Data Science

17 September 2019

Class Summary

[OPENING]: This class is formally titled, “Analyse et Traitment de l’Information,” but the class is essentially applications of Data Science.

[GRADING]: The final exam will be an oral exam. The structure is as follows: you will be invited to pick one/two questions out of a hat. These questions are essentially chapters of our studies. You will be given 20 minutes to prepare with notes, at the end of which you will give a presentation outlining the chapter as well as answer professor’s questions. You will be given the whiteboard, and are encouraged to demonstrate knowledge of the mathematical process of the material you cover.

[WARNING]: Pay heavy attention the mathematical representation/presentation of the theories/algorithms we talk about in class. This class will reward formal mathematical understandings very well, and will punish you for not doing so.

[SUMMARY]: A vector space is essentially just a set of values that adhere to a certain set of rules. In the practical application sense, they are just restraints/constraints on how we organize data in order to make efficient computations with computers. The reason why we want to be in a vector space when we do computations, is because we care about computations such as averages, deviations, etc. Those calculations are valuable, and therefore we design the organization of our data to adhere to certain patterns in order to guarantee that those computations still hold true for vectors the same way we think about them if we were using just simple numbers (like in math). A basis is what ends up allowing us to represent functions as a vector of numbers (also keep in mind that for a subspace, there is not just one basis. You can have many. That’s why choosing a good basis is hard and also relevant). Linear Independence means that none of the vectors we have collected can be written as a combination of any other vectors, in a practical application sense, it means that each vector of data we have brings unique information. We care about this before we want our data to give us meaningful insight on how to solve our problem. A basis is the maximum family of vectors in a span (which makes sense because if we had any more they’d presumably be able to be made by a combination of vectors in our span which would break independence), and also the minimum, since we wish our vectors to span all the values in our space (this makes sense because we are able to represent a function as a vector of numbers. If we had any less, we would not be representing all the values that define our function). A “norm” is just a way for us to talk about size or in the two-dimensional way of saying it, the “magnitude.” Mathematically it is a function that assigns a size to a vector in a vector space. A norm is established if you can satisfy the conditions, much like how we had to satisfy conditions to establish a vector space. A linear map, also called a linear transformation, is just some relation that preserves operations of addition and scalar multiplication.