Data Science Project Topic and Questions

Unifying Data Science

January 21, 2020

For this assignment, you and your group must decide on a topic area for your class data science project, identify a problem you think needs to be solved in this domain, and write three questions (one descriptive, one causal, and one predictive) that you think would help address your question.

The goal of this exercise is for you and your team to start the process of picking a topic for your project. You will not be held to answering one of these questions as your final project, but you should take this as an opportunity to get feedback on questions you may seek to answer.

Note that while you should keep your questions specific and actionable, you do not need to present a detailed plan for how you would answer your questions (e.g. I don't need a list of variables you need form specific datasets). Your actual project will need to answer a causal question, but since we haven't covered causal inference yet, you aren't expected to present a method for answering your causal question.[1] With that in mind focus on picking a topic area that excites you, a problem data science can help address, and good, well-motivated, answerable questions.

In writing your questions, please pick a predictive question that is *not* the direct extension of your causal question. A good way to do this is to focus on a situation where you wish to predict something *without* manipulation – i.e. "can I predict behavior X using [existing data]" or "can I predict outcome Y using [existing data]" rather than "what would be the result of action Z," or "what would be the result of policy change A." (As we'll discuss later in the class, predictive questions that don't involve manipulations are ones that often lend themselves to supervised machine learning).

Due: Wednesday, January 29th, 10am

---

[1]Many of you are also at the very start of machine learning, so it also wouldn't make sense to ask you to explain how you'd answer your predictive question.