September 6, 2018

DATA 601

50 Years of Data Science – A Summary of the first Ten Pages

In this article, author David Donoho discusses the history of data science, more importantly the inception and evolution of data science. It all started approximately fifty years ago, when “John Tukey called for a reformation of academic statistics” (Donoho). The article is written in portions, split into the different parts of data science. The solid case for data science to be created, as Donoho states, is for facing essential questions and using scientifically backed techniques to answer those questions.

The article starts off by discussing the relationship between statistics and data science. When data science was emerging, many statisticians felt that what they did – which was “collecting and analyzing numerical data in large quantities” (Donoho) was very similar (if not the same) to data science. There are many ways these professions differ. One big difference between the two fields is the handling of “big data”.

Data science deals with the massive amounts of data, which rationally could not have been handled before the emergence of the new field. Skills include but not limted to Hadoop, Map/Reduce, MatLab, R, Python etc. Also, the ability to create algorithms that handle such big data in an organized manner are needed to thrive in this field.

Donoho also discusses the hiring bump from companies such as Google or Amazon that will need to be filled in the next couple of years by data scientists or “big data talent”. These companies deal heavily in big data, from creating more affective search engines to providing a better online shopping service.