

The Matchmakers

Gloria D'Azevedo
Pihu Yadav

Analysis Goal

- Want to create a model that takes in relevant information from two people and computes the probability that they will match
- Find characteristics that men and women most prefer in a partner
 - Most likely will be different between genders
 - Use information based on whether or not the participant “likes” their partner

Data and Survey Question Types

- Categorical variables:
 - Age, field type, career type
- Ordinal variables:
 - “How interested are you in these activities?”
- Weighted variables:
 - “Divide up 100 points on the following 6 attributes based on importance to you in a partner.”

Supervised Model Techniques

- Best subset selection
- Step AIC (forward and backward linear model selection)
- Choosing model predictors by hand based on scatter plots

Unsupervised Model Techniques

- K-Nearest-Neighbors (KNN) with Imputed Data
 - Low correct classification rate with $K=3$ to 103
 - Used data from all surveys
- KNN with complete data
 - No instances of complete participants for all questions
 - Poor results when using responses from initial survey

Next Steps

- Use Resampling methods such as bootstrap and cross validation on the data to estimate test error for each model
- Add regularization for linear models for robustness

Questions?