

Machine Learning - Lab 3

Exploratory Data Analysis

Souhaib BEN TAIEB

24 February 2020

"Data was gathered from participants in experimental speed dating events from 2002-2004. During the events, the attendees would have a four minute "first date" with every other participant of the opposite sex. At the end of their four minutes, participants were asked if they would like to see their date again. They were also asked to rate their date on six attributes: Attractiveness, Sincerity, Intelligence, Fun, Ambition, and Shared Interests.

The dataset also includes questionnaire data gathered from participants at different points in the process. These fields include: demographics, dating habits, self-perception across key attributes, beliefs on what others find valuable in a mate, and lifestyle information."

- Load the Speed Dating data (available at this link)
- Read the associated documentation (available at this link)
- Confirm the number of males and females in each wave given in the documentation is correct
- How many people have participated to the speed dating experiment?
- How many dates each person has participated to? Compute a summary of these numbers.
- Use the function *ggplot* in package *ggplot2* to visualize ten variables you think are important in dating.

Exercise

Write code to answer the following questions:

1. What are the least desirable attributes in a male partner? Does this differ for female partners?
2. How important do people think attractiveness is in potential mate selection vs. its real impact?
3. Are shared interests more important than a shared racial background?
4. Can people accurately predict their own perceived value in the dating market?
5. In terms of getting a second date, is it better to be someone's first speed date of the night or their last?
6. Write two other interesting observations using this data set.