Wide and Deep bandits

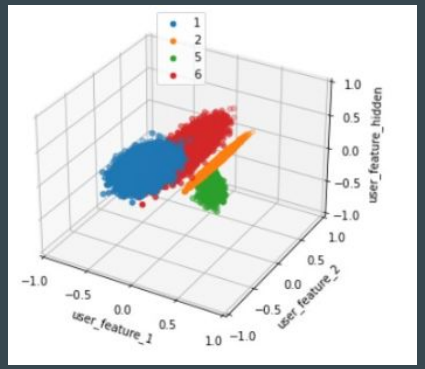
demo Mar 12, 2021

Alexsey, Jenny, Nirmal, Tengfei

Summary of this week

Investigate on 3 other datasets:

- Starbucks dataset (Tengfei and Jenny)
- ADS-16 dataset (Nirmal)
- Generated dataset (Alexsey) **



Starbucks Dataset

- 10 unique offers (offer_id, reward, channel, difficulty, duration, offer_type)
- 17000 unique users (user_id, gender, age, reg_date, income)
- 306k transcripts (user_id, event, value, time)

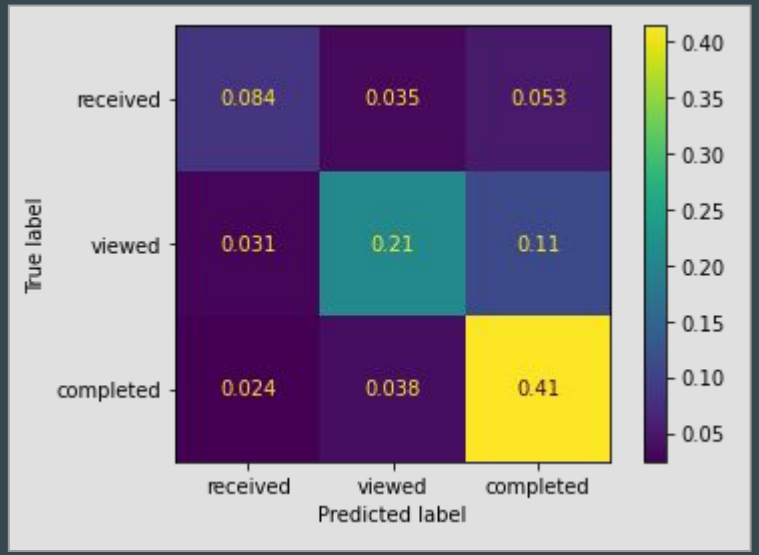
transcript

	person	event	value	time
0	78afa995795e4d85b5d9ceeca43f5fef	offer received	{'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}	0
1	a03223e636434f42ac4c3df47e8bac43	offer received	{'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}	0
2	e2127556f4f64592b11af22de27a7932	offer received	{'offer id': '2906b810c7d4411798c6938adc9daaa5'}	0
3	8ec6ce2a7e7949b1bf142def7d0e0586	offer received	{'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}	0
4	68617ca6246f4fbc85e91a2a49552598	offer received	{'offer id': '4d5c57ea9a6940dd891ad53e9dbe8da0'}	0
...	...	...	...	...
306529	b3a1272bc9904337b331bf348c3e8c17	transaction	{'amount': 1.5899999999999999}	714

event	value
offer received	{'offer_id': xxxxxxxxxx}
offer viewed	{'offer_id': xxxxxxxxxx}
offer complete	{'offer_id': xxxxxxxxxx}
transaction	{'amount': 1.589999999}

Starbucks Dataset

- predictive accuracy ~ 71.4%



similar accuracy:

Decision Tree vs Wide and Deep

reason:

The dataset is relatively small considering the number of users, so it means that the wide part won't learn/memorize too much for each user.

ADS-16 dataset

- 300 ads in total: 20 Ad categories x 15 Ads in each category
- 120 users in total

	Name	Last Name	Gender	Age	Paypal	Type of Job	Weekly working hours	Income	Home country	Home town	C C
0	Hidden	Hidden	F	62	hidden@comcast.net	Housewife/Househusband	Full Time	1	United States of America	Apollo	1

	Most visited websites	Most listened musics	Most watched movies	Most watched tv programmes	Most read books
0	Media (Books, DVD, CD/DVD Music) sites, Grocer...	Classical Music, Easy Listening, Jazz	Action, Thriller, Drama, Comedy, Mystery	Comedy, Drama	Mystery

	fave1	fave2	fave3	fave4	fave5
0	U0001-IM-POS/1.png	U0001-IM-POS/2.png	U0001-IM-POS/3.png	U0001-IM-POS/4.png	U0001-IM-POS/5.png
1	my cats	my cats	movie we are in	tv show we are in	movie we are in

	Cat0	Cat1	Cat2	Cat3	Cat4
0	Clothing & Shoes	Automotive	Baby Products	Health & Beauty	Media (E
1	1,1,1,1,1,3,1,1,1,1,1,1,1,1,1	1,1,1,1,1,1,1,1,1,1,1,1,1,1,1	2,1,1,1,3,1,1,1,1,1,1,1,1,1,1	1,1,1,3,1,1,1,1,1,1,1,2,2,2,1	1,1,1,1,

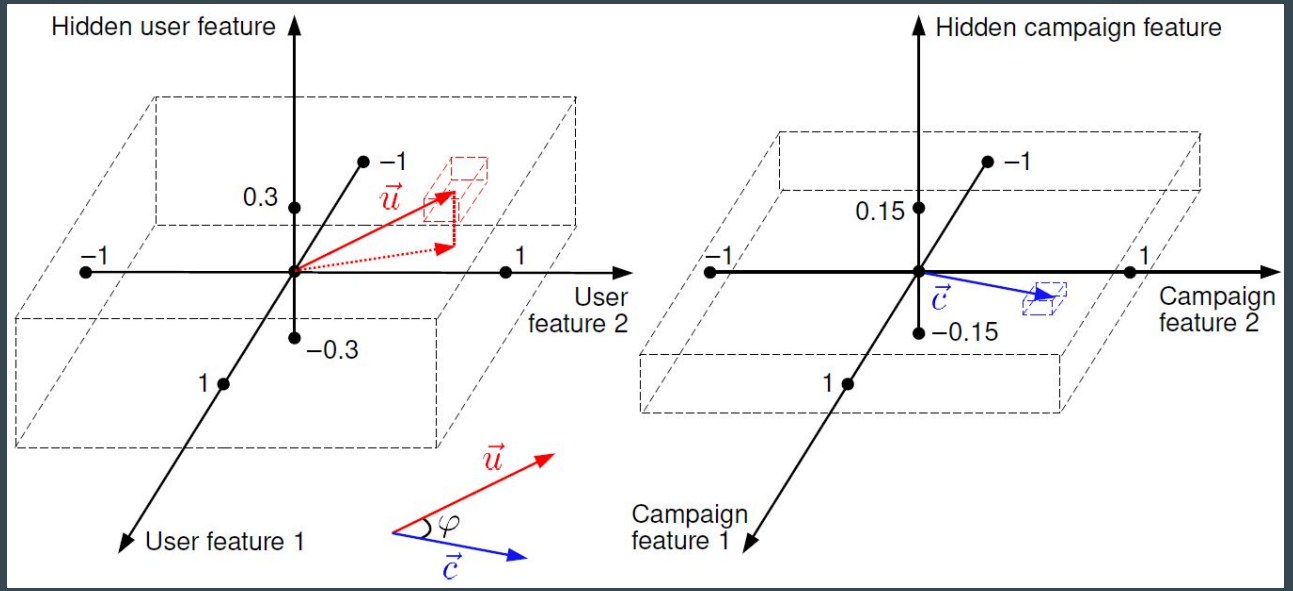
Too many features for each user, and at the same time, we have a very small number of data points for the model to learn. So in this case, we decided to **stop** using this dataset.

Generated dataset

10,000,000 samples:

- 1000 unique users: feature_1, feature_2 and feature_hidden
- 100 unique campaigns: feature_1, feature_2 and feature_hidden
- optimal action (10 possible actions)
- reward

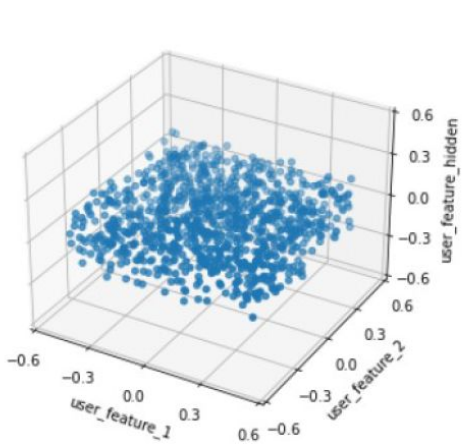
Generated dataset



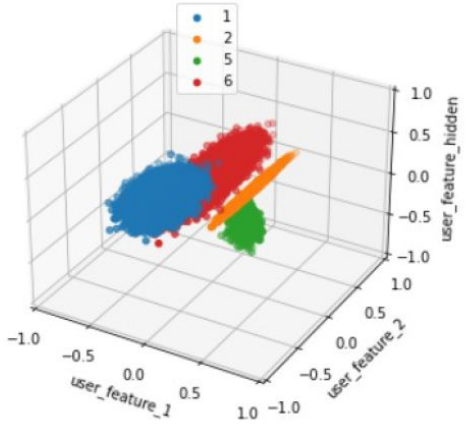
$$a = \begin{cases} 1, & \text{if } u \parallel c \\ \text{ceil}\left[\frac{10}{\pi}\arccos\left(\frac{\vec{u}\cdot\vec{c}}{|\vec{u}||\vec{c}|}\right)\right], & \text{otherwise} \end{cases}$$

Generated dataset

Mean values of user parameters for each of the 1000 users

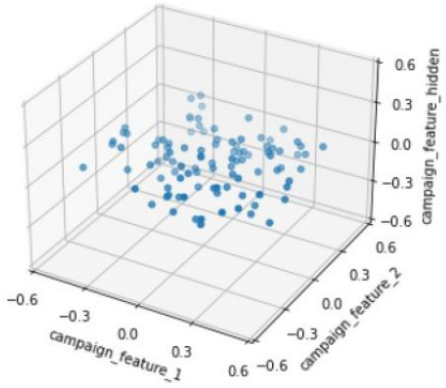


Values of user parameters for all occurrences of select users (1,2,5,6)

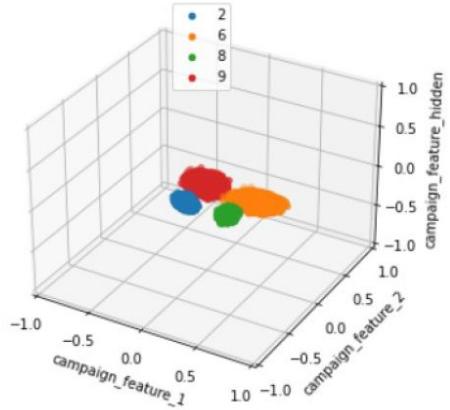


Generated dataset

Mean values of campaign parameters for each of the 100 campaigns

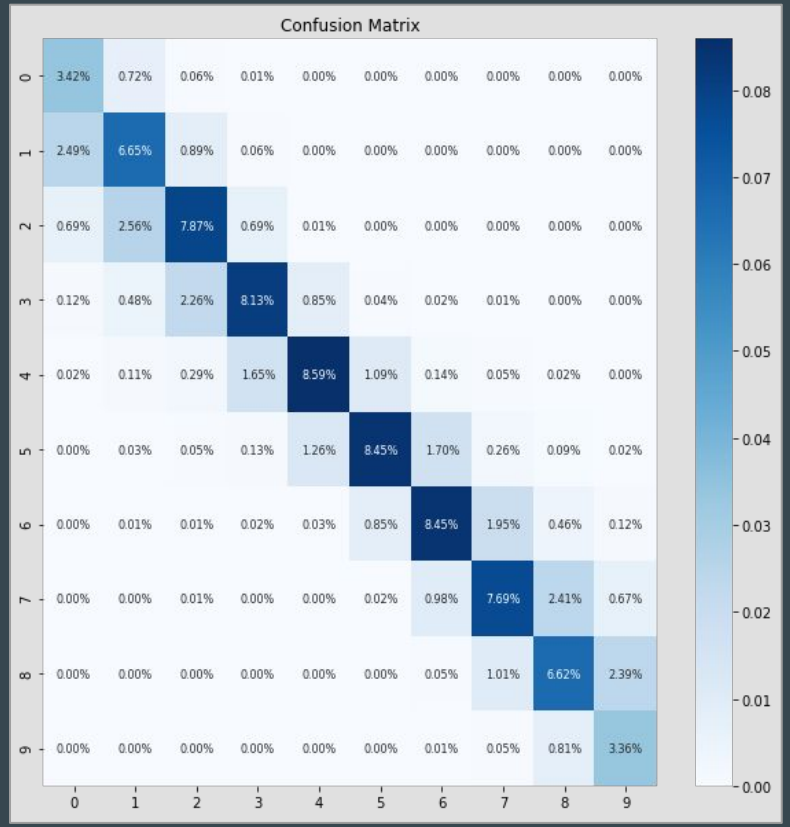


Values of campaign parameters for all samples in select campaigns (2,6,8,9)



Generated dataset

- LGBMdataset ~ 55.8%
- Grid sorting ~ 59.1%
- Deep-only ~ 67.9%
- W&D with user_id in wide ~ 68.3%
- W&D with user_id and campaign_id in wide ~ 68.4%
- W&D with cross products ~ on-going
- W&D with bandits ~ on-going



Next week

- Complete the testing on the generated dataset, hopefully to demonstrate the advantages of W&D bandits.
- Upload the latest version of codes
- If everything goes well, finalize this track(with a recap blog?).