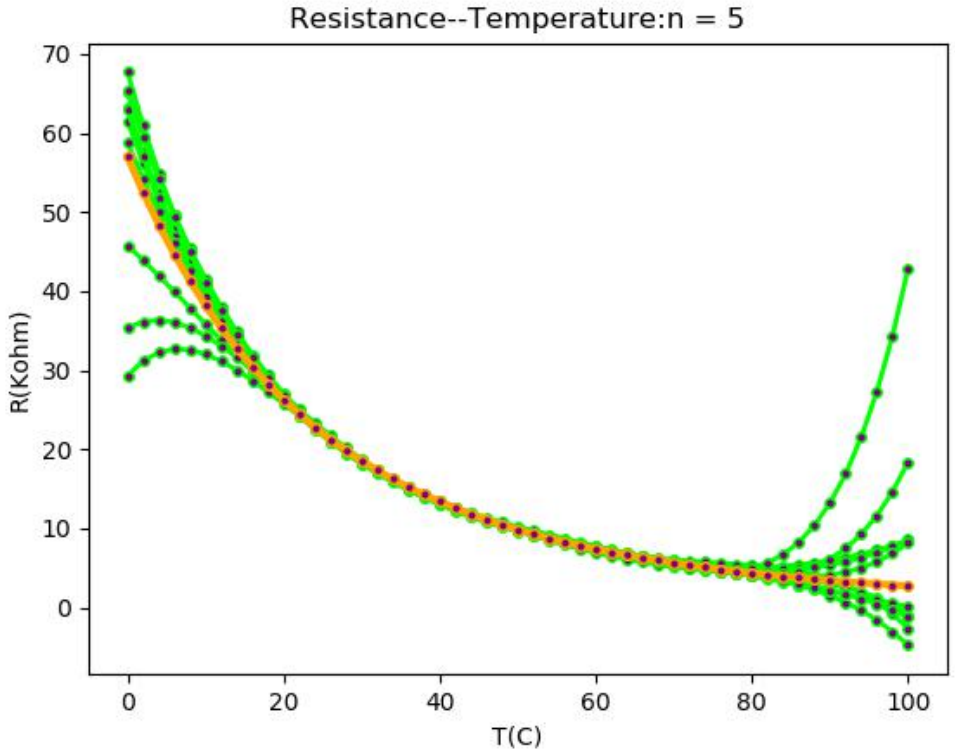
N=3


Resistance--Temperature:n = 3
R(Kohm)
T(C)
0
10
20
30
40
50
60
0
20
40
60
80
100

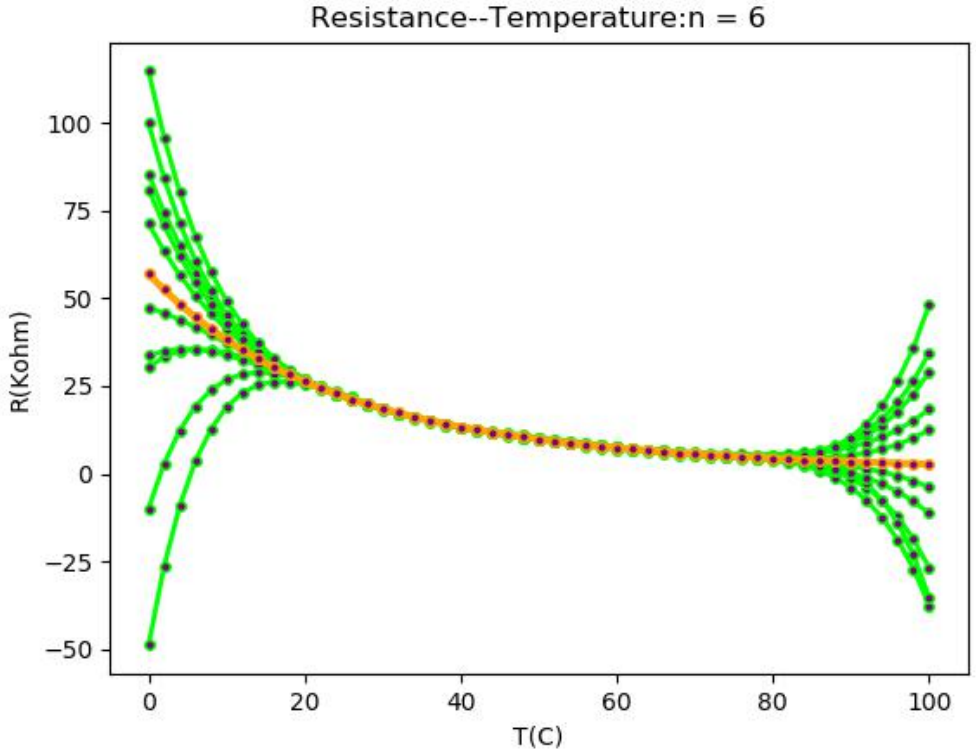
N=4


Resistance--Temperature:n = 4
R(Kohm)
T(C)
0
10
20
30
40
50
60
0
20
40
60
80
100

N=5



N=6

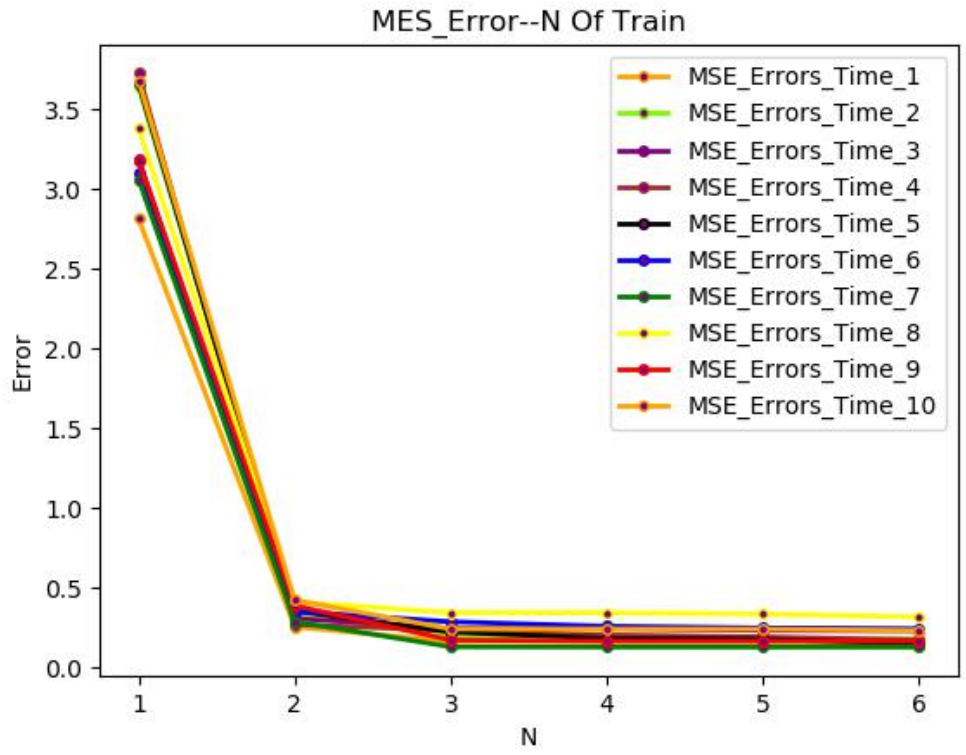


通过观察上面的拟合效果不难发现，随着阶次的升高，多次实验的拟合结果

在测试集上的方差逐渐增大。比如：当阶次为 1 时，10 条拟合曲线方差较小；然后，当阶次升高时，虽然在训练集上的表现逐渐提高，但是 10 条曲线在测试集上的表现方差较大。

为了减小高阶数据对模型的敏感性，可以多次实验取平均，这样可以减小高阶模型在测试集上的误差。

在训练数据集上的误差如下图所示 ：

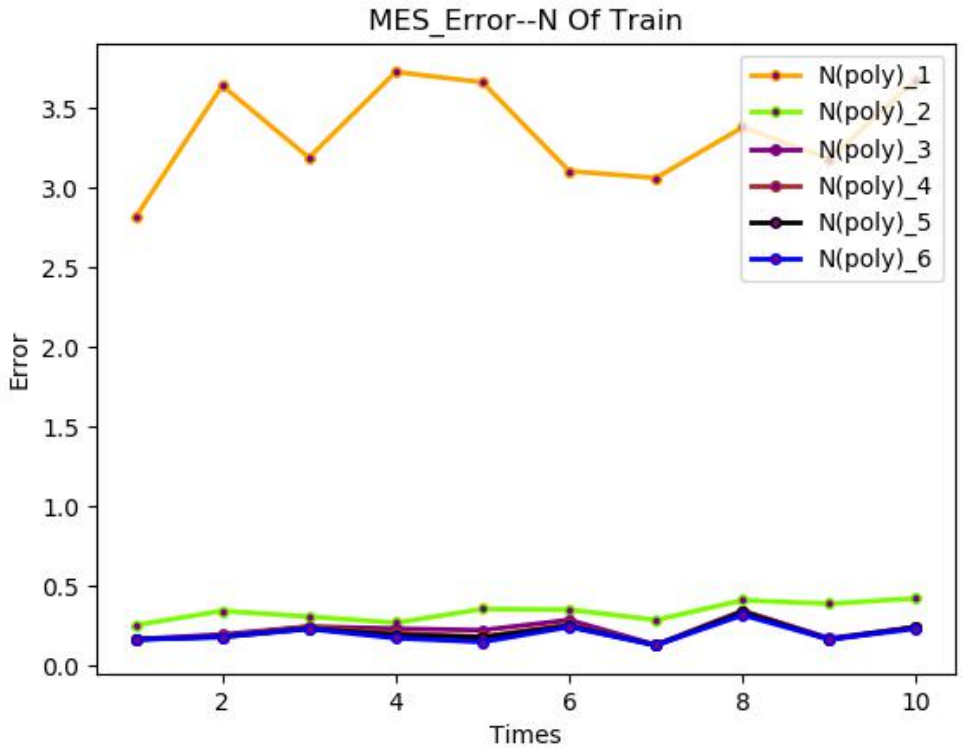


(*横轴表示拟合次数)

可见，在相同强度的不同的噪声影响下，训练误差有类似的下降过程。阶次越高，误差越小，且本次实验，在阶次为 2 时，训练误差已经比较小了。

各曲线的位置标明，不同组的实验结果会有差异，且拟合好过也会有好有坏，但是整体表现相似，不会因为实验次序而受到较大影响。

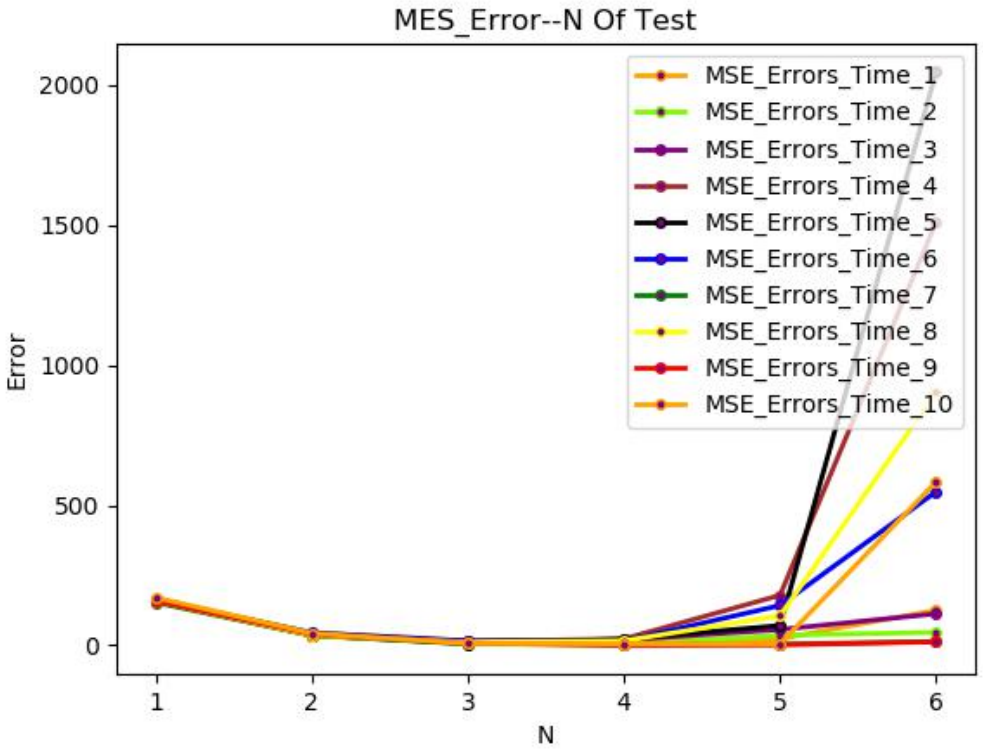
横向比较 10 次实验不同的阶次的训练集 MES 结果如下：



（*横轴表示 10 次实验，不同的曲线表示不同的拟合阶次）。

可以看到，相同阶次的拟合过程在不同的训练数据集上有类似的表现。阶次越高，曲线越低，即误差越小。

对应的 10 次实验在训练集上的结果如下：



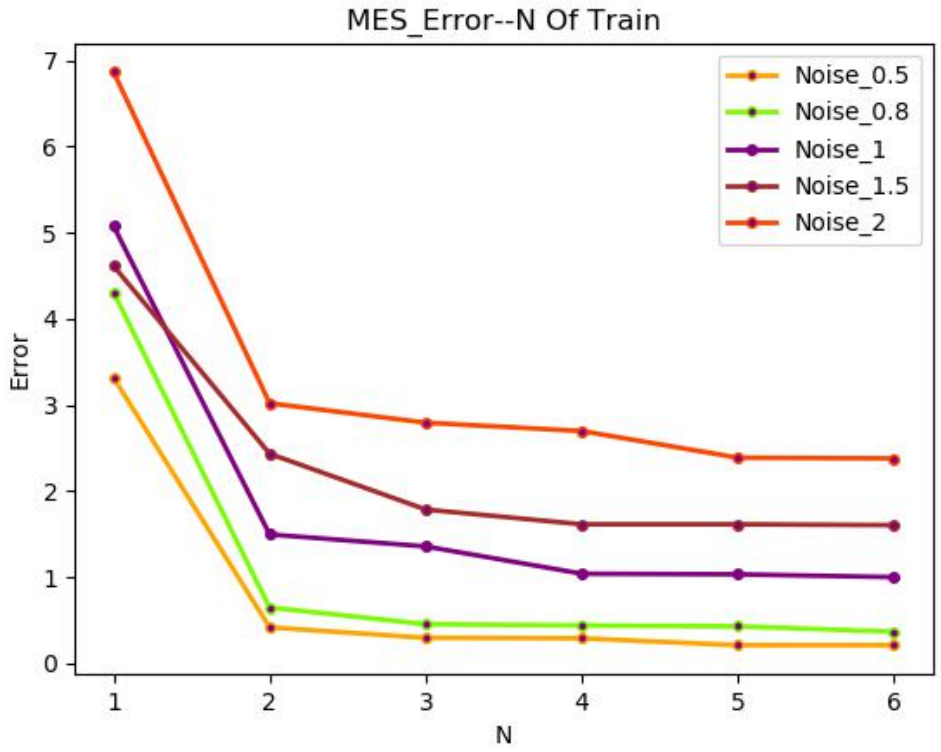
每一条误差都是先减小，后增大。且当几次为 6 次时，出现了明显的过拟合，导致测试误差显著增大。

对于不同次的实验，有一定的误差，但与实验次数没有肉眼可观的关系，这也符合一般的实验规律，即不同的实验得出的数据可能有差异，但是总体走势相同。

4）改变噪声强度（通过改变所加噪声的标准偏差实现），重复 2），3）内容，观察并讨论数据中不同噪声强度给拟合（学习）带来的影响；

4.1）添加不同噪声重复任务 2:

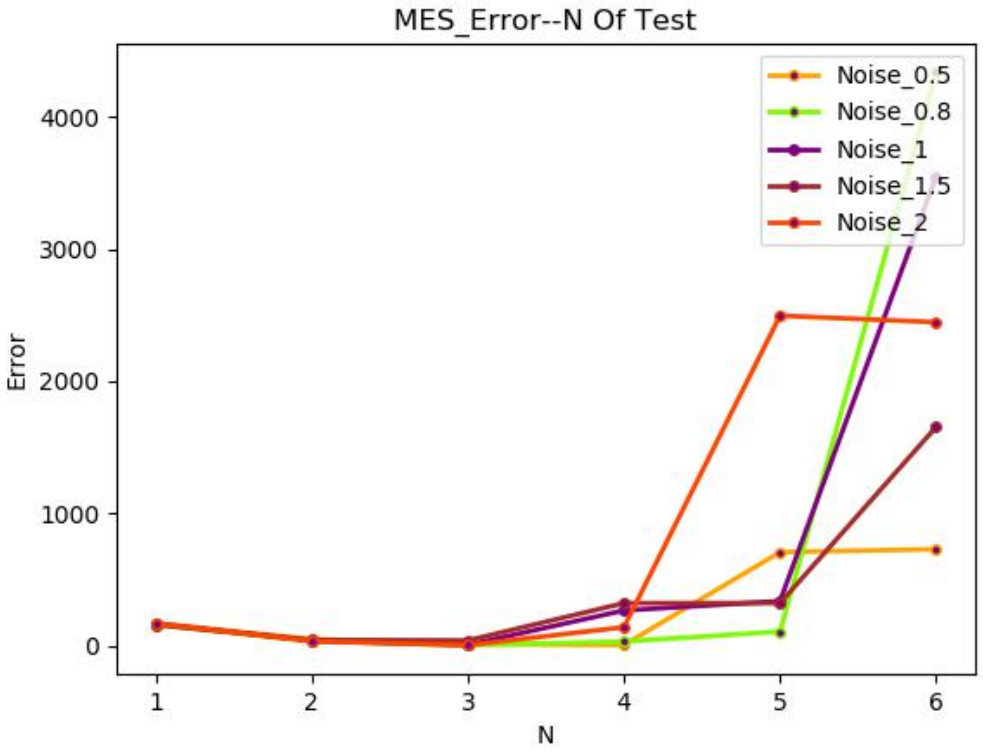
在训练集上的表现。



（*横坐标表示拟合多项式的最高项系数；图例中，数字表示添加噪声的方差）

从结果可以看到，噪声的方差越大，拟合结果在训练集上的均方差越大。各个均方差随阶次的变化相似，醉着阶次的升高，误差减小。

在测试集上的均方差如下图所示。



（*横坐标表示拟合多项式的最高项系数；图例中，数字表示添加噪声的方差）

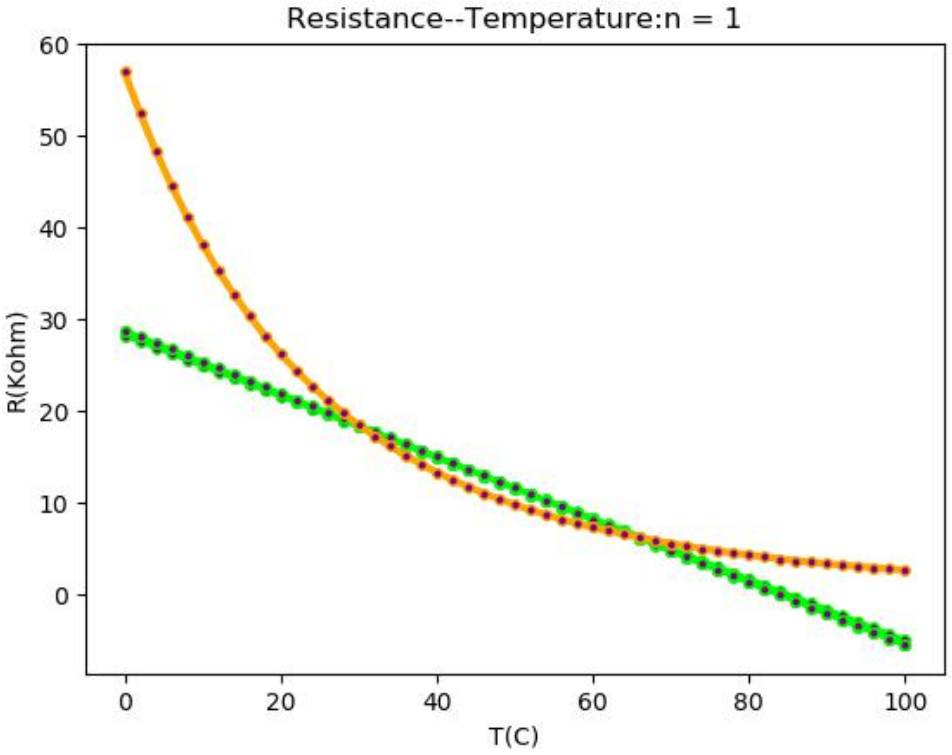
从实验结果可以看出，拟合结果在测试集上的结果的差异不是很明显。在拟合次数比较高时，差异性出现，并且经过多次实验方向，其波动性性也很大。

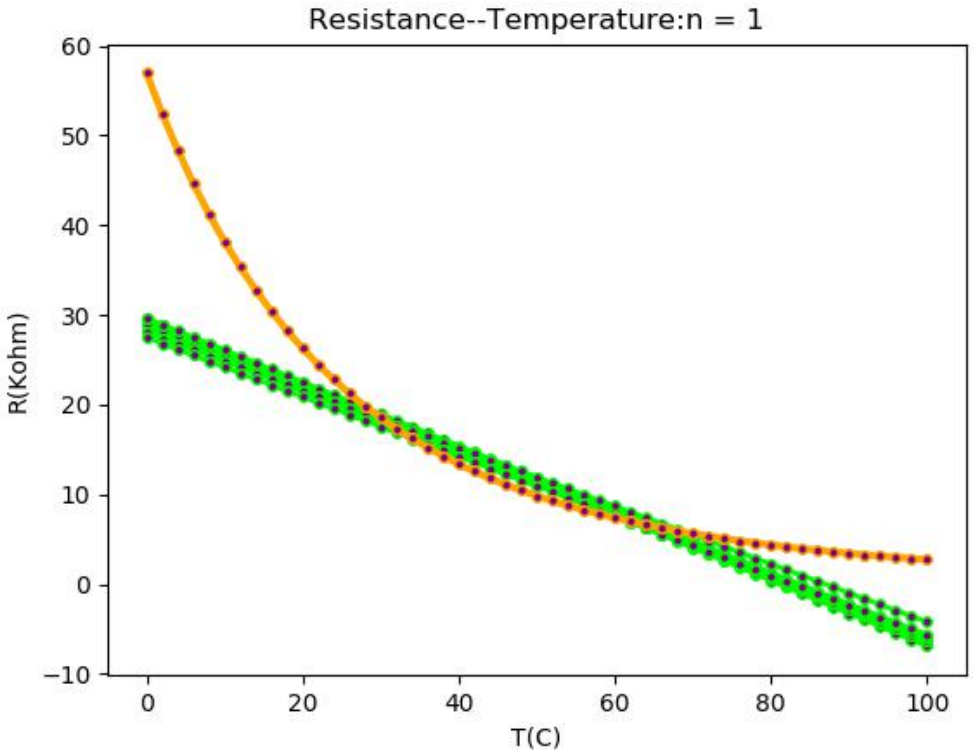
各个曲线的走势同之前的实验。随着阶次的升高，误差先减小，后因出现过拟合，测试误差增大。

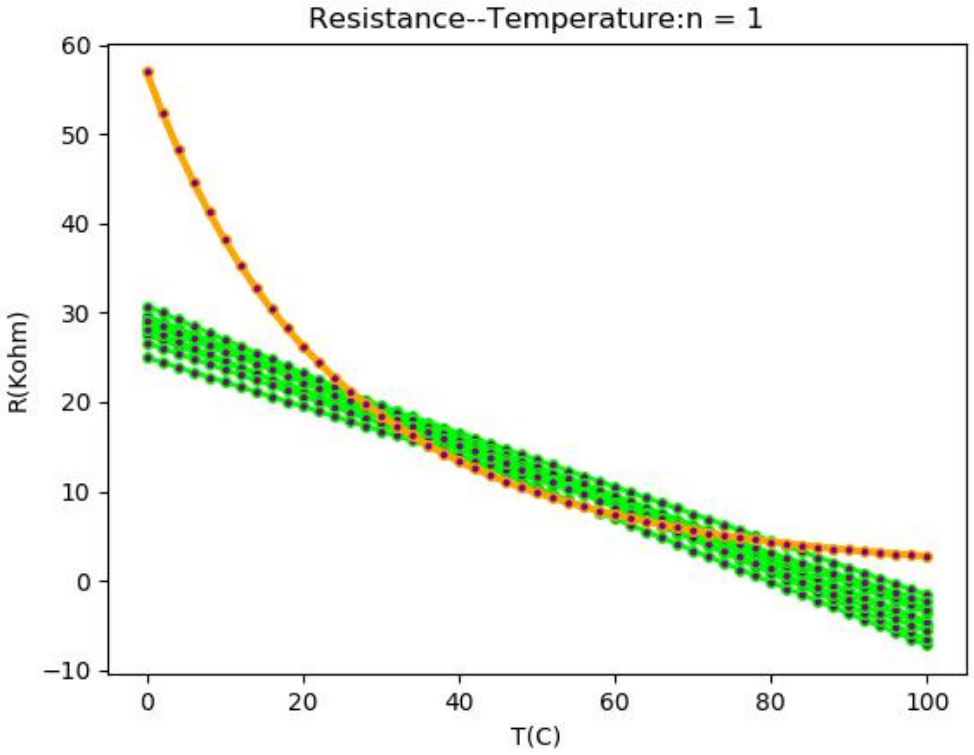
4.2）添加不同噪声重复任务 3

为了便于观察，此处选取了 3 个方差进行模拟：分别是 0.5 , 2，5

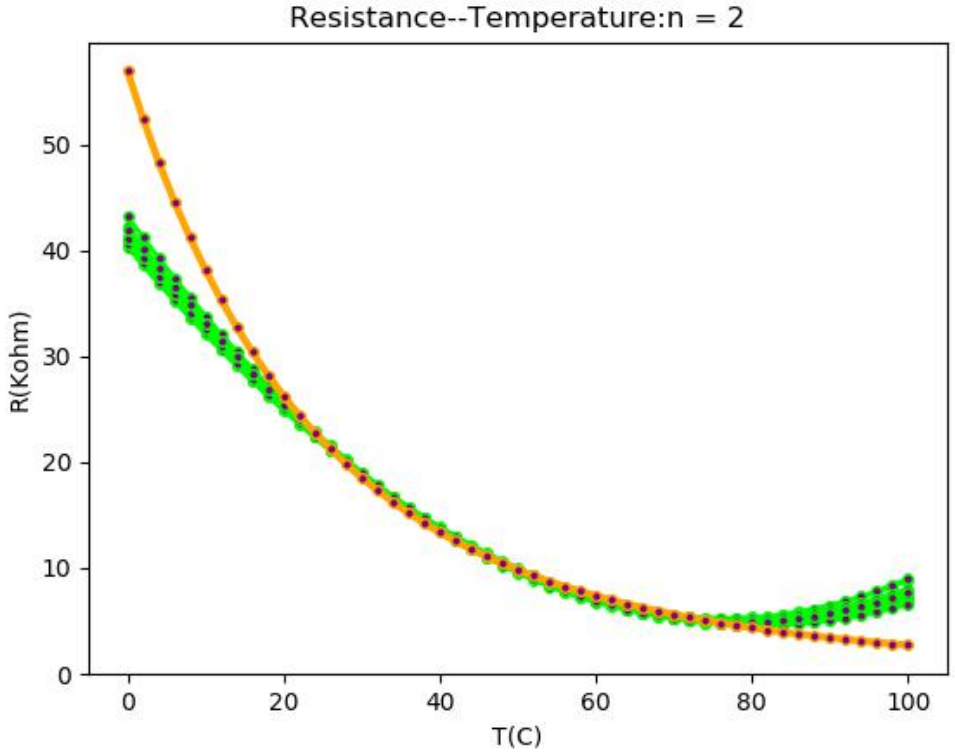
N=1（噪声方差依次是 0.5,2,2.5）

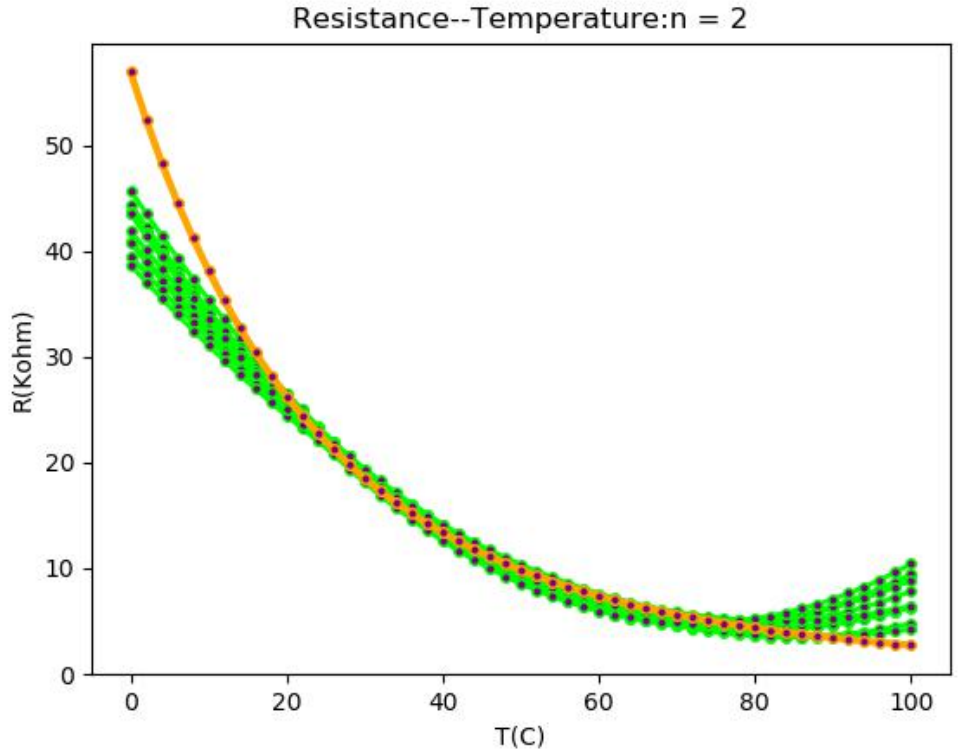

Resistance--Temperature:n = 1
R(Kohm)
T(C)
60
50
40
30
20
10
0
0
20
40
60
80
100

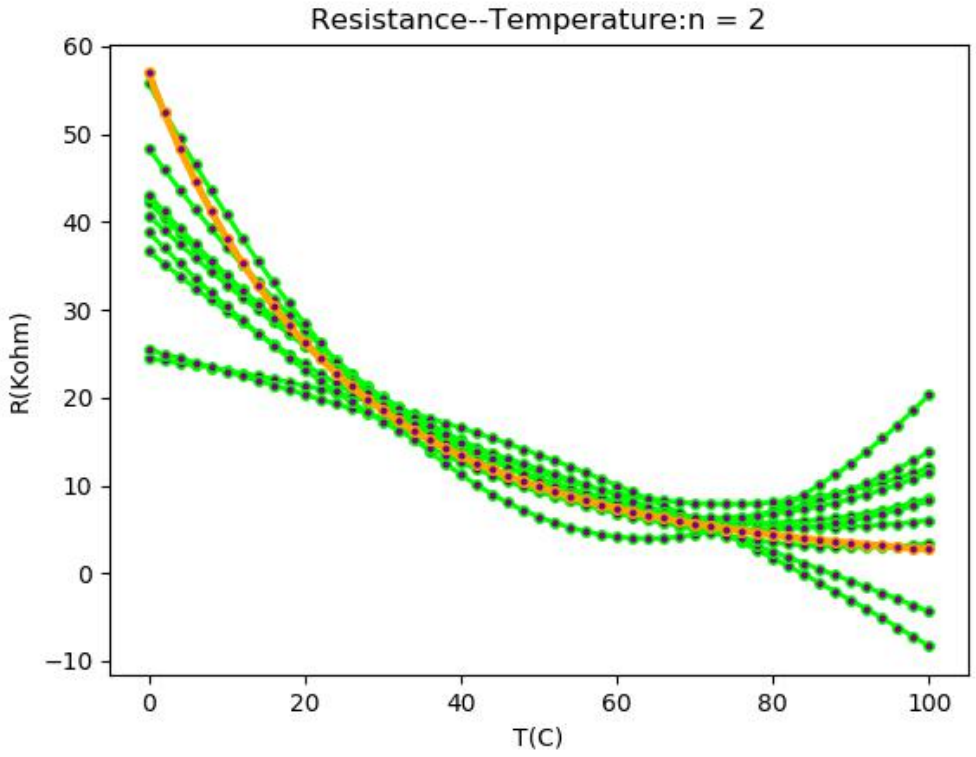

Resistance--Temperature:n = 1
R(Kohm)
T(C)
60
50
40
30
20
10
0
−10
0
20
40
60
80
100



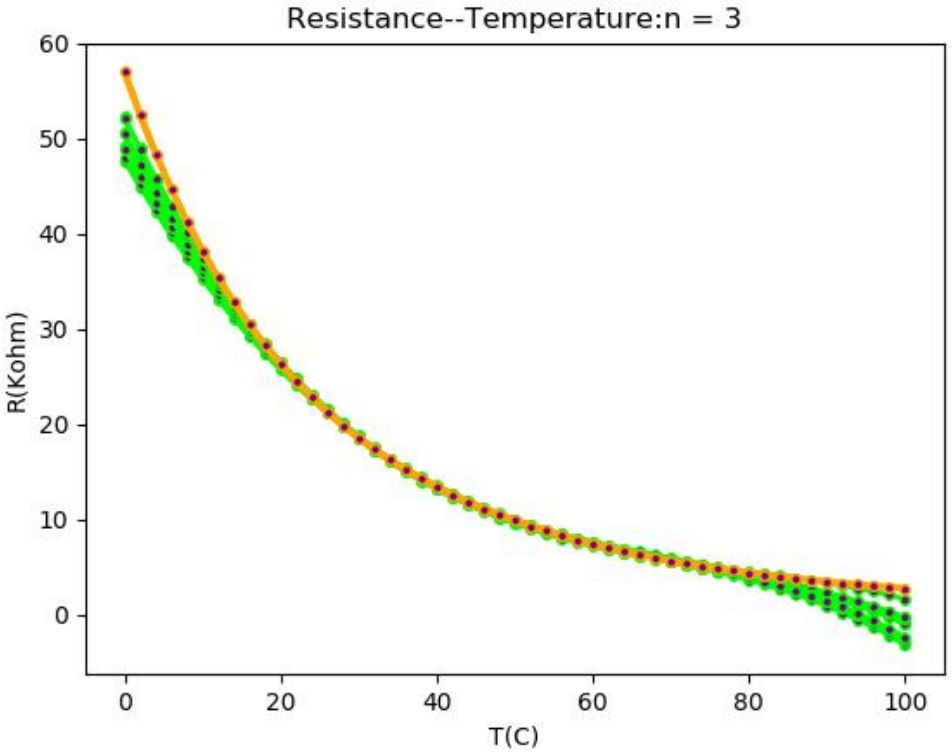
N=2（噪声方差依次是 0.5,2,2.5）

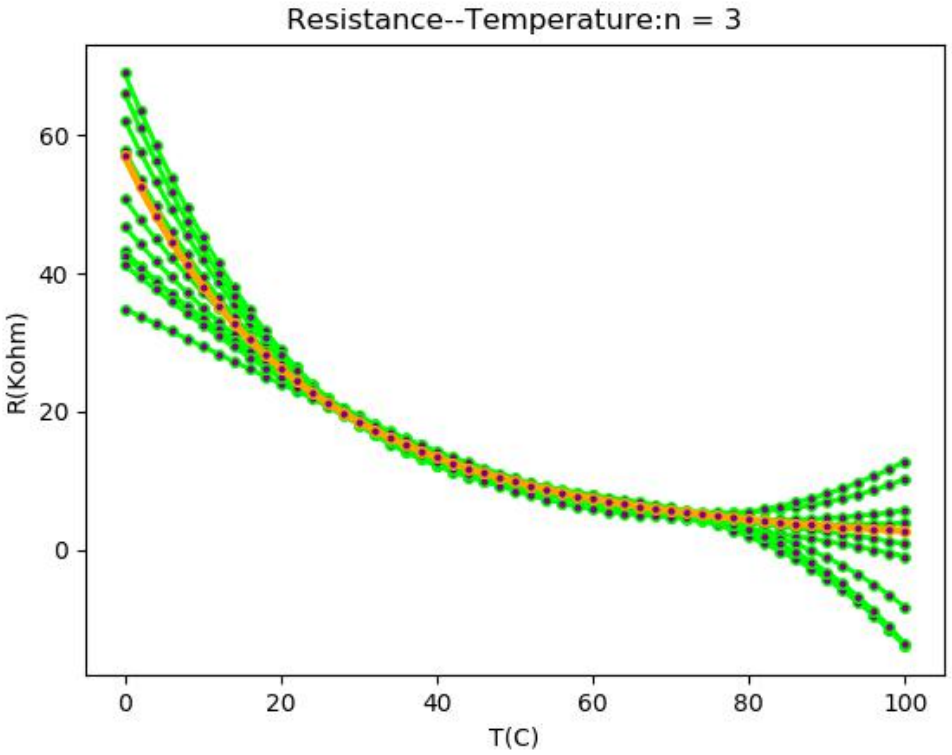


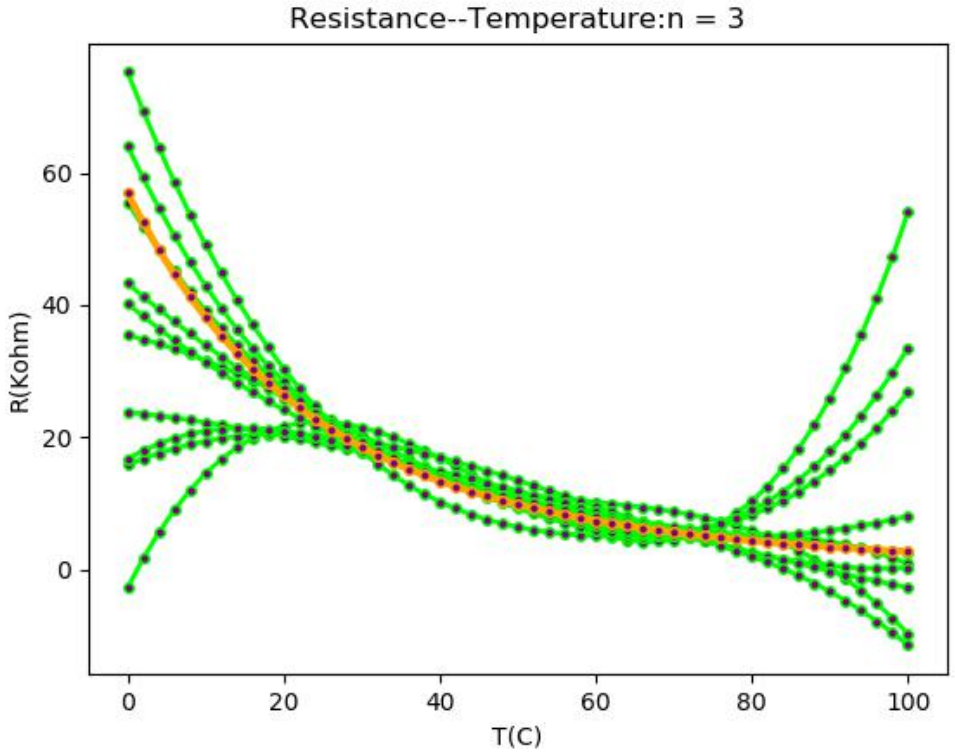




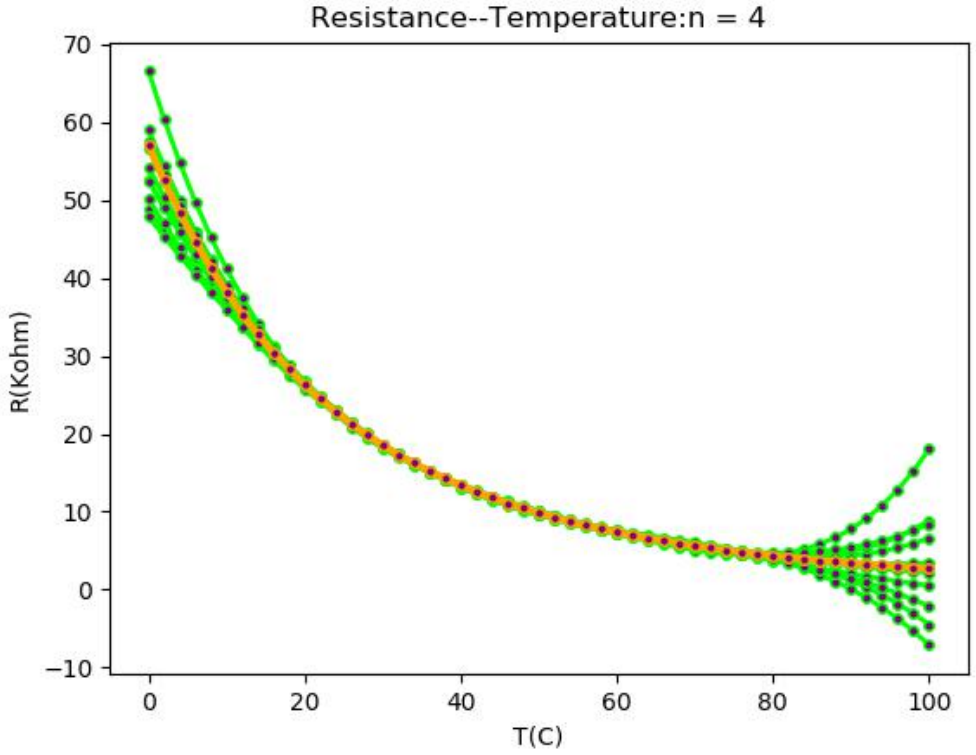
N=3（噪声方差依次是 0.5,2,2.5）

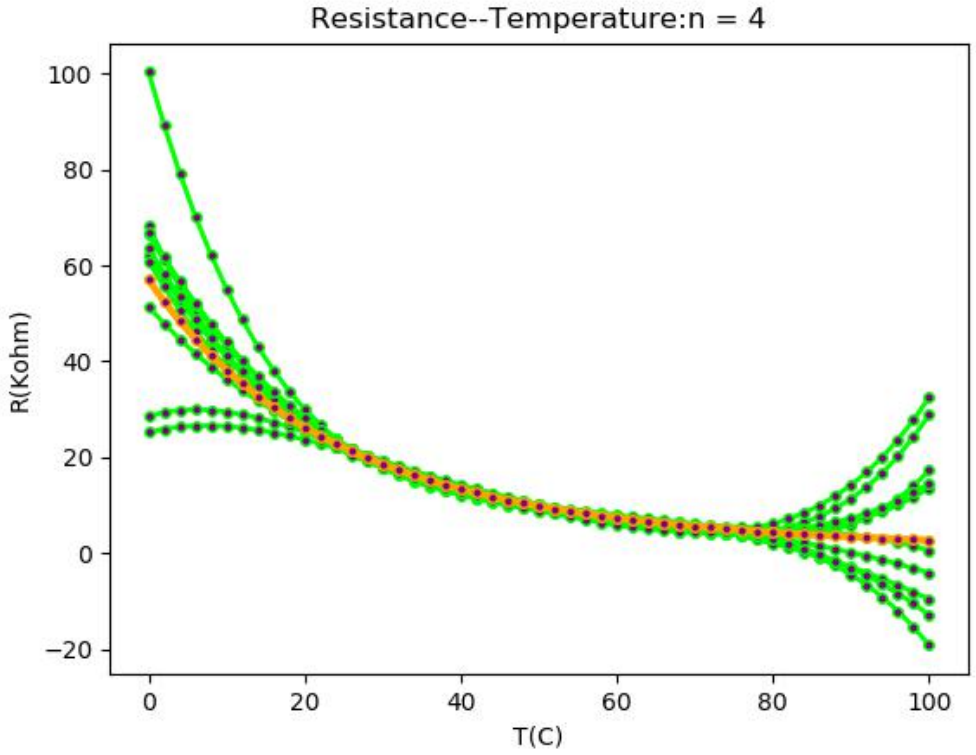

Resistance--Temperature:n = 3
R(Kohm)
T(C)
0
10
20
30
40
50
60
0
20
40
60
80
100

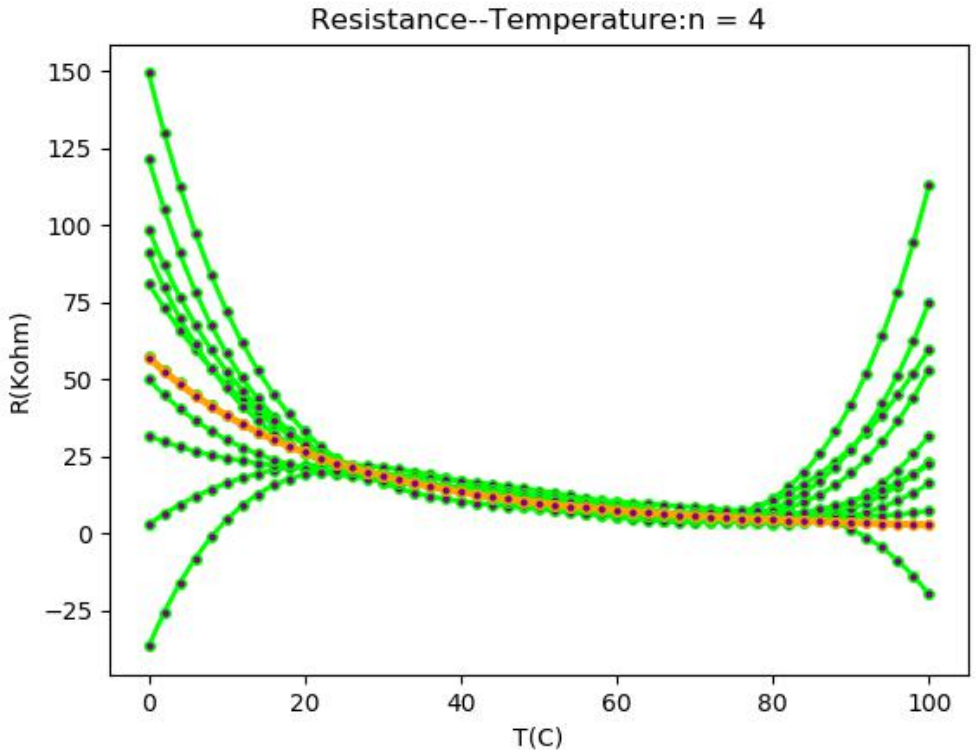

Resistance--Temperature:n = 3
R(Kohm)
T(C)
0
20
40
60
0
20
40
60
80
100



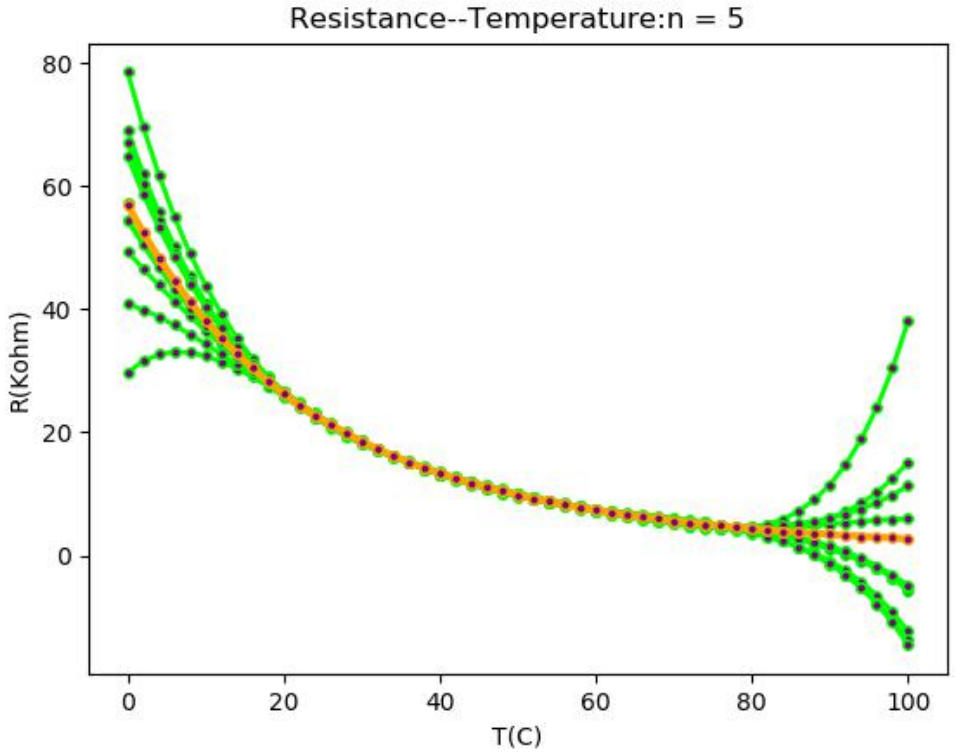
N=4（噪声方差依次是 0.5,2,2.5）

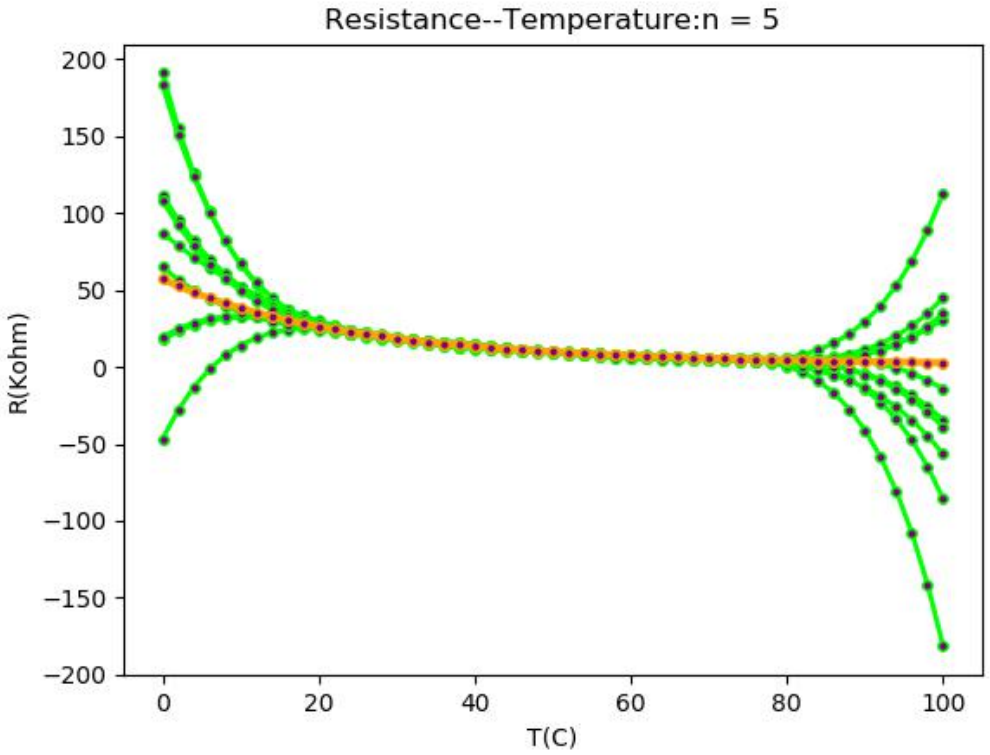


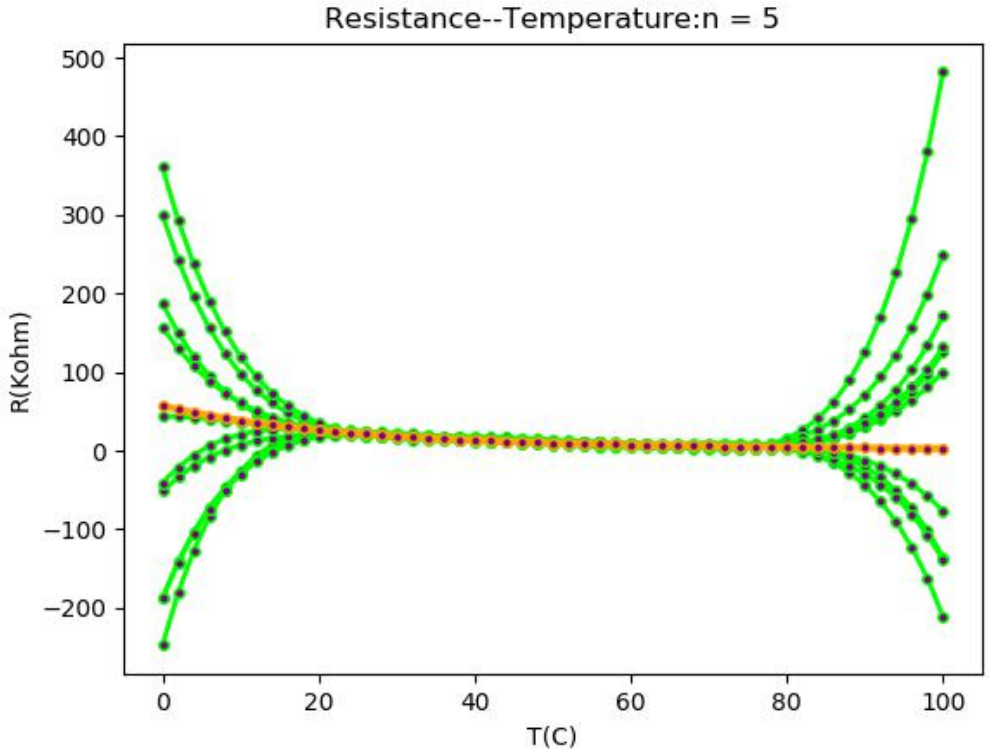




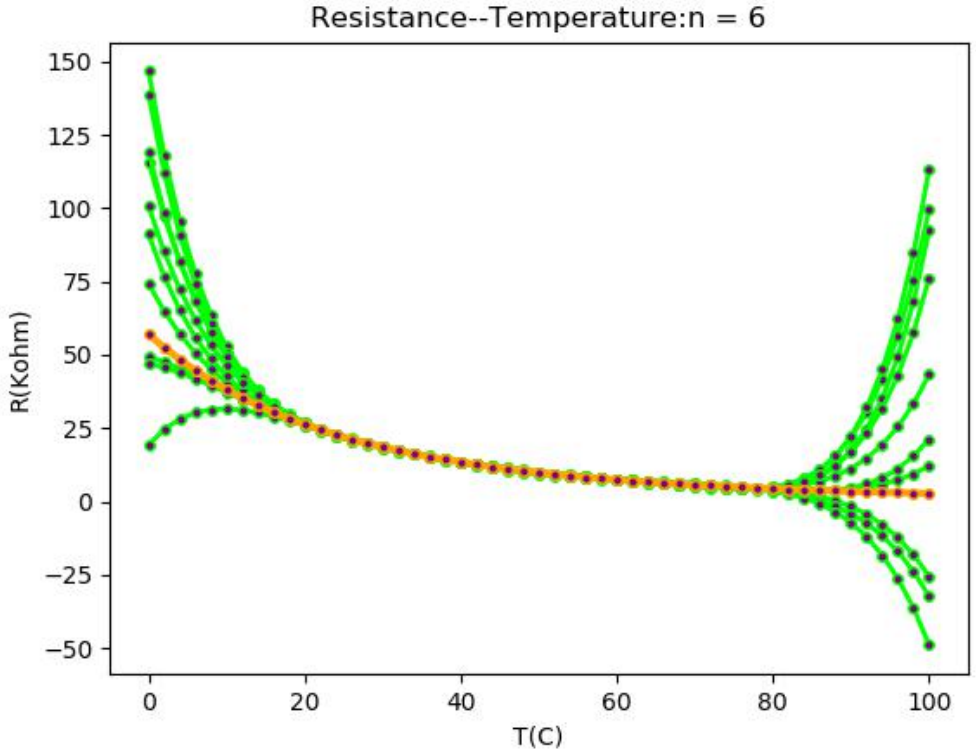
N=5（噪声方差依次是 0.5,2,2.5）

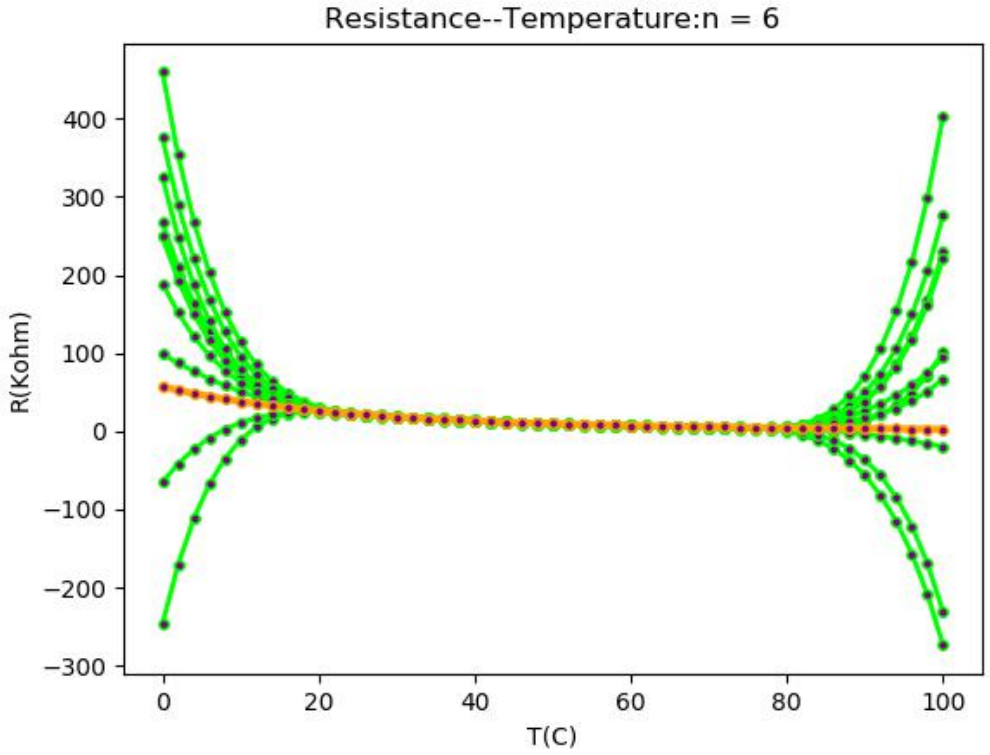

Resistance--Temperature:n = 5
R(Kohm)
T(C)
80
60
40
20
0
0
20
40
60
80
100

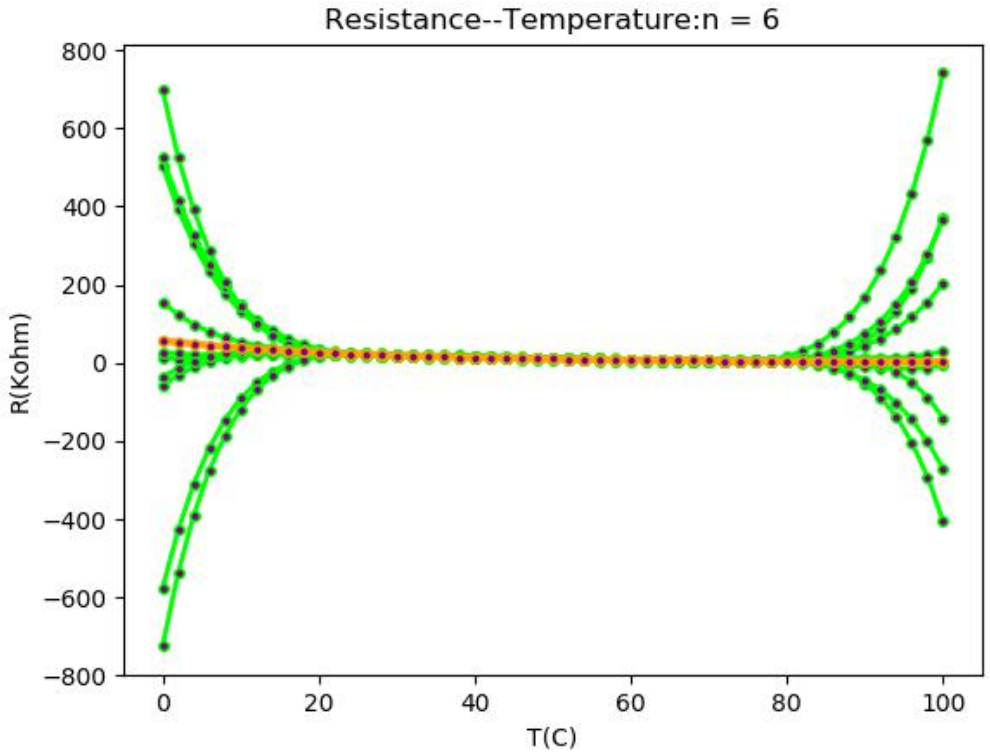

Resistance--Temperature:n = 5
R(Kohm)
T(C)
200
150
100
50
0
−50
−100
−150
−200
0
20
40
60
80
100



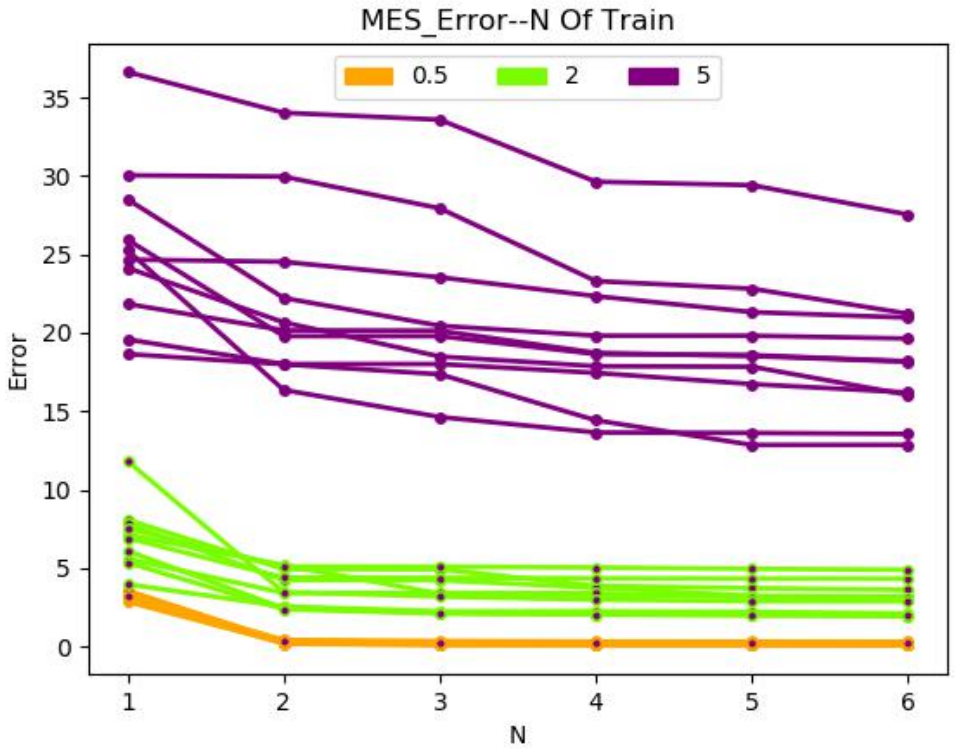
N=6（噪声方差依次是 0.5,2,2.5）







在训练集上的结果如下图所示：



（*横坐标是拟合多项式的阶次。纵坐标是均方差。每一个方差的实验进行 10 次。图中共有 3*10=30 组实验的结果。图例对应的是添加的噪声的方差）

由实验结果可以明显的看出，噪声的方差越大，拟合模型在训练集上的均方差越大。对于同一个均方差，在噪声比较小时，各次实验结果**波动小**。随着噪声的增大，拟合结果在各次实验当中的波动也逐渐增大。

同时噪声越大，无论是训练集或者测试集，所得模型的误差也越大。因此，在实际实验当中，精确测量实验数据有利于我们得到更准确的结论。

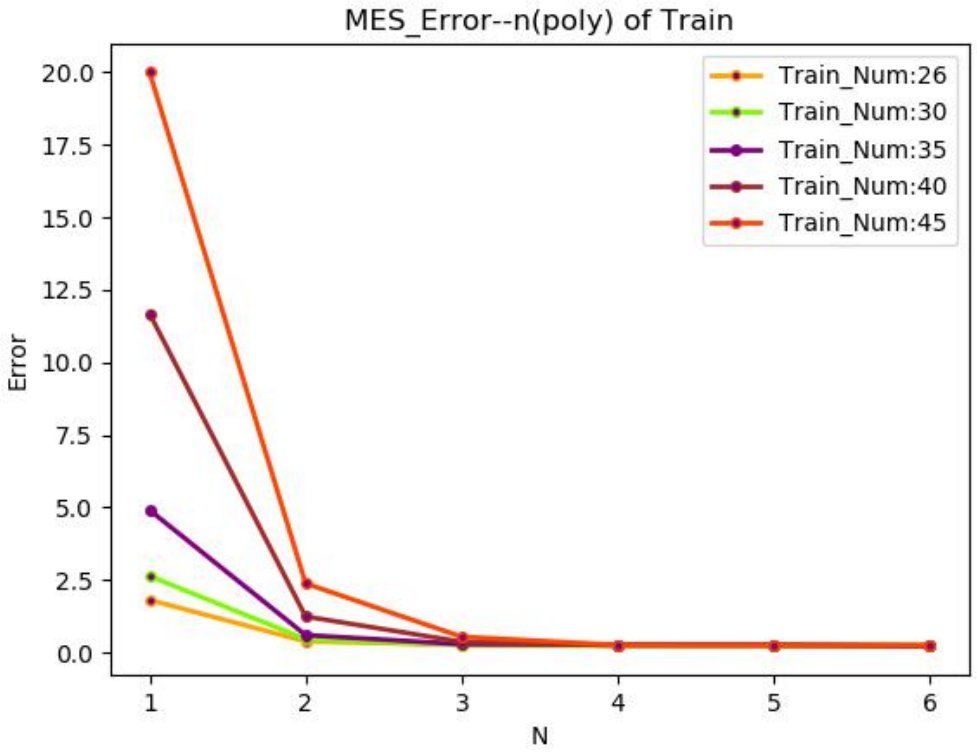
5）将实验数据温度 20℃~80℃ 范围进行调整（扩大或缩小），重复 2），3）内容（需要对训练集及测试集范围进行对应调整），观察并讨论由于采用不同规模训练数据给拟合（学习）结果带来的影响；

解：原始数据有 51 组，分别选取 26，30,35,40,45 组数据作为训练集，进行试

验。

5.1）重复任务 2:

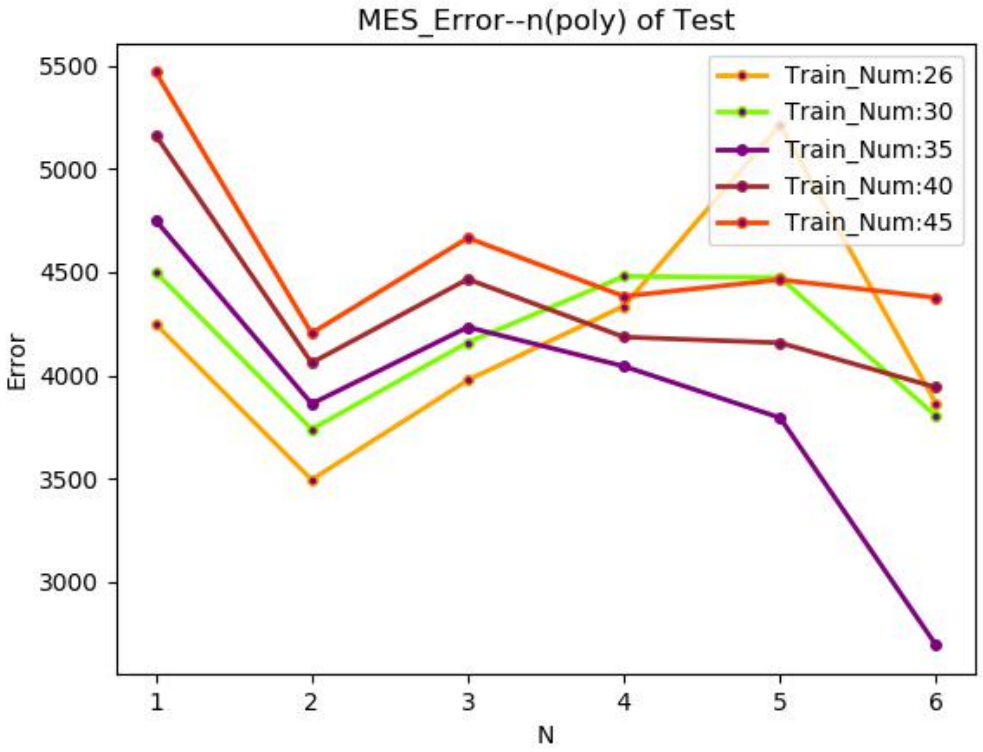
训练集误差如下:



（*横坐标是拟合多项式的阶次。纵坐标是均方差。图例为所选用的训练数据集的数目）

可以看出，在阶次相对较低时，数据集越大，对应的均方差越大。

对应的测试集误差结果如下:



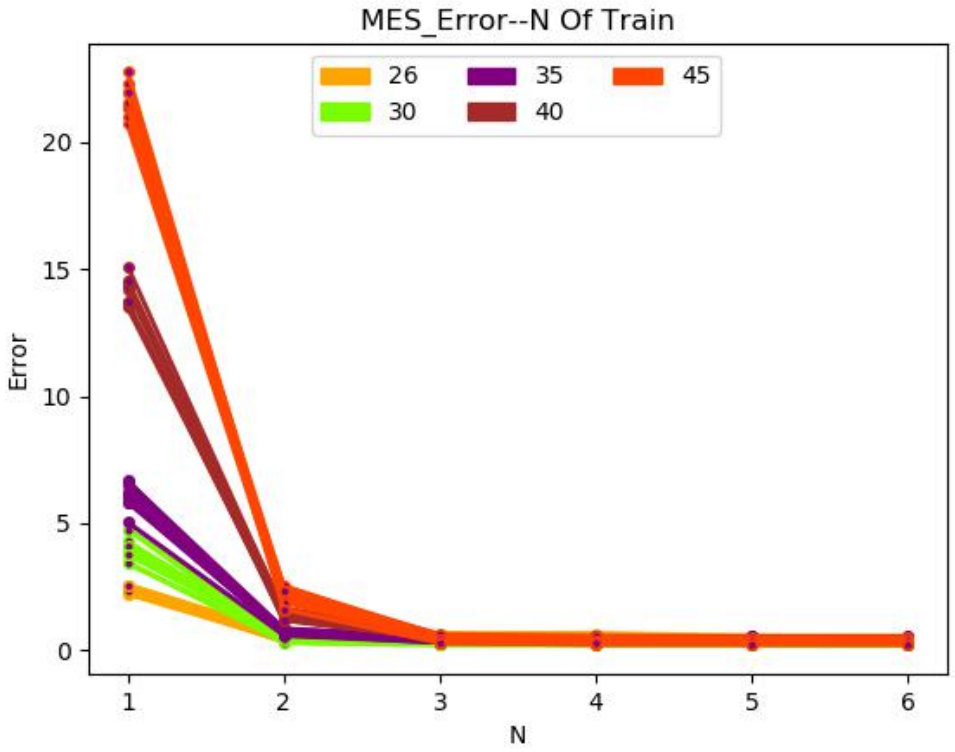
（*横坐标是拟合多项式的阶次。纵坐标是均方差。图例为所选用的训练数据集的数目）

可见，训练集的占比并不是越多或者越少越好。就本次实验结果来看，当训练集占比 35/50=70%（紫色）时，有较好的模型泛化能力，即在 测试集上的误差较小。

5.2） 重复任务 3

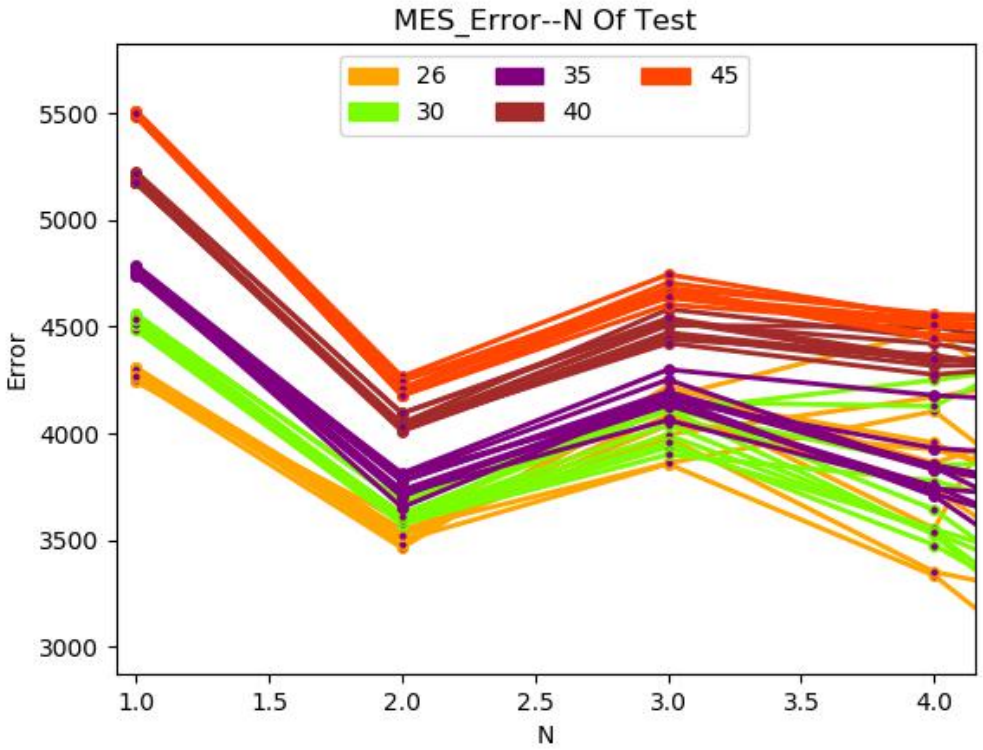
各组实验分别进行 10 组。

训练集：



（*横坐标是拟合多项式的阶次。纵坐标是均方差。图例为所选用的训练数据集的数目）

可见。同只进行一次实验（任务 2）的结果类似。**组内有差异，组间有区别。**测试集结果如下：



（*横坐标是拟合多项式的阶次。纵坐标是均方差。图例为所选用的训练数据集的数目）

就实验结果来看，训练即越少（如 26），其对应的测试集在阶次较低时表现较好，但是阶次升高之后，**其鲁棒性不够，波动很大**。综合来看，当选择训练集为 35 个数据时，**测试集误差较小，鲁棒性也较好**。实验结论同。即 70%的测试集相对较好。

6）选做：采用梯度下降算法，重复 2），3）内容，探讨模型参数初值、学习率对结果的影响。

答：使用梯度下降法进行多项式回归。

①为了防止梯度爆炸，首先将数据归一化，即将数据化为均值为 0，方差为 1 的数据。

②选取中间的 20 到 80 摄氏度的数据进行训练，其余数据进行预测。

③使用多项式模型回归。

④训练次数 10000 次，训练数据较少，所以决定每次将所有数据计算结束之后在更新梯度，这样保证了损失能够单调下降。

⑤学习率定位 0.0001。

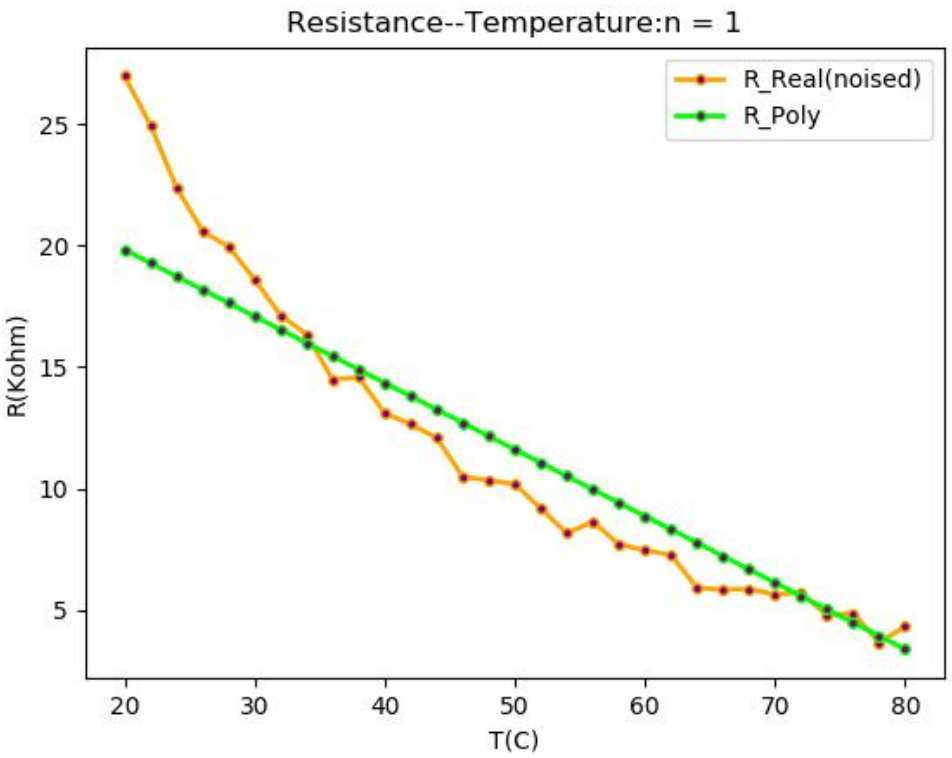
⑥损失函数为均方误差。

⑦使用随机初始化。

6.1）重复任务 2

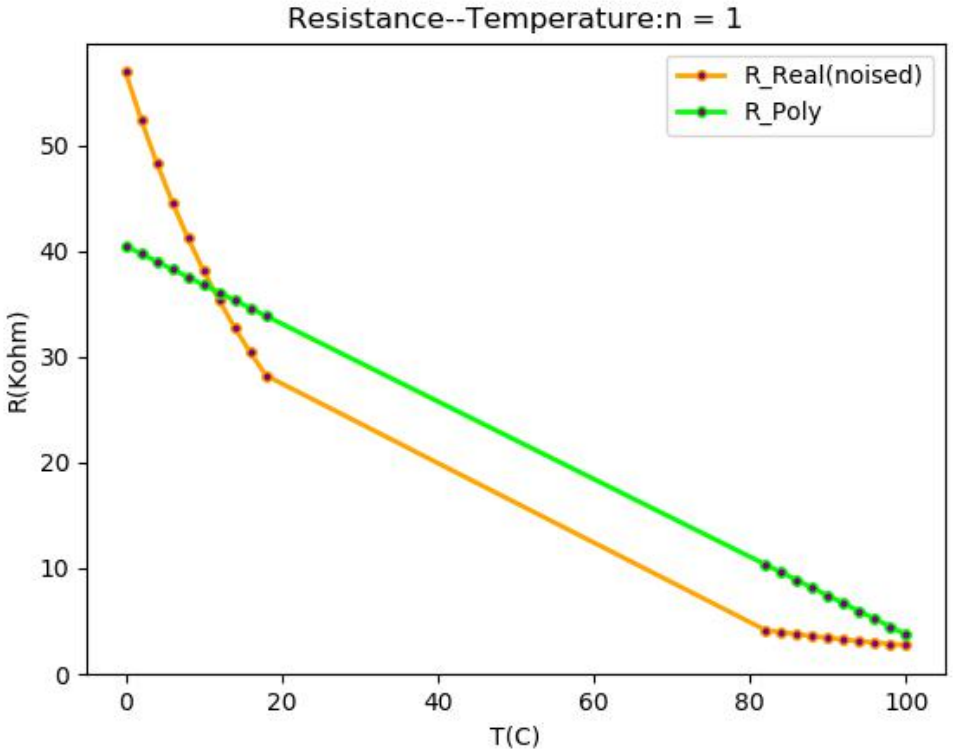
N=1

【训练集】



(*黄线：真实数据；绿线：拟合数据)

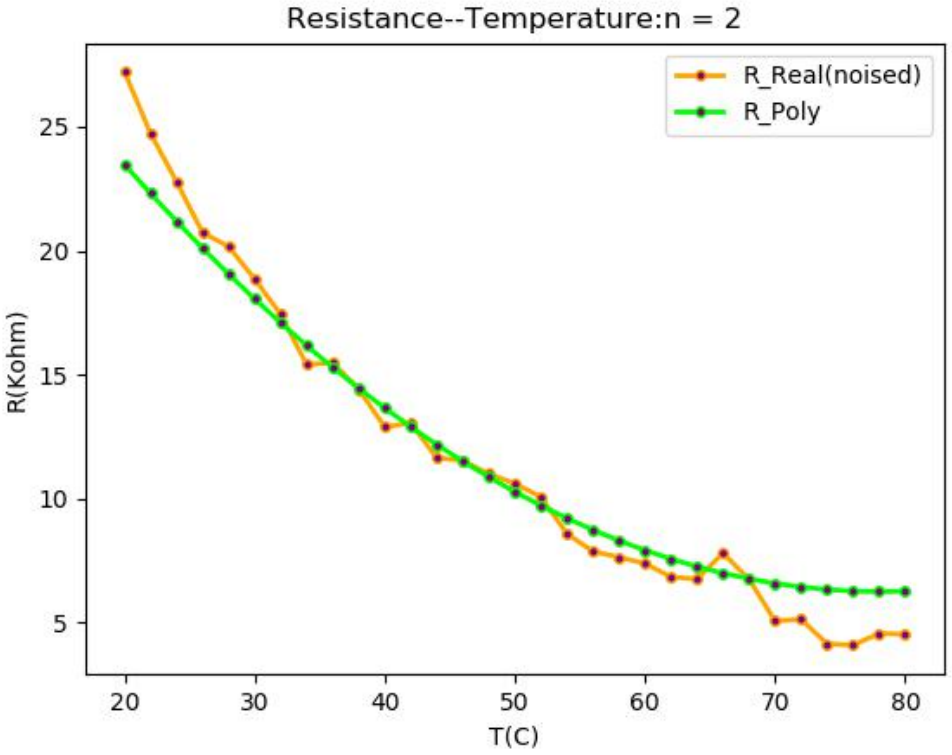
【测试集】



（*黄线：真实数据；绿线：拟合数据）

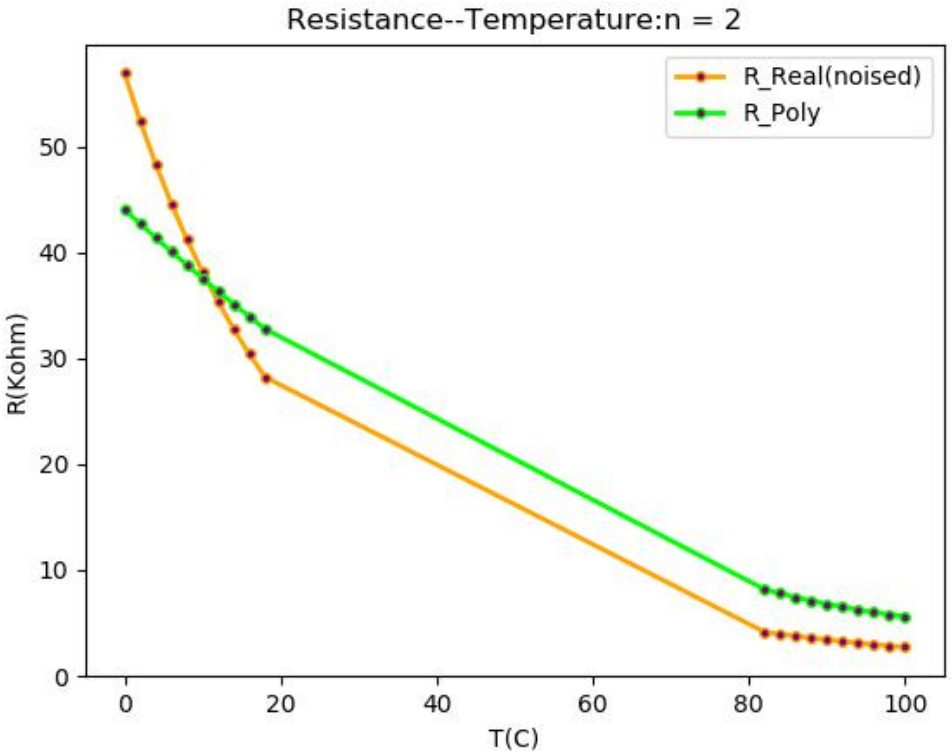
N=2

【训练集】



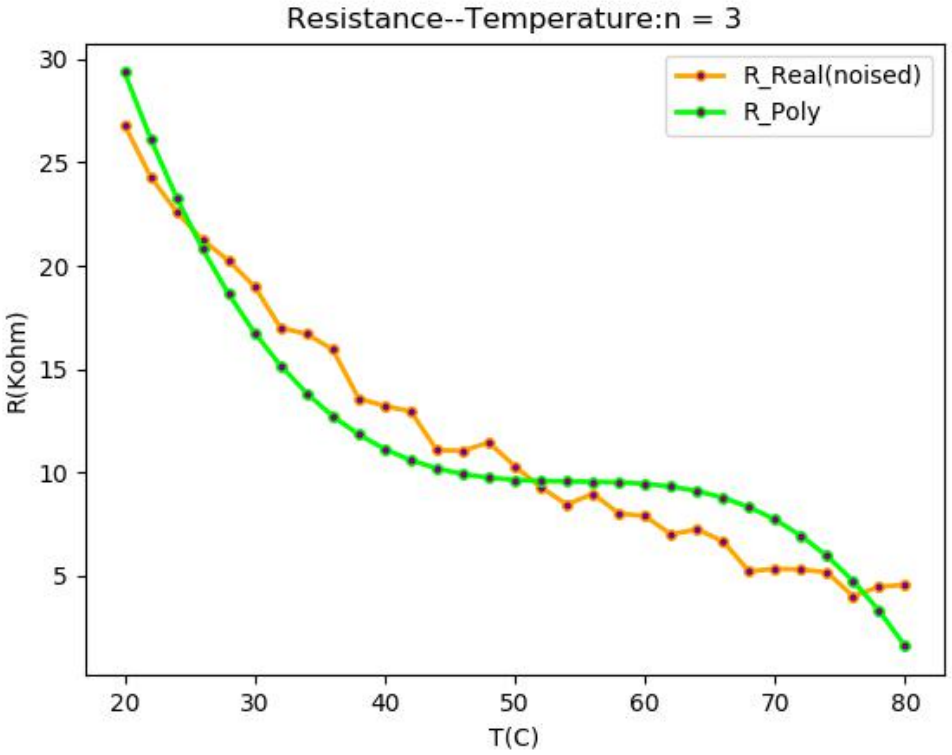
(*黄线：真实数据；绿线：拟合数据)

【测试集】

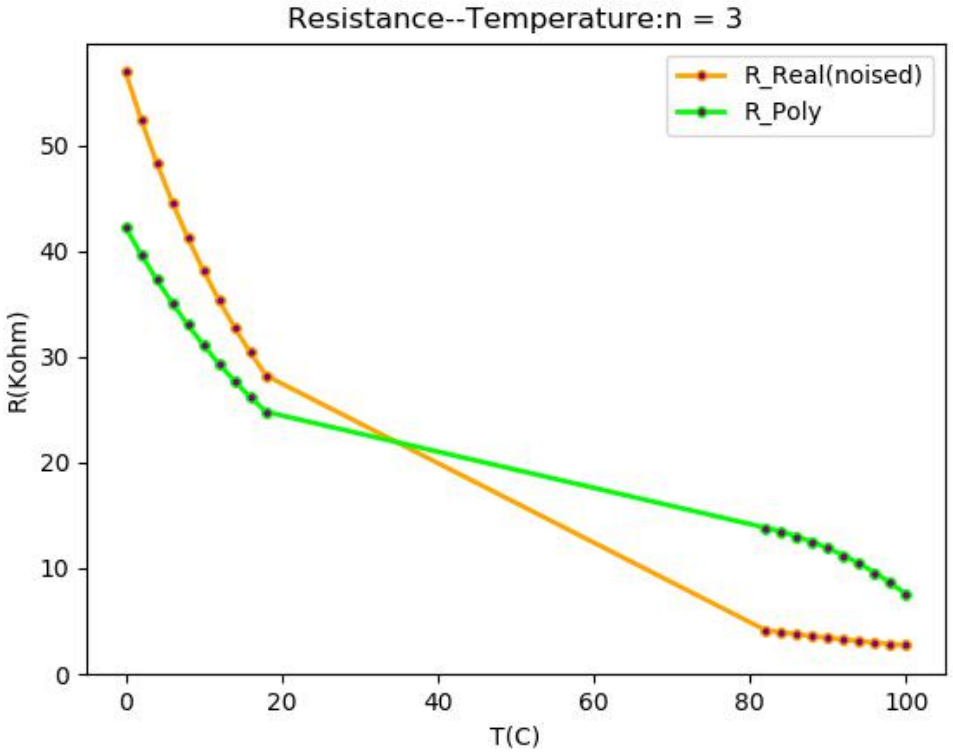


N=3

【训练集】

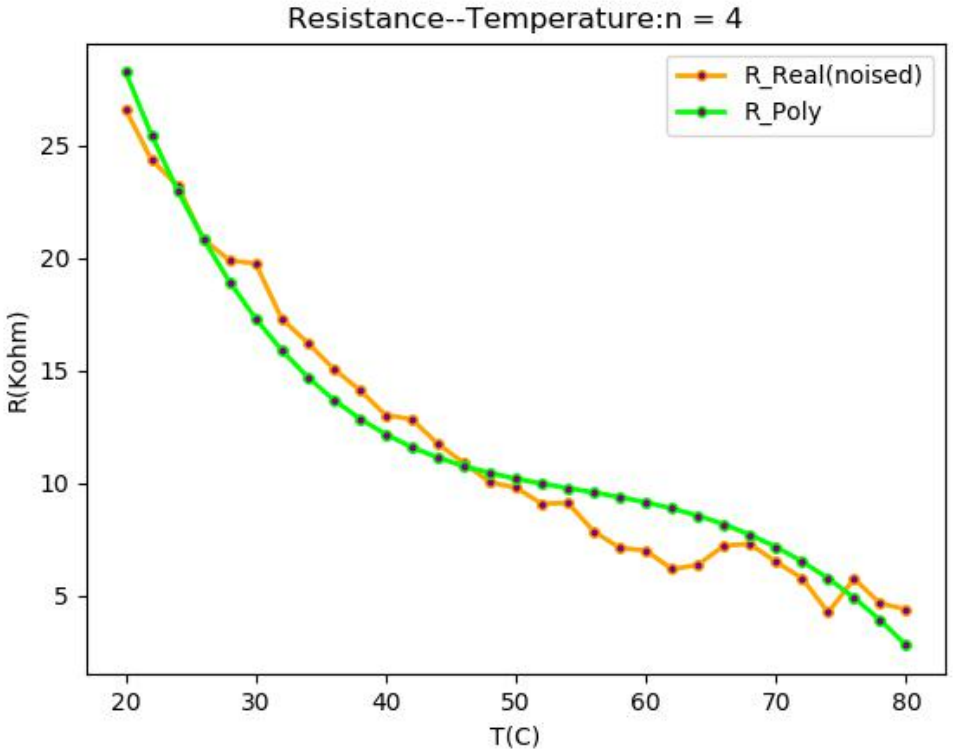

Resistance--Temperature:n = 3
R_Real(noised)
R_Poly
R(Kohm)
T(C)
30
25
20
15
10
5
20
30
40
50
60
70
80

【测试集】

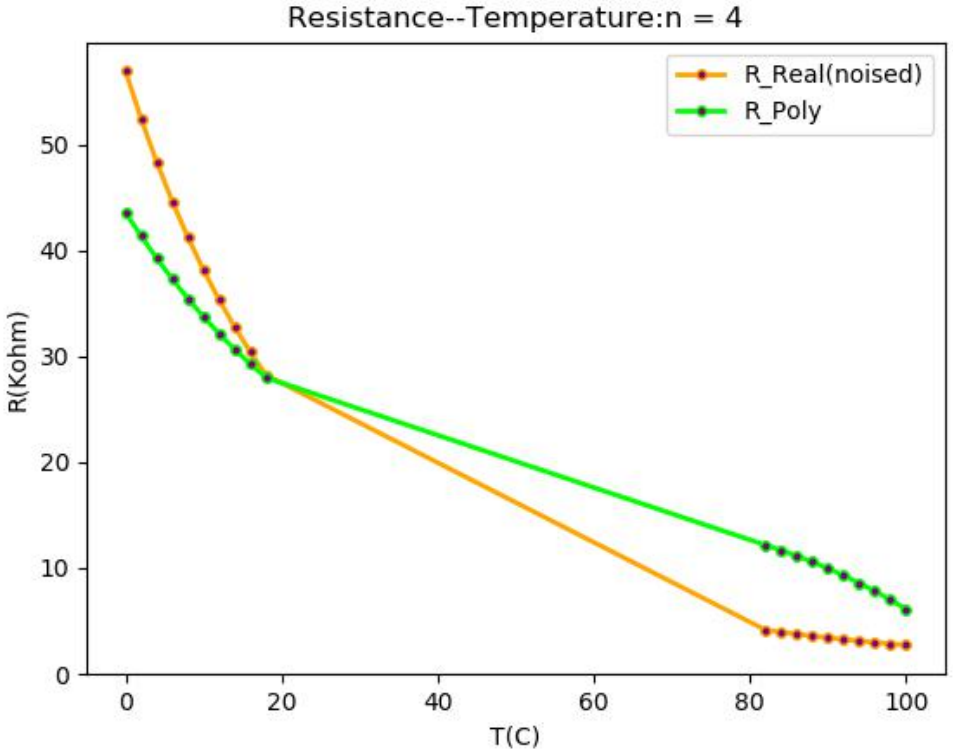

Resistance--Temperature:n = 3
R_Real(noised)
R_Poly
R(Kohm)
T(C)
0
10
20
30
40
50
0
20
40
60
80
100

N=4

【训练集】

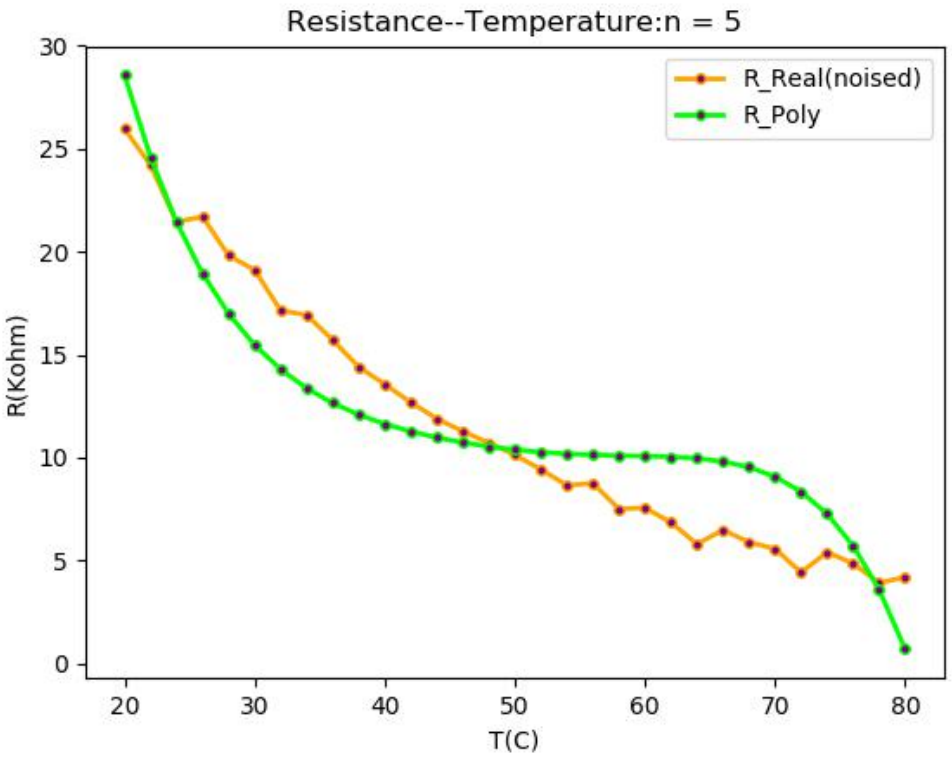

Resistance--Temperature:n = 4
R_Real(noised)
R_Poly
R(Kohm)
T(C)
5
10
15
20
25
20
30
40
50
60
70
80

【测试集】

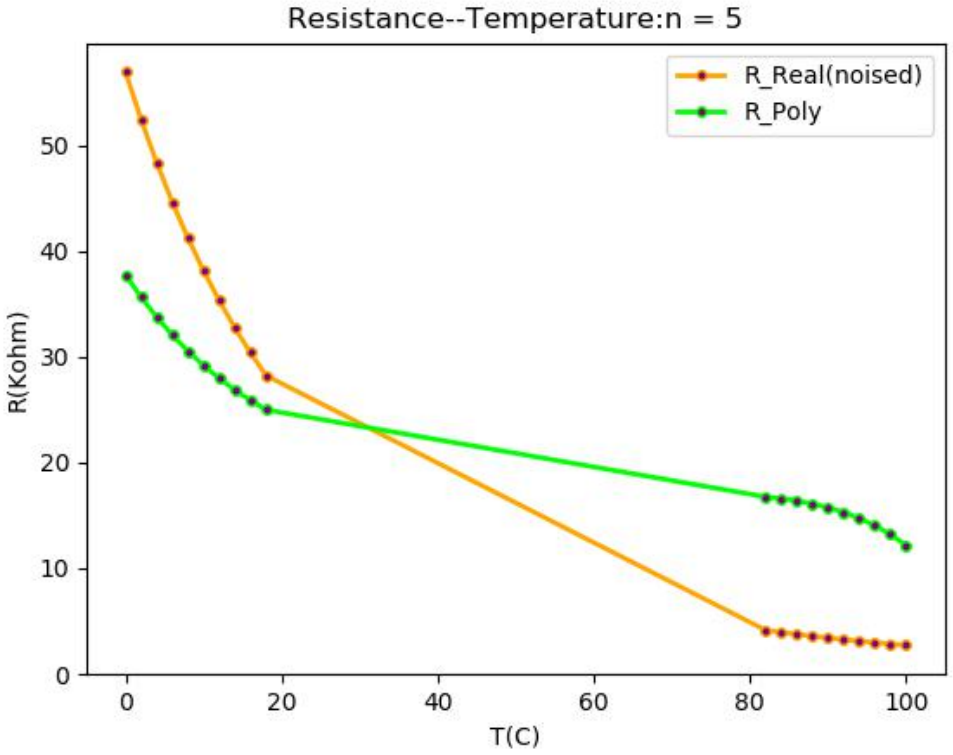

Resistance--Temperature:n = 4
R_Real(noised)
R_Poly
R(Kohm)
T(C)
0
10
20
30
40
50
0
20
40
60
80
100

N=5

【训练集】

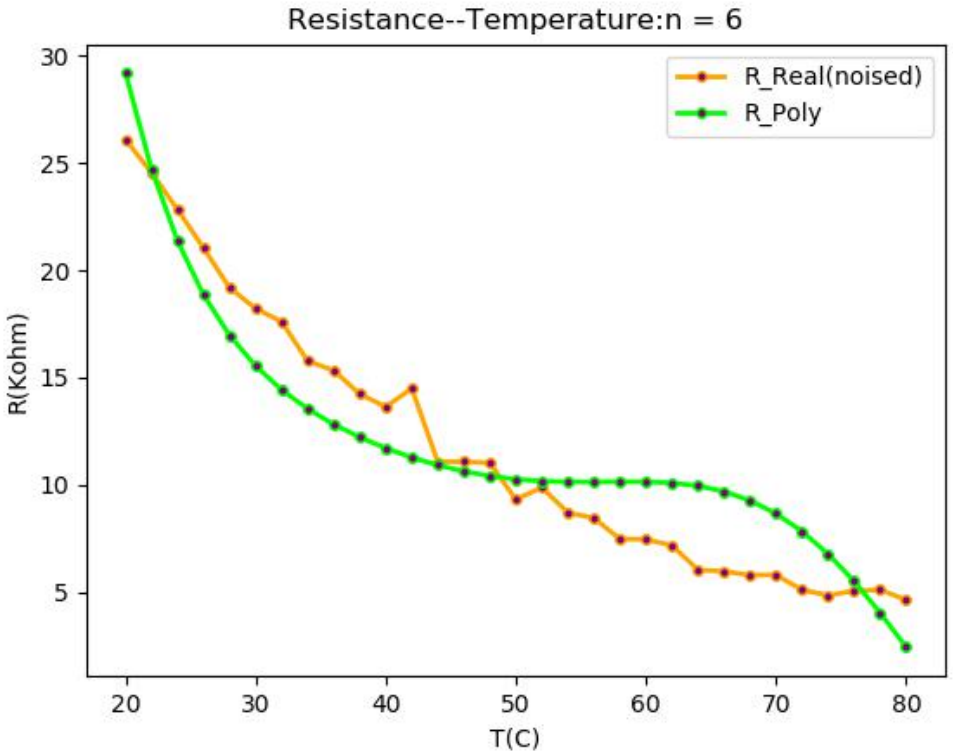

Resistance--Temperature:n = 5
R_Real(noised)
R_Poly
R(Kohm)
T(C)
0
5
10
15
20
25
30
20
30
40
50
60
70
80

【测试集】

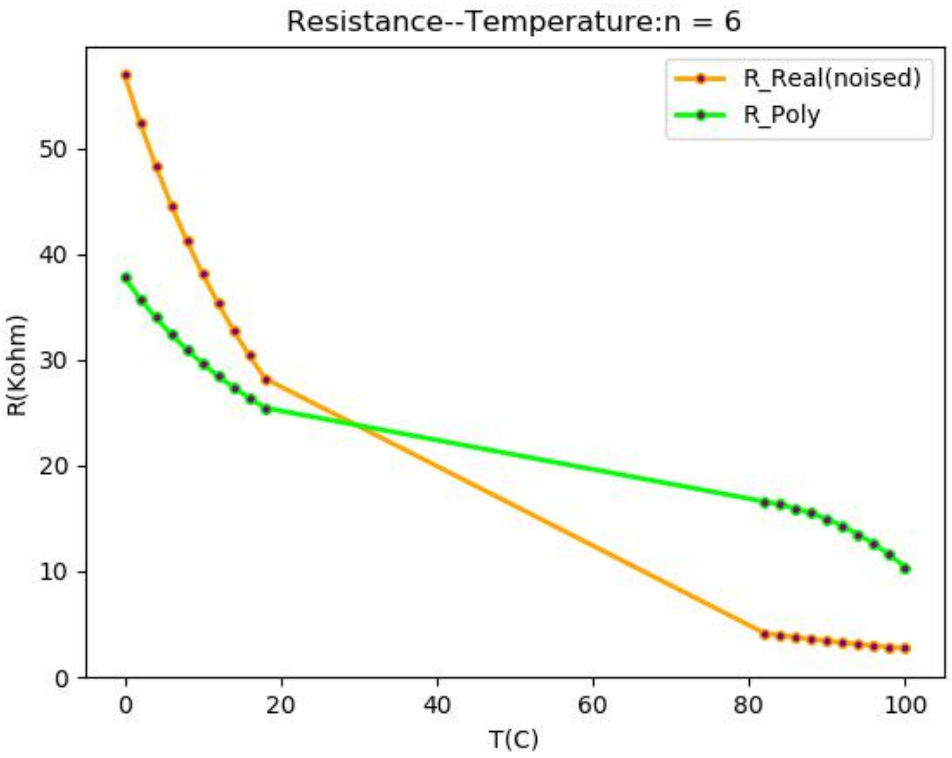

Resistance--Temperature:n = 5
R_Real(noised)
R_Poly
R(Kohm)
T(C)
0
10
20
30
40
50
0
20
40
60
80
100

N=6

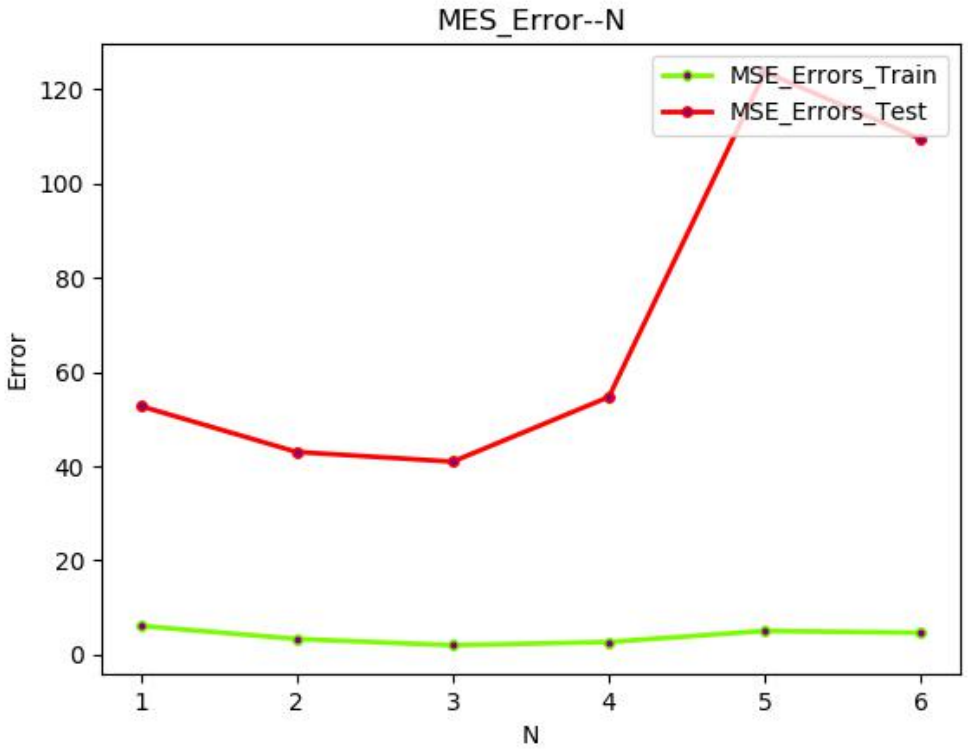
【训练集】


Resistance--Temperature:n = 6
R_Real(noised)
R_Poly
R(Kohm)
T(C)
5
10
15
20
25
30
20
30
40
50
60
70
80

【测试集】



【n 从 1 到 6，MES 的变化情况】



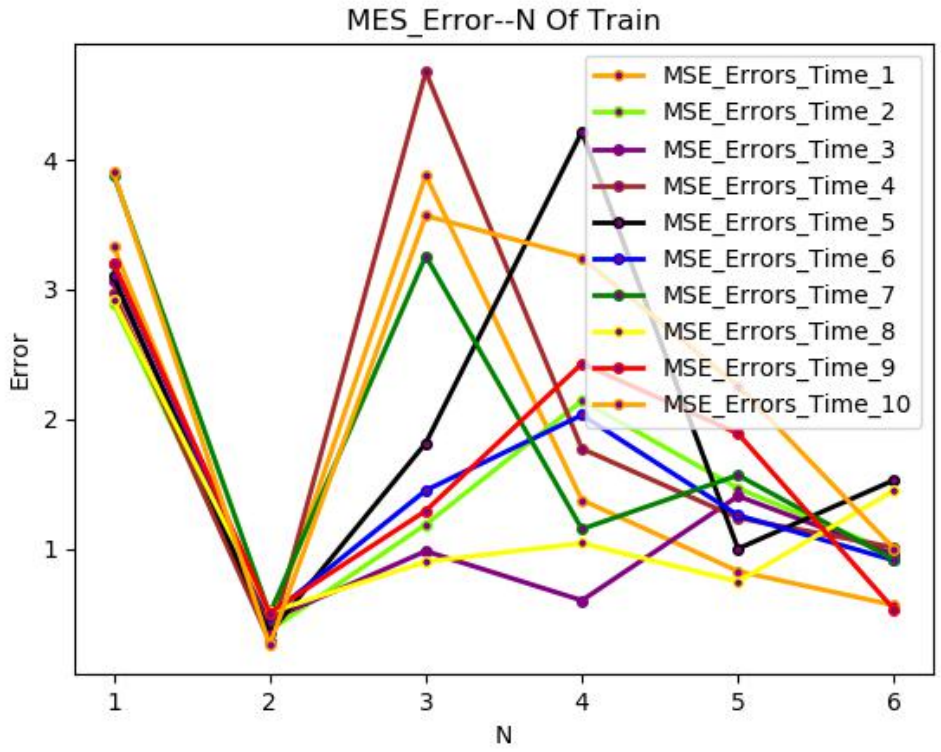
（*其中：横坐标是所选用多项式模型的阶次，纵坐标是 MSE 的大小，绿线是训练数据，红线是测试数据）

同样，使用梯度下降多项式拟合过程也出现了相同的“红线”走势，即随着模型的阶次升高，拟合效果首先变好，然后在训练数据集上出现过拟合，使得模型的泛化能力下降，在测试集上的 MSE 变大。

6.2）重复任务 3

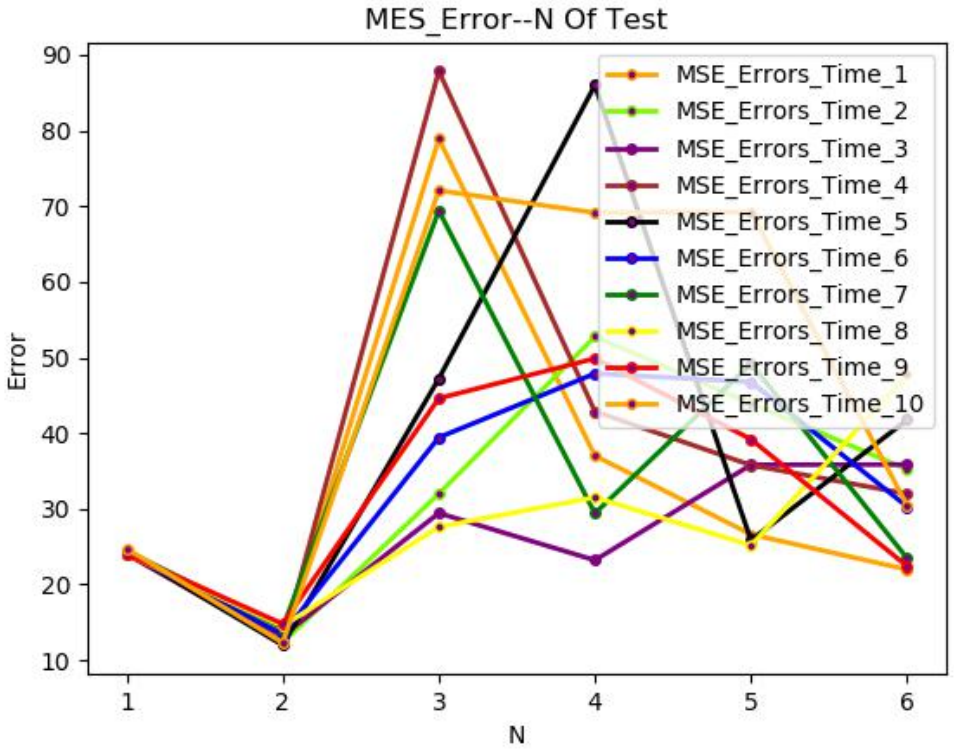
添加相同均值和方差的不同噪声进行模拟 10 次的 MSE 变化如下。

【训练集】



(*横坐标：模型阶次，每一条曲线对应一组数据)

【测试集】

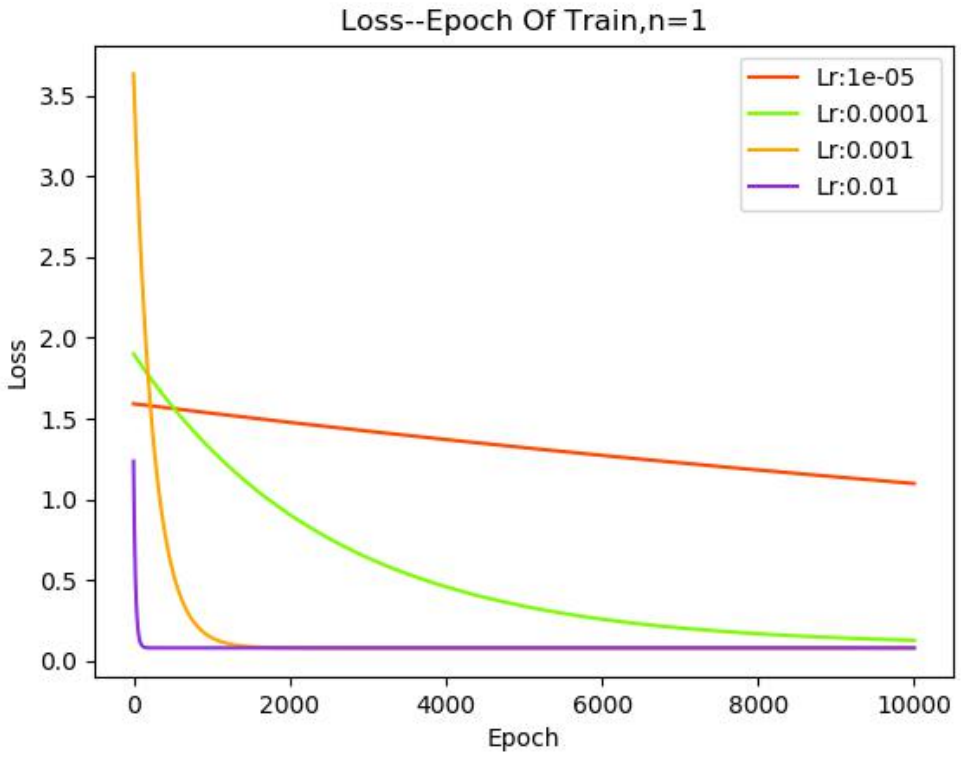


（*横坐标：模型阶次，每一条曲线对应一组数据）

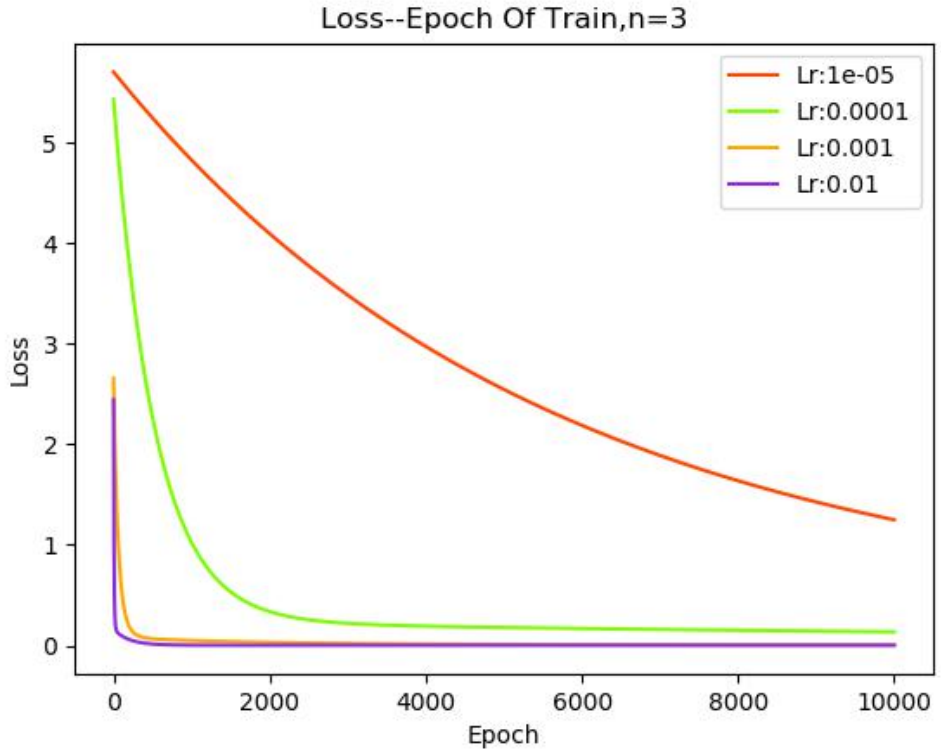
6.3）学习率对实验结果影响

选取 4 个数量级的学习率进行学习：　[0.00001,0.0001,0.001,0.01]

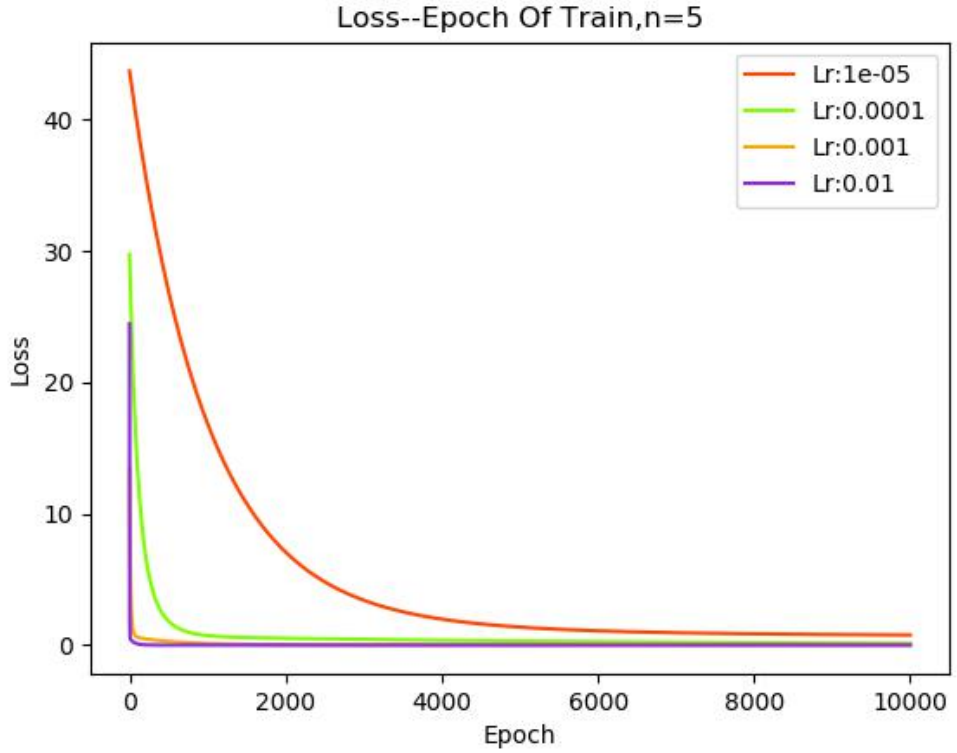
对于 1 次拟合：



对于 3 次拟合：



对于 5 次拟合：



可见，对于学习率，其值越大 loss 下降越快。但在实验过程中也发现，这样

的现象是有条件的，即我对数据进行了归一化处理，否则，较大的学习将会出现梯度爆炸的现象。这是我们不想要的。综合来看，对于该实验，学习率选取为 0.001 是比较合适的选择。

6.4）参数初值对拟合的影响。

实验过程当中，都参数的初始化主要仿真了两种，0 初始化和随机初始化。

0 初始化使得模型的初始误差相等，全局搜索能力减弱，使得模型的多次实验达到相同的局部极小。随机初始化（由于对数据进行了归一化，即使是随机初始化，也是比较小的值）使得模型具有较强的全局有所能力，多次实验将获得不同的局部极小。

但由于本事拟合的数据较为简单，两者的差异并不是很明显，同时拟合模型的差异也不是很大。

一般来讲，随机初始化是较好的选择。

7）思考：假如实验前已事先了解热敏电阻测温机理并掌握其阻值与温度的关系符合（1）式所描述的模型，你将如何考虑从实验数据获得热敏电阻的阻值与温度关系模型？

答：如果已知模型，则主要任务便是确定参数的大小。参数主要是确定 R0，T0，以及β。而对于本模型，由于有指数项，是确定参数的难点所在。为了避免指数项的存在，可以将模型等式两边同时取 ln，然后对于 T，可以要用 1/T 作为模型参数。此时，该模型已经等效的变为一个一元一次线性模型，简化了参数的确定。

然后获取训练数据。首先，可控制温度 T 不变，多次测量电阻取平均值，得到一对 T0，R0.。然后改变温度（与 T0 相差较大，比如 30K）为 T1，再次多次测量电阻，取平均值 R1，带入即可确定β，此时该模型已经完全确定。