

Week 10 homework

Week 10 homework

Let's assume the company that wants to monetize from a single source dataset is a major bookstore company, for example Barnes and Noble or Amazon. The dataset they'll use to get value is Alumni Magazine. Since alumni magazine gives us information about a subject's College/University, Year of Graduation, Majors, Current City etc. Based on the frequency of certain Majors and year of graduation we can use multinomial logistic regression to indicate which category of books may sell in coming years. For example, if we cluster samples of graduation year ,2016, with a specific major, and saw an increase in that same major in the following year 2017 graduation, then we can expect an increase for a book that may relate to that specific major. And that's one possible way a book company can monetize from the alumni magazine data.

Like dataset 1 problem, we're still a bookstore company. But now we'll use 2 different datasets instead of one. We can still use the alumni magazine along with the company website. Assuming that some samples from one dataset matches exactly with the other dataset. We can record a person's most search item on the company site. We'll then look at the person's income on the alumni magazine and based on the income we can make suggestive items for sale on that particular subject. So, a high income, a person with more disposable income may get suggestive items that are higher in price, and these items also relate to what he/she search on the company's site.