User Choice Prediction

Ayush Kanodia (akanodia@stanford.edu), Sakshi Ganeriwal *(sakshig@stanford.edu)*

https://youtu.be/NgvhkAReN4g

Motivation & Objectives

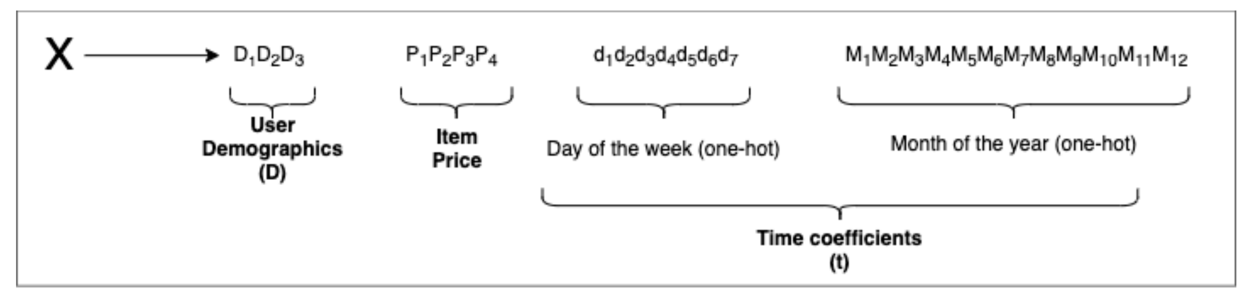
- Goal is to predict consumer choice given consumer purchase history along with some observables of consumers and items and price, time of day, day of week information and so on.
- We explored multinomial logit models, basic Neural net models, followed by multi-task learning models with and without embedding and also Bayesian embeddings.
- Building Neural networks with embeddings for these problems was challenging due to data constraints.
- The best model gives nearly 92.69% accuracy, with 2% improvements over the state-of-the-art Bayesian embedding model.

Datasets

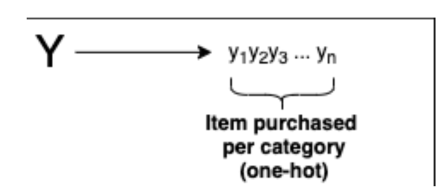
- We are using a novel dataset (~10 GB) from Susan Athey's Lab at Stanford GSB, on **user shopping choice** derived from multiple grocery stores in a large geographical region.
- The data contains complete billing information for a store on each day including information about transaction id, date, user id (from loyalty card logs), item id, price and quantity.
- We have partial user observables which includes generic personal attributes.
- And, also some information about items and information about their hierarchical classification into groups (typical examples are fruit, milk, detergent).

Features

- Our model belongs to the class of basic collaborative filtering models, except that our problem comes with **repeated choice** and more structure.
- We use a demographic **user feature** which is whether the user is married only present for 10% of users.
- Another feature we derive from the data is the tercile of every user for **the number of items** they purchase given a visit.
- modeling choice per category **encodes category information** much like a feature vector would to incorporate the item features.
- We use the following time features: month of year, week of year and day of week.

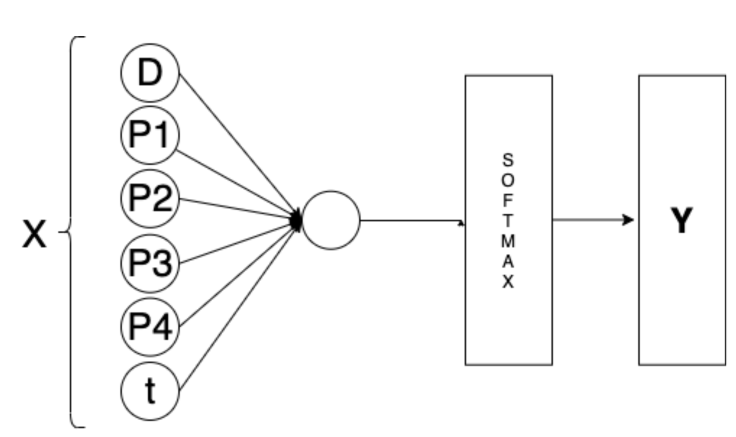


Input Feature Vector

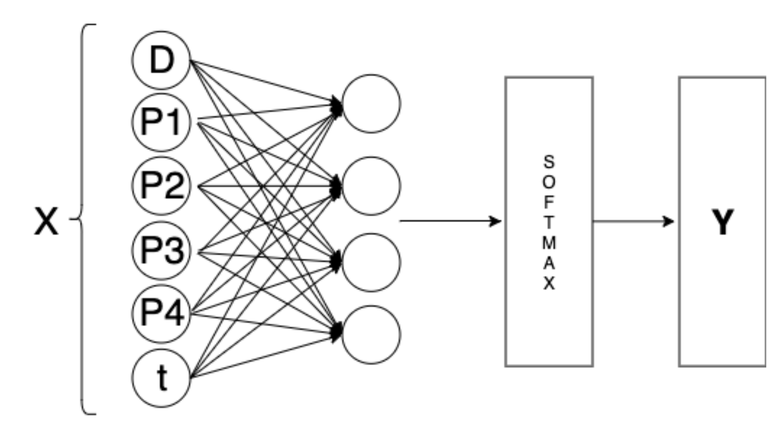


Output Feature Vector

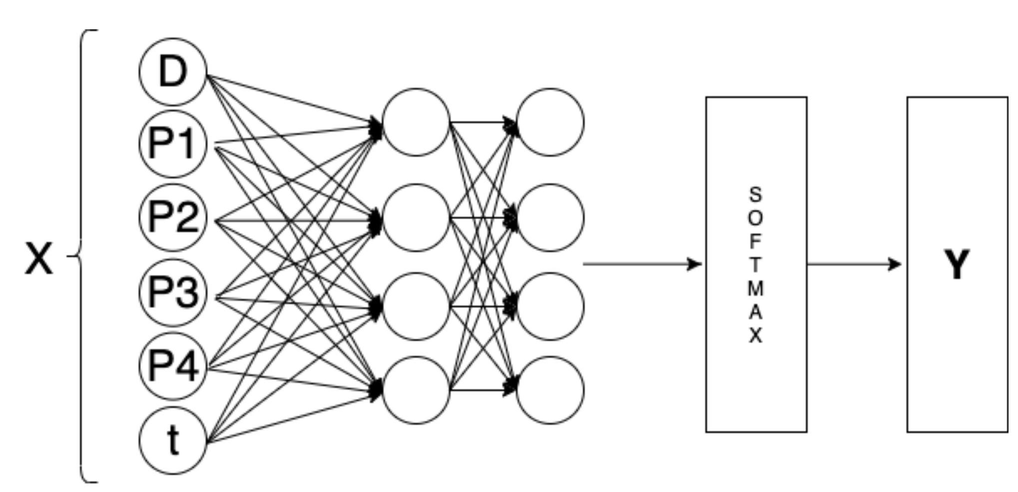
Model Architecture



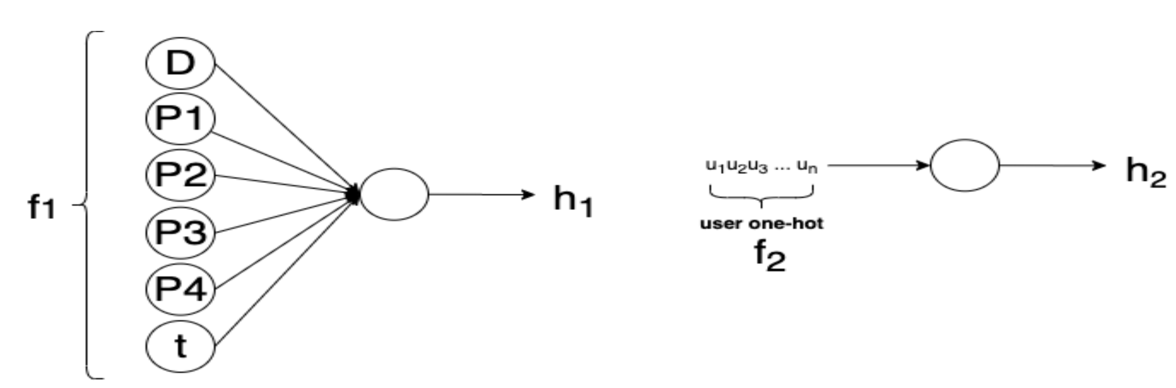
Basic Multinomial logit



MLP with 1 hidden layer

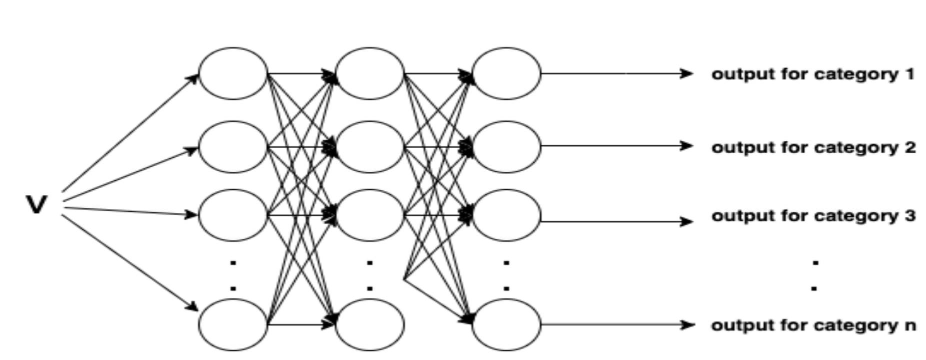


MLP with 2 hidden layer



$h_3 = h_1 \,.\, h_2$ $\quad$ $v = \text{concat}(h_1, h_2, h_3, h_4)$

$h_4 = \text{concat}(f_1, f_2)$



Multi-Task Learning with Embeddings

Results

The numbers are calculated per category and weighted by number of purchases for each category.

Model	Log Likelihood	Accuracy	Precision	Recall	F1
baseline_item_freq	-0.284942	0.917270	0.855519	0.917270	0.882245
baseline_user_item_freq	-0.282097	0.923744	0.902912	0.923744	0.908003
basic_mnl	-0.379898	0.917270	0.507557	0.534546	0.519109
multilayer_mnl	-0.328202	0.918198	0.519576	0.535474	0.525645
unrestricted_mnl	-0.264767	0.924011	0.689506	0.718780	0.698052
multilayer_mnl_two_layer	-0.245823	0.925648	0.691126	0.720417	0.705042
multilayer_mnl_one_layer	-0.240314	0.926139	0.701658	0.720907	0.705855
multitask_basic	-0.274243	0.918732	0.879776	0.918732	0.889628
multitask_user_onehot	-0.237934	0.926906	0.904628	0.926906	0.909831
multitask_embeddings	-0.236758	0.926886	0.904936	0.926886	0.910115
one level bayesian embeddings	-0.244352	0.925208	-	-	-
two level bayesian embeddings	-0.243596	-	-	-	-

Table 1: Test Set Results

Discussion & Future Work

- There had been no deep learning models being used for this problem so far.
- We also observe that the baseline model on user-item frequency performs better than the basic and unrestricted MNL models.
- In the series of models we designed, the best performing multi-task model with user embeddings beats state of the art embedding models which is a non-deep learning model.
- We are currently investigating the utility of this model in the real world.
- An exciting point of difference is how counterfactual accuracy would compare for our deep learning models versus current state of the art models.

References

[1]Francisco JR Ruiz, Susan Athey, and David M Blei. Shopper: A probabilistic model of consumer choicewith substitutes and complements.arXiv preprint arXiv:1711.03560, 2017.

[2]Susan Athey, David Blei, Robert Donnelly, Francisco Ruiz, and Tobias Schmidt. Estimating heterogeneousconsumer preferences for restaurants and travel time using mobile location data. InAEA Papers andProceedings, volume 108, pages 64–67, 2018.

[3]Robert Donnelly, Francisco J. R. Ruiz, David M. Blei, and Susan Athey. Counterfactual inference forconsumer choice across many product categories.CoRR, abs/1906.02635, 2019.

[4]Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborativefiltering. InProceedings of the 26th international conference on world wide web, pages 173–182, 2017.