

Feature Discovery in Small-Sized Experiments in Early Drug Development

Steven Wallaert, Promotor: Prof. Dr. Ir. Olivier Thas



Introduction

As we currently reside in the era of big data, artificial intelligence and machine learning methods are on a rising tide. Interest in these methods are ever increasing since constraints in terms of computing power, storage capacity and connectivity are less of a concern. However, in certain technological settings data gathering is still limited in sample size due to high costs and limited resources. More specifically, in early drug development research high throughput techniques such as and are often used in small-sized experiments. The high dimensional data sets produced by these experiments often include tens of thousands of features and only 10 to 50 observations. Researchers in this field hope they can exploit the power of machine learning and AI to make more and better discoveries that would eventually lead to new and better drugs. Yet, since their main application is in situations where data sets consist of thousands or even millions of observations and only a fraction of features, little is known about how these ML techniques empirically perform. As such, it is unclear how these data sets in early drug discovery are best approached.

Objectives

1. Gain insight in how researchers within this field analyzed these kinds of data in the recent years.
2. Perform a simulation study comparing the most relevant methods in varying realistic conditions.
3. Formulate clear data analysis guidelines
4. Exemplify by means of a data analysis of real data

Methods

1. Literature review
2. Simulation study

- Aim:
- Data generating mechanisms
- Estimands
- Methods
- Performance metrics

Results

Usually you want to have a nice table displaying some important results that you have calculated. In `posterdown` this is as easy as using the `kable` table formatting you are probably use to as per typical R Markdown formatting.

You can reference tables like so: Table [1](#). Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam placerat augue at velit tincidunt semper. Donec elementum porta posuere. Nullam interdum, odio at tincidunt feugiat, turpis nisi blandit eros, eu posuere risus felis non quam. Nam eget lorem odio. Duis et aliquet orci. Phasellus nec viverra est.

Table 1: Table caption.

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
5.1	3.5	1.4	0.2
4.9	3.0	1.4	0.2
4.7	3.2	1.3	0.2
4.6	3.1	1.5	0.2
5.0	3.6	1.4	0.2
5.4	3.9	1.7	0.4
4.6	3.4	1.4	0.3
5.0	3.4	1.5	0.2
4.4	2.9	1.4	0.2
4.9	3.1	1.5	0.1

Or with figures: [Figure 1](#), or [Figure 2](#).

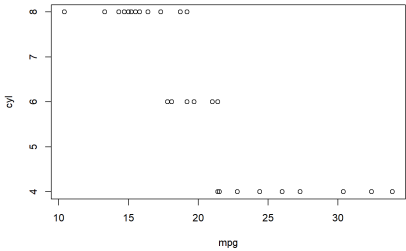


Figure 1: Great figure!

```
data <- iris

plot(x = data$Sepal.Length,
     y = data$Sepal.Width,
     col = data$Species,
     pch = 19,
     xlab = "Sepal Length (cm)",
     ylab = "Sepal Width (cm)")
```

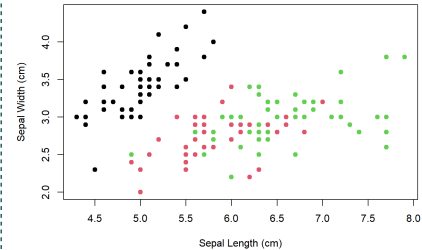


Figure 2: Amazing, right?!

Next Steps

Aliquam sed faucibus risus, quis efficitur erat. Vestibulum semper m quis tempus eleifend. Aliquam sagittis dictum ipsum, quis viverra l eleifend ut. Curabitur sagittis vitae arcu eget faucibus. In non eleme felis. Duis et aliquam nunc. Nunc pulvinar sapien nunc, vel pretium efficitur in. Fusce fringilla maximus leo et maximus. Fusce at l laoreet, iaculis mi at, auctor odio. Praesent sed elementum justo. Ae consetetur risus rhoncus tincidunt efficitur. Praesent dictum maur diam maximus maximus (Thorne 2019).

Conclusion

Try posterdown out! Hopefully you like it!

References

Thorne, Brent. 2019. Posterdown: Generate Pdf Conference Posters Using R Markdown. <https://github.com/brentthorne/posterdown>.