Economics and Probabilistic Machine Learning

David M. Blei
Departments of Computer Science and Statistics
Columbia University
Modern probabilistic modeling
An efficient framework for discovering meaningful patterns in massive data.
Many available packages.
Typically fast and scalable.

Generative Prob Models
Domain-specific knowledge and assumptions.
Challenging to implement.
May not be fast or scalable.

How to use traditional machine learning and statistics to solve modern problems
Probabilistic machine learning: tailored models for the problem at hand.
Probabilistic machine learning: tailored models for the problem at hand.

- Compose and connect reusable parts
- Driven by disciplinary knowledge and its questions
- Large-scale data, both in terms of data points and data dimension
- Focus on discovering and using structure in unstructured data
- Exploratory, observational, causal analyses
Many software packages available; typically fast and scalable
More challenging to implement; may not be fast or scalable
The probabilistic pipeline

- **Design** models that reflect our domain expertise and knowledge
- Given data, **compute** the approximate posterior of hidden variables
- Use the computation to **predict** the future or **explore** the patterns in your data.
Population analysis of 2 billion genetic measurements
Communities discovered in a 3.7M node network of U.S. Patents
Topics found in 1.8M articles from the New York Times
Neuroscience analysis of 220 million fMRI measurements
Figure 1: Measure of “eventness,” or time interval impact on cable content (Eq. 2). Grey background indicates the number of cables sent over time. This comes from the model fit we discuss in Section 3. Capsule successfully detects real-world events from National Archive diplomatic cables.

The intuition behind Capsule is this: embassies write cables throughout the year, usually describing typical business such as the visiting of a government official. Sometimes, however, there is an important event, e.g., the fall of Saigon. When an event occurs, it pulls embassies away from their typical business to write cables that discuss what happened and its consequences. Thus Capsule effectively defines an “event” to be a moment in history when embassies deviate from what each usually discusses, and when each embassy deviates in the same way.

Capsule embeds this intuition into a Bayesian model. It uses hidden variables to encode what “typical business” means for each embassy, how to characterize the events of each week, and which cables discuss those events. Given a corpus, the corresponding posterior distribution provides a filter on the cables that isolates important moments in the diplomatic history. Figure 1 illustrates the mean of this posterior.

Capsule can be used to explore any corpora with the same underlying structure: text (or other discrete multivariate data) generated over time by known entities. This includes email, consumer behavior, social media posts, and opinion articles.

We present the model in Section 2, providing both a formal model specification and guidance on how to use its posterior to detect and characterize real-world events. In Section 3, we evaluate Capsule and explore its results on a collection of U.S. State Department cables and on simulated data.

Related work. We first review previous work on automatic event detection and other related concepts. In both univariate and multivariate settings, the goal is often that analysts want to predict whether or not rare events will occur (Weiss and Hirsh, 1998; Das et al., 2008). Capsule, in contrast, is designed to help analysts explore and understand the original data: our goal is interpretability, not prediction.

Events uncovered from 2M diplomatic cables
Our perspective:

- Customized data analysis is important to many fields.
- This pipeline separates assumptions, computation, application.
- It facilitates solving data science problems.
What we need:

- **Flexible** and **expressive** components for building models
- **Scalable** and **generic** inference algorithms
- **New applications** to stretch probabilistic modeling into new areas
Figure S2: Population structure inferred from the TGP data set using the TeraStructure algorithm at three values for the number of populations $K$. The visualization of the ✓’s in the Figure shows patterns consistent with the major geographical regions. Some of the clusters identify a specific region (e.g. red for Africa) while others represent admixture between regions (e.g. green for Europeans and Central/South Americans). The presence of clusters that are shared between different regions demonstrates the more continuous nature of the structure. The new cluster from $K=7$ to $K=8$ matches structure differentiating between American groups. For $K=9$, the new cluster is unpopulated.
Here I discuss two threads of research with Susan Athey’s group.

- Build probabilistic models to analyze large-scale consumer behavior; many consumers choosing among many items
- (Caveat: I’m not an economist.)
- also joint with Francisco Ruiz
Vision: a utility model for baskets of items:

$$U(\text{basket}) = \text{[subs/comps]} + \text{[shopper]} + \text{[prices]} + \text{[other]} + \epsilon$$

Goals
- design, fit, check, and revise this model
- answer counterfactual questions about purchase behavior
Economic embeddings
Identifying substitutes and co-purchases in large-scale consumer data.
Word embeddings are a powerful approach for analyzing language.

- Discovers a *distributed representation* of words
  - Distances appear to capture semantic similarity.

- Many variants, but each reflects the same main ideas:
  - Words are placed in a low-dimensional latent space
  - A word’s probability depends on its distance to other words in its context

Image: Paul Ginsparg
- **Exponential family embeddings** generalize this idea to other types of data.

- Use *generalized linear models* and *exponential families*

- Examples:
  - neuroscience; recommender systems; networks; shopping baskets

Zebrafish brain activity

Interactions between countries
Consider a vacation-town deli; it has six items.

Customers either buy [pizza, soda] or [peanut butter, jam, bread].
Customers only buy one type of peanut butter at a time.

Items bought together (or not) are co-purchased (or not).
The peanut butters are substitutes.
(For now we ignore many issues, e.g., formal definitions, price, causality.)

We would like to capture this purchase behavior.
We endow each item with two (unknown) locations in a real space $\mathbb{R}^k$: an embedding $\rho$ and context vector $\alpha$.

The conditional probability of each item depends on its embedding and the context vectors of the other items in the basket,

$$x_{b,i} \mid x_{b,-i} \sim \text{Poisson} \left( \exp \left\{ \rho_i \top \sum_{j \neq i} \alpha_j x_{b,j} \right\} \right).$$

$\alpha_i$ are latent product attributes
$\rho_i$ indicate how product $i$ interacts with other products’ attributes
Pizza and soda are never bought with bread, jam, and PB; and vice versa
Bread, jam, and PB are bought together
PB #1 is never bought with PB #2
Embedding Context → interaction / α. attributes

PB #1 is bought with similar items as PB #2
The goal of an EF-EMB is to discover a useful representation of data.

Observations \( x = x_{1:n} \), where \( x_i \) is a \( D \)-vector.

Examples:

<table>
<thead>
<tr>
<th>DOMAIN</th>
<th>INDEX</th>
<th>VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>position in text ( i )</td>
<td>word indicator</td>
</tr>
<tr>
<td>Neuroscience</td>
<td>neuron and time ( n, t )</td>
<td>activity level</td>
</tr>
<tr>
<td>Network</td>
<td>pair of nodes ( s, d )</td>
<td>edge indicator</td>
</tr>
<tr>
<td>Shopping</td>
<td>item and basket ( d, b )</td>
<td>number purchased</td>
</tr>
</tbody>
</table>
Three ingredients:

- *context*, *conditional exponential family*, *embedding structure*

- Two latent variables per data index, an embedding and a context vector

- Model each data point conditioned on its context and latent variables.

- The latent variables interact in the conditional. How depends on which indices are in context and which one is modeled.
Each data point $i$ has a context $c_i$, a set of indices of other data points.

We model the conditional of the data point given its context, $p(x_i | x_{c_i})$.

Examples

<table>
<thead>
<tr>
<th>DOMAIN</th>
<th>DATA POINT</th>
<th>CONTEXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>word</td>
<td>surrounding words</td>
</tr>
<tr>
<td>Neuroscience</td>
<td>neuron activity</td>
<td>activity of surrounding neurons</td>
</tr>
<tr>
<td>Network</td>
<td>edge</td>
<td>other edges on the two nodes</td>
</tr>
<tr>
<td>Shopping</td>
<td>purchased item</td>
<td>other item counts on the same trip</td>
</tr>
</tbody>
</table>
Context

- Each data point \( i \) has a context \( c_i \), a set of indices of other data points.
- We model the conditional of the data point given its context, \( p(x_i \mid x_{c_i}) \).
- Examples

<table>
<thead>
<tr>
<th>DOMAIN</th>
<th>DATA POINT</th>
<th>CONTEXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>word</td>
<td>surrounding words</td>
</tr>
<tr>
<td>Neuroscience</td>
<td>neuron activity</td>
<td>activity of surrounding neurons</td>
</tr>
<tr>
<td>Network</td>
<td>edge</td>
<td>other edges on the two nodes</td>
</tr>
<tr>
<td>Shopping</td>
<td>purchased item</td>
<td>other item counts on the same trip</td>
</tr>
</tbody>
</table>
Each data point $i$ has a context $c_i$, a set of indices of other data points. We model the conditional of the data point given its context, $p(x_i \mid x_{c_i})$.

Examples

<table>
<thead>
<tr>
<th>DOMAIN</th>
<th>DATA POINT</th>
<th>CONTEXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>word</td>
<td>surrounding words</td>
</tr>
<tr>
<td>Neuroscience</td>
<td>neuron activity</td>
<td>activity of surrounding neurons</td>
</tr>
<tr>
<td>Network</td>
<td>edge</td>
<td>other edges on the two nodes</td>
</tr>
<tr>
<td>Shopping</td>
<td>purchased item</td>
<td>other item counts on the same trip</td>
</tr>
</tbody>
</table>
The **EF-EMB** has latent variables for each data point's index:

- an *embedding* $\rho[i]$ and a *context vector* $\alpha[i]$

These are used in the conditional of each data point,

$$x_i \mid x_{c_i} \sim \text{exp-fam}(\eta(x_{c_i}; \rho[i], \alpha[c_i]), t(x_i)).$$

(Poisson for counts, Gaussian for reals, Bernoulli for binary, etc.)
The natural parameter combines the embedding and context vectors,

\[
\eta_i(x_{ci}) = f \left( \rho[i]^\top \sum_{j \in c_i} \alpha[j] x_j \right),
\]

E.g., an item’s embedding (interaction) helps determine its count; its context vector (attributes) helps determine other item’s counts.
Embedding Structure

- The embedding structure determines how parameters are shared.
- E.g., \( \rho[i] = \rho[j] \) for \( i = (\text{Oreos}, t) \) and \( j = (\text{Oreos}, u) \).
- Sharing enables learning about an object, such as a neuron, node, or item.
The embedding structure determines how parameters are shared.

- E.g., $\rho[i] = \rho[j]$ for $i = (\text{Oreos}, t)$ and $j = (\text{Oreos}, u)$.

- Sharing enables learning about an object, such as a neuron, node, or item.
The embedding structure determines how parameters are shared.

E.g., $\rho[i] = \rho[j]$ for $i = (\text{Oreo}, t)$ and $j = (\text{Oreo}, u)$.

Sharing enables learning about an object, such as a neuron, node, or item.
We model each data point, conditional on the others.

Combine these ingredients in a “pseudo-likelihood” (i.e., a utility)

\[ \mathcal{L}(\rho, \alpha) = \sum_{i=1}^{n} \left( \eta_i^\top t(x_i) - a(\eta_i) \right) + \log f(\rho) + \log g(\alpha). \]

Fit with stochastic optimization; exponential families simplify the gradients.
The objective resembles a collection of GLM likelihoods.

The gradient is

\[ \nabla_{\rho[j]} \mathcal{L} = \sum_{i=1}^{I} \left( t(x_i) - \mathbb{E}[t(x_i)] \right) \nabla_{\rho[j]} \eta_i + \nabla_{\rho[j]} \log f(\rho[j]). \]

(Stochastic gradients give justification to NN ideas like “negative sampling.”)
Market basket analysis

- Data | Purchase counts of items in shopping trips at a large grocery store
  - Category-level | 478 categories; 635,000 trips; 6.8M purchases
  - Item-level | 5,675 items; 620,000 trips; 5.6M purchases
- Context | Other items purchased at the same trip
- Structure | Embeddings for each item are shared across trips
- Family | Poisson (and we downweight the zeros)
Recall the conditional probability

\[ x_i \mid \mathbf{x}_{-i} \sim \text{Poisson} \left( \exp \left\{ \rho_i^\top \sum_{j \neq i} \alpha_j x_j \right\} \right). \]

- \( \alpha_i \) reflects attributes of item \( i \)
- \( \rho_i \) reflects the interaction of item \( i \) with attributes of other items.
A 2D representation of category attributes $\alpha_i$
- infant formula
- disposable diapers
- disposable pants
- baby accessories

- baby/youth wipes
- infant toiletries

- childrens/infants analgesics
A 2D representation of item attributes $\alpha_i$
### Category level

<table>
<thead>
<tr>
<th>MODEL</th>
<th>$K = 20$</th>
<th>$K = 50$</th>
<th>$K = 100$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson embedding</td>
<td>-7.497</td>
<td>-7.284</td>
<td>-7.199</td>
</tr>
<tr>
<td>Poisson embedding (downweighting zeros)</td>
<td>-7.110</td>
<td>-6.994</td>
<td><strong>-6.950</strong></td>
</tr>
<tr>
<td>Additive Poisson embedding</td>
<td>-7.868</td>
<td>-8.191</td>
<td>-8.414</td>
</tr>
<tr>
<td>Hierarchical Poisson factorization</td>
<td>-7.740</td>
<td>-7.626</td>
<td>-7.626</td>
</tr>
<tr>
<td>Poisson PCA</td>
<td>-8.314</td>
<td>-9.51</td>
<td>-11.01</td>
</tr>
</tbody>
</table>

### Item level

<table>
<thead>
<tr>
<th>MODEL</th>
<th>K=50</th>
<th>K=100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson embedding</td>
<td>-7.72</td>
<td>-7.64</td>
</tr>
<tr>
<td>Hierarchical Poisson factorization</td>
<td>-7.86</td>
<td>-7.87</td>
</tr>
</tbody>
</table>

We want to use this fit to understand purchase patterns.

- **Exchangeables** have a similar effect on the purchase of other items.

- **Same-category items** tend to be exchangeable and rarely purchased together (e.g., two types of peanut butter).

- **Complements** are purchased (or not purchased) together (e.g., hot dogs and buns).
PB #1 and PB #2 induce similar distributions of other items

But they are rarely purchased together.

Define the sigmoid function between two items,

\[ \sigma_{ki} \triangleq \frac{1}{1 + \exp\{-\rho_k^T \alpha_i\}} ; \quad \bar{\sigma}_{ki} \triangleq 1 - \sigma_{ki} \]
The “substitute predictor” is

\[
- \left( \sum_{k \neq \{i, j\}} \sigma_{ki} \log \left( \frac{\sigma_{ki}}{\bar{\sigma}_{kj}} \right) + \bar{\sigma}_{ki} \log \left( \frac{\bar{\sigma}_{ki}}{\bar{\sigma}_{kj}} \right) \right) - \sigma_{ji} \log \left( \frac{\sigma_{ji}}{1 - \sigma_{ji}} \right).
\]
The “complement predictor” is the negative of the last term

\[
\sigma_{ji} \log \left( \frac{\sigma_{ji}}{1 - \sigma_{ji}} \right).
\]

Notes:
- We use the symmetrized version of both quantities
- These quantities generalize to other exponential families
### Example co-purchases at the category level

<table>
<thead>
<tr>
<th>ITEM 1</th>
<th>ITEM 2</th>
<th>SCORE (RANK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>organic vegetables</td>
<td>organic fruits</td>
<td>6.18 (01)</td>
</tr>
<tr>
<td>vegetables (&lt;10 oz)</td>
<td>beets (&gt;=10 oz)</td>
<td>5.64 (02)</td>
</tr>
<tr>
<td>baby food</td>
<td>disposable diapers</td>
<td>3.43 (32)</td>
</tr>
<tr>
<td>stuffing</td>
<td>cranberries</td>
<td>3.30 (36)</td>
</tr>
<tr>
<td>gravy</td>
<td>stuffing</td>
<td>3.23 (37)</td>
</tr>
<tr>
<td>pie filling</td>
<td>evaporated milk</td>
<td>3.09 (42)</td>
</tr>
<tr>
<td>deli cheese</td>
<td>deli crackers</td>
<td>2.87 (55)</td>
</tr>
<tr>
<td>dry pasta/noodles</td>
<td>tomato paste/sauce/puree</td>
<td>2.73 (63)</td>
</tr>
<tr>
<td>mayonnaise</td>
<td>mustard</td>
<td>2.61 (69)</td>
</tr>
<tr>
<td>cake mixes</td>
<td>frosting</td>
<td>2.49 (78)</td>
</tr>
<tr>
<td>ITEM 1</td>
<td>ITEM 2</td>
<td>SCORE (RANK)</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>bouquets</td>
<td>roses</td>
<td>0.20 (01)</td>
</tr>
<tr>
<td>frozen pizza 1</td>
<td>frozen pizza 2</td>
<td>0.18 (02)</td>
</tr>
<tr>
<td>bottled water 1</td>
<td>bottled water 2</td>
<td>-0.07 (03)</td>
</tr>
<tr>
<td>carbonated soft drinks 1</td>
<td>carbonated soft drinks 2</td>
<td>-0.12 (04)</td>
</tr>
<tr>
<td>orange juice 1</td>
<td>orange juice 2</td>
<td>-0.37 (05)</td>
</tr>
<tr>
<td>bathroom tissue 1</td>
<td>bathroom tissue 2</td>
<td>-0.58 (06)</td>
</tr>
<tr>
<td>bananas 1</td>
<td>bananas 2</td>
<td>-0.61 (07)</td>
</tr>
<tr>
<td>salads-convenience 1</td>
<td>salads-convenience 2</td>
<td>-0.63 (08)</td>
</tr>
<tr>
<td>potatoes 1</td>
<td>potatoes 2</td>
<td>-0.66 (09)</td>
</tr>
<tr>
<td>bouquets</td>
<td>blooming</td>
<td>-1.18 (10)</td>
</tr>
</tbody>
</table>

Top ten potential substitutes at the category level
<table>
<thead>
<tr>
<th>ITEM 1</th>
<th>ITEM 2</th>
<th>SCORE (RANK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ygrt peach ff</td>
<td>ygrt mxd berry ff</td>
<td>19.83 (0001)</td>
</tr>
<tr>
<td>s&amp;w beans garbanzo</td>
<td>s&amp;w beans red kidney</td>
<td>14.42 (0002)</td>
</tr>
<tr>
<td>whiskas cat fd beef</td>
<td>whiskas cat food tuna/chicken</td>
<td>8.45 (0149)</td>
</tr>
<tr>
<td>parsnips loose</td>
<td>rutabagas</td>
<td>8.32 (0157)</td>
</tr>
<tr>
<td>celery hearts organic</td>
<td>apples fuji organic</td>
<td>4.36 (0995)</td>
</tr>
<tr>
<td>85p ln gr beef patties 15p fat</td>
<td>sesame buns</td>
<td>4.35 (1005)</td>
</tr>
<tr>
<td>kiwi imported</td>
<td>mangos small</td>
<td>3.22 (1959)</td>
</tr>
<tr>
<td>colby jack shredded</td>
<td>taco bell taco seasoning mix</td>
<td>2.89 (2472)</td>
</tr>
<tr>
<td>star magazine</td>
<td>in touch magazine</td>
<td>2.87 (2497)</td>
</tr>
<tr>
<td>seasoning mix fajita</td>
<td>mission tortilla corn super sz</td>
<td>2.87 (2500)</td>
</tr>
</tbody>
</table>

Example co-purchases at the UPC level
<table>
<thead>
<tr>
<th>ITEM 1</th>
<th>ITEM 2</th>
<th>SCORE (RANK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>coffee drip grande</td>
<td>coffee drip venti</td>
<td>-0.33 (001)</td>
</tr>
<tr>
<td>sandwich signature reg</td>
<td>sandwich signature lrg</td>
<td>-1.17 (020)</td>
</tr>
<tr>
<td>market bouquet</td>
<td>alstromeria/rose bouquet</td>
<td>-2.89 (186)</td>
</tr>
<tr>
<td>sushi shoreline combo</td>
<td>sushi full moon combo</td>
<td>-3.76 (282)</td>
</tr>
<tr>
<td>semifreddis bread baguette</td>
<td>crusty sweet baguette</td>
<td>-7.65 (566)</td>
</tr>
<tr>
<td>orbit gum peppermint</td>
<td>orbit gum spearmint</td>
<td>-7.96 (595)</td>
</tr>
<tr>
<td>snickers candy bar</td>
<td>3 musketeers candy bar</td>
<td>-7.97 (598)</td>
</tr>
<tr>
<td>cheer Indry det color guard</td>
<td>all Indry det liquid fresh rain</td>
<td>-7.99 (602)</td>
</tr>
<tr>
<td>coors light beer btl</td>
<td>coors light beer can</td>
<td>-8.12 (621)</td>
</tr>
<tr>
<td>greek salad signature</td>
<td>neptune salad signature</td>
<td>-8.15 (630)</td>
</tr>
</tbody>
</table>

Example potential substitutes at the UPC level
Word embeddings have become a staple in natural language processing. We distilled its essential elements, generalized to consumer data.

Compared to classical factorization, good performance in many data:
- movie ratings, neural activity, scientific reading, shopping baskets.
Summary and Questions

- How can we capture higher-order structure in the embeddings?
- Why downweight the zeros?
- How can we include price and other complexities?

Poisson factorization
A computationally efficient method for discovering correlated preferences
Economics

- Look at items within one category (e.g. yoghurt)
- Try to estimate the effects of interventions (e.g., coupons, price, layout)

Machine learning

- Look at all items
- Estimate user preferences and make predictions (recommendations)
- Ignore causal effects of interventions
Implicit data is about users interacting with items
- clicks
- “likes”
- purchases

Less information than explicit data (e.g. ratings), but more prevalent
Poisson factorization

- **Assumptions**
  - Users (consumers) have *latent preferences* $\theta_u$.
  - Items have *latent attributes* $\beta_i$.
  - How many items a shopper purchased comes from a Poisson.

- The posterior $p(\theta, \beta \mid y)$ reveals purchase patterns.

\[
\begin{align*}
\theta_{uk} & \sim \text{Gam}(\cdot, \cdot) \\
\beta_{ik} & \sim \text{Gam}(\cdot, \cdot) \\
y_{ui} & \sim \text{Poisson} \left( \theta_u^\top \beta_i \right)
\end{align*}
\]

ITEM ATTRIBUTES   USER PREFERENCES

✓ u ✓ yu ✓ (ui
✓ uk ⇠ Gam.; /
✓ ik ⇠ Gam.; /
✓ yui ⇠ Poisson

θuk ~ Gam(·, ·)
βik ~ Gam(·, ·)
yui ~ Poisson(θuT βi)

Advantages

- captures heterogeneity of users
- implies a distribution of total consumption
- efficient approximation, only requires non-zero data

News articles from the New York Times

“Business Self-Help”
Stay Focused And Your Career Will Manage Itself
To Tear Down Walls You Have to Move Out of Your Office
Self-Reliance Learned Early
Maybe Management Isn't Your Style
My Copyright Career

“Personal Finance”
In Hard Economy for All Ages Older Isn't Better It's Brutal
Younger Generations Lag Parents in Wealth-Building
Fast-Growing Brokerage Firm Often Tangles With Regulators
The Five Stages of Retirement Planning Angst
Signs That It's Time for a New Broker

“All Things Airplane”
Flying Solo
Crew-Only 787 Flight Is Approved By FAA
All Aboard Rescued After Plane Skids Into Water at Bali Airport
Investigators Begin to Test Other Parts On the 787
American and US Airways May Announce a Merger This Week

Scientific articles from Mendeley

“Biodiesel”
Biodiesel from microalgae.
Biodiesel from microalgae beats bioethanol
Commercial applications of microalgae
Second Generation Biofuels
Hydrolysis of lignocellulosic materials for ethanol production

“Political Science”
Social Capital: Origins and Applications in Modern Sociology
Increasing Returns, Path Dependence, and Politics
Institutions, Institutional Change and Economic Performance
Diplomacy and Domestic Politics
Comparative Politics and the Comparative Method

“Astronomy”
Theory of Star Formation
Error estimation in astronomy: A guide
Astronomy & Astrophysics
Measurements of Omega from 42 High-Redshift Supernovae
Stellar population synthesis at the resolution of 2003
Figure 4 shows the normalized mean precision at 20 recommendations as a function of user activity for different datasets.

- **Mendeley**: Shows a consistent improvement in precision for users with higher activity levels, with a slight decline for the lowest activity users.
- **New York Times**: Demonstrates a gradual increase in precision with increasing activity levels.
- **Echo Nest**: Exhibits a steady increase in precision for moderately active users, but a slight decline for the least active users.
- **Netflix (implicit)**: Displays a significant improvement in precision for all activity levels, with the highest performance for the middle activity range.
- **Netflix (explicit)**: Consistent across all activity levels, indicating a robust performance in all scenarios.

The charts also include mean recall at 20 recommendations, showing a similar trend in performance across different datasets and user activity levels.
**“FRUIT”**
- stone fruit
- pears
- tropical fruit
- apples
- grapes

**“CAT CARE”**
- cat food wet
- cat food dry
- cat litter & deodorant
- canned fish
- paper towels

**“BABY ESSENTIALS”**
- baby food
- starbucks coffee
- disposable diapers
- infant formula
- baby/youth wipes

**“HEALTHY”**
- health and milk substitutes
- organic vegetables
- organic fruits
- cold cereal
- vegetarian / organic frozen

---

**Consumer #1 : “Cats and Babies”**

**Consumer #2 : “Healthy and Cats”**
Consider a utility model of a single purchase with Gumbel error,

\[ U(y_{ui}) = \log(\theta_u^T \beta_i) + \epsilon. \]

Suppose a shopper \( u \) buys \( N \) items. Then

\[ y_u \mid N \sim \text{Multi}(N, \pi_u) \]

\[ \pi_{ui} \propto \exp\{\theta_u^T \beta_i\}. \]

Thus, the unconditional distribution of counts is Poisson factorization,

\[ y_{uj} \sim \text{Poisson}(\theta_u^T \beta_j). \]
Poisson factorization and economics

- With this connection, we can devise new utility models, e.g.,

\[ U(y_{ui}) = \log(\theta_u^T \beta_i + \alpha_u \exp(-c \cdot \text{price}_i)) + \epsilon. \]

- ...and other factors
  - time of day
  - in stock
  - date
  - observed item characteristics & category
  - demographic information about the shopper

- Inference is still efficient.
  With assumptions, we can answer counterfactual questions.
Poisson factorization efficiently analyzes large-scale purchase behavior.

Next steps
- include notions of co-purchases and substitutes
- include time of day at a level that is unconfounded
- include price and stock out; answer counterfactual questions

Research in recommendation systems can help economic analyses.
**Probabilistic machine learning**: design expressive models to analyze data.

- Tailor your method to your question and knowledge
- Use generic and scalable inference to analyze large data sets
- Form predictions, hypotheses, inferences, and revise the model
Opportunities for economics and machine learning

- Push economics to high-dimensional data and scalable computation
- Push ML to explainable models, applied causal inference, new problems
- Develop new modeling methods together