LECTURE 5 – FROM PREDICTIONS TO PRESCRIPTIONS

Universal AI

Opportunity

- Availability of data (often big data) in electronic form.
- Can we develop a theory that unifies OR/MS and ML/S that goes:

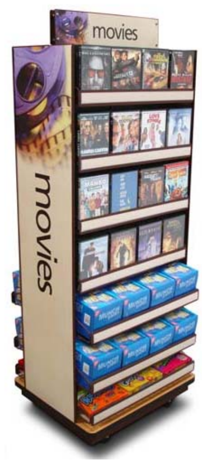
  From data to prescriptions?

A Real World Problem

- A Global Fortune 100 multimedia company.
- 1 billion units of entertainment media shipped per year
- Sells 1/2 million different titles on CD/DVD/Bluray at over 50,000 retailers worldwide

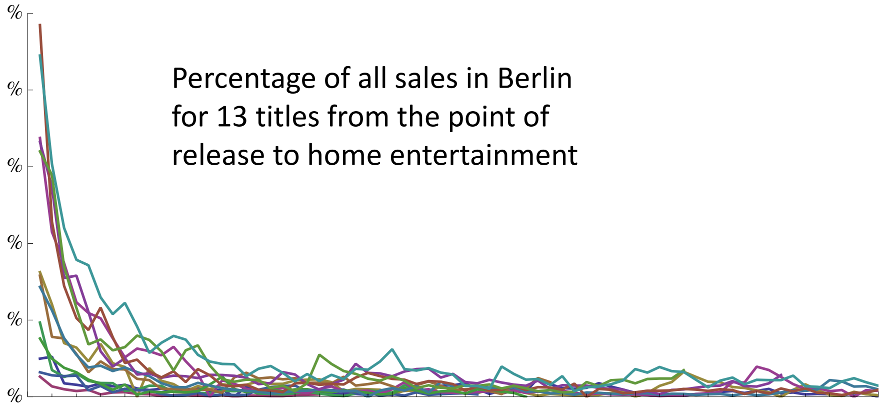
Key Issues

- Limited shelf space at retail locations
- Huge array of potential titles
- Highly uncertain demand for new releases
- Which titles to order and in what quantities?
- Maximize number media sold



Internal Company Data

- Sales by item/location, 2010 to present
- ~50GB *after* aggregating transaction records by week



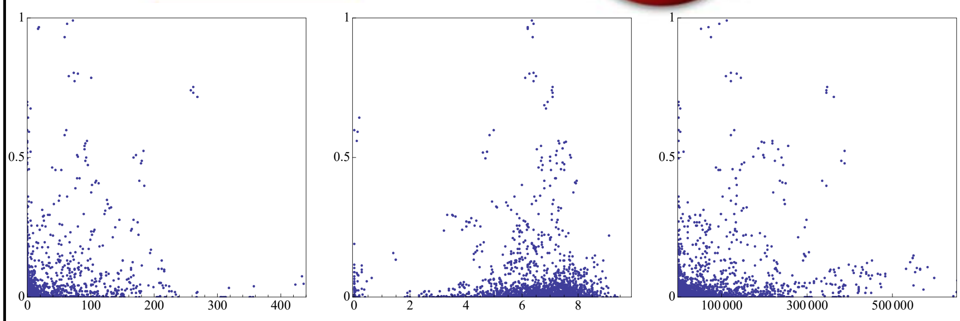
Internal Company Data

- Sales by item/location, 2010 to present
- ~50GB *after* aggregating transaction records by week
- Location info:
  - Address
    - Google Geocoding API
- Item info:
  - Medium (DVD/BLU)
  - Obfuscated title
    - Disambiguation

Beyond internal company data: Harvesting public data (more X)

- **Movie/series**
- **Actors** (find actor communities; Blondel et al 2008)
- **Plot summary** (cosine similarities, hierarchically clustered)
- **Box office gross, US**
- **Oscar wins and nominations** and other awards
- **Professional (meta-)ratings**, **user ratings**
- **Num of user ratings**
- **Genre** (can be multiple)
- **MPAA rating**

Beyond internal company data: Harvesting public data

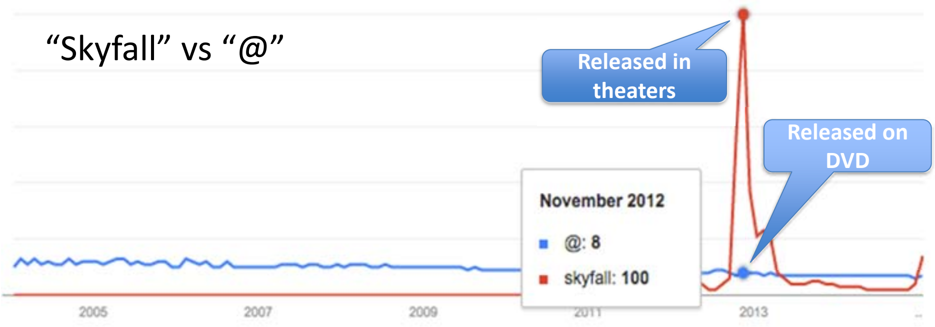


Box office gross
$\rho = 0.32$

IMDb rating
$\rho = 0.02$

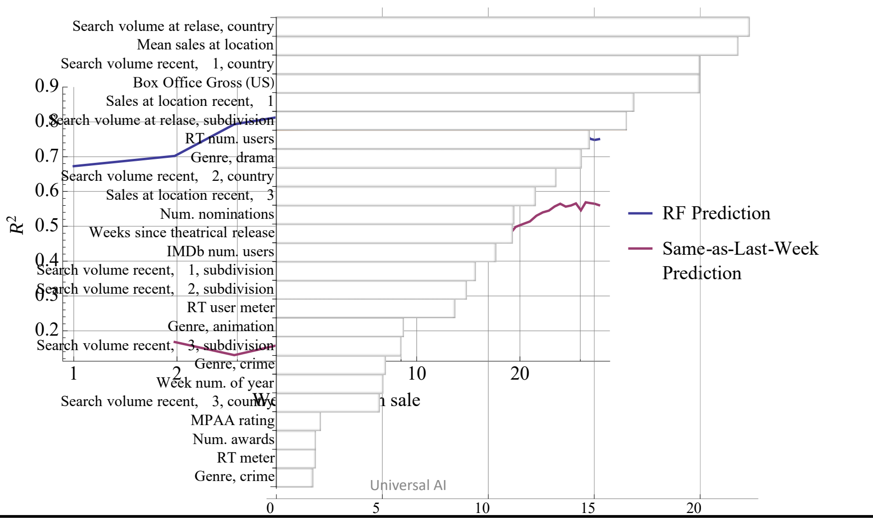
Number user votes
$\rho = 0.25$

Beyond internal company data: Harvesting public data



Predicting Demand

- Random forest regressor
- ***New*** titles: out-of-sample R^2 = 0.67



Problem

- Maximize number of items sold.
- Focus on video media, Europe

- Suppose we order z=10 DVDs and the demand is y=6
- How many do we sell?

- Suppose we order z=6 DVDs and the demand is y=9.
- How many do we sell?

- Formula: Number of DVDs sold: min (y,z)

Problem

- Maximize number of items sold.
- Focus on video media, Europe
- r index locations, t index periods, j index products.
- Y_j demand for j, z_{trj} order, x_{tr} auxiliary data.

$$\max \quad \mathbb{E}\left[\sum_{j=1}^{d} \min\{Y_j,\, z_{trj}\} \,\middle|\, X = x_{tr}\right]$$

$$\text{s.t.} \quad \sum_{j=1}^{d} z_{trj} \leq K_r$$

$$z_{trj} \geq 0 \qquad \forall j = 1, \ldots, d$$

The general problem

- Data $y^1, \ldots, y^N$ on quantities of interest Y
  E.g. demand for movies
- Data $x^1, \ldots, x^N$ on associated covariates X
  E.g. prior sales, google trends
- Decision z, how many DVDs to order to maximize total sales
  - Formally: $\max E[c(Y, z)|X]$

Solution Approach

- Bertsimas and Kallus (2016) have developed
  a general methodology to address the problem.
- Key idea From Predictions to Prescriptions.
- Shown that the method is asymptotically optimal, that is
  As the amount of data increases, approach is stronger.
- Analogous to R square in prediction, we defined
  P coefficient of prescriptiveness.

Existing Approaches

- **Predict first, optimize second.**

  Predict y(x,z)

  Optimize C(y(x,z), z)

  Issue: variability not taken into account

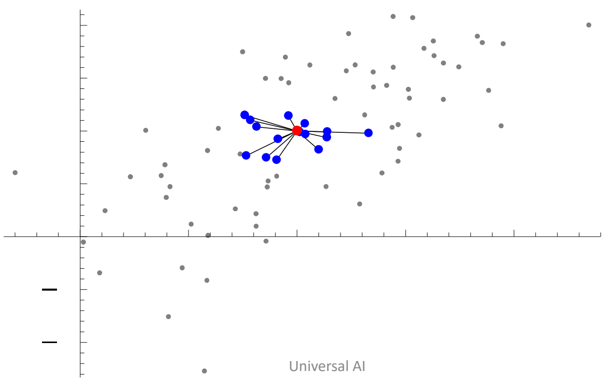
Predict first, optimize second.

Optimize 1/N Sum C(y^i, z)

Issue: Does not address predictability

*k*NN

$$\hat{z}_N^{k\mathrm{NN}}(x) \in \arg\min_{z\in\mathcal{Z}} \sum_{x^i \text{ is } k\mathrm{NN} \text{ of } x} c(z; y^i)$$



CART

$$\left\{(x^1,\, y^1), (x^2,\, y^2), (x^2,\, y^2), (x^3,\, y^3), (x^4,\, y^4), (x^5,\, y^5), (x^6,\, y^6), (x^7,\, y^7), (x^8,\, y^8), (x^9,\, y^9), (x^{10},\, y^{10})\right\}$$

$$\hat{m}(x) = \frac{1}{10}\left(y^1 + y^2 + y^3 + y^4 + y^5 + y^6 + y^7 + y^8 + y^9 + y^{10}\right)$$

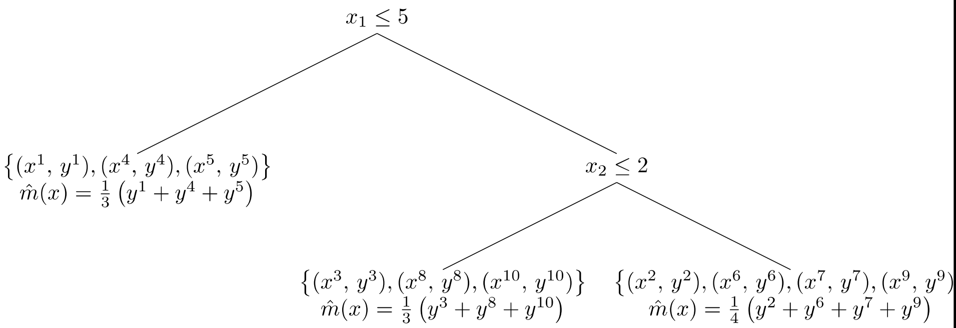
CART

$$x_1 \leq 5$$

$$\left\{(x^1,\, y^1), (x^4,\, y^4), (x^5,\, y^5)\right\} \qquad \left\{(x^2,\, y^2), (x^3,\, y^3), (x^6,\, y^6), \ldots, (x^{10},\, y^{10})\right\}$$

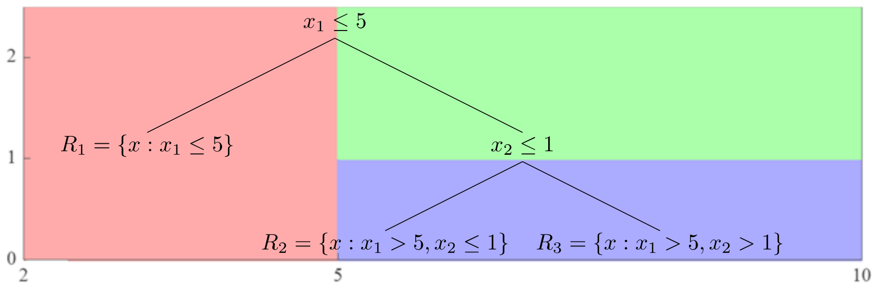
$$\hat{m}(x) = \frac{1}{3}\left(y^1 + y^4 + y^5\right) \qquad \hat{m}(x) = \frac{1}{7}\left(y^2 + y^3 + y^6 + \cdots + y^{10}\right)$$

CART



Universal AI

CART



Implied binning rule $R(x) = (j \text{ s.t. } x \in R_j)$

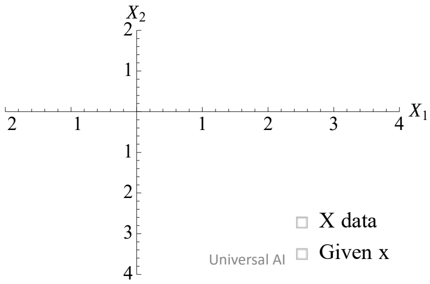
$$\hat{z}_N^{\mathrm{CART}}(x) \in \arg\min_{z \in \mathcal{Z}} \sum_{\mathcal{R}(x^i) = \mathcal{R}(x)} c(z; y^i)$$

Universal AI

Random Forest

- Train T trees on bootstrapped samples and randomly selected feature subsets
- Get T binning rules $R^t(x) = \left(j \text{ s.t. } x \in R_j^t\right)$

$$\hat{z}_N^{\mathrm{RF}}(x) \in \arg\min_{z \in \mathcal{Z}} \sum_{t=1}^{T} \frac{1}{|\{j : R^t(x^j) = R^t(x)\}|} \sum_{R^t(x^i) = R^t(x)} c(z; y^i)$$



Value of a Prescription

- *Coefficient of Prescriptiveness*

$$P = \frac{\min_{z \in \mathcal{Z}} \sum_{i=1}^{N} c(z; y^i) - \sum_{i=1}^{N} c(\hat{z}_N(x^i); y^i)}{\min_{z \in \mathcal{Z}} \sum_{i=1}^{N} c(z; y^i) - \sum_{i=1}^{N} \min_{z \in \mathcal{Z}} c(z; y^i)} \quad \begin{array}{l} \leq 1 \\ \to [0, 1] \end{array}$$

- Measures the prescriptive value of X and of the prescription trained
- Contrast with R^2.

Back to the media Example

- Proposed Approach improved profitability in the European stores by 12%
- P ranged from 0.8 to 0.92 in various European stores.

Conclusions

- **A new framework from predictions to prescriptions**
  - General purpose
- **Theory**
  - Computational tractability
  - Asymptotic optimality
- **Performance metric**
  - Coefficient of prescriptiveness
- **Practice**
  - Material Improvement for a Fortune 100 company