

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：

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

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

- 1) 以 2°C 作为间隔（步长），画出该种热敏电阻在温度范围为 $0^\circ\text{C} \sim 100^\circ\text{C}$ 间阻值随温度变化的特性曲线；
- 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 $20^\circ\text{C} \sim 80^\circ\text{C}$ 。现采用如下多项式模型描述热敏电阻阻值与温度关系

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

其中， t 为温度（单位为 $^\circ\text{C}$ ）， R_t 为温度为 $t^\circ\text{C}$ 时热敏电阻阻值（单位为 $k\Omega$ ）， n 为模型阶次， a_n 为不同阶次项系数。

在 1) 中获得的 $20^\circ\text{C} \sim 80^\circ\text{C}$ 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 $20^\circ\text{C} \sim 80^\circ\text{C}$ （训练集）上和温度范围 $0^\circ\text{C} \sim 100^\circ\text{C}$ 刨除 $20^\circ\text{C} \sim 80^\circ\text{C}$ 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；

（注：可能用到的 matlab 函数：polyfit; randn）

- 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；
- 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；

- 5) 将实验数据温度 $20^{\circ}\text{C}\sim 80^{\circ}\text{C}$ 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；
- 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。
- 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？

Reference:

- [1] 李宏毅，机器学习，Video Lectures, Regression, Where does error come from, Gradient Descent, 相关内容。
- [2] 李航，《统计学习方法》第一章，清华大学出版社，2012.
- [3] Christopher Bishop, Pattern Recognition and Machine Learning, Chapter 1 Introduction, Springer, 2006.