



3 messages

Wed, Feb 17, 2016 at 3:24 AM

To: Ali Bagherpour <ali.bagherp@gmail.com>, Andrew Ho <Kironide@gmail.com>, Chad Groft <clgroft@gmail.com>, David Bolin <david@bolin.at>, Jacob Pekarek <jpekarek@trinity.edu>, Jaiwithani <jaiwithani@gmail.com>, James Cook <cookjw@gmail.com>, Linchuan Zhang <email.linch@gmail.com>, Matthew Gentzel <magw6270@terpmail.umd.edu>, Olivia Schaefer <taygetea@gmail.com>, Sam Eisenstat <sam.eisenst@gmail.com>, Tom Guo <tomguo4@gmail.com>, Trevor Murphy <trevor.m.murphy@gmail.com>

Yesterday was exhausting in part because of the heat. Fortunately the next 5 days will be much cooler :-).

Pair assignments

Ali <---> Jacob
Olivia <---> Chad
Jai <---> Trevor
Tom <---> Andrew
David <---> James
Matthew <---> Linch

Some reading

Today we'll be looking at the data from [Gender Differences in Mate Selection: Evidence from a Speed Dating Experiment](#). Take a look at Sections I and II for their main results and methodology. Read the [documentation](#) for the dataset to get a clear sense for what it contains. The dataset itself can be downloaded [as a CSV file](#).

I used the dataset for my very first data science project. I have links to my blog posts on the subject at <http://jonahsinick.com/speed-dating-project/>. Take a look at the first post *How subjective is attractiveness?*, which gives an example of a presentation of an analysis.

If you would like to see a coherent project done in R, you can check out my code at <https://github.com/JonahSinick/speedDating>. It's far from a model of good R code, as I was just learning at the time, but the functionality of the code is at least well documented in the readme.

The project today is around **predicting attractiveness, intelligence, sincerity, and other traits as measured by group consensus in terms of variables collected on participants before the events**. We will be using shrinkage methods as described in section 6.2 of [Introduction to Statistical Learning](#).

Starter code to follow.

Wed. Feb 17. 2016 at 10:40 AM

To: Ali Bagherpour <ali.bagherp@gmail.com>, Andrew Ho <Kironide@gmail.com>, Chad Groft <clgroft@gmail.com>, David Bolin <david@bolin.at>, Jacob Pekarek <jpekarek@trinity.edu>, Jaiwithani <jaiwithani@gmail.com>, James Cook <cookjw@gmail.com>, Linchuan Zhang <email.linch@gmail.com>, Matthew Gentzel <magw6270@terpmail.umd.edu>, Olivia Schaefer <taygetea@gmail.com>, Sam Eisenstat <sam.eisenst@gmail.com>, Tom Guo <tomguo4@gmail.com>, Trevor Murphy <trevor.m.murphy@gmail.com>

...and here is some starter code (also uploaded to "Code Samples" in the curriculum.)

We'll be using cross validation at the level of speed dating events. For information about this, see the section titled "I used cross validation" under http://lesswrong.com/lw/lfa/methodology_for_predicting_speed_dating/.

[Quoted text hidden]

 **day3Starter.R**
3K

Jonah Sinick <jsinick@gmail.com>

Wed, Feb 17, 2016 at 12:19 PM

To: Ali Bagherpour <ali.bagherp@gmail.com>, Andrew Ho <Kironide@gmail.com>, Chad Groft <clgroft@gmail.com>, David Bolin <david@bolin.at>, Jacob Pekarek <jpekarek@trinity.edu>, Jaiwithani <jaiwithani@gmail.com>, James Cook <cookjw@gmail.com>, Linchuan Zhang <email.linch@gmail.com>, Matthew Gentzel <magw6270@terpmail.umd.edu>, Olivia Schaefer <taygetea@gmail.com>, Sam Eisenstat <sam.eisenst@gmail.com>, Tom Guo <tomguo4@gmail.com>, Trevor Murphy <trevor.m.murphy@gmail.com>

Aside from the expositions of regularized linear regression linked above (from ISLR and Applied Predictive Modeling) you might find Andrew Ng's lecture 7 from his machine learning course (linked [here](#)) to be useful.