Recommendation Systems || Lecture 8

University of Washington
Dr. Karthik Mohan

July 25, 2022

Logistics

Loss not going down (RMSE is high) → Tomorrow noon

1. Assignment 2 due wednesday night
2. Project deliverables doc up on canvas - Breakdown of project to help you stay on track! (Syllabus)
3. Assignment 3 will be assigned wednesday and you get 2 weeks
   ↳ Last Assignment (Kaggle Contest) News Recs

Project Deliverables

1. **Baseline Model:** For your project, try couple of baseline models such as truncated SVD, etc
2. **Neural Network Model:** Try at least one NN model - E.g. AutoRec (AutoEncoder for Recommendations) or DeepRec or MLP (Multi-layer perceptron), etc
3. **Report:** Your report should be well formated and organized just like a conference paper - With a) Abstract b) Introduction c) Reference to literature d) Your unique contribution in the paper e) Description of models used f) Description of data sets g) Descrption of metrics and key results comparing different models over the metrics (use a table for this) h) Conclusions and scope for future work
4. **Code:** Publish your complete code on github and link to it in the report. Keep your code organized (modular, use of classes, etc) (unit tests)
5. **Presentation:** One presentation per team

Project Timeline

1. **July 29:** Finish your 1:1 review of the project proposal

Project Timeline

1. **July 29:** Finish your 1:1 review of the project proposal
2. **August 8:** Checkpoint for project - Submit baseline models, results and code | Report

Project Timeline

1. **July 29:** Finish your 1:1 review of the project proposal
2. **August 8:** Checkpoint for project - Submit baseline models, results and code
3. **August 10:** Assignment 3 deadline

Project Timeline

1. **July 29:** Finish your 1:1 review of the project proposal
2. **August 8:** Checkpoint for project - Submit baseline models, results and code
3. **August 10:** Assignment 3 deadline
4. **August 17:** Final project presentation (10 minutes per team)

Last Lecture

Project Timeline

1. **July 29:** Finish your 1:1 review of the project proposal
2. **August 8:** Checkpoint for project - Submit baseline models, results and code
3. **August 10:** Assignment 3 deadline
4. **August 17:** Final project presentation (10 minutes per team)
5. **August 19:** *Neurips style* white paper submission (check Neurips conference page for details)

↳ Latex format for conference paper

Project Timeline

1. **July 29:** Finish your 1:1 review of the project proposal
2. **August 8:** Checkpoint for project - Submit baseline models, results and code
3. **August 10:** Assignment 3 deadline
4. **August 17:** Final project presentation (10 minutes per team)
5. **August 19:** *Neurips style* white paper submission (check Neurips conference page for details)
6. **August 19:** Final code submission and link to github project.

Today!

1. Industry Case Study on Recommendations

ICE #1

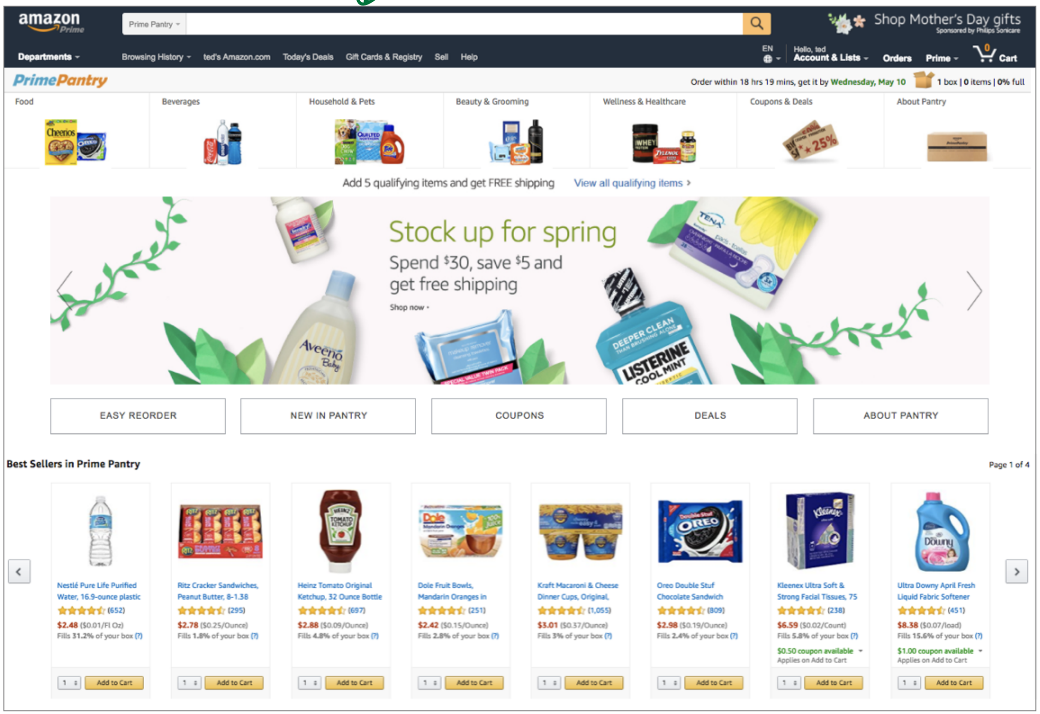
Let's say you have a loss function $l(X, y; \theta)$ where X, y is the data and θ are the parameters. You want to optimize (minimize) the loss for θ. You take the gradient of the loss l at the current pointe, θ^1. What does the direction of the gradient point to?

1. Direction of steepest ascent
2. Direction of steepest descent
3. It's just a collection of partial derivatives. Can't really attribute meaning to it.
4. None of the above

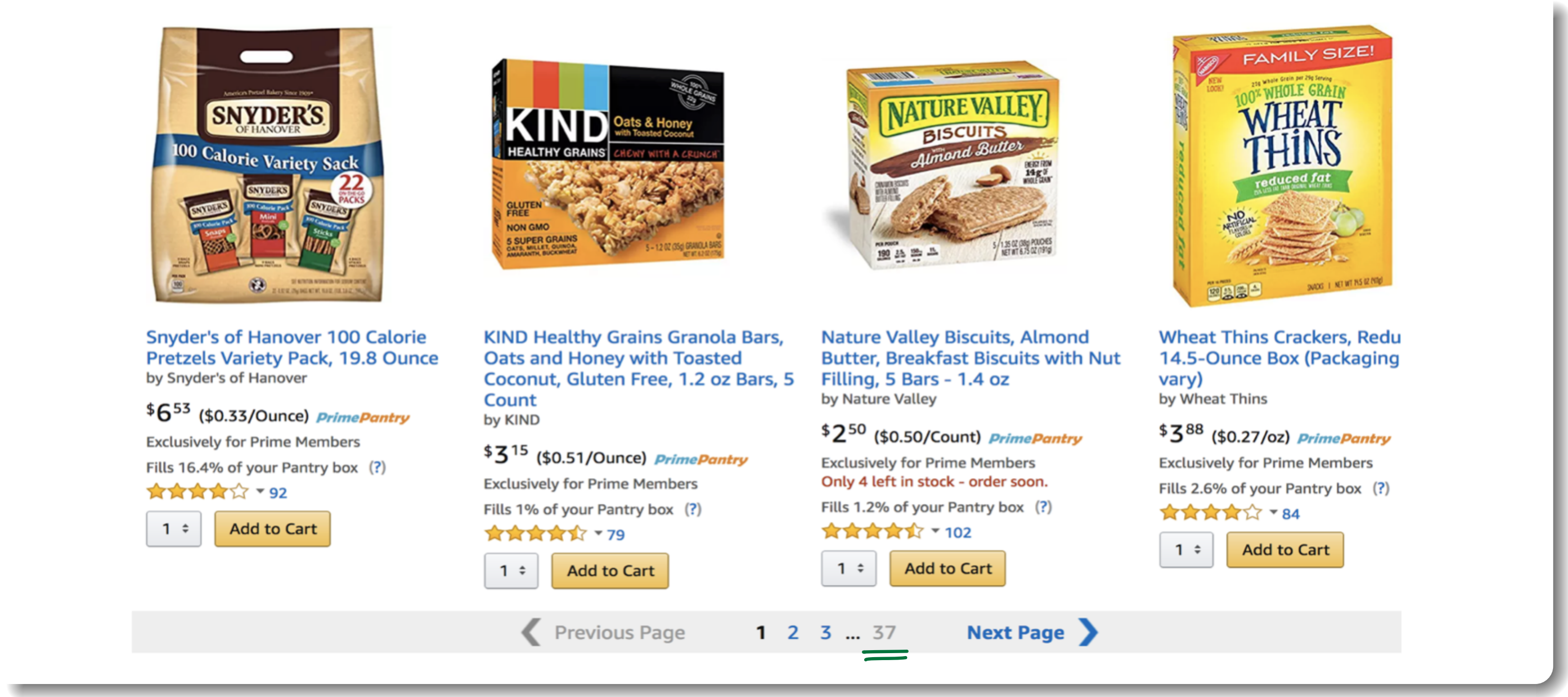
Case Study: Stocking your Pantry

Variety

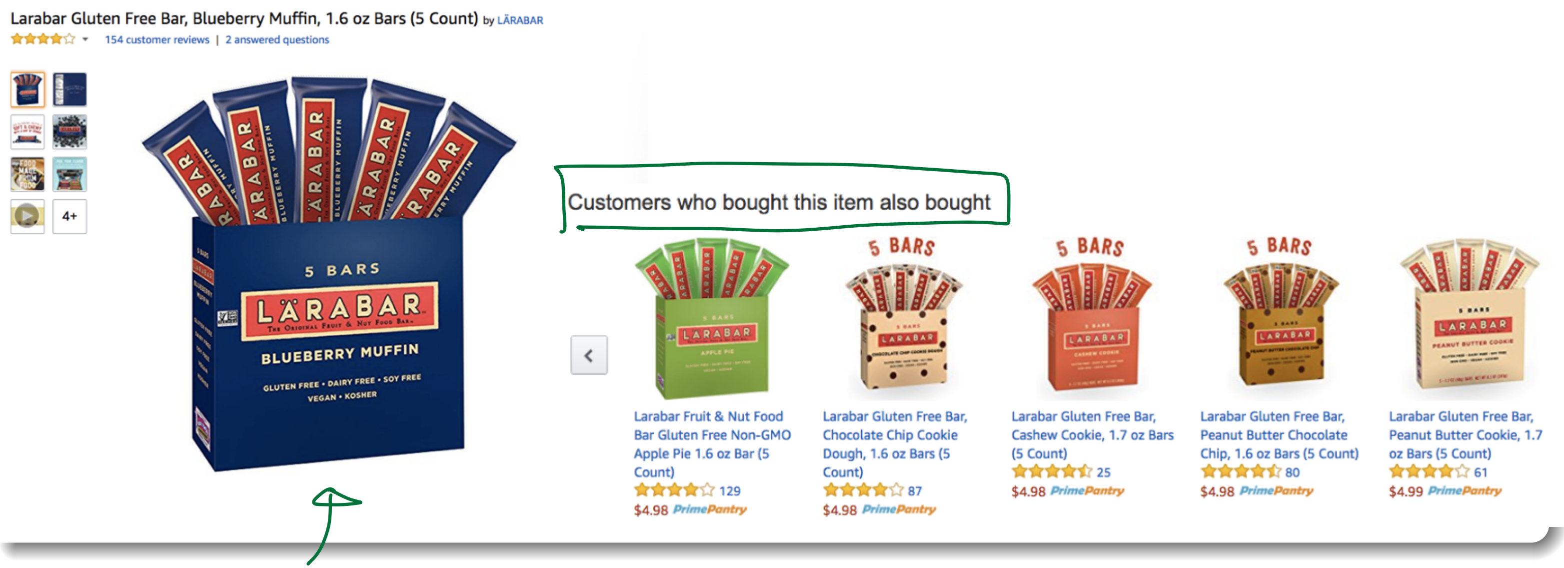
Case Study: Online Pantry



Discovery and Browsing



Recommendations are not basket oriented



Business problem

Make it easier for a customer to fill a basket with a personalized carousel of relevant and diverse products.

1 2

How to measure?

- Increase in units purchased per order
- Decrease in cart abandonments

Business metrics

Precision/ Recall F-Score RMSE } offline metrics

online metrics

Contributions

- **Provide** a joint relevance and diversity model for personalized basket recommendations
- **Demonstrate** the effectiveness of this model in online and offline experiments

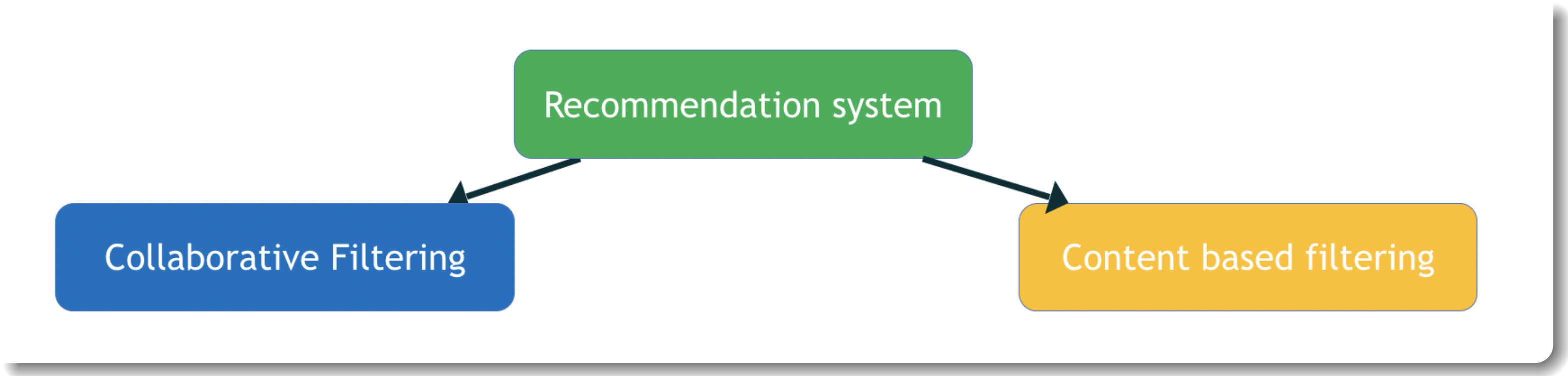
A/B testing

Random Assignment

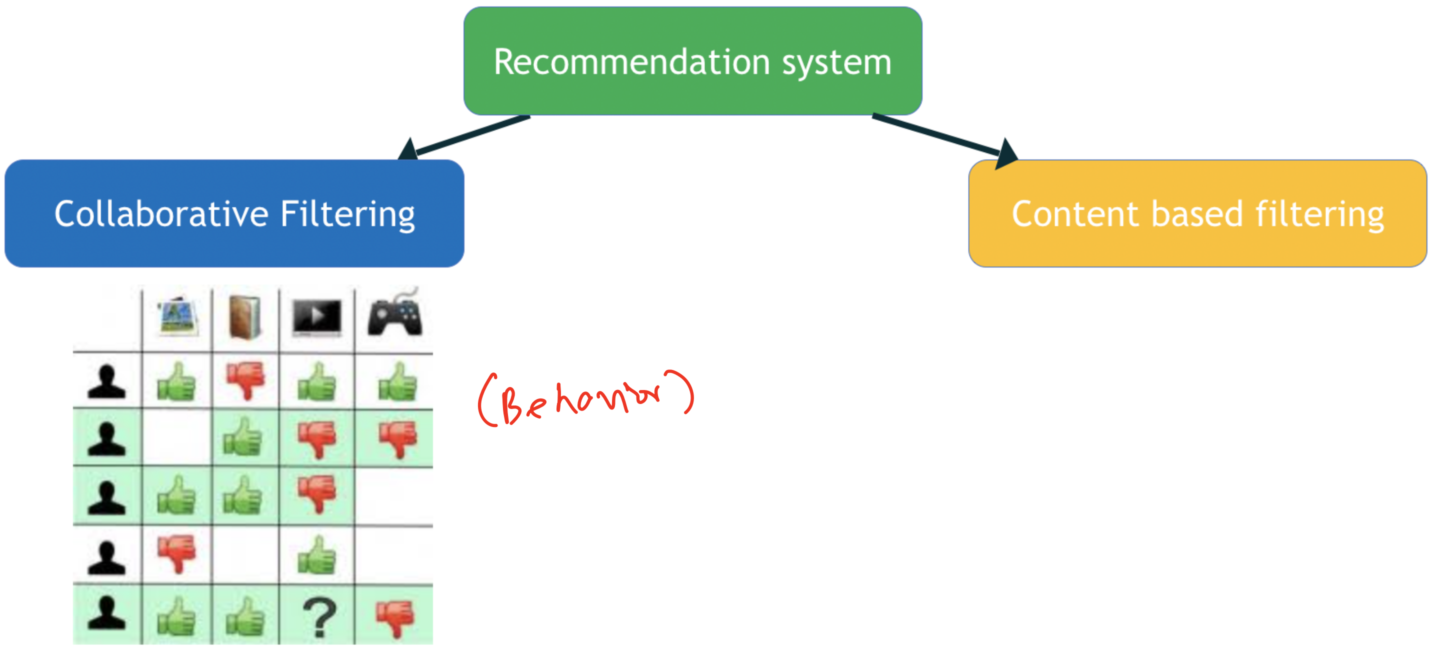
A – Control → webpage has no changes

B – Treatment → Your change in the webpage

Traditional Recommendation Systems

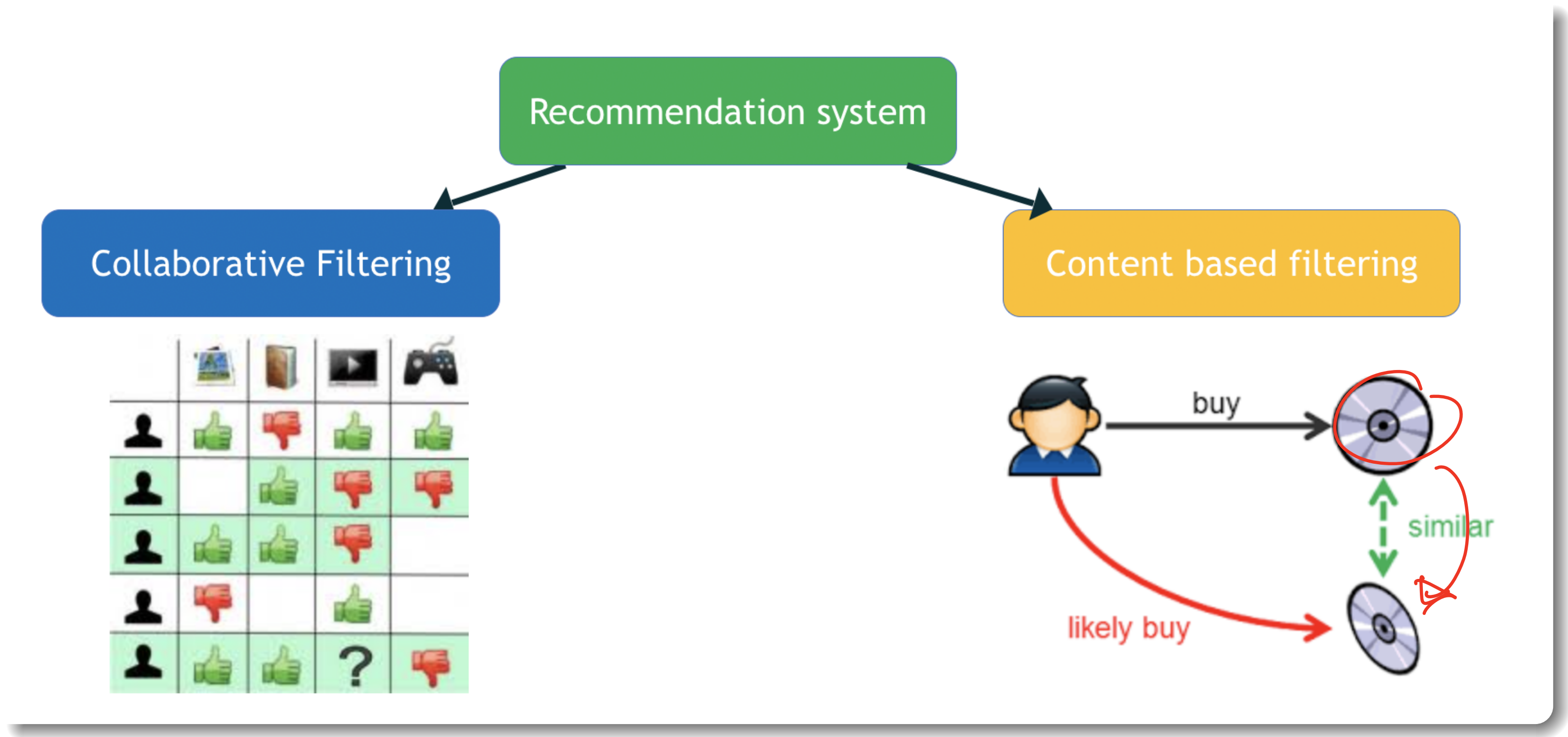


Traditional Recommendation Systems



(Behavior)

Traditional Recommendation Systems



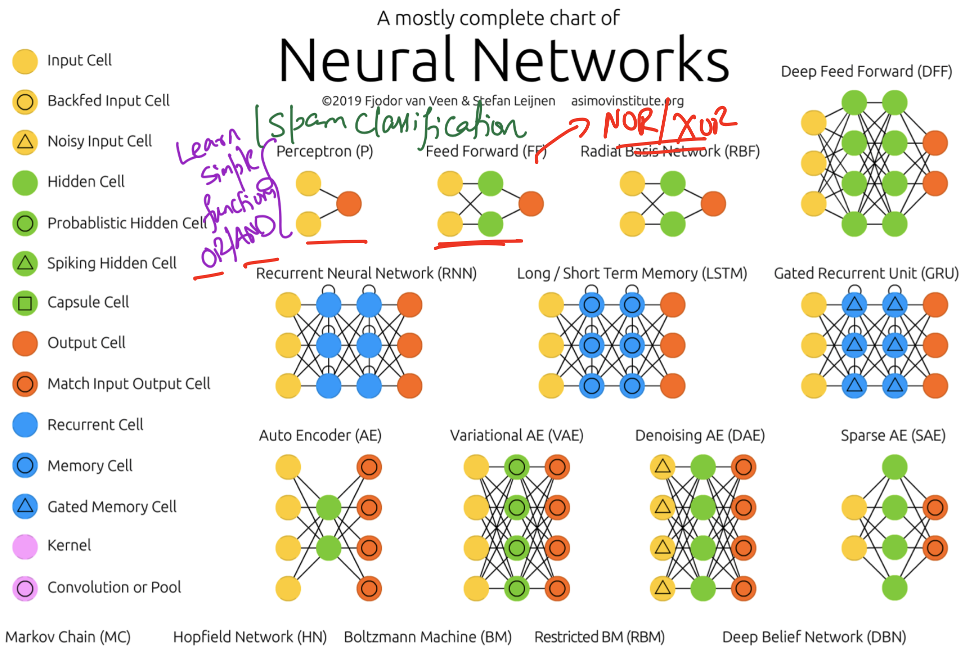
Deep Learning for Recommendation Systems

- YouTube
- Netflix
- Yahoo News

- Amazon Videos
- Pantry Basket recommendations
- Fresh recommendations
- More!?

Quick tour of DL Architectures

Neural Networks Zoo



Quick tour of DL Architectures

Neural Networks Zoo

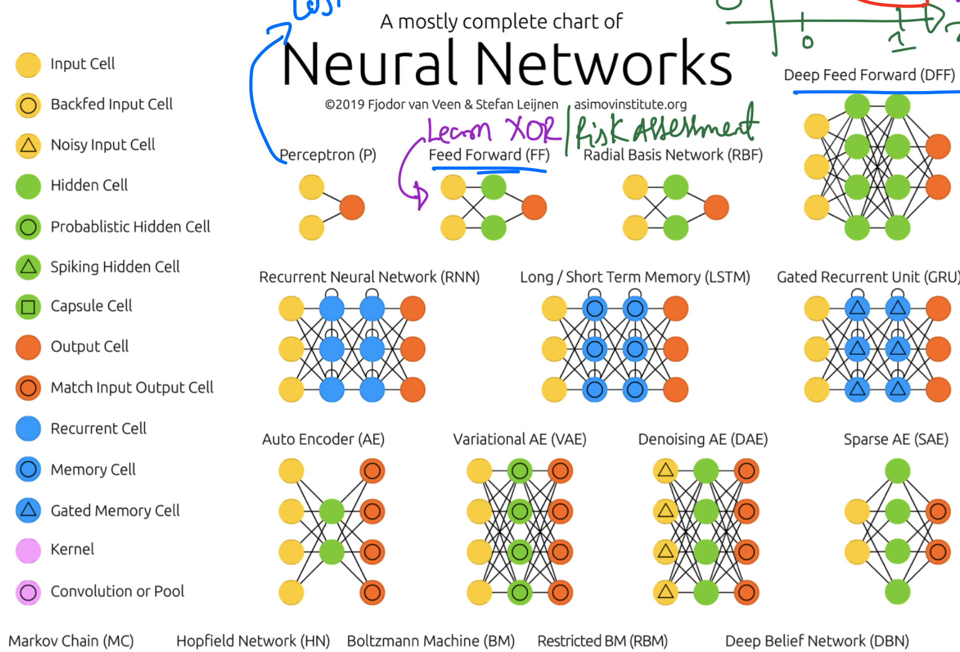
Logistic Regression (Learn Linear Decision boundary)

Decision boundary is non-linear

XOR

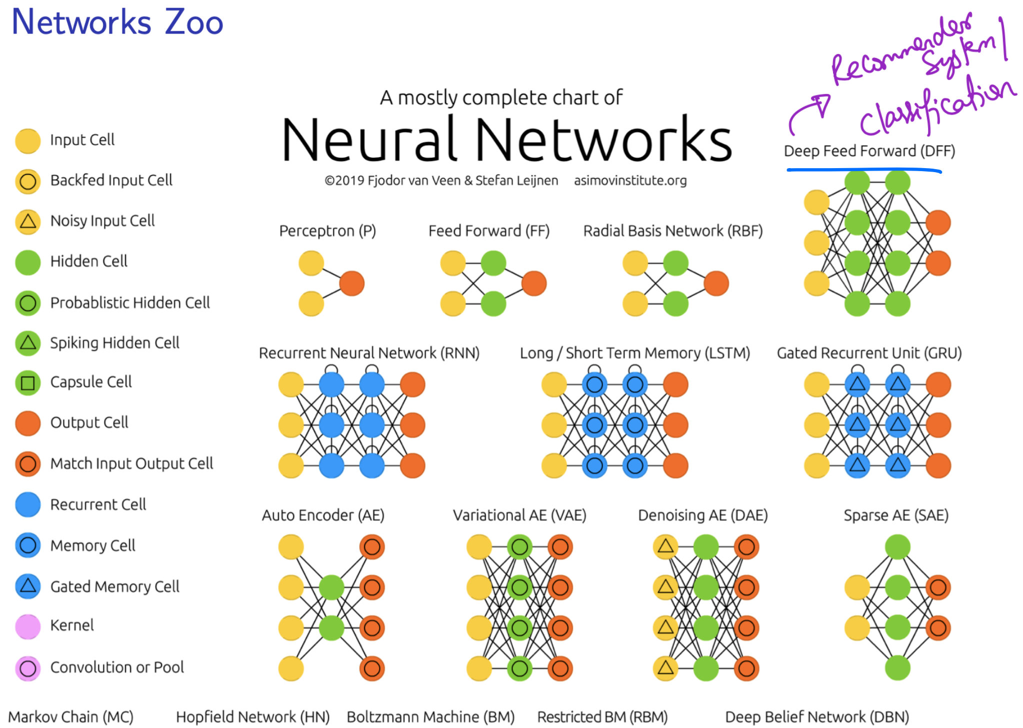
x_2 1 0 0 1 x_1

Learn XOR | Risk Assessment



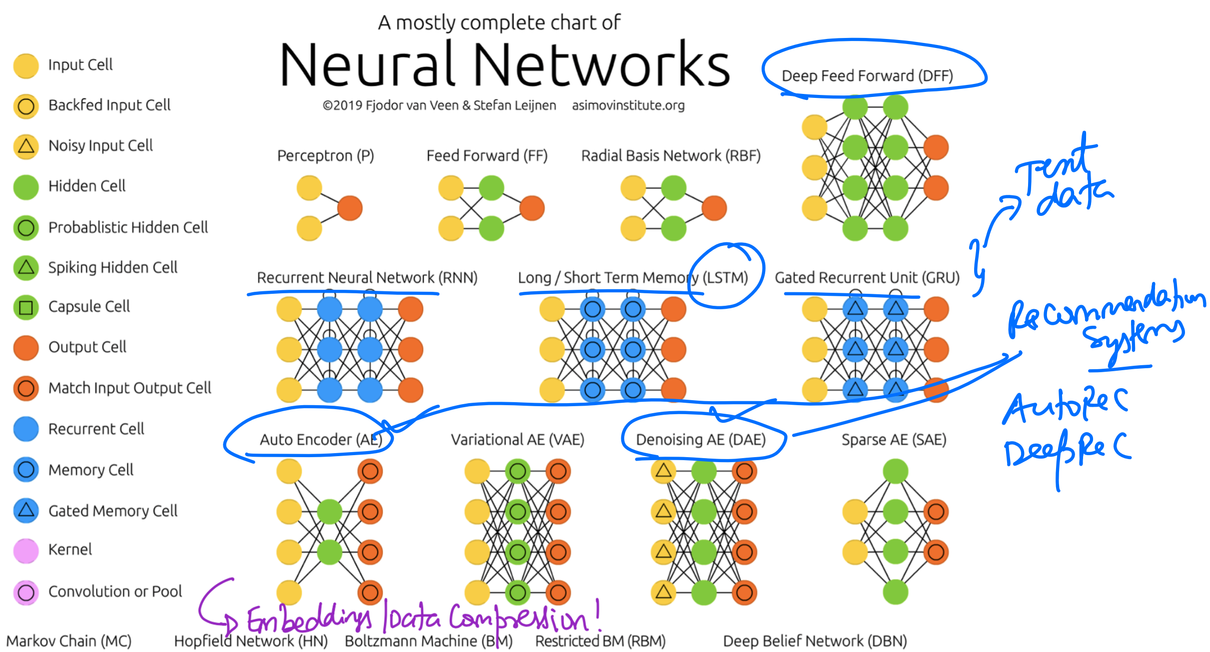
Quick tour of DL Architectures

Neural Networks Zoo



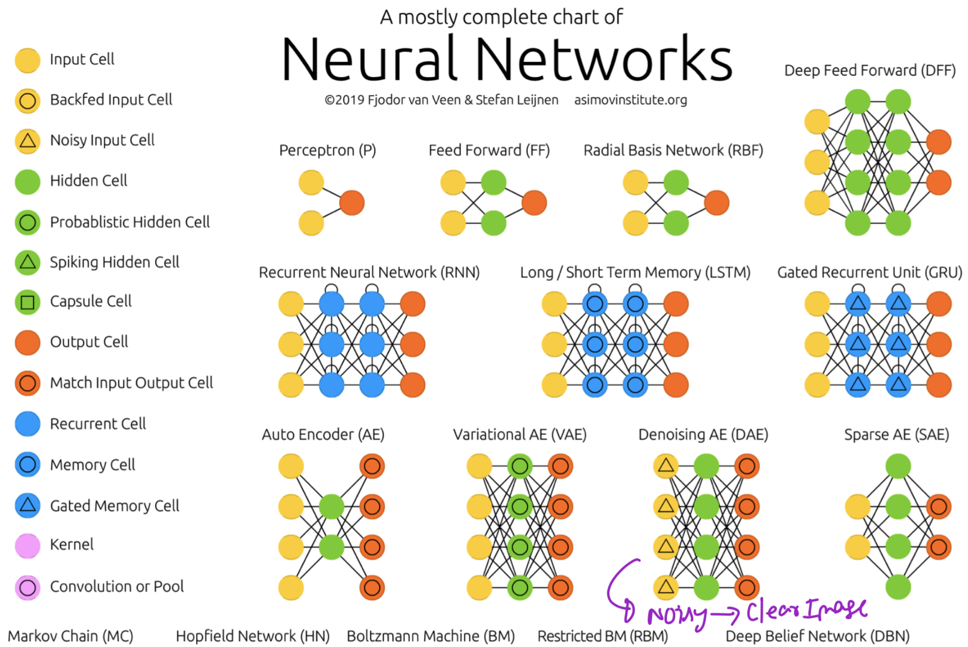
Quick tour of DL Architectures

Neural Networks Zoo



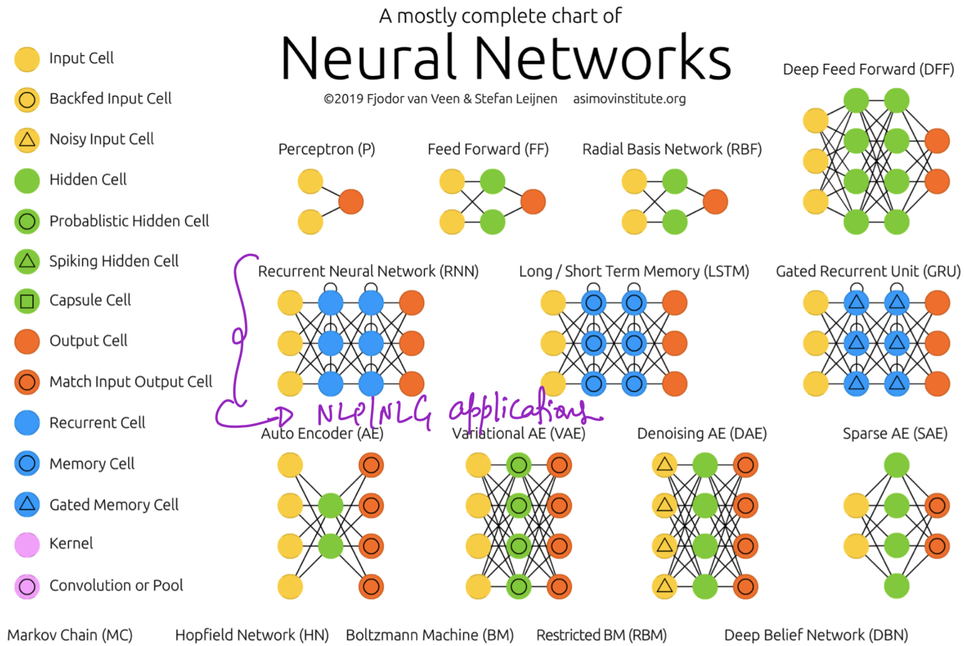
Neural Networks Zoo

Neural Networks Zoo



Neural Networks Zoo

Neural Networks Zoo



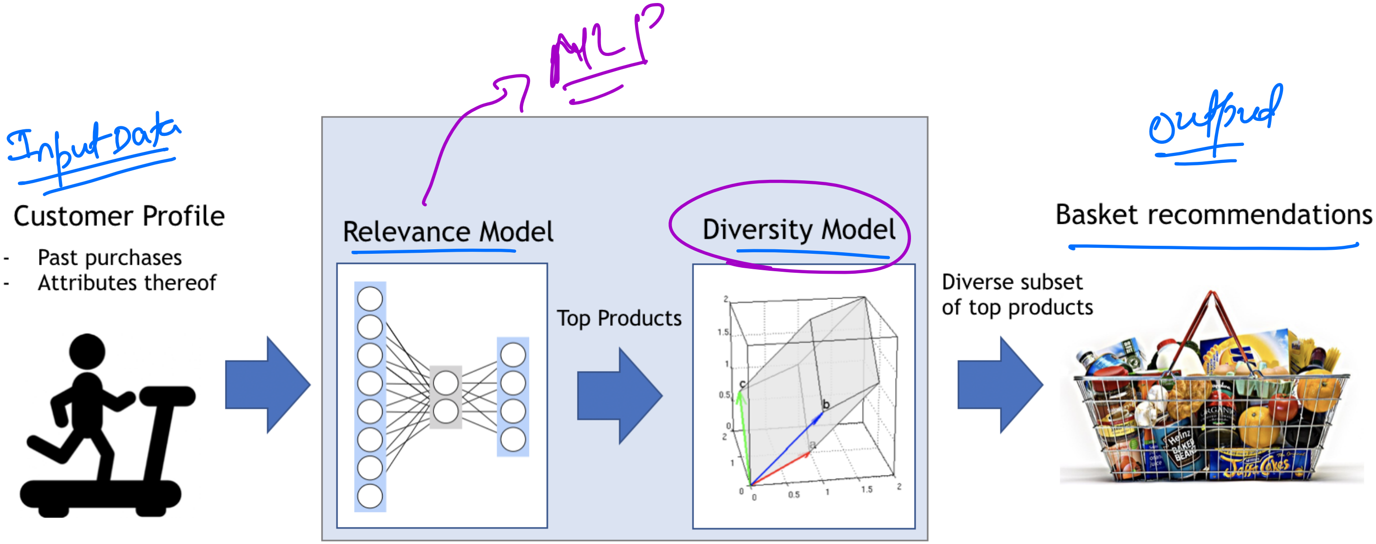
Deep Learning for Recommendation Systems

- Auto-Encoders (Non-linear MF)
- Multi-layer perceptron (MLP) [YouTube] (FFN) → Today!
- Deep semantic structured models (DSSM) (next Lec)
- Sequence models based on RNN (recs within session) (Text Data)
- Other models (e.g. wide and deep learning) [Google]

Deep Learning for Recommendation Systems

- Auto-Encoders (Non-linear MF)
- Multi-layer perceptron (MLP)
- Deep semantic structured models (DSSM)
- Sequence models based on RNN (recs within session)
- Other models (e.g. wide and deep learning) [1]

Joint Relevance and Diversity Model for Pantry Recommendations

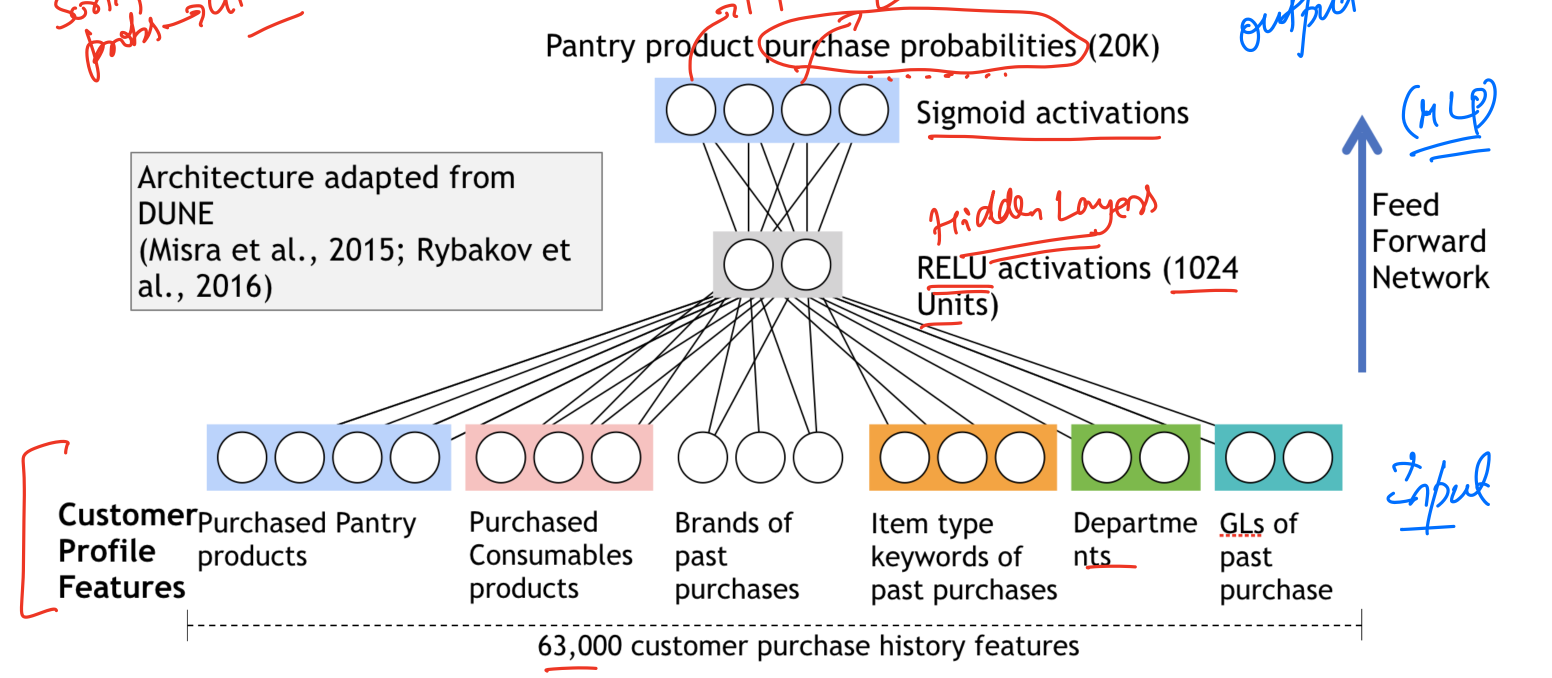


Balancing relevance and diversity for prime pantry recommendations. Mohan et al. AMLC 2017

Relevance Model

Relevance Model

Sorting probs → Given a ranking!



Input → BOW (0's & 1's)

Precision @ K vs Precision

Recs

[1 3 5]

productIds

Purchase (Truth)

[1 5 4 9]

productIds

$$\text{Precision} = \frac{\text{Recs} \cap \text{Purchase}}{|\text{Recs}|} = \frac{2}{3}$$

$$\text{Recall} = \frac{\text{Recs} \cap \text{Purchase}}{|\text{Purchase}|} = \frac{2}{4}$$

$$\text{Precision@}2 = \frac{1}{2}$$

$$\text{Recall@}2 = \frac{1}{4}$$

Precision High, Recall is Low $\Rightarrow$ Model is not having Coverage

E.g.

Recs

[1 3]

Purchase

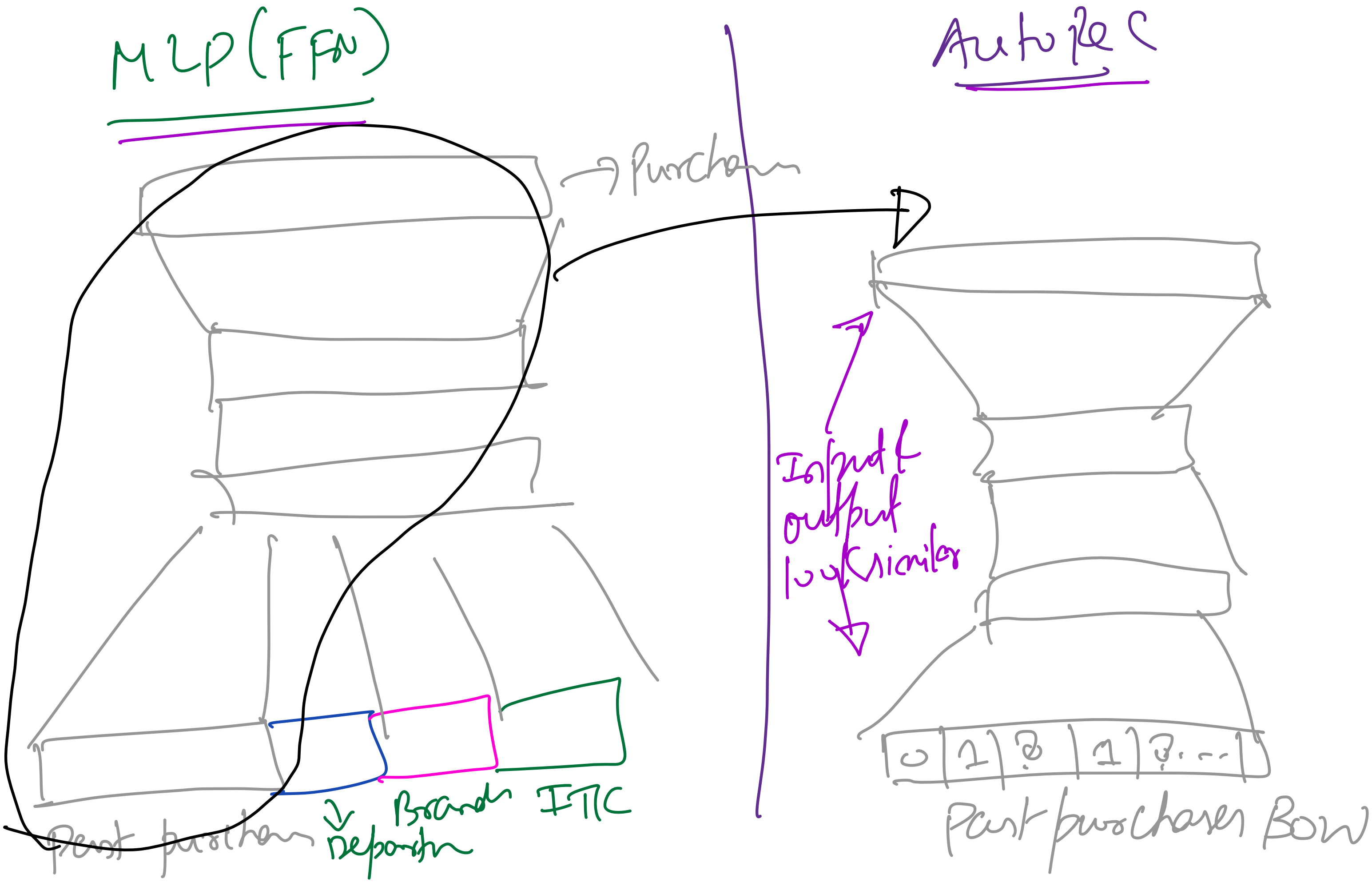
[1 3 5 7 9 10 11 15]

$$\text{Precision} = \frac{2}{2} = 100\% \quad (\text{Accuracy})$$

$$\text{Recall} = \frac{2}{8} = 25\% \quad (\text{Coverage})$$

$$\text{F-Score} = \frac{2PR}{P+R} \quad (\text{Harmonic Mean})$$

MLP vs AutoEncoder (AutoRec)



Diversity Model

Embeddings for Products

2D Represented

Represent products in product space with a large matrix of embedding coordinate vectors "***L***"

$$L = \begin{pmatrix} 1.5 & 1.9 & 1.8 & 1.4 & \cdots & 0.4 \\ 0.6 & 0.1 & 1.0 & 1.6 & \cdots & 1.9 \\ 0.6 & 1.6 & 1.6 & 1.6 & \cdots & 1.8 \\ 0.6 & 1.0 & 0.1 & 1.6 & \cdots & 0.6 \\ 0.8 & 1.4 & 1.9 & 0.8 & \cdots & 0.7 \end{pmatrix}$$

Emb

We obtain these embedding vectors from the Product2Vec service [London et al, 2017]

Diversity Subset Selection through DPP

$S_{ij} \rightarrow$ Similarity between product i & product j

$= L_i^T L_j$

Similarity matrix

Let L be the Product2Vec embeddings. $S = L^T L$ is the similarity matrix of the asins.

Determinantal point processes (DPP)

DPP [3] is a probability distribution over subsets of a ground set.

$$\mathcal{P}(J) \propto \log(\det(S_J))$$

Diversity and DPP

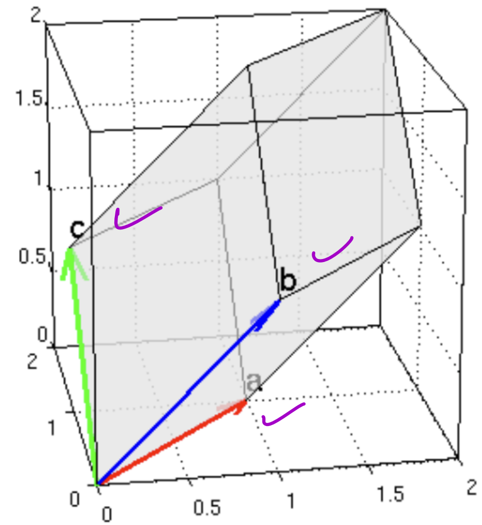
DPP favors sets that are more diverse.

[2] London et al. Product2Vec: A multi-task learning framework for cold-start product recommendation. AMLC 2016
[3] Kulesza et al. Determinantal point process for machine learning. Foundations and trends in ML, 2012

Visualizing Determinants

Larger the volume $\Leftrightarrow$ more diversity!

$$\det(LL^T) = \mathrm{Vol}^2$$

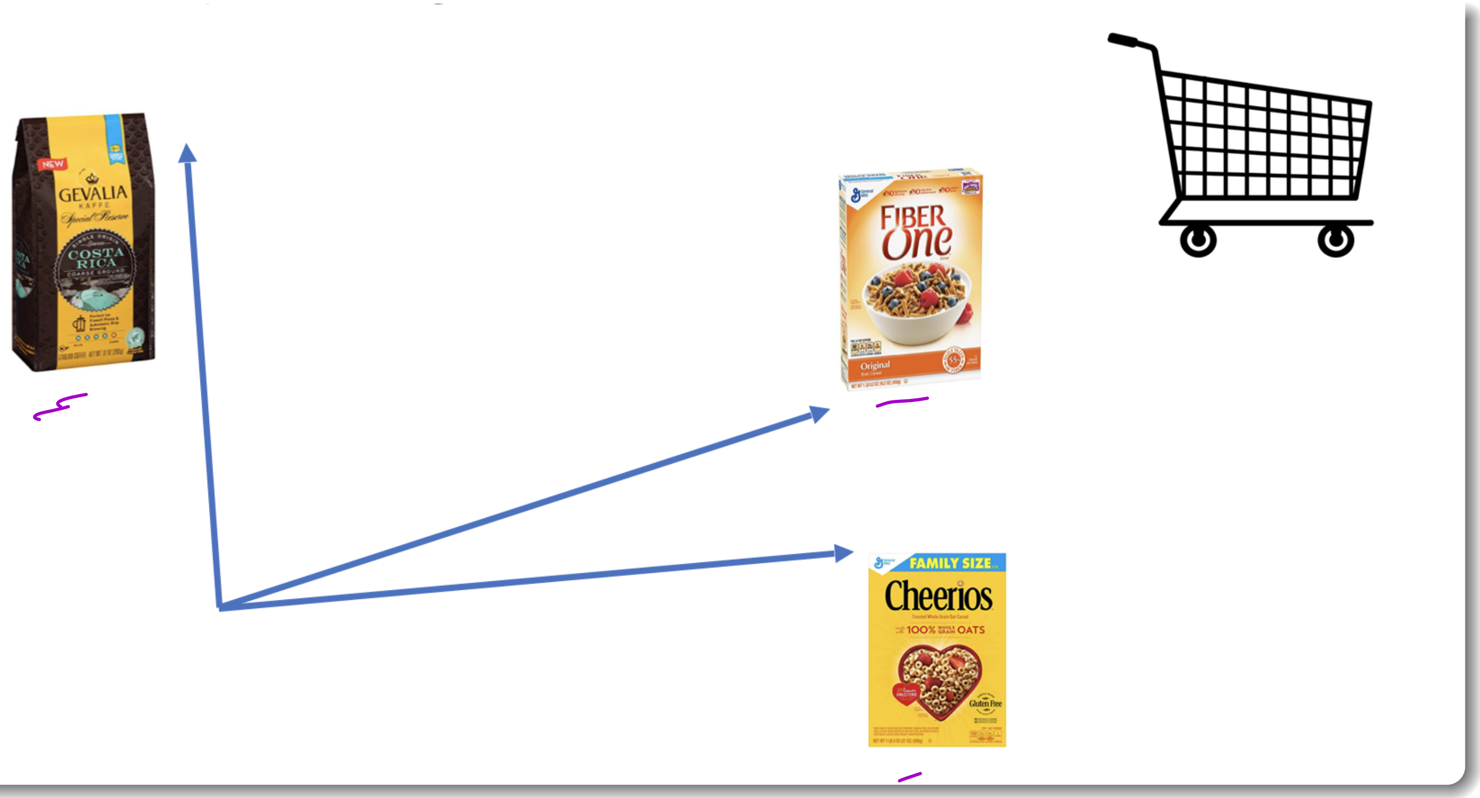


Given product vectors { **a**, **b**, **c** } and matrix

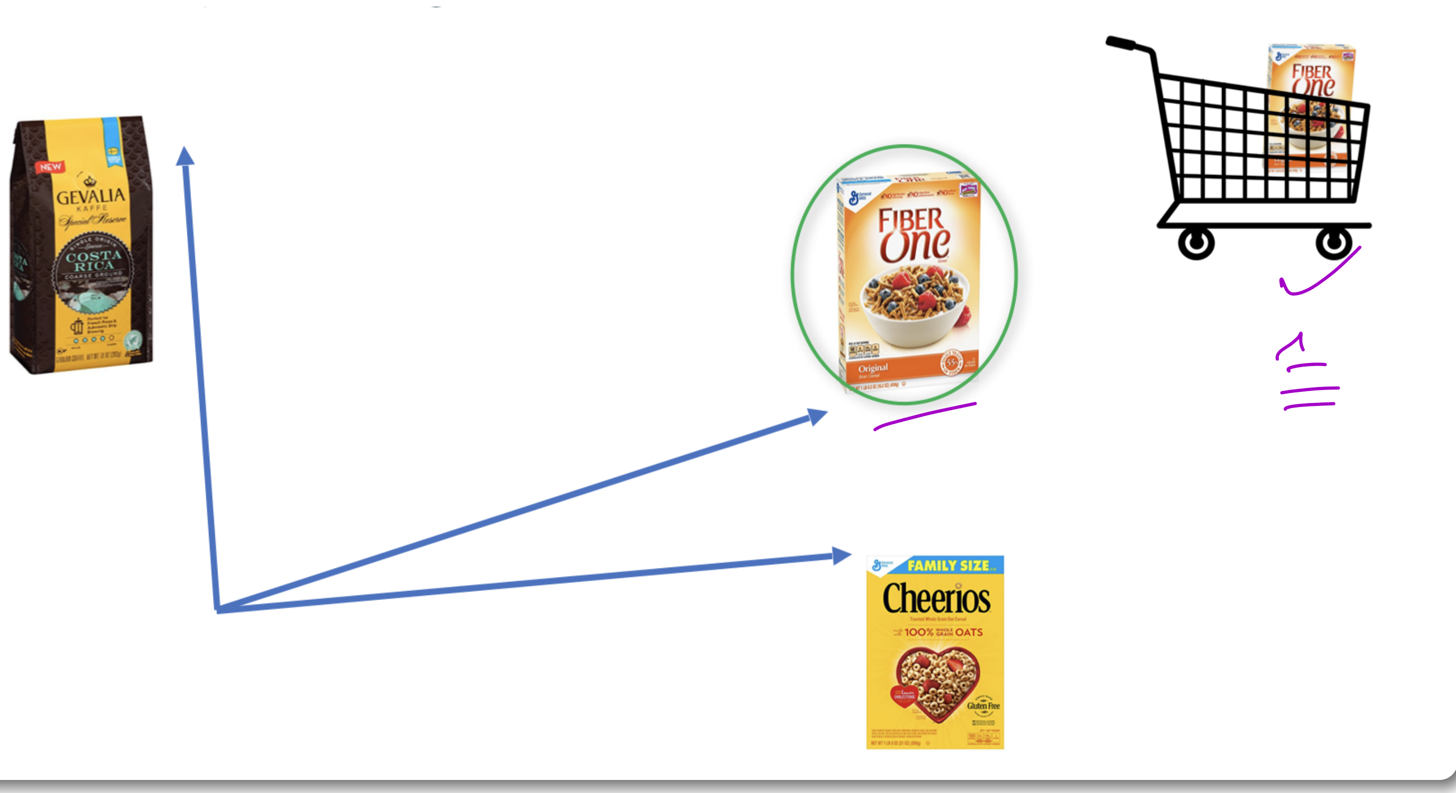
$$L = \begin{pmatrix} | & | & | \\ \mathbf{a} & \mathbf{b} & \mathbf{c} \\ | & | & | \end{pmatrix}$$

the determinant of $\boldsymbol{LL}^T$ is the squared volume of the parallelogram.

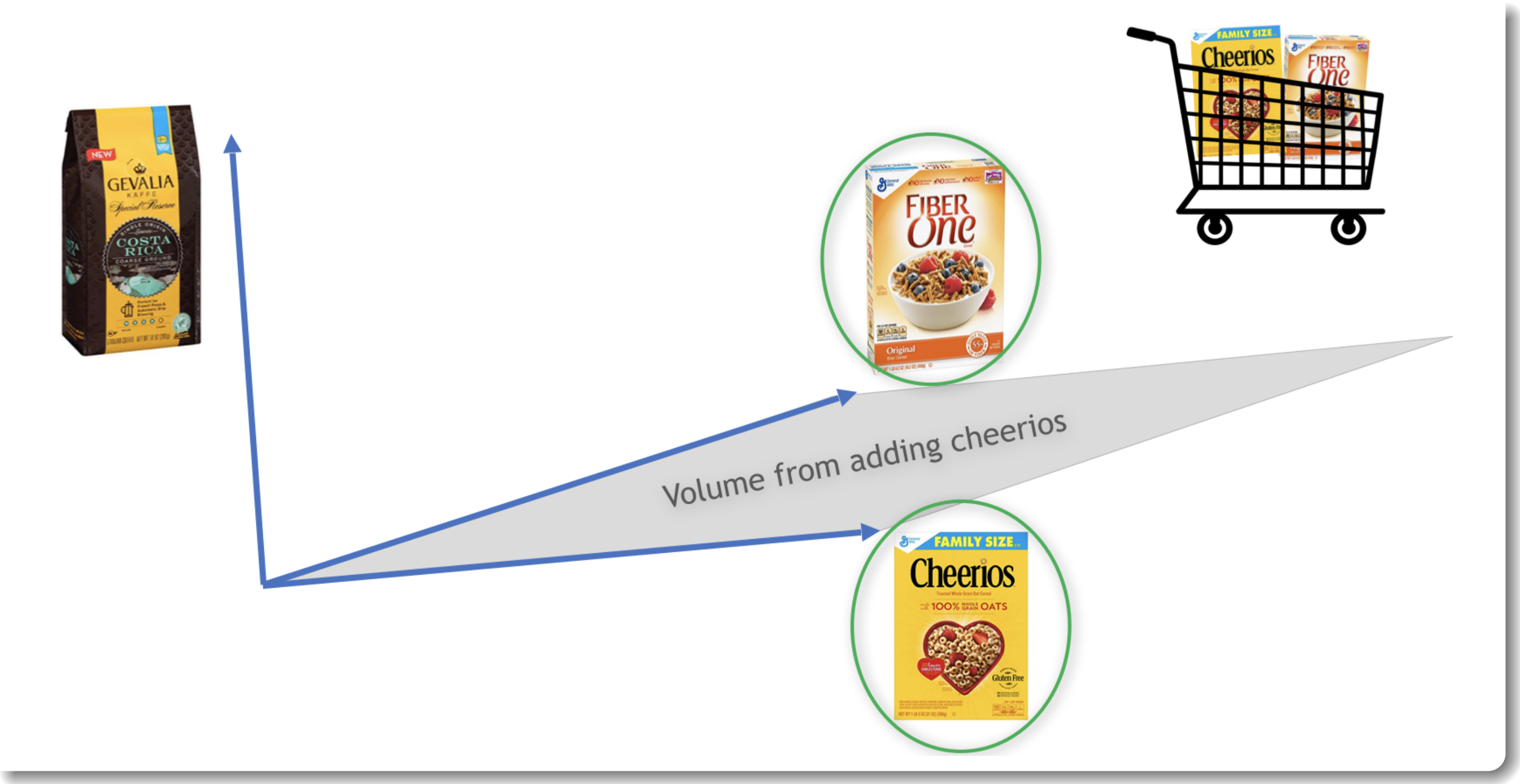
Diversity through Determinants



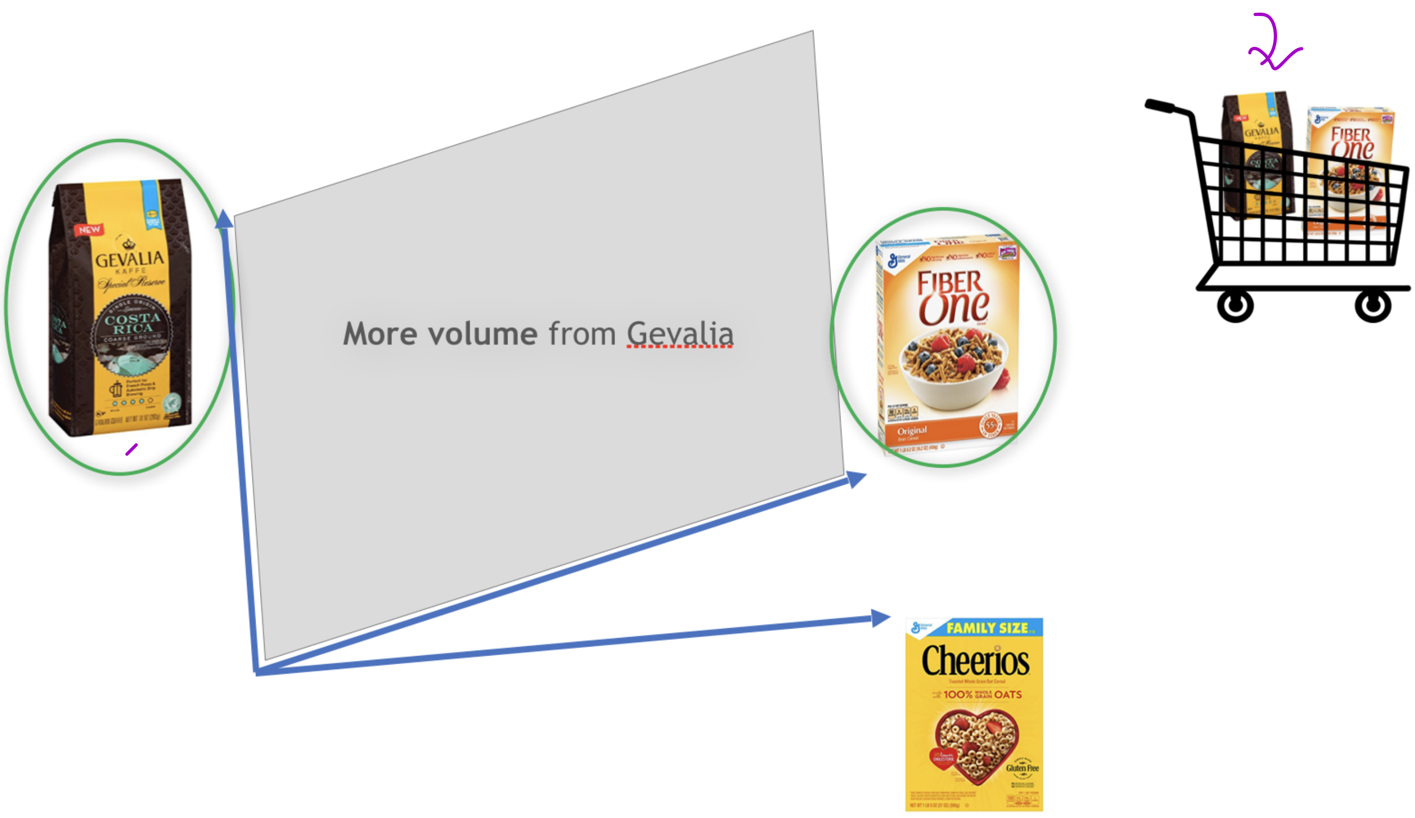
Diversity through Determinants



Diversity through Determinants



Diversity through Determinants



Trading-off Relevance and Diversity

Optimization objective strikes a balance	$\max_{J:\lvert J\rvert=k} \sum_{j\in J} r_j$

The r_j are scores from the relevance model

(prob. scores)

Trading-off Relevance and Diversity

Trade-off parameter

Optimization objective strikes a balance	$\max_{J:\|J\|=k} \underbrace{\sum_{j\in J} r_j}_{} + \lambda \underbrace{\log\det\left(\sum_{j\in J} L_j L_j^T + \gamma\mathcal{I}\right)}_{\text{Diversity}}$

The r_j are scores from the relevance model

log-det measures diversity of the set

Trading-off Relevance and Diversity

Optimization objective strikes a balance	$$\max_{J:\lvert J\rvert=k} \underbrace{\sum_{j\in J} r_j}\ + \lambda \underbrace{\log\det\left(\sum_{j\in J} L_j L_j^T + \gamma\mathcal{I}\right)}$$

Trade-off parameter

Smoothing parameter

The r_j are scores from the relevance model

Hyper-parameter

log-det measures diversity of the set

Optimizing the trade-off

Hardness of mode-finding

Mode-finding in DPP is NP-hard [4]. Hence, maximizing the joint relevance-diversity objective is also NP-hard.

Approximation algorithm

We provide a $(1 - 1/e)$-approximation algorithm to find the optimal sub-set of the joint relevance-diversity optimization problem.

Greedy Algorithm for Subset Selection (GDPP)

Initialize

Initialize set $\mathbf{J}$ with the most relevant ~~asin~~. product

Iterate until $|\mathbf{J}| = K$

- Pick asin j that has the appropriate balance between relevance and diversity:

$$j \leftarrow \arg\max_{k \notin \mathbf{J}} \left(r_k + \lambda \log\det(L_{\mathbf{J}} L_{\mathbf{J}}^T + L_k L_k^T + \gamma \mathcal{I} \right)$$

- Add asin j to set $\mathbf{J}$.

Greedy Algorithm for Subset Selection (GDPP)

Initialize

Initialize set $\mathbf{J}$ with the most relevant asin.

Iterate until $|\mathbf{J}| = K$

- Pick asin j that has the appropriate balance between relevance and diversity:

$$j \leftarrow \arg\max_{k \notin \mathbf{J}} \left(\mathbf{r_k} + \lambda \log(1 + \underbrace{\mathbf{L_k^T}(L_{\mathbf{J}} L_{\mathbf{J}}^T + \gamma \mathcal{I})^{-1} \mathbf{L_k})}_{\text{Efficient computation via rank-1 updates}} \right)$$

- Add asin j to set $\mathbf{J}$.

Relevance of ASIN *k*

Efficient computation via rank-1 updates

Greedy Algorithm for Subset Selection (GDPP)

Theorem (GDPP approximation)

GDPP enjoys a $(1 - 1/e)$-approximation guarantee in optimizing the joint relevance diversity objective.

Proof.

We note that the diversity part of the objective is submodular and monotone increasing. Since the relevance sum, $\sum_{k \in \mathbf{J}} r_k$ is modular and monotone increasing, the joint relevance-diversity objective is submodular and monotone increasing. The result thus follows from Nemhauser et al [4] □

Nemhauser et al. An analysis of approximation algorithms for maximizing submodular set functions. 1978.

Diversity Algorithm Walkthrough

Candidate products

Similar

Similar

Visualizing selection process

Visualizing selection process

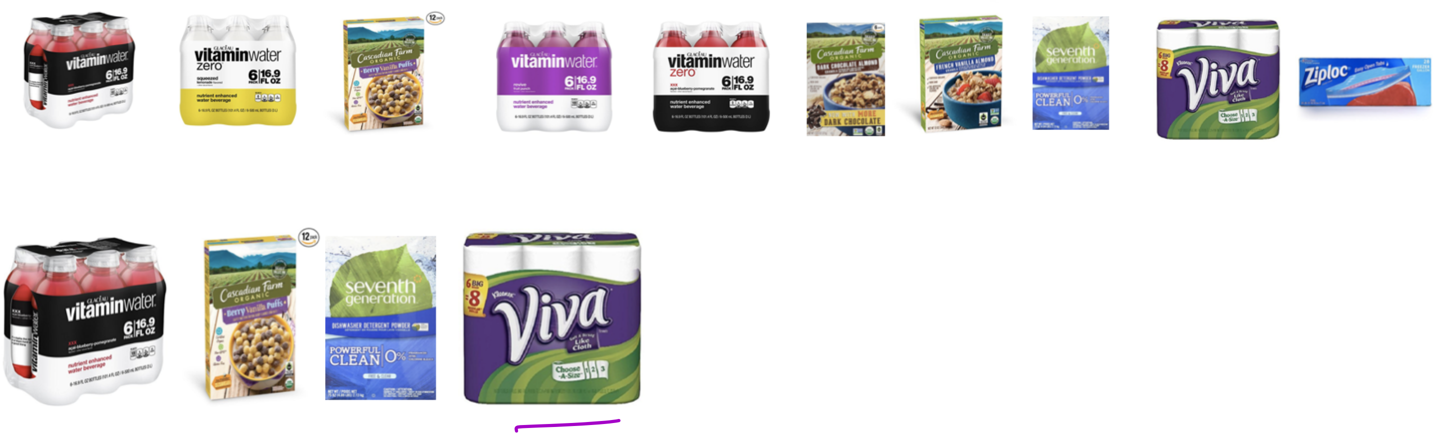
Most relevant product (1)

Visualizing selection process

close

Visualizing selection process

Visualizing selection process



Visualizing selection process

Visualizing selection process

Showcase to customer

– Relevant yet diverse!

Free-shipping A/B test

- Prime Pantry charges $5.99 per box ordered
- Pantry offers sponsored products which qualify customers for free shipping

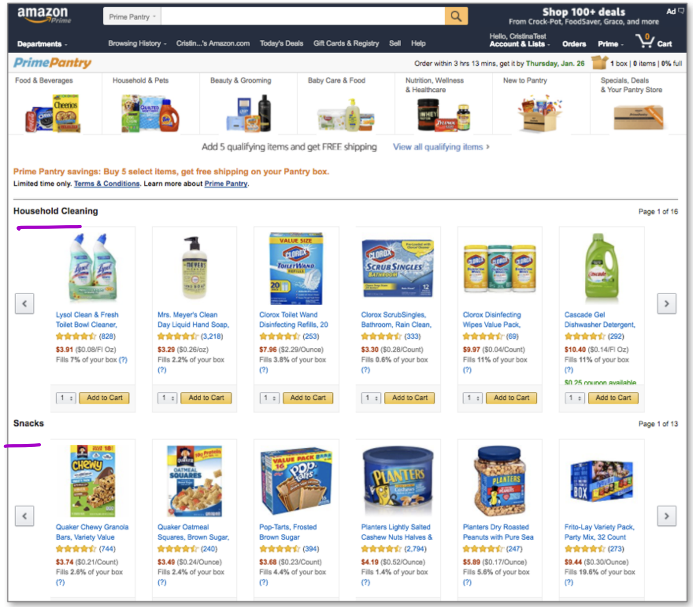
"Buy 5 Select Items Get Free Shipping"

- Help Prime Pantry customers find qualifying products through personalized recommendations

Control

Control Experience

- Free ship product carousels by category
- Categories ordered by popularity
- No personalized carousels

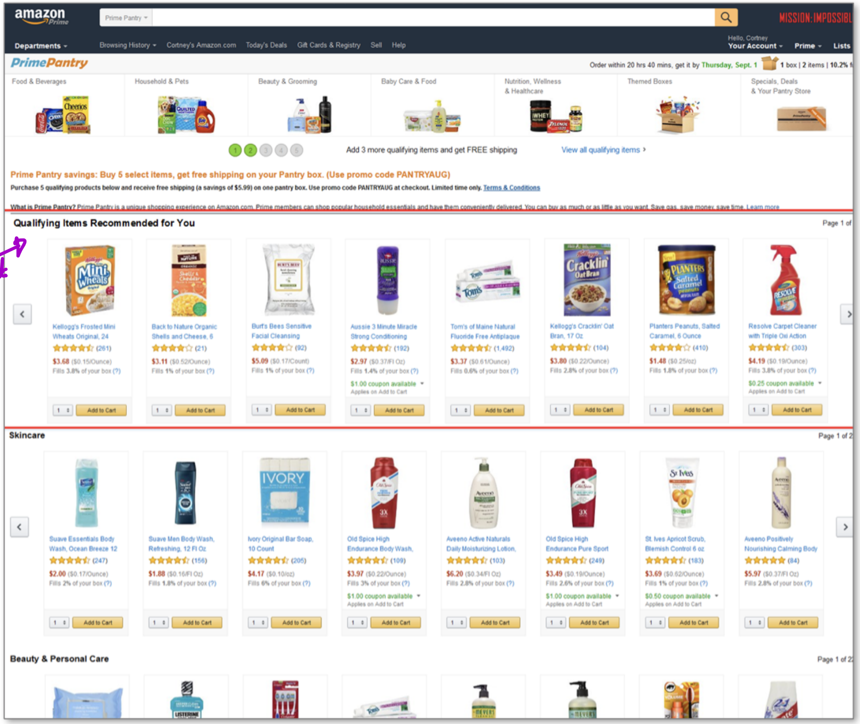


Treatment

Treatment

- Personalized carousel at the top of
  the page with qualifying items Perfect
  recommended by model

"Qualifying Items Recommended for You"



A/B test outcome

Statistical
Significant

Metric	Percentage change	P-value
Units metric	**x% increase**	**0.001 ***
Session metric	**y% decrease**	**0.004 ***
Pantry metric	**z% increase**	**0.001 ***

Business
metrics

Summary

- MLP/FFN (Extension of AutoRec)
- Relevance & Diverse Model (Jointly)
- Offline Testing (Train/val/Test)
  & online Testing (A/B Test)
- Improvement in offline metrics
  Precision/Recall/F-score
  & business metrics

Other facets

1. Training Schedule
2. Training Time Cost
3. Inference time optimization
4. Cold-Start Customers