

Lee Panter  
Methods Project Proposal

### **What is the Method?**

I am proposing a project on Self Organizing Maps (SOMs).

*Before I proceed with a description of this method, I'd like to state that I am not completely sure I understand this method. Having said that, I believe SOMs will be both useful and relevant to my project.*

I believe that this method can be simply characterized as: a way of inferring a relationship between a supervised outcome and the results of an unsupervised learning process.

So...what does that mean?

Suppose you have a high-dimension data set (lots of variables) with an outcome variable of interest. A SOM collapses the information contained in the predictor space into smaller, more easily managed groups. This condensing of information is performed in such a way that optimizes representation of the original predictor space, then the new (reduced) space is "mapped" to the outcome space.

The method is algorithmically similar to fitting and mapping an Artificial Neural Network (ANN), but SOMs are distinct from ANNs in the method by which they calculate a fitted value's loss and compensate for this loss.

I think this method can be applied to regression or classification, but as of now I have only seen it applied in the context of classification.

### **Why do you want to cover this method?**

It is a belief of mine that: supervised learning approaches can often be improved by implementing an unsupervised exploratory data analysis step. This conclusion is not particularly insightful, but the concept of a SOM is an example of why the conclusion is still valid.

I think that this methods presentation is going to be useful to others. I have seen applications of this method ranging from Biology to Finance, which means that this method would be a good candidate for presentation in the Consultation context.

I am extremely interested in the theoretical and applied aspects of this method. Even in the short amount of reading I have completed on SOMs, there has been lengthy discussion of

“topological preservation” and “competitive processes”. So, I’m guessing that an introduction into the theory would be sufficient. As for the application; I am hoping that I can learn to train an ANN at the same time I am learning how to fit a SOM. I have learned a couple of packages for ANNs in R, and I know there are a couple that work for SOMs, but I am hoping this can all be done in KERAS, or other recommended program.

### **How does it benefit you consulting project?**

Throughout the course of my meetings with Alan I have had to limit expectations to results on the depression health surveys. In reality, the method that (I think) Alan is attempting to implement might be relatable to a variety of psychological traits. He even gave me data that would make this feasible.

I have the ultimate goal of producing visual representations of Alan’s data using SOMs. This would be distinct from the Methods Project itself. My intention (thus far) would be to essentially just *swap outcomes*. Instead of using the PHQ9 data to model a depression outcome, I would use a SOM to relate it to a different psychological trait as quantified by the Quick Diagnostic Panel (QDP). This is still a process I am thinking through, but the motivation behind using the SOM is the ability to relate the QDP result of interest back to the PHQ9 predictor space, and provide usable visualizations in the process.