**Part 1: Components Transformation Temperature Analysis**

Boiling point of alcohol and paraffin with different C atoms:

|  |  |  |
| --- | --- | --- |
| Number of C | Alcohol/℃ | Paraffin/℃ |
| 1 | 64.7 | -161.49 |
| 2 | 78.29 | -88.6 |
| 3 | 97.2 | -42.04 |
| 4 | 118.75 | -0.5 |
| 5 | 137.75 | 36.07 |
| 6 | 156.75 | 68.73 |
| 7 | 175.45 | 98.43 |
| 8 | 193.95 | 125.68 |
| 9 | 212.05 | 150.82 |
| 10 | 229.85 | 174.155 |
| 11 | 247.15 | 195.928 |
| 12 | 263.95 | 216.323 |
| 13 | 280.25 | 235.466 |
| 14 | 295.85 | 253.577 |
| 15 | 310.75 | 270.685 |
| 16 | 324.85 | 286.864 |
| 17 | 338.15 | 302.15 |
| 18 | 350.45 | 316.71 |
| 19 | 361.95 | 329.9 |
| 20 | 372.35 | 343.78 |

Melting point of alcohol and paraffin with different C atoms:

|  |  |  |
| --- | --- | --- |
| Number of C | Alcohol/℃ | Paraffin/℃ |
| 1 | -97.68 | -182.456 |
| 2 | -114.1 | -182.798 |
| 3 | -126.2 | -187.68 |
| 4 | -89.3 | -138.29 |
| 5 | -77.59 | -129.73 |
| 6 | -44.6 | -95.32 |
| 7 | -34 | -90.58 |
| 8 | -15.5 | -56.77 |
| 9 | -5 | -53.49 |
| 10 | 6.9 | -29.64 |
| 11 | 15.3 | -25.579 |
| 12 | 23.8 | -9.582 |
| 13 | 30.6 | -5.39 |
| 14 | 37.5 | 5.86 |
| 15 | 43.9 | 9.922 |
| 16 | 49.2 | 18.158 |
| 17 | 53.9 | 21.984 |
| 18 | 57.9 | 28.16 |
| 19 | 61.7 | 31.89 |
| 20 | 65.4 | 36.43 |

**（All data from Aspen Properties database）**

**Part 2 Hansch Model**

为方便操作未采用Excel，使用python编了一个小程序实现拟合。

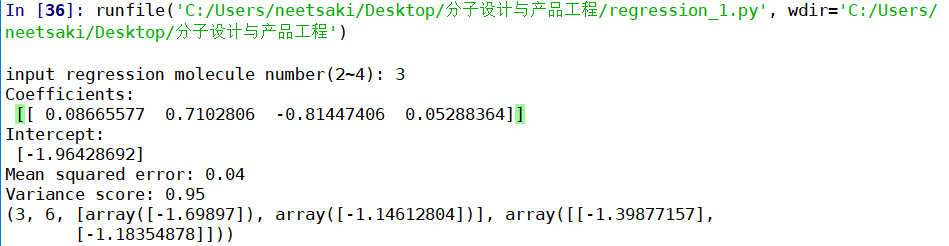
以选取第二个分子为例：

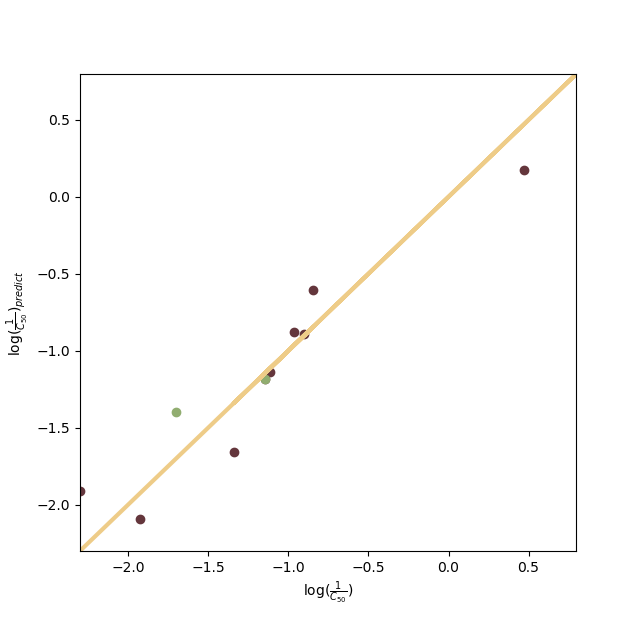
原始数据：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Substituents** | IC50 (M) | 1/IC50 | log(1/IC50) | **** | **** | **m** | **MR** |
| H | **200** | 0.01 | -2.30 | 0.00 | 0.00 | 0.00 | 1.03 |
| F | **85** | 0.01 | -1.93 | 0.02 | 0.14 | 0.34 | 0.92 |
| Cl | **50** | 0.02 | -1.70 | 0.50 | 0.71 | 0.37 | 6.03 |
| Br | **13** | 0.08 | -1.11 | 0.74 | 0.86 | 0.39 | 8.88 |
| I | **7.0** | 0.14 | -0.85 | 1.25 | 1.12 | 0.35 | 13.94 |
| CH3 | **14** | 0.07 | -1.15 | 0.31 | 0.56 | -0.07 | 5.65 |
| CH2CH3 | **4.8** | 0.21 | -0.68 | 1.04 | 1.02 | -0.07 | 10.30 |
| CH(CH3)2 | **0.34** | 2.94 | 0.47 | 2.34 | 1.53 | -0.07 | 14.98 |
| CH2CH(CH3)2 | **0.16** | 6.25 | 0.80 | 3.92 | 1.98 | -0.10 | 19.62 |
| OCH3 | **22** | 0.05 | -1.34 | 0.00 | -0.02 | 0.12 | 7.87 |
| OCH(CH3)2 | **9.2** | 0.11 | -0.96 | 0.13 | 0.36 | 0.10 | 17.06 |
| N(CH3)2 | **8.0** | 0.13 | -0.90 | 0.03 | 0.18 | -0.15 | 15.55 |

记录为“3-Substituted.mat”

运行记录：





说明：input regression molecule number(2~4)对应于选择“2-Substituted.mat”, “3-Substituted.mat”,“4-Substituted.mat”其中一个母体；

Coefficients对应拟合的系数

Intercept对应拟合的截距

Mean squared error对应拟合的均方差

Variance score对应拟合的R方

最后一列前两个数代表抽出来做TEST集的两个集团的编号，此处对应Br和CH2CH3

后面两个向量代表此两个数的真值与预测值；

输出图片如右图所示：

对角线为y=x，点出的两个绿点为TEST集预测与实际对比，点出的若干红点为TRAIN集预测与实际对比。

据此可知结果为：

|  |  |
| --- | --- |
| **Experimental Activity** | **Predicted Activity** |
| **log(1/IC50)** | **log(1/IC50)** |
| -2.30 | -1.87502535 |
| -1.93 | -1.756052112 |
| -1.70 | -1.730566482 |
| -1.11 | -1.579111992 |
| -0.85 | -1.353589508 |
| -1.15 | -1.963921122 |
| -0.68 | -1.897200228 |
| 0.47 | -1.838266128 |
| 0.80 | -1.740484668 |
| -1.34 | -1.180760918 |
| -0.96 | -0.701275092 |
| -0.90 | -0.868228758 |

注：其中TEST集合为避免人为选取影响，采用随机种子产生两个数来选取。（由于随机选取可能与其他同学作业相重复，不过此处理已覆盖所有可能性）

本程序成功实现所提供所有分子随机选取一~两个分子作为TEST集，其余作为TRAIN集最小二乘法多元拟合的任务。

所有结果均已上传至：<https://github.com/neetsaki/Molecule-design-homework>

其他运行结果截图如下：



