A Short Course in Chemometrics

Python tools to standardize and accelerate chemometric modeling and data analysis.

Welcome

🎯 The goal of this short course is to introduce and explain elementary chemometric analysis methods. We will also touch on more advanced ML approaches. The course will cover the use of python-based tools that can accelerate your workflow and improve reproducibility.

We will assume **no prior knowledge or familiarity** with any of these methods, tools, or mathematical background. We will review only as much mathematics as is necessary to ground an understanding of the methods discussed since a deep understanding is not necessary for application, which is the focus of this course.

Why Are We Here?

🚀 What we hope to achieve:

1. Give you a new set of tools to help you do your job better

2. Create a coherent and more consistent approach to chemometric analysis across NIST by introducing you to a standard library for these tasks

3. Improve reproducibility and transparency

Example

Given the following sequence: [1, 2, 3, 4]

? Q1: What is the median?

? Q2: What is the 75th percentile?

Example

A: There are, in fact, up to 13 different ways to compute percentile (median = 50th percentile)!

🙋 A1: "Correct" answers include: [2.0, 2.5, 3.0]

🙋 A2: "Correct" answers include: [3.0, 3.25, 3.5, 3.5625, 3.583, 3.75, 4.0]

The main 9 are reviewed here: R. J. Hyndman and Y. Fan, "Sample quantiles in statistical packages," The American Statistician, 50(4), pp. 361-365, 1996, but numpy has implemented 4 more!

Example

" There are a large number of different definitions used for sample quantiles in statistical computer packages. Often within the same package one definition will be used to compute a quantile explicitly, while other definitions may be used when producing a boxplot, a probability plot, or a QQ plot. ... We argue that there is a **need to adopt a standard definition** for sample quantiles so that the same answers are produced by different packages and within each package. "
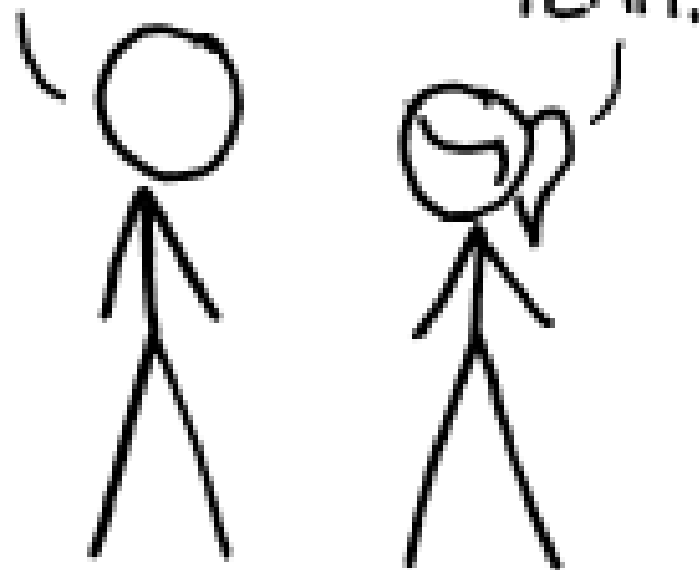
-Hyndman & Fan

What Should I Get Out of This?

📚 In the end you will be able to go to a library of standardized example notebooks, select the one you need, enter your data, then run it from start to finish.

This course will also teach you to modify and expand things as needed.

This course is currently structured in a lecture-based format, but may change in the future to include more interactive aspects.

Instructors

646.04 - Nate Mahynski, nathan.mahynski@nist.gov

Thanks 👏 also goes to:

646.04 - Bill Krekelberg, william.krekelberg@nist.gov

646.04 - Tom Allison, thomas.allison@nist.gov

Feel free to reach out with

❓ Questions
💬 Comments
😕 Concerns
👍 Want to collaborate

Course Layout

All course material can be found at
https://github.com/mahynski/chemometric-carpentry.

Much of it is based on the 🐍 Python package PyChemAuth available
at https://github.com/mahynski/pychemauth

The syllabus (README) on the course GitHub site has links to all the
notebooks and materials so you can use it to navigate.

In The Future

❓ You can ask questions, provide feedback, and find community support on the GitHub Discussions page for this course.
❌ If you find a mistake please submit a Bug Report.
🔭 If you would us to cover new area(s) or have an idea to improve this course, please submit a Feature Request!
💡 Is you have requests or ideas specific to PyChemAuth you can find similar options on its Issues page.
👬 Please consider contributing to PyChemAuth examples!

Getting Started

You will need:

1. A Google account, and

2. 🧠 An open mind!

All code is free and open source. It is run in the cloud on [Google Colab](#) so you do not need to install anything, and you can access it anywhere, anytime as long as you have an internet connection.

Let's get started: https://github.com/mahynski/chemometrics_short_course