2024 Undergraduate Reseach Conference

Using Simple Python to Help Workflows in the Sciences

Dean Walck
Advisor: Dr. Yonghong Tong

NIAGARA UNIVERSITY

Abstract

As a student in the natural sciences, there is an emphasis on the collection, review, critique, and parsing of experimental data. After being collected and recorded, data is typically processed using numerous programs. These programs are powerful and ubiquitous, but not always accessible. Furthermore, desired analytical outputs of data in undergraduate study are often relatively benign: these include mean, median, linear regression, variance, and standard deviation.

This study employs Python, the most widely used programming language, to develop a tool aimed at expediting processing of experimental data. In turn, this tool potentially allows for time to be freed up for other forms of analysis, provide immediate data results for calculations, and provide early detection of errors and anomalies in data sets. This may also provide a basis for future features expansion under Python's open-source framework. Furthermore, availability of the widget means errors in real data analysis are likely to decrease due to the reduced time between data collection and data processing, which often relies upon the integrity of experimental records and an understanding of experimental processes as they've physically happened. Thus, this tool is not an effort to reinvent the wheel, but a simple effort to intuitively increase workload throughput.

Introduction

Typically, the curriculum of the natural sciences does not favor extensive work with coding or advanced studies in applied mathematics. These subjects are usually subsumed into the mathematics curriculum, whereas chemists and biologists tend to employ software suites to process data gathered in the laboratory.

Therefore, an effort to develop a Python module was motivated by the desire to tailor a program for specifically tackling data analysis typically seen in an undergraduate-level chemistry or biochemistry laboratory, by exploiting data analysis packages innate to Python. This approach allows for a focus on obtaining information that is relevant to the nature of such studies being conducted in a more accessible medium (i.e. a cel phone). This also allows the possibility of expanding the feature base as the need arises given the program's modularity.

Result

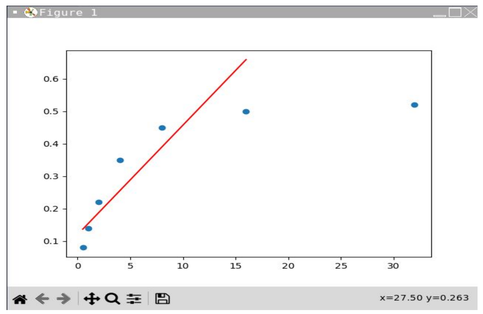
Developing the Python program was mostly a matter of importing multiple packages, including numpy, matplotlib, and sklearn, which were relevant to the types of data analysis being performed. These packages provide mathematical and analytical functions which contributors have added at some point to Python's open-source directory. These functions may be selectively implemented in each program to achieve the specific objectives of the author.

An example data set of the enzyme kinetics of Lactase Dehydrogenase obtained through spectrophotometry was used as a test case for the program's functions. This type of data set produces a plot of a non-linear Michaelis-Menten hyperbolic curve. This incorporates a set of two variables, the Lactose concentration (in molar), and the initial reaction rate (mols/time).

Data:

Lactose Concentration (mM)	Initial Rate (µmol/min)
0.5	0.08
1.0	0.14
2.0	0.22
4.0	0.35
8.0	0.45
16.0	0.50
32.0	0.52

```
Enter the x-values (separated by spaces): 0.5 1 2 4 8 16 32
Enter the y-values (separated by spaces): 0.08 0.14 0.22 0.35 0.45 0.50 0.52
Mean squared error: 0.014235516673216191
R^2 score: 0.677199168407796
x-values:
Mean: 9.071428571428571
Median: 4.0
Range: 31.5
Variance: 112.74489795918369
Standard deviation: 10.6181400423607
y-values:
Mean: 0.32285714285714284
Median: 0.35
Range: 0.44
Variance: 0.02716326530612245
Standard deviation: 0.16481281899816666
```



This experiments typically produce a curve which is used to determine characteristics about the enzyme being studied. Linear regression and the line of best fit allow for extrapolation of the maximum velocity of the catalyzed reaction, the initial velocities at a given concentration of enzyme, as well as the type of activity around active sites. Therefore, this was a good model to use to extract general statistical data (which is valid for any data set), as well as data to make predictions about the object being studied.

In this case, data was sufficiently processed without error, and a line of best fit was produced. Unfortunately, the line of best fit was linear, which would only be valid in the case of a Lineweaver-Burke transformation of the data. Likewise, the equation of best fit should be provided. These can be corrected with a few modifications.

Conclusion

This Python program was able to calculate the desired data analysis in a matter of seconds, requiring very little computational power. Therefore, this program should be useable on any device able to run it, including cell phones.

Modifications of the code should provide more nuanced interpretations of the data, including non-linear best-fit curves, as well as their explicit expressions. However, for linear data sets the projects remain valid. Overall, the utility of such a widget is apparent, particularly in the ability to represent data immediately which may bring to light peculiarities in the data set. However, this does not replace more formal representations of the data, as would be provided in a lab report.

Future Work

It's possible to greatly expand functionality of the widget through the use of existing libraries, as well as user-defined functions.

- More extensive statistical analysis may be implemented through additional function packages.
- More advanced data modeling may be implemented, with a choice of which model best fits the data present.
- A sufficiently complex UI may streamline features to make the program more accessible.

Acknowledgement

This research has been supported by Niagara University.

References

Materials and data used information from Python and Data Analysis class materials, StackOverflow, Google API and reference documentation, and Python documentation from Replit and Python community forums.