| | |
|---|---|
| **CIS 700/4: Machine Learning and Econometrics** | **Jan 17, 2017** |

Lecture 1: #

*Professor Shivani Agarwal* | *Zach Schutzman*

## 1.1 Intro, Overview, Administrivia

This course will deal with problems about data in the form of rankings or choices.

Think retail recommendations, search rankings, preference surveys

Computing sports rankings from game data and outcomes (incomplete, noisy)

Expression of rankings and preferences

Voting, social choice, political polls

Focus of the course: **analyzing** and **modeling** rankings and choice data!

### Outline

**Part 1:** Lectures

Introductory Material

**Part 2-4:** Paper discussions

Ranked data

Comparisons

Choice Models

**Part 5:** Project presentations

### Project

Important, 'hands-on', complementary to course material, culminates in a report/paper and presentation.

Projects done in small teams (size 2?), topic and teams by Feb. 10, proposal Feb. 17, midterm report Mar. 17, Final report Apr. 14.

## What's Ahead

### Ranked data

Suppose we have an election with a bunch of candidates. Each voter provides top 3 choices. How do we decide who wins from this partially ranked data? Are there natural groupings?

**Definition 1.1** *A* ***probabilistic permutation model*** *over $n$ items is a probability distribution over the outcomes $\sigma \in S_n$.*

***Example:*** *Random utility models (RUMs)*
*$X_1 \ldots X_n$ are random variables with arbitrary distributions. Assign utility $u_i$ to item $i$ equal to a random draw from $X_i$, then sort the $u_i$. This is a permutation, which we can calculate the probability of observing. This idea extends to partial orderings on a $k$ subset, where we consider the order on $k$ and that these $k$ are ranked higher than all others.*

***Example:*** *Plackett-Luce Model*
*Parametrized by $n$ positive numbers, representing a score for each of the $n$ items. Assign probability $p(\sigma)$ for the first item, its score over the sum of all the scores, given that the probability that the item in the second place is its score over the sum of the remaining scores, etc. Stop after the product of the first $k$ terms for partial.*

*We can use these, plus inference methods (max likelihood, etc), to estimate parameters.*

*We can mix Plackett-Luce models to determine partitions and groupings.*

***Example:*** *Recursive Inversion Models*
*Hierarchical ordering to determine groupings. Can be learned from fully ranked data. Project idea: learn this structure from binary choice data.*

### Ranking from pairwise comparisons

***Example:*** *Suppose we have chess games - data is pairwise, noisy, and incomplete. We can have 3 more matches. How do we choose who plays? What's the objective (maximize info gain? find best player?). For every pair of items $i, j$, there is a probability $p_{i,j} \in [0, 1]$ that $i$ is better than $j$ (if there are no ties, $p_{j,i} = 1 - p_{i,j}$).*

*Bradley-Terry-Luce model: $p_{i,j} = \frac{w_i}{w_i + w_j}$. This can be viewed as a special case of a random utility model. Project idea: which $m$ comparisons do we sample next?*

### Discrete Choice models and Assortment Optimization

***Example:*** *Suppose you have a bunch of kinds of camera and a limited amount of space (7 cameras, space for 4, e.g.). You change the display occasionally and observe which choice customers make. What is the optimal subset of size 4 to show, as to maximize expected revenue/profit?*

***Example:*** *Discrete choice model over $n$ items. For each possible subsets and items in that subset, specify a choice probability $p(i|S)$ as the probability of item $i$ being chosen from assortment $S$. In terms of a RUM, we can assign $p(i|S)$ as the probability that $u_i$ is the max over all items in $S$.*

***Example:*** *Multinomial logit model. Specifies $p(i|S) = \frac{w_i}{\sum w_j}$.*