**Lesson 04 Discussion**

As you have seen in class and in your lab, unsupervised learning models are useful in finding relationships in unlabeled data. Since much, and perhaps most, of the available data in the real world is unlabeled, unsupervised learning methods are often quite useful.

Think about a machine learning problem (real, or hypothetical) where you might apply unsupervised learning. Discuss the following points:

1. Why is unsupervised learning useful for this problem?
2. Which unsupervised learning methods would you plan to try, and why?
3. How will you evaluate and compare the results of your models?

  To receive credit, post you thought and then comment on at least one other post.

As I mentioned before in a previous discussion, one of the ways I have applied unsupervised learning was to explore thermodynamic data on commercial engines to find clusters/patterns as well as a way to determine feature importance on a prediction model.

<https://www.ibm.com/cloud/learn/unsupervised-learning>

<https://towardsdatascience.com/unsupervised-learning-and-data-clustering-eeecb78b422a>

*While working on the thermodynamics’ side for commercial turbofan engines very quickly we realized that if we wanted to have any prediction model in production looking at “on-wing” and/or “test-cell” data providing real time predictions we needed to somehow reduce the number of features entering the model while maintaining certain degree of control and accuracy. As you can imagine for turbofan engines we have many sensors while flying, plus extra measurements from extra sensors while testing the engine on the ground (“shop-test cell”), which in turn might lead to thousands of inputs for every single data point. During this time many different methodologies were implemented to try to solve this problem of dimensionality reduction and one of those techniques implemented in some of the models (especially for those with a high degree of complexity) was PCA to find a lesser set of features that could provide a comparable (or even better) performance, while improving the time response. Using this methodology we were able to find great insights on the contribution of the kind of sensors we were using, but ultimately for most of our models this technique was just the first step for feature selection and not the final answer to our high dimensionality issue. This mainly because our customers needed not only the prediction but a model that was easier to interpret.*