# 检测原理作业：曲线拟合

林嘉成 2016011498
(自动化系 自 66)

## 目录

# 1 热敏电阻温度变化特性曲线

以 2 $°C$ 作为间隔（步长），画出该种热敏电阻在温度范围为 0 $°C$C 100 $°C$ 间阻值随温度变化的特性曲线如下。

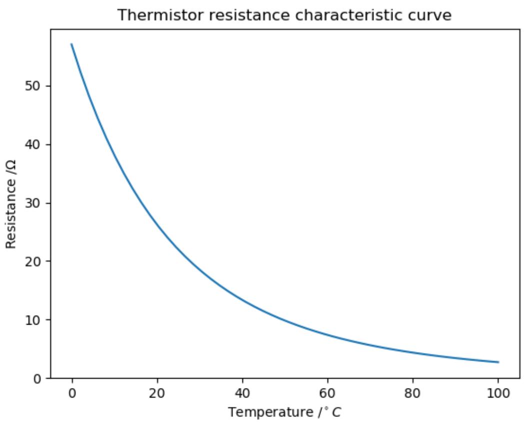



图 1: 热敏电阻温度变化特性曲线

# 2 不同阶次多项式模型拟合

使用不同阶次多项式模型进行拟合，得到拟合结果如下。

```
----------degree = 1----------
Linear Regression coef: [[ 0.         -0.33373393]]
Linear Regression intercept: [28.32049754]
Train Mean Square Error: 3.47676470265524
Test Mean Square Error: 163.43344548690638
----------degree = 2----------
Linear Regression coef: [[ 0.         -0.95817934  0.00624445]]
Linear Regression intercept: [41.93340729]
Train Mean Square Error: 0.2924235350133179
Test Mean Square Error: 39.475276971433665
----------degree = 3----------
Linear Regression coef: [[ 0.00000000e+00 -1.37604660e+00  1.52959919e-02 -6.03435862e-05]]
Linear Regression intercept: [47.74087402]
Train Mean Square Error: 0.2196280750253749
Test Mean Square Error: 11.411887897892424
----------degree = 4----------
Linear Regression coef: [[ 0.00000000e+00 -1.81483440e+00  3.01811857e-02 -2.70289902e-04
   1.04973158e-06]]
Linear Regression intercept: [52.23197842]
Train Mean Square Error: 0.21434106859150429
Test Mean Square Error: 6.174048851190565
----------degree = 5----------
Linear Regression coef: [[ 0.00000000e+00 -3.39767435e+00  1.03614177e-01 -1.88164265e-03
   1.78767239e-05 -6.73079707e-08]]
Linear Regression intercept: [65.0756307]
Train Mean Square Error: 0.20920338546388084
Test Mean Square Error: 14.242583287009817
----------degree = 6----------
Linear Regression coef: [[ 0.00000000e+00 -4.24073338e+00  1.53295218e-01 -3.37157176e-03
   4.19492898e-05 -2.66781775e-07  6.64904481e-10]]
Linear Regression intercept: [70.74771594]
Train Mean Square Error: 0.20908673907120595
Test Mean Square Error: 21.047836155351657
```

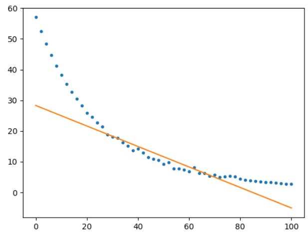
图 2: 不同阶次多项式模型拟合结果

训练集和测试集的均方误差如下。

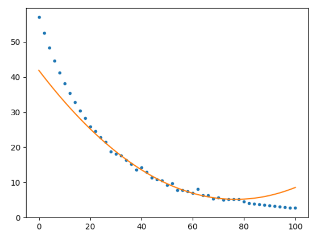
表 1: 训练集与测试集的均方误差

| 阶次 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 训练集 MSE | 3.476 | 0.2924 | 0.2196 | 0.2143 | 0.2092 | 0.2091 |
| 测试集 MSE | 163.433 | 39.475 | 11.412 | 6.174 | 14.243 | 21.048 |

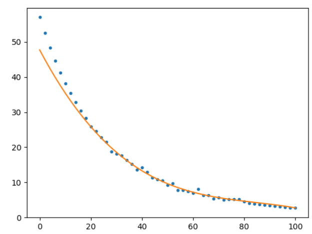
不同阶次的拟合曲线如下，其中 20-80 为训练集，其他为测试集。



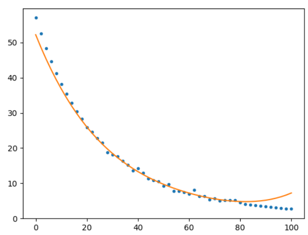
(a) 阶次 n=1



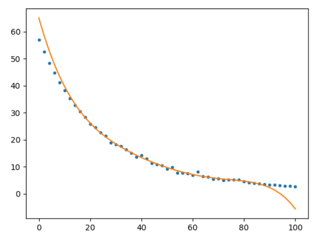
(b) 阶次 n=2



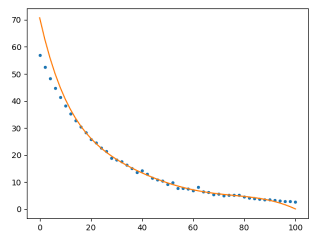
(c) 阶次 n=3



(d) 阶次 n=4



(e) 阶次 n=5



(f) 阶次 n=6

图 3: 不同阶次模型的拟合曲线

由实验结果可以看出，随着模型阶次的增加，训练集均方误差逐渐减小，即训练集拟合更好，但是随之出现了测试集会出现过拟合的现象，表现为随着阶次的增加，MSE 先减小后增加。由该次实验结果可以看出，当阶数为 4 时，测试集的 MSE 最小。

# 3 使用不同训练数据进行拟合

表 2: 训练集的均方误差

| 实验次数 | 不同阶次训练集 MSE | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 实验 1 | 3.1411 | 0.3878 | 0.2605 | 0.2435 | 0.2090 | 0.2066 |
| 实验 2 | 3.0446 | 0.3589 | 0.2166 | 0.2006 | 0.1942 | 0.1863 |
| 实验 3 | 3.1584 | 0.2616 | 0.1757 | 0.1756 | 0.1741 | 0.1734 |
| 实验 4 | 3.2370 | 0.3273 | 0.1552 | 0.1551 | 0.1545 | 0.1540 |
| 实验 5 | 3.4011 | 0.3699 | 0.2019 | 0.1940 | 0.1921 | 0.1553 |
| 实验 6 | 3.3566 | 0.2533 | 0.1172 | 0.1170 | 0.1168 | 0.1000 |
| 实验 7 | 3.6759 | 0.2714 | 0.1367 | 0.1296 | 0.1291 | 0.1283 |
| 实验 8 | 3.0219 | 0.2391 | 0.1800 | 0.1488 | 0.1341 | 0.1307 |
| 实验 9 | 2.7394 | 0.4088 | 0.2675 | 0.2528 | 0.2405 | 0.2307 |
| 实验 10 | 2.6865 | 0.3232 | 0.2129 | 0.1824 | 0.1618 | 0.1553 |

表 3: 测试集的均方误差

| 实验次数 | 不同阶次测试集 MSE | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 实验 1 | 163.4904 | 42.8715 | 10.4209 | 3.7581 | 216.4723 | 385.5329 |
| 实验 2 | 159.9231 | 40.0714 | 9.7733 | 50.8343 | 86.3589 | 826.7745 |
| 实验 3 | 162.4685 | 40.9177 | 11.7928 | 10.3340 | 7.4264 | 70.7243 |
| 实验 4 | 158.5656 | 37.3490 | 7.1646 | 6.0429 | 8.7330 | 11.9451 |
| 实验 5 | 157.9089 | 38.1685 | 5.9595 | 0.7162 | 10.0821 | 2064.0711 |
| 实验 6 | 162.4518 | 39.7450 | 7.1660 | 9.6810 | 8.2023 | 795.243 |
| 实验 7 | 161.8748 | 38.9078 | 5.5009 | 2.6347 | 11.2767 | 68.1057 |
| 实验 8 | 159.6508 | 37.3259 | 14.0395 | 13.1711 | 171.9643 | 442.8708 |
| 实验 9 | 156.6311 | 42.5548 | 12.5768 | 1.0554 | 110.6795 | 643.8349 |
| 实验 10 | 160.5384 | 43.3430 | 14.5231 | 6.5551 | 125.6581 | 415.0086 |

由实验结果可知，采用不同训练数据，最佳的模型阶数在 3,4,5 之间，其中最佳模型阶数为 4 的次数多于 3，而 5 只有 1 次。随着阶数的增加，训练集 MSE 依旧逐渐减小，训练集拟合逐渐增强，但是测试集随着阶数的增加依旧会出现过拟合现象。

## 4 使用不同噪声强度的训练数据进行拟合

改变噪声强度（通过改变所加噪声的标准偏差实现），得到实验结果如下。

表 4: 训练集的均方误差

| 噪声标准差 | 不同阶次训练集 MSE | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 0 | 3.0850 | 0.1418 | 0.0046 | 0.0001 | 2.4770e-06 | 4.5870e-08 |
| 0.5 | 3.3394 | 0.3129 | 0.1787 | 0.1772 | 0.1743 | 0.1736 |
| 1 | 4.3773 | 0.8322 | 0.6596 | 0.6459 | 0.6431 | 0.6029 |
| 1.5 | 4.2383 | 1.8116 | 1.8082 | 1.8070 | 1.7937 | 1.7935 |
| 2 | 8.0887 | 2.7528 | 2.7518 | 2.7015 | 2.7006 | 2.5403 |
| 2.5 | 10.5398 | 6.8154 | 6.2588 | 6.2264 | 6.1070 | 6.0075 |
| 3 | 14.1771 | 11.8897 | 10.7364 | 10.7126 | 10.5526 | 10.5344 |
| 3.5 | 18.0781 | 14.8889 | 14.8482 | 14.8480 | 13.9359 | 13.9353 |
| 4 | 27.9821 | 18.8151 | 18.7464 | 18.2440 | 18.2376 | 17.6096 |
| 4.5 | 23.9634 | 21.7743 | 21.2559 | 19.9082 | 19.9075 | 19.8617 |
| 5 | 20.5398 | 12.7291 | 12.5796 | 12.5426 | 9.1137 | 9.0330 |

表 5: 测试集的均方误差

| 噪声标准差 | 不同阶次测试集 MSE | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 0 | 160.6696 | 39.1552 | 7.9034 | 1.3696 | 0.2091 | 0.0286 |
| 0.5 | 157.1043 | 35.8063 | 6.8196 | 2.3106 | 15.8353 | 37.2995 |
| 1 | 158.7565 | 36.8888 | 3.2727 | 5.3015 | 32.2028 | 2518.0672 |
| 1.5 | 166.5114 | 47.7973 | 57.0723 | 50.6043 | 19.2555 | 15.4296 |
| 2 | 158.2544 | 30.9710 | 35.6636 | 92.9572 | 131.4827 | 10655.3453 |
| 2.5 | 148.5287 | 32.7288 | 16.0124 | 36.9658 | 717.3811 | 7065.7058 |
| 3 | 144.7143 | 27.9128 | 110.3996 | 98.5737 | 653.2484 | 1604.6964 |
| 3.5 | 191.0265 | 73.6570 | 38.0776 | 40.3788 | 5972.9378 | 5978.5856 |
| 4 | 140.7443 | 69.5932 | 40.0968 | 790.1645 | 876.5187 | 46830.3544 |
| 4.5 | 132.7368 | 25.2731 | 170.8551 | 1830.0535 | 1891.9596 | 8614.2170 |
| 5 | 189.4303 | 35.5476 | 4.3548 | 22.3991 | 25475.7926 | 30698.5085 |

整体上，仍然有：在噪声存在的情况下，随着模型阶次的增加，训练集 MSE 逐渐减小，而测试集会出现过拟合的现象。

当没有噪声时，随着模型阶次的增加，训练集的 MSE 逐渐减小，而高阶模型拟合得效果很好，导致测试集的 MSE 也很小。当加入噪声后，则最佳的模型阶数在 3,4 之间，即出现过拟合现象。噪声标准差越大，则可能造成高阶模型的测试集 MSE 很大，即出现了严重过拟合。

此外，随着噪声的增加，训练集拟合效果整体逐渐变差。

# 5 采用不同规模的训练数据进行拟合

选取不同规模的训练数据，由于多次实验结果类似，故该部分实验均做多次观察，但只记录一次的实验数据。

表 6: 不同规模的训练数据拟合的 MSE(阶次 1-3)

| 训练集分布 | 不同阶次训练集及测试集 MSE | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| | 训练集 | 测试集 | 训练集 | 测试集 | 训练集 | 测试集 |
| 10-90 | 12.6864 | 184.6485 | 1.0455 | 34.8853 | 0.2647 | 5.5561 |
| 20-80 | 2.6169 | 152.2669 | 0.3093 | 41.6213 | 0.1566 | 12.7397 |
| 30-70 | 0.7997 | 136.7053 | 0.1567 | 35.6833 | 0.0931 | 24.7797 |
| 40-60 | 0.2003 | 105.6302 | 0.1898 | 52.6393 | 0.1887 | 52.5567 |

表 7: 不同规模的训练数据拟合的 MSE(阶次 4-6)

| 训练集分布 | 不同阶次训练集及测试集 MSE | | | | | |
|---|---|---|---|---|---|---|
| | 4 | | 5 | | 6 | |
| | 训练集 | 测试集 | 训练集 | 测试集 | 训练集 | 测试集 |
| 10-90 | 0.2221 | 1.3434 | 0.2056 | 0.8192 | 0.1997 | 3.9736 |
| 20-80 | 0.1554 | 6.2330 | 0.1526 | 19.5850 | 0.1517 | 104.8849 |
| 30-70 | 0.0903 | 64.7231 | 0.0889 | 541.9848 | 0.0888 | 859.5608 |
| 40-60 | 0.1886 | 382.1701 | 0.1748 | 6997349 | 0.1745 | 23817910 |

由实验结果可以看出，训练数据规模越大（测试集规模越小），则测试集 MSE 较小，而由于训练数据增多，故训练集 MSE 稍微增加（拟合难度增加）。在训练数据规模较大时，高阶次模型的过

拟合的现象减弱了很多。

在训练数据量较小时，随着阶次的升高，训练集拟合效果逐渐增强，而过拟合现象越发严重，最终导致测试集的 MSE 非常大。

## 6 梯度下降法进行计算

由于多次实验结果类似，故该部分实验均做多次观察，但只记录一次的实验数据。

使用批量梯度下降，Loss 函数使用 MSE，迭代 50000 次，则有

$$\theta_j := \theta_j - \alpha\frac{1}{m}(h_\theta(x^i) - y^i)x_j^{(i)}$$

则相同数据相同初值下，不同学习率的拟合结果如下。

表 8: 不同学习率进行梯度下降拟合的 MSE(阶次 1-3)

| 学习率 | 不同阶次训练集及测试集 MSE | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| | 训练集 | 测试集 | 训练集 | 测试集 | 训练集 | 测试集 |
| 1 | 3.4767 | 156.0053 | 0.2924 | 37.6335 | NaN | NaN |
| 0.1 | 3.4767 | 156.0053 | 0.2924 | 37.6335 | 0.2482 | 25.3916 |
| 0.01 | 3.4767 | 156.0053 | 0.2924 | 37.7171 | 0.3601 | 54.4912 |
| 0.001 | 3.4767 | 156.0053 | 1.3442 | 87.5778 | 0.4782 | 65.2145 |
| polyfit | 3.4767 | 156.0053 | 0.2924 | 37.6335 | 0.2196 | 10.8729 |

表 9: 不同学习率进行梯度下降拟合的 MSE(阶次 4-6)

| 学习率 | 不同阶次训练集及测试集 MSE | | | | | |
|---|---|---|---|---|---|---|
| | 4 | | 5 | | 6 | |
| | 训练集 | 测试集 | 训练集 | 测试集 | 训练集 | 测试集 |
| 1 | NaN | NaN | NaN | NaN | NaN | NaN |
| 0.1 | 0.2307 | 22.7173 | 0.2443 | 33.6120 | 0.2495 | 40.5549 |
| 0.01 | 0.3673 | 58.3898 | 0.3014 | 45.0716 | 0.2692 | 55.7288 |
| 0.001 | 0.5162 | 86.6505 | 0.6144 | 99.9083 | 0.3625 | 54.6786 |
| polyfit | 0.2143 | 5.8902 | 0.2092 | 13.5691 | 0.2090 | 20.0509 |

在迭代次数相同的情况下，学习率较大时，会发生模型训练不收敛的情况（如 $\alpha = 1$）；学习率较小时，则对比 polyfit 的结果，可以看出模型并没有收敛到最优解。实验中效果较好的模型的学习率为 $\alpha = 0.1$，但与 polyfit 对比之后发现其并没有完全收敛至最优解。故可以得知：**梯度下降法进行拟合很依赖于学习率**，如果学习率太大，则模型无法收敛；如果学习率太小，则需要更多的迭代次数、更小的学习率才能收敛至最优解，这显然没有 polyfit 得到的拟合效果好。

此外，对于低阶次模型（1 阶、2 阶），梯度下降法的拟合效果与 polyfit 几乎相同，即简单模型（如低阶次）使用梯度下降法拟合会更容易。

选取学习率 $\alpha = 0.1$，前三次初始值由 (0,0.01) 正态分布随机初始化确定，后三次初值由 (0,1) 正态分布随机初始化确定。则得到实验结果如下。

表 10: 不同初值进行梯度下降拟合的 MSE(阶次 1-3)

| 实验次数 | 不同阶次训练集及测试集 MSE | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| | 训练集 | 测试集 | 训练集 | 测试集 | 训练集 | 测试集 |
| 1 | 3.4767 | 156.0053 | 0.2924 | 37.6335 | 0.2583 | 28.5624 |
| 2 | 3.4767 | 156.0053 | 0.2924 | 37.6335 | 0.2580 | 28.4788 |
| 3 | 3.4767 | 156.0053 | 0.2924 | 37.6335 | 0.2581 | 28.5038 |
| 4 | 3.4767 | 156.0053 | 0.2924 | 37.6335 | 0.2564 | 27.9965 |
| 5 | 3.4767 | 156.0053 | 0.2924 | 37.6335 | 0.2796 | 34.6577 |
| polyfit | 3.4767 | 156.0053 | 0.2924 | 37.6335 | 0.2196 | 10.8729 |

表 11: 不同初值进行梯度下降拟合的 MSE(阶次 4-6)

| 实验次数 | 不同阶次训练集及测试集 MSE | | | | | |
|---|---|---|---|---|---|---|
| | 4 | | 5 | | 6 | |
| | 训练集 | 测试集 | 训练集 | 测试集 | 训练集 | 测试集 |
| 1 | 0.2280 | 20.1404 | 0.2392 | 29.5011 | 0.2469 | 36.5257 |
| 2 | 0.2280 | 20.1535 | 0.2392 | 29.4425 | 0.2470 | 36.5920 |
| 3 | 0.2280 | 20.1331 | 0.2393 | 29.5408 | 0.2469 | 36.5815 |
| 4 | 0.2290 | 21.1124 | 0.2432 | 32.7401 | 0.2508 | 40.2011 |
| 5 | 0.2367 | 28.6713 | 0.2567 | 45.7335 | 0.2562 | 45.3721 |
| polyfit | 0.2143 | 5.8902 | 0.2092 | 13.5691 | 0.2090 | 20.0509 |

由于整个问题是凸问题，故不同初始值对最终收敛的解影响不大。观察到，不同的初始值对较低阶次收敛至的最优解没有太大影响。对于较高阶次的模型，若初始值的标准差越小，则相同迭代次数收敛越接近最优解，且标准差较小时，相同分布初始化相同迭代次数后收敛的最优解得到的 MSE 差别不是很大。

由于拟合前，各维数据规模相差较大，故如果直接训练则很难拟合，故首先对数据进行标准化处理。

# 7 思考题

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下式所描述的模型，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？

$$R_T = R_{T_0} e^{\beta(1/T - 1/T_0)}$$

**答**：将左右取对数，则有

$$\ln R_T = \ln R_{T_0} + \beta(\frac{1}{T} - \frac{1}{T_0})$$

整理的

$$\ln R_T = \beta \frac{1}{T} + b$$

其中有 $b = \ln R_{T_0} - \dfrac{\beta}{T_0}$。则线性回归得到 $\ln R_T$ 与 $\dfrac{1}{T}$ 之间的关系，反变换即可得到 $R_T$ 与 $T$ 的关系。