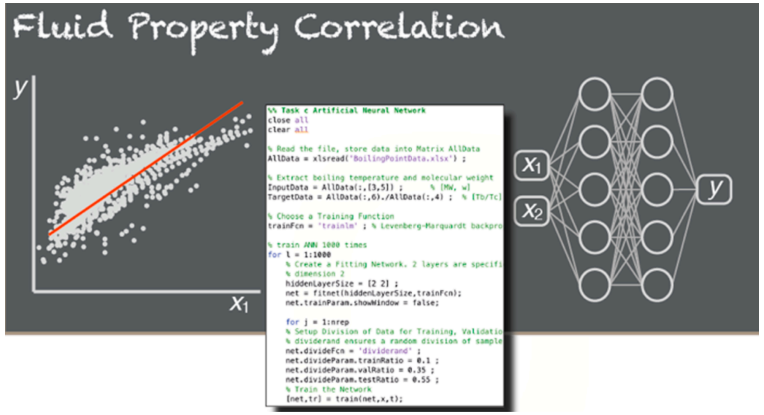
Imperial College London

# Machine Learning for Fluid Property Correlations

Classroom Examples with MATLAB

**Lisa Joss & Erich A. Müller**

*Department of Chemical Engineering*
*Imperial College London  U.K.*

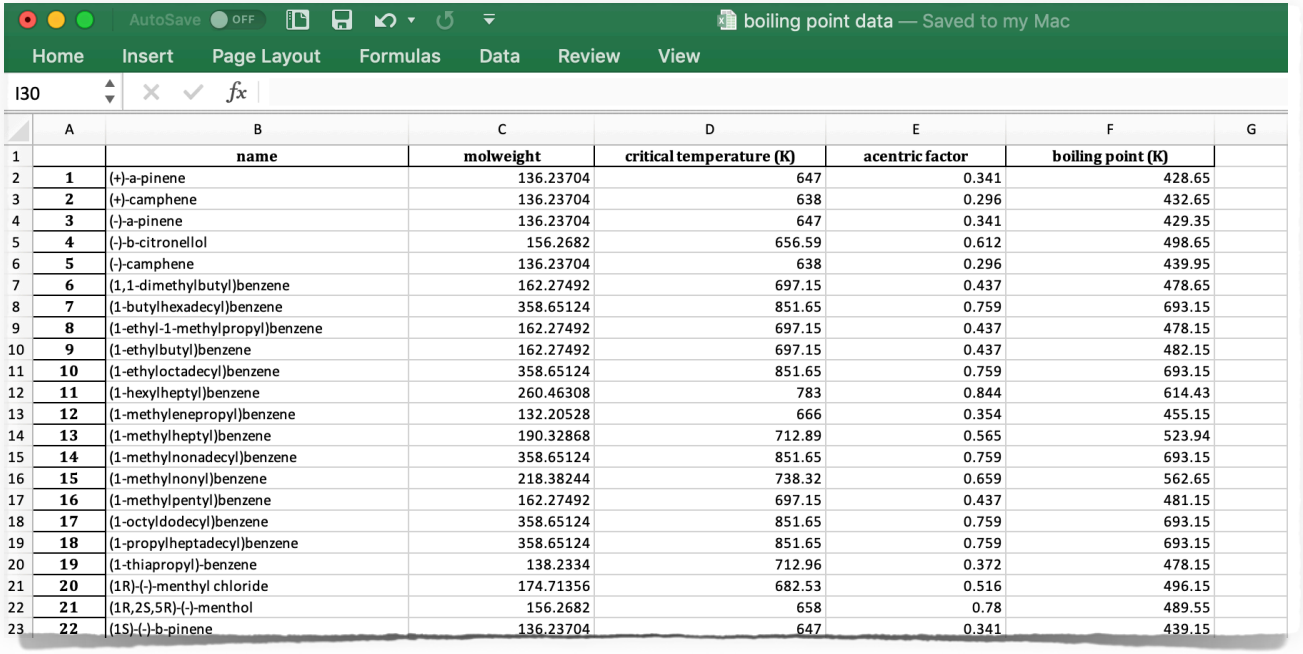


## Motivation & Background

- 1st year Chem Eng undergraduates at Imperial College London
- Students are enrolled in a mandatory 6-week hands-on “Introduction to programming and MATLAB” course. The course is practical pass/fail module.
- The course covers the basics of programming in the first three weeks and then starts focusing on skillsets needed for the rest of the Chem Eng curriculum including plotting, solution of linear sets of equations, ordinary differential equations, etc.
- There was a pressing need (and request from students) to be introduced to some (basic) notions of Machine Learning.

# The challenge:

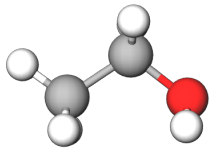
**Given a (very large) table of physical properties for many organic substances, can you produce an engineering-quality correlation for the boiling point?**




| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | | name | molweight | critical temperature (K) | acentric factor | boiling point (K) | |
| 2 | 1 | (+)-a-pinene | 136.23704 | 647 | 0.341 | 428.65 | |
| 3 | 2 | (+)-camphene | 136.23704 | 638 | 0.296 | 432.65 | |
| 4 | 3 | (-)-a-pinene | 136.23704 | 647 | 0.341 | 429.35 | |
| 5 | 4 | (-)-b-citronellol | 156.2682 | 656.59 | 0.612 | 498.65 | |
| 6 | 5 | (-)-camphene | 136.23704 | 638 | 0.296 | 439.95 | |
| 7 | 6 | (1,1-dimethylbutyl)benzene | 162.27492 | 697.15 | 0.437 | 478.65 | |
| 8 | 7 | (1-butylhexadecyl)benzene | 358.65124 | 851.65 | 0.759 | 693.15 | |
| 9 | 8 | (1-ethyl-1-methylpropyl)benzene | 162.27492 | 697.15 | 0.437 | 478.15 | |
| 10 | 9 | (1-ethylbutyl)benzene | 162.27492 | 697.15 | 0.437 | 482.15 | |
| 11 | 10 | (1-ethyloctadecyl)benzene | 358.65124 | 851.65 | 0.759 | 693.15 | |
| 12 | 11 | (1-hexylheptyl)benzene | 260.46308 | 783 | 0.844 | 614.43 | |
| 13 | 12 | (1-methylenepropyl)benzene | 132.20528 | 666 | 0.354 | 455.15 | |
| 14 | 13 | (1-methylheptyl)benzene | 190.32868 | 712.89 | 0.565 | 523.94 | |
| 15 | 14 | (1-methylnonadecyl)benzene | 358.65124 | 851.65 | 0.759 | 693.15 | |
| 16 | 15 | (1-methylnonyl)benzene | 218.38244 | 738.32 | 0.659 | 562.65 | |
| 17 | 16 | (1-methylpentyl)benzene | 162.27492 | 697.15 | 0.437 | 481.15 | |
| 18 | 17 | (1-octyldodecyl)benzene | 358.65124 | 851.65 | 0.759 | 693.15 | |
| 19 | 18 | (1-propylheptadecyl)benzene | 358.65124 | 851.65 | 0.759 | 693.15 | |
| 20 | 19 | (1-thiapropyl)-benzene | 138.2334 | 712.96 | 0.372 | 478.15 | |
| 21 | 20 | (1R)-(-)-menthyl chloride | 174.71356 | 682.53 | 0.516 | 496.15 | |
| 22 | 21 | (1R,2S,5R)-(-)-menthol | 156.2682 | 658 | 0.78 | 489.55 | |
| 23 | 22 | (1S)-(-)-b-pinene | 136.23704 | 647 | 0.341 | 439.15 | |

**Excel sheet with over 5000 entries**

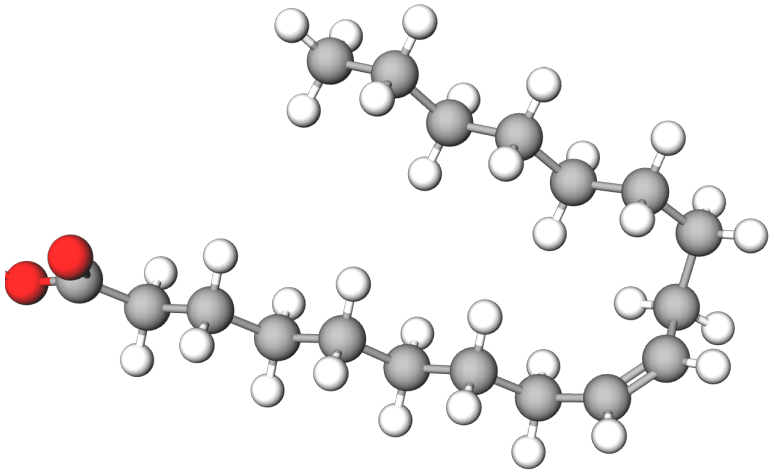
**A first empirical observation is that the boiling point is proportional to the molecular weight**



**Ethanol**

**MW = 46.07**

**$T_b$ = 78.4 ºC**

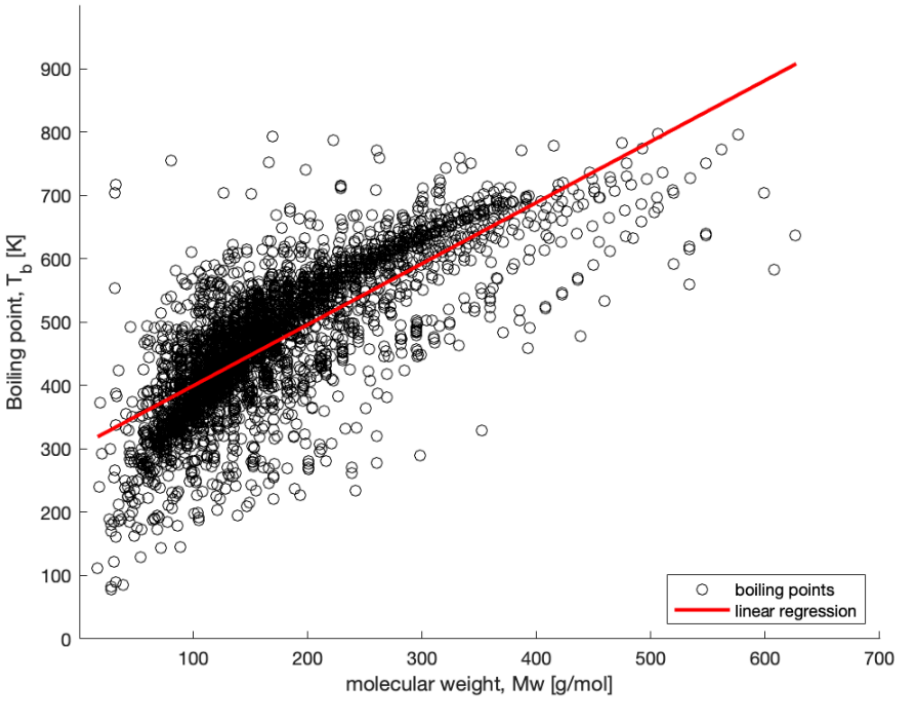


**Oleic acid**

**MW = 282.47**

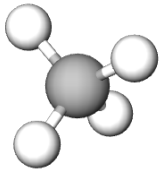
**$T_b$ = 360 ºC**

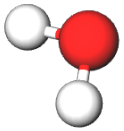
# TASK 1 : Linear fit of the boiling point to the molecular weight




- The correlation is rather poor ($R^2 = 0.76$)
- There is *some* trend, but obviously there are other parameters which are also of importance.
- Other fits ( logarithmic, quadratic, etc. ) will clearly not be successful.

**Second ansatz : the boiling point has some relation with the acentric factor**





**Methane**

**MW = 16.04**

**$T_b$ = -161.5 ºC**

**ω = 0**

**Water**

**MW = 18.01**

**$T_b$ = 100 ºC**

**ω = 0.344**

- The acentric factor is an empirical number

$$\omega = -\log_{10}(p_r^{sat}) - 1, \quad \text{at} \quad T_r = 0.7$$

- Its value is close to zero for noble gases and increases as the molecule becomes non-spherical and/or polar.
- It is commonly tabulated (along critical properties)

## TASK 2 : Multivariate correlation

- Assume that the boiling point is a linear function of both the molecular weight and the acentric factor

$$T_b = \theta_0 + \theta_1 \omega + \theta_2 MW$$

- Scale the boiling temperature with the appropriate critical temperature. This scales the $T_b$ values from 0.7 to 1

$$y = T_b / T_c$$

- Solve the problem by matrix manipulation

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{100} \end{pmatrix} = \begin{pmatrix} 1 & \omega_1 & MW_1 \\ 1 & \omega_2 & MW_2 \\ \vdots & \vdots & \vdots \\ 1 & \omega_{100} & MW_{100} \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{pmatrix} = \mathbf{X}\theta$$

$$\theta = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T y$$

**Classical procedure**

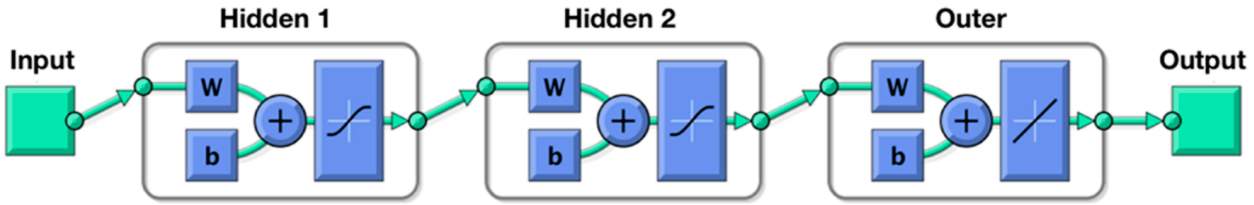
Determine a priori the mathematical (plausible) relationship

Employ physical insights

Solve the minimum likelihood problem

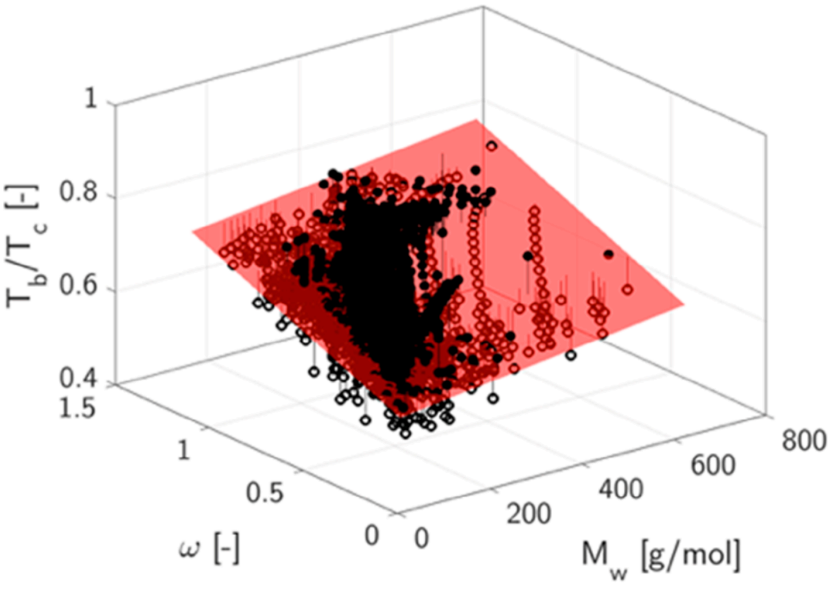
## TASK 3 : Employ an Artificial Neural Network (ANN)

- No assumption is made with respect to the mathematical structure of the “correlation”
- The *features* are ω and MW ( by simple inference)
- Solve the problem *training* with 100 randomly selected data points
- MATLAB has a built-in ANN encoder



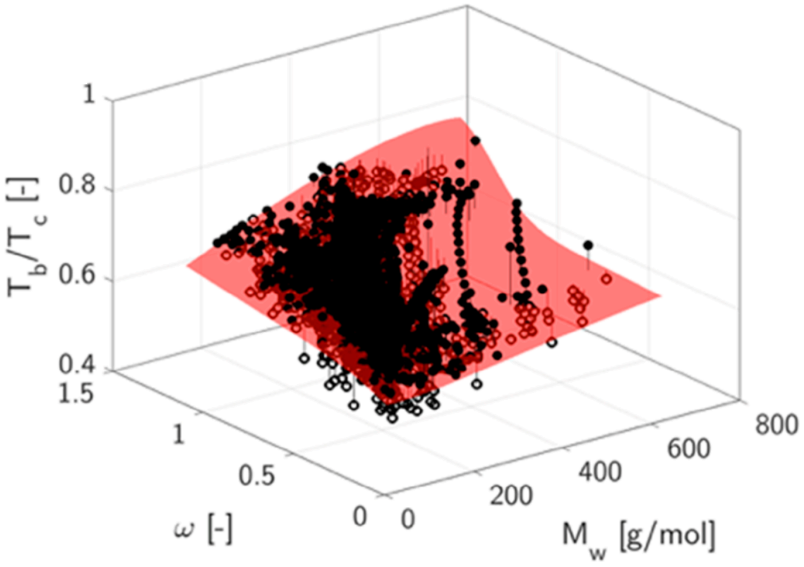

- The ANN is composed of two hidden layers with a tan–sigmoid transfer function and an outer layer with a linear transfer function (gray boxes).
- The weights (W) and biases (b) are optimized using the Levenberg–Marquardt algorithm.
- Green boxes represent the algorithm input (left) and output (right)

**Multivariate**




**$R^2$ = 0.84**
**AAD = 2.7 %**

**ANN**




**$R^2$ = 0.89**
**AAD = 2.2 %**

# Example using live scripts

## Conclusions

- The exercise has been extremely well received by the students, who come back asking for more material to expand their understanding of the topic.
- The example can be expanded and improved easily, although the ML correlation is already quite good. NOTE: Converting the exercise to Python is rather straightforward.
- Machine learning has crept up in a large number of the final year research projects and has proven to be an extremely popular topic (and a skill requested by employers).
- A final year elective on Machine learning in Process Engineering is now being developed.

**Please direct the questions to**


- ☆ Lisa Joss lisa.lj.joss@gmail.com
- ☆ Erich A. Müller e.muller@imperial.ac.uk


**More details**


L. Joss and E. A. Müller, “Machine Learning for Fluid Property Correlations: Classroom Examples with MATLAB,” *J. Chem. Educ.*, **96**(4), 697–703, 2019.


MOLECULAR SYSTEMS ENGINEERING

http://www.molecularsystemsengineering.org