Dimensionality Reduction: Through the Eyes of Our Working Example

OUR STORY

High dimensional data are becoming more prevalent in enterprise settings, and data are becoming more mature, especially in larger, established companies. AAVAiL is no exception. The *internet of things* and an increasing amount of unstructured text data are two contributing factors to this effect. AAVAiL is saturated with data from tens of thousands of mobile devices, its streaming media infrastructure, and text communications from customers. One of the most essential strategies for dealing with this type of complexity is dimensionality reduction.

Physics and the biological sciences have dealt with high-dimensional data for a long time, and the same techniques used for dealing with high-dimesionality in those fields are now becoming commonplace in industry. Dimensionality reduction techniques are commonly used in many areas of data science including topic modeling and image segmentation. In these materials we will summarize some of matrix decomposition and manifold learning methods that can be readily incorporated into you data transformation pipelines. We will apply some of these methods in the context of topic modeling in the next case study.

THE DESIGN THINKING PROCESS

Along with managing bias in your data, you'll also be expected to manage the complexity of your data via dimensionality reduction as you pass through the *Ideate* and *Prototyping* phases of the design thinking process. One key area where dimensionality reduction is very useful is in data visualization. You *will* be asked to generate visuals of critical data distributions, and it will almost always be necessary to employ various dimensionality reduction techniques to arrive at a data set that can be easily used in a data plot.