Prediction Customer Behaviour: SVM and Neural Network

Yuhao Ding, Bilal Munawar, Kenton Blacutt
May 2019

NYU Shanghai
Hack the Pearl

Viewing the dataset

- Unbalanced Dataset:
  1-2% of customers

Viewing the dataset

- Unbalanced Dataset:
  1-2% of customers
- A focus on the dataset:
  Demographics
  Account Balance
  Change in Account Balance
  TMD Expiration Date
  Previous Wealth Purchases

First Step:

Identifying Pairwise Connections

Intuition:

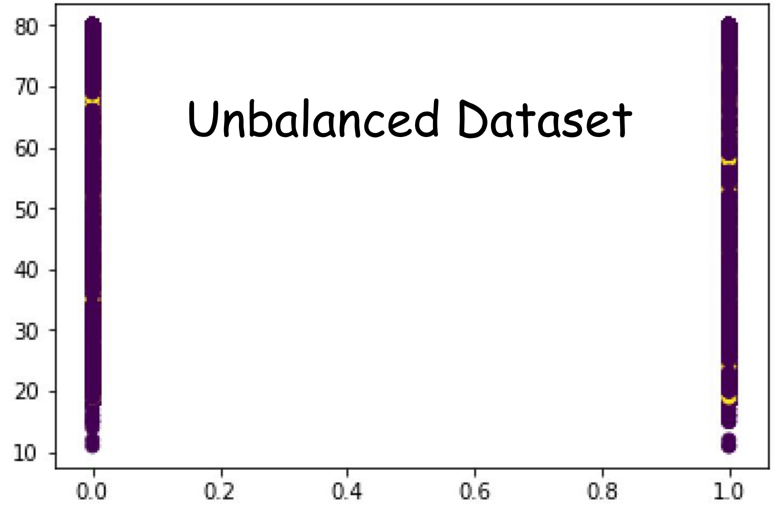
Some of the features, might be the key factor of determining wealth purchases.

Expectation:

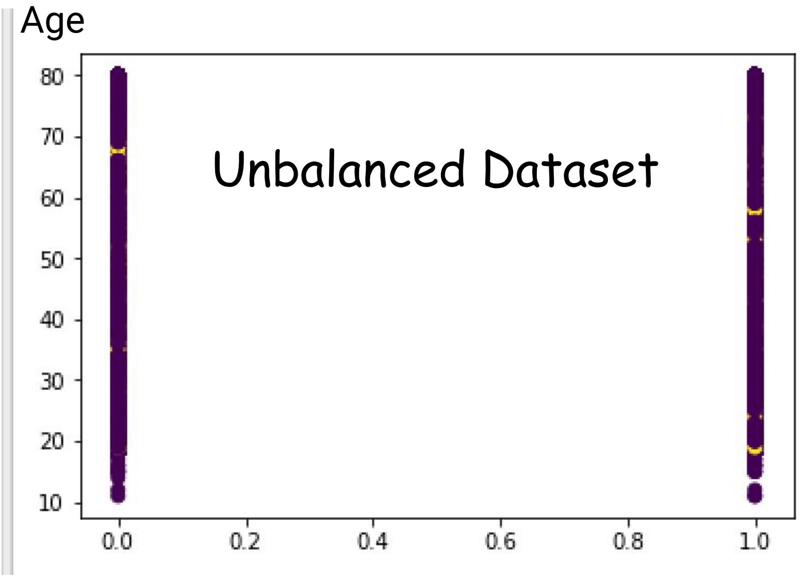
We expect that some of the factors may be positively deterministic and some are vice versa.

For example, it may be of high probability that a person travelling abroad will buy insurance.

Age, gender and Purchases?

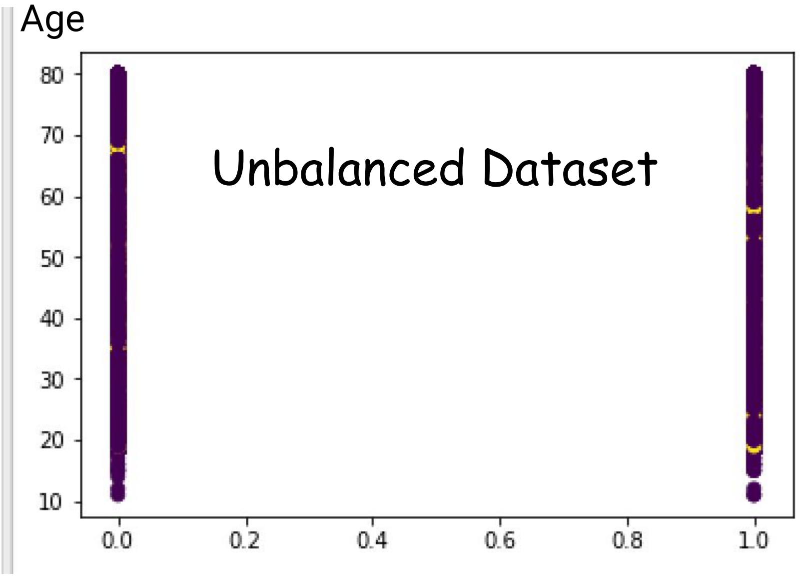


Age and Purchases?

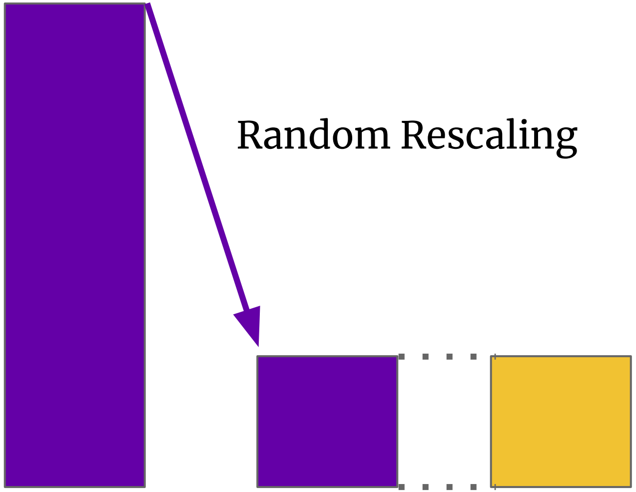


- We use random rescaling to remove the unbalance

Age and Purchases?

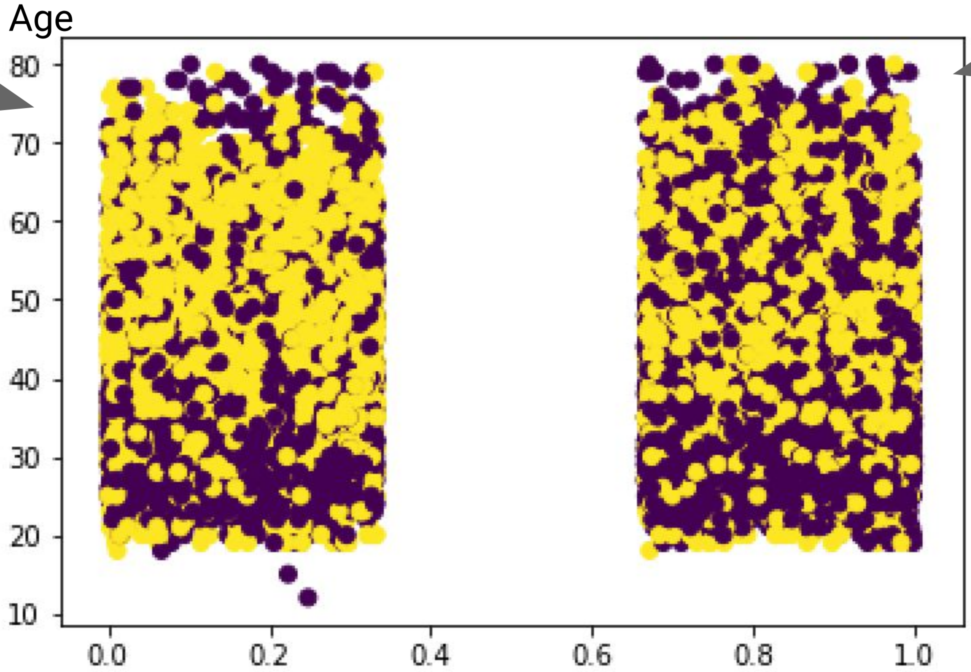


We use random rescaling to remove the unbalance



Age and Purchases?

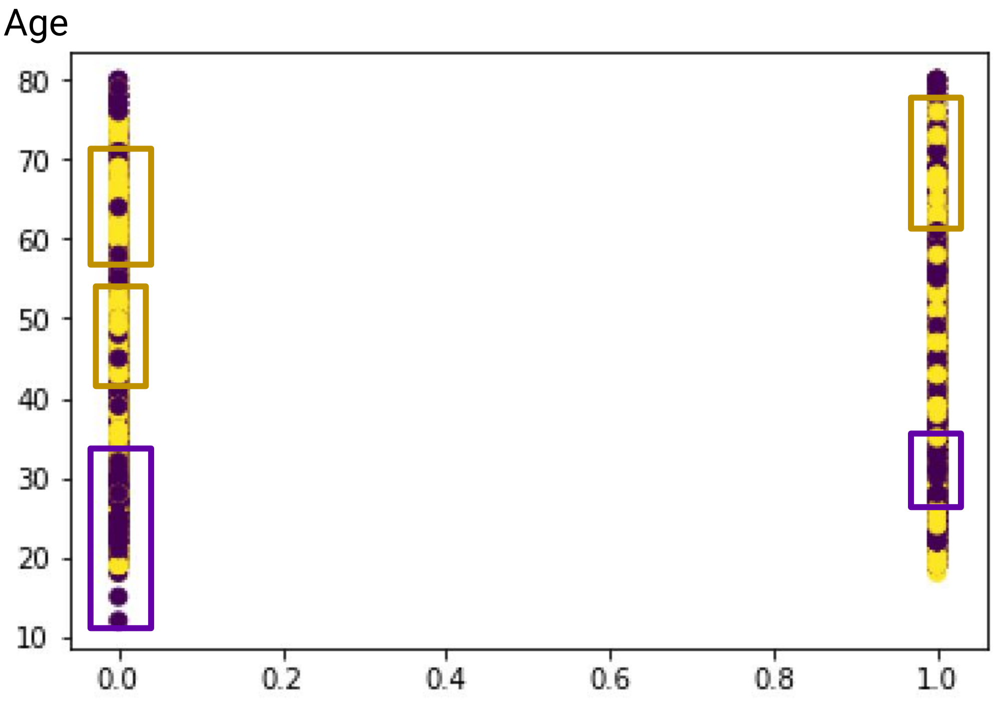
Female



Male

≈ 6000 data points

Intuition: Key areas



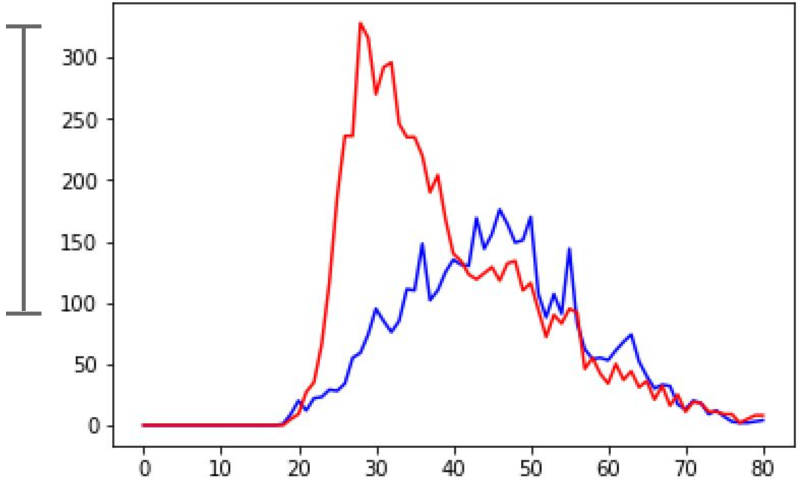
Finally,

People who purchased

People who did not purchase

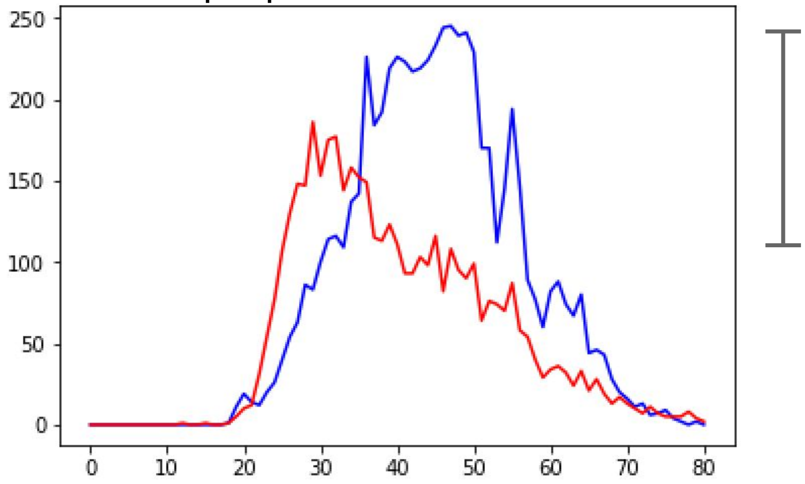
Male:

Number of people



Female:

Number of people



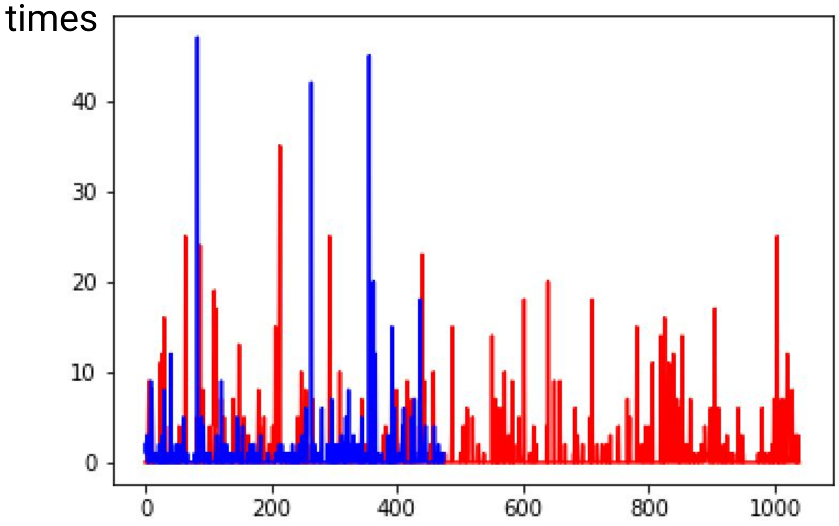
Similarly, we can identify these relations between other pairwise features

Transaction?

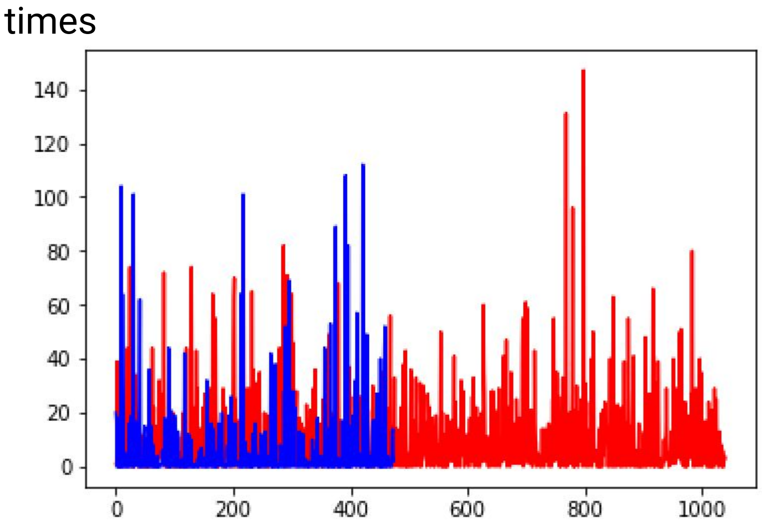
People who did not purchase

People who purchased

Internet banking



Credit card



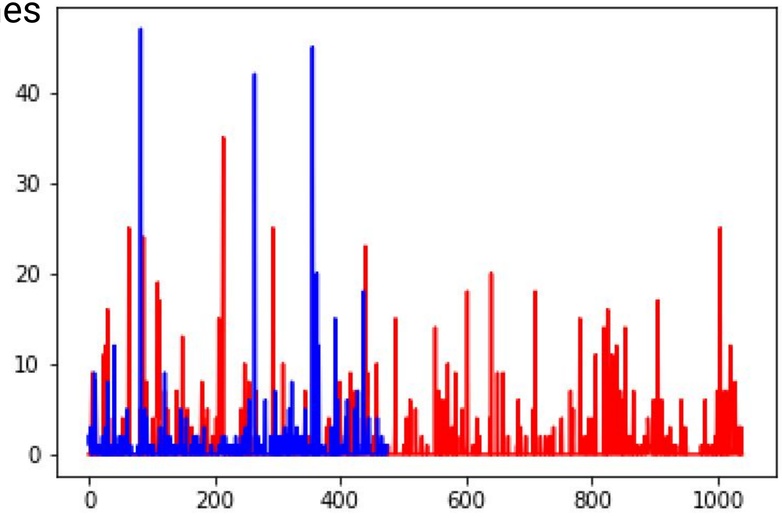
Transaction?

People who did not purchase

People who purchased

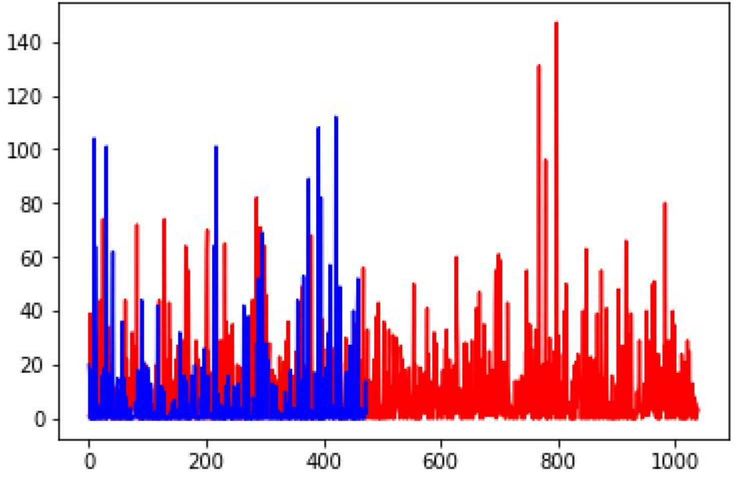
Internet banking

times



Credit card

times



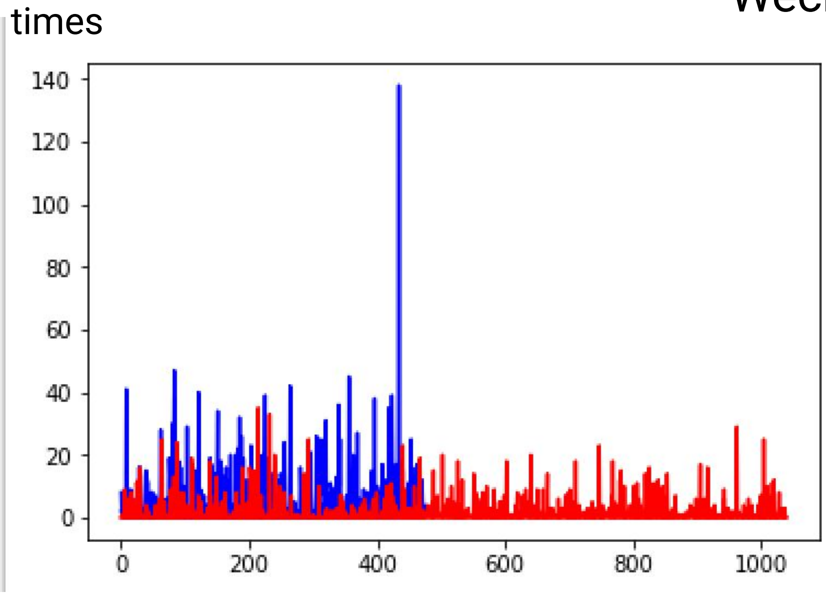
Not very helpful

Transaction?

People who did not purchase

People who purchased

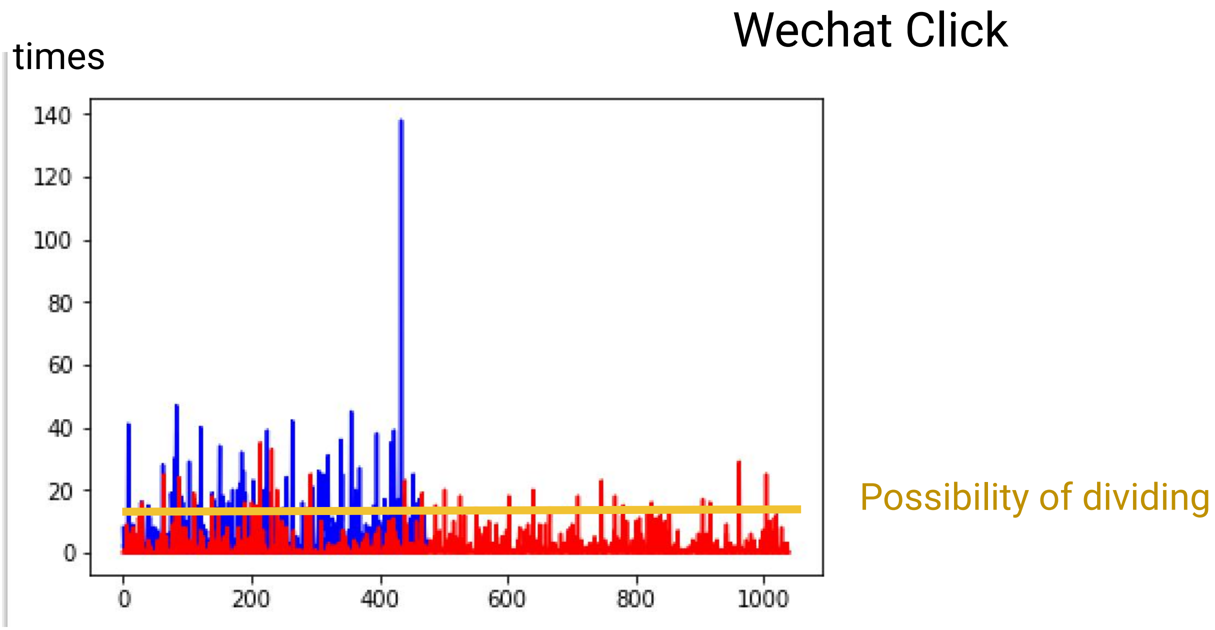
Wechat Click



Transaction?

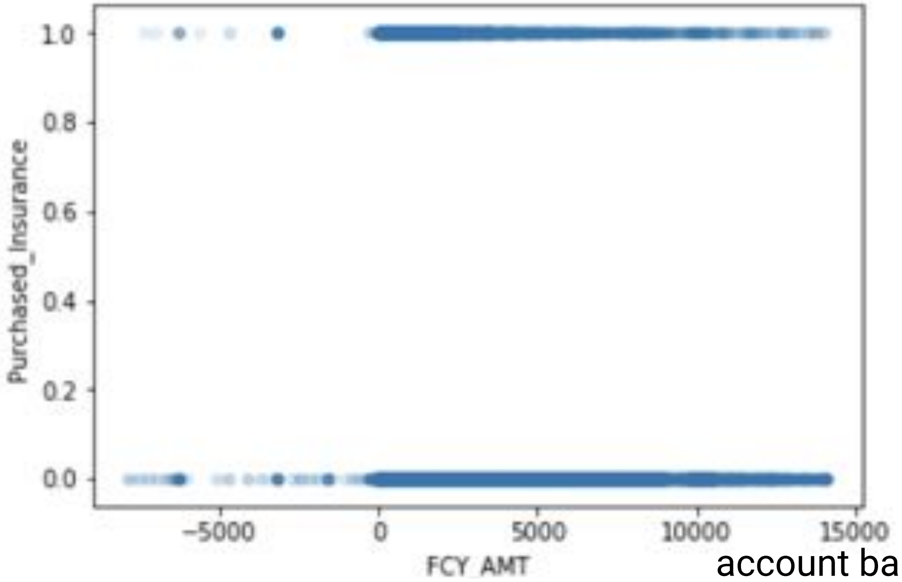
People who did not purchase

People who purchased



Account Balance

whether they purchased (0/1)



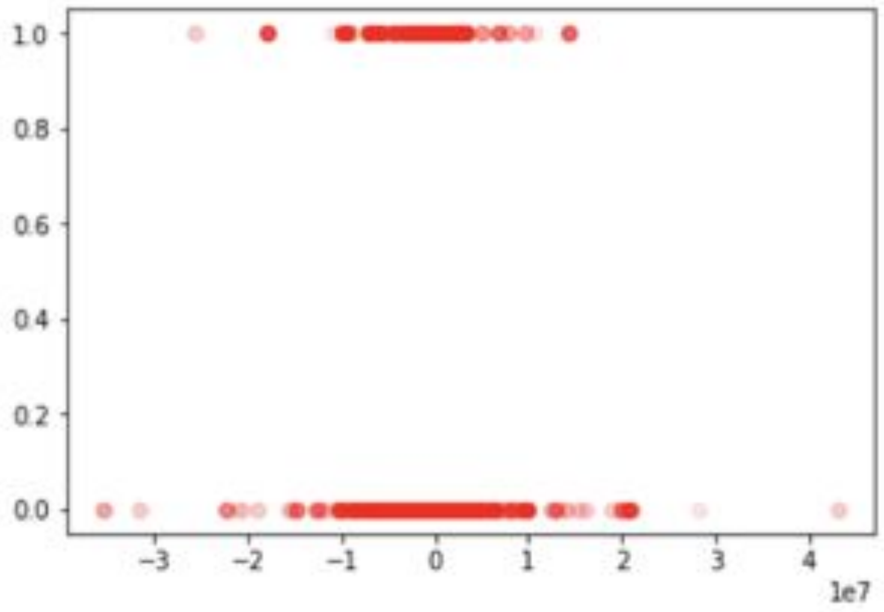
account balance

Intuition:

- People with a negative balance are unlikely to purchase a wealth product.

Change in Account Balance

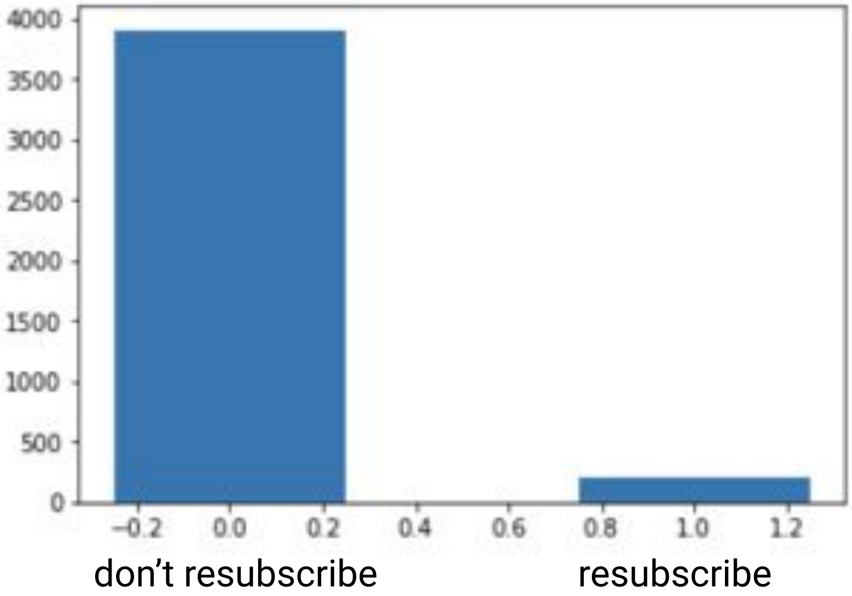
whether they purchased (0/1)



Intuition:

- People with a relatively constant balance are the likeliest to make wealth purchases.

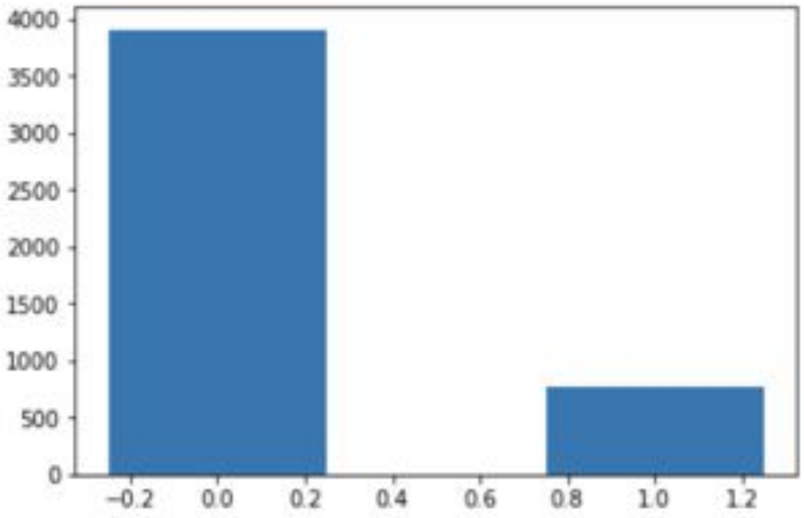
TMD Expiration



Intuition:

- Of the 4,000 customers whose TMD expired, only about 5% purchase a wealth product within the next three months.
- **Not useful**

TMD Expiration – how about next year?



don’t resubscribe resubscribe

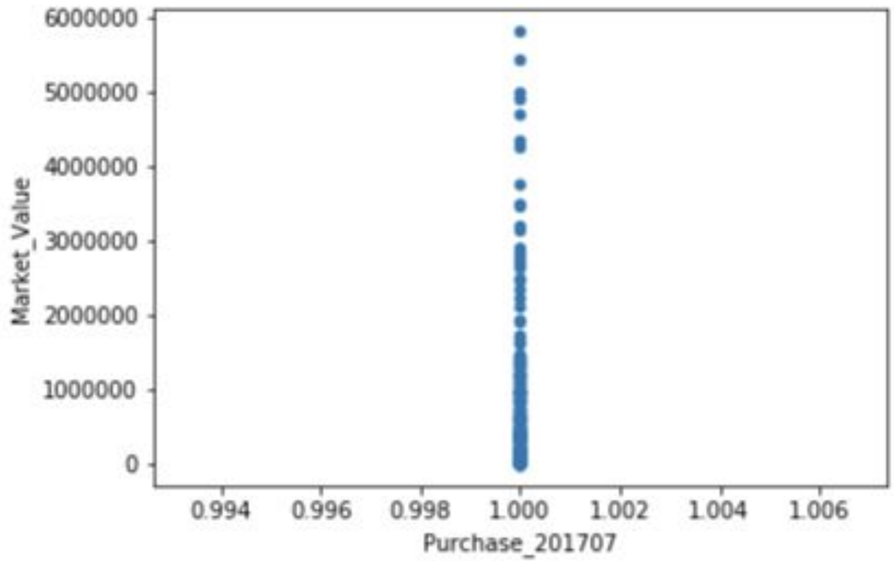
Intuition:

- Of the 4,000 customers whose TMD expired, only about 20% purchase a wealth product within the next year.
- **Not useful also**

More insights in resubscription

Why do only 20% of customers reinvest in HSBC?

market value of previous insurance

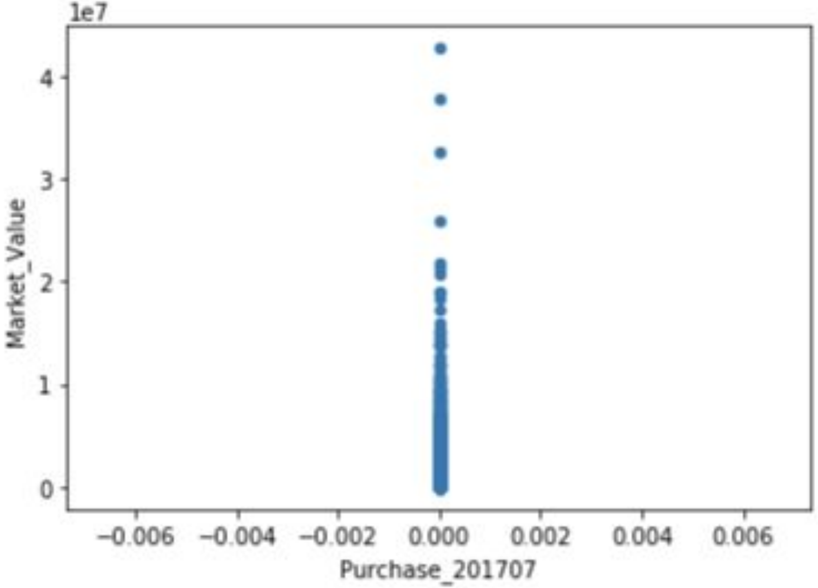


Market Value of Customer's Insurance & Investment a month prior to resubscription.

More insights in resubscription

Why do only 20% of customers reinvest in HSBC?

market value of previous insurance



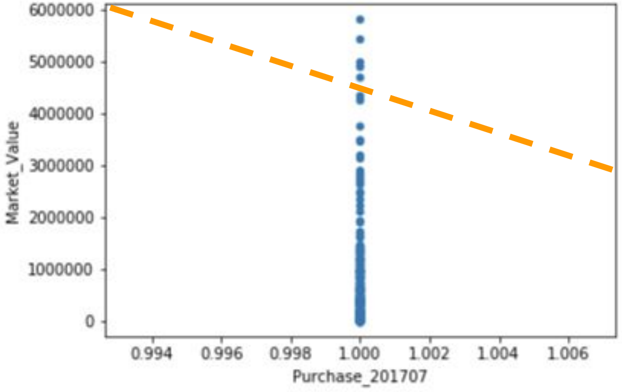
Market Value of Customer's Insurance & Investment a month prior to not resubscribing.

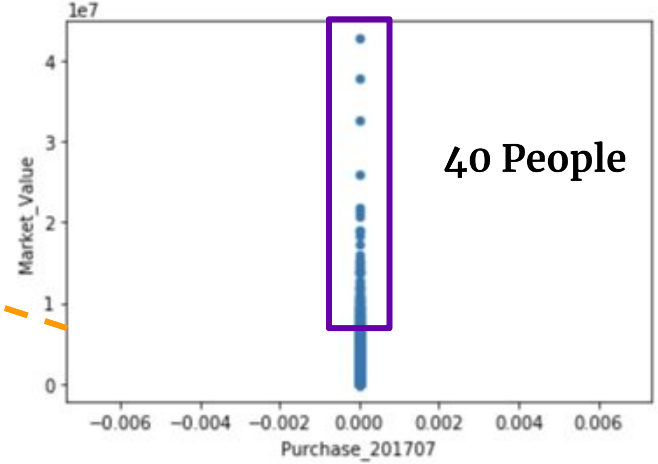
More insights in resubscription

Why do only 20% of customers reinvest in HSBC?

resubscribe

don't resubscribe





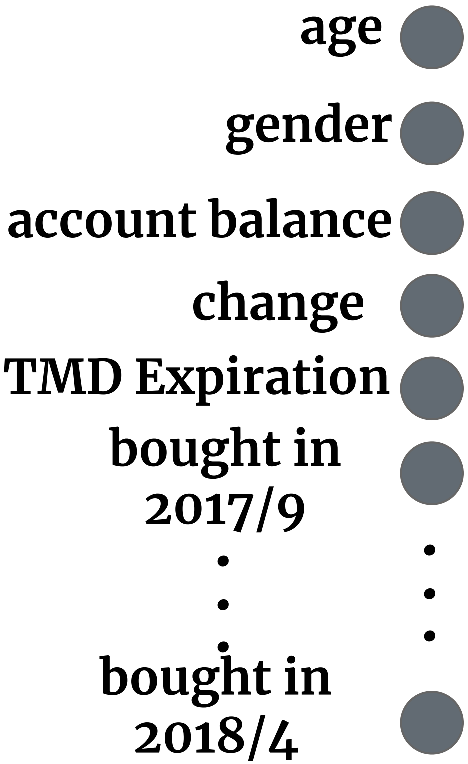
So now we have gained some useful intuitions in pairwise features, it’s time to train a neural network……

Before that, we are taking into account the data with the strongest correlation: previous purchases

The best measure we could find of whether someone is going to purchase a wealth product is if they purchased one in the past. Leveraging this data we can probably get a far better prediction.

500999801104896	0	0	0	0	0	1	0	0	1	0	0	0
500999800909269	0	0	0	0	0	0	1	0	0	0	0	0
500999801050439	0	0	0	0	0	0	0	1	0	0	0	0
500999800637274	0	0	0	0	0	0	0	0	1	0	0	0
500999800860166	0	0	0	1	0	0	1	0	0	0	0	0
500999800159173	0	0	1	0	0	0	0	1	0	0	0	0
500999800519798	0	0	0	0	0	0	0	0	0	0	0	0
500999800733232	0	0	0	1	0	1	0	0	0	1	0	0
500999800744795	0	0	0	1	0	0	0	0	0	0	0	0
500999800765170	0	0	1	0	0	0	0	0	1	0	0	0
500999800438290	0	0	0	0	0	1	0	0	0	0	0	0
500999801028232	0	0	1	0	0	0	0	0	0	0	0	0
500999800455348	0	0	0	0	0	0	0	0	0	0	0	0
500999801101078	0	0	1	1	1	0	0	0	1	0	0	0
500999800913964	0	0	0	0	0	0	0	0	0	0	0	0
500999800639073	0	0	1	1	1	0	1	0	0	1	0	0
500999800998986	0	0	0	1	0	0	0	0	0	0	0	0
500999800969192	0	0	0	0	0	1	0	0	0	0	0	0
500999800740393	0	0	0	0	0	0	1	0	0	0	0	0
500999800909089	0	0	0	0	0	0	1	0	0	0	0	0

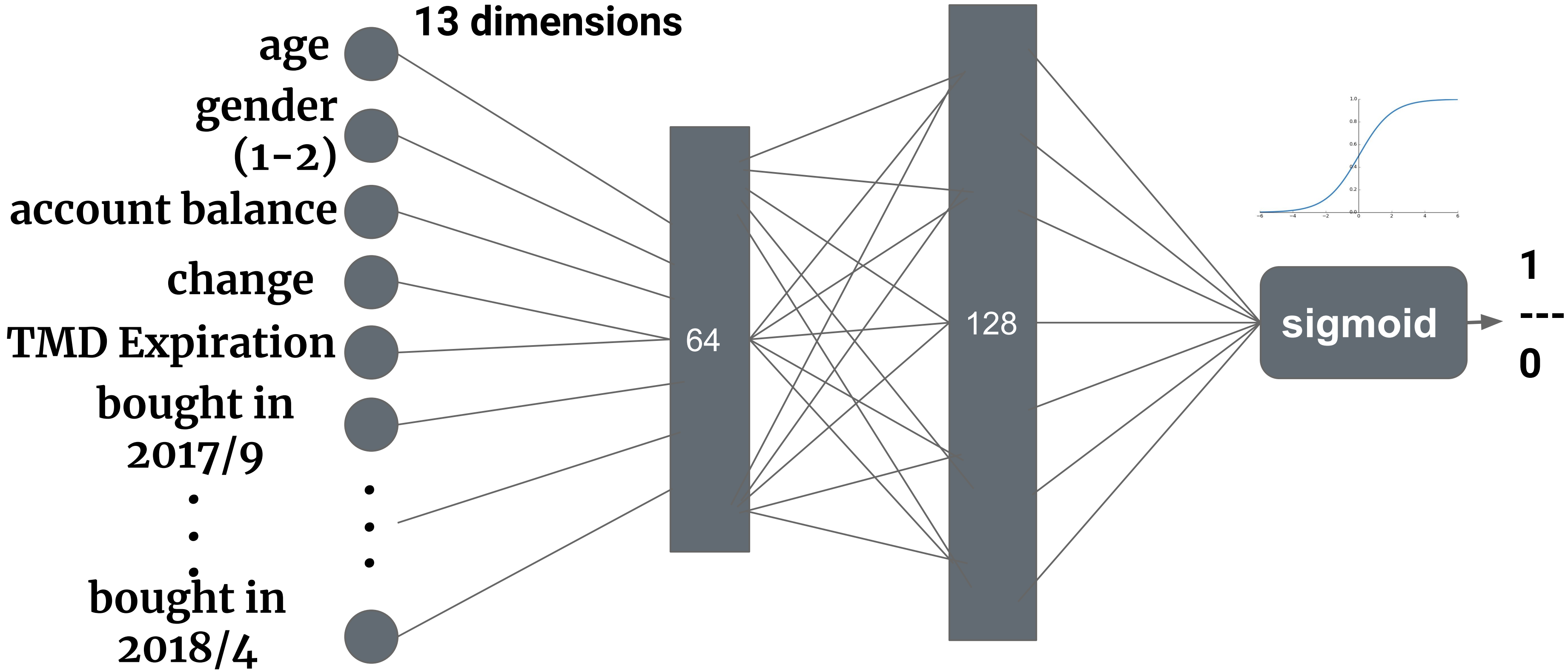
Finally, a neural network



Finally, a neural network

age	45
gender	1
account balance	3000000
change	-200000
TMD Expiration	1
bought in 2017/9	1
⋮	0
	0
	1
	0
	1
bought in 2018/4	0

Finally, a neural network



Training

We provide the randomly rescaled training set to avoid problem of unbalanced data.

Training

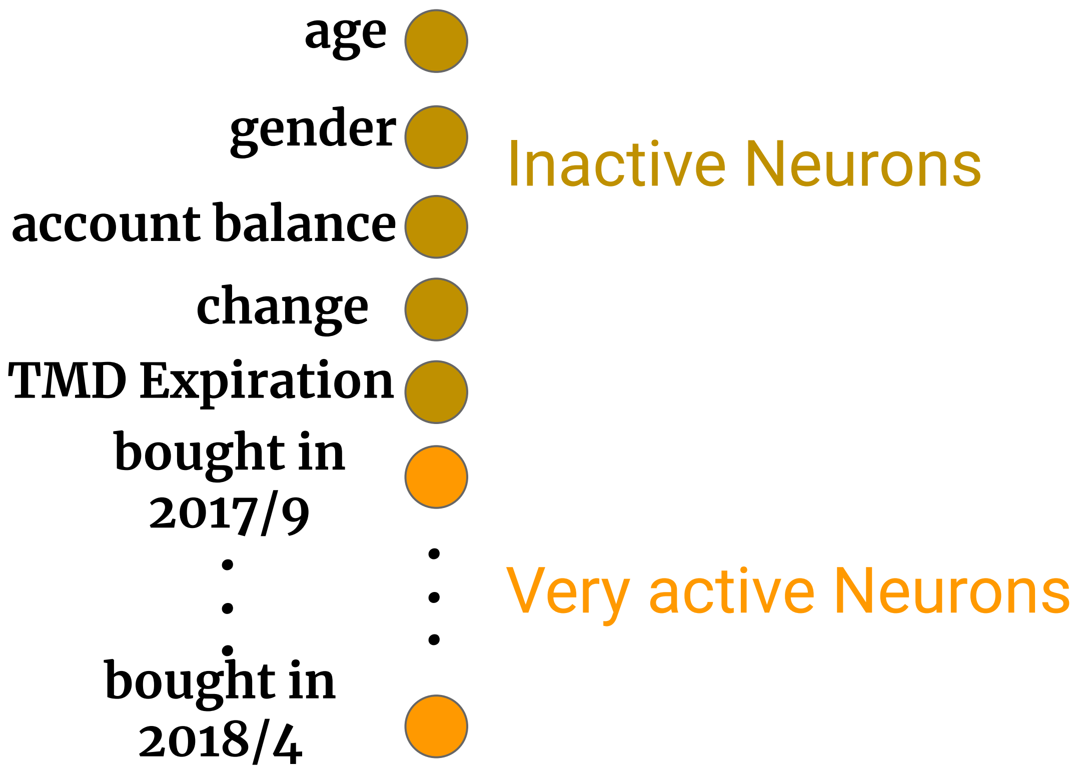
We provide the randomly rescaled training set to avoid problem of unbalanced data. The best case validation accuracy is 56%.

Test error

The model will provide around 0.6% of predictions of 1.

In our testset, 473 are accurately predicted with the correct amount to be 2769

What the problem might be?



Potential problems:

- unbalanced dataset
- Inactive neurons (very small weight)
- The program cannot meet a satisfying training error

Future Improvements

- As we do find some very useful pairwise connection with people who purchased and not, we believe a model with promising accuracy is highly possible
- More dimensions of data
- A bigger network
  - or possibly, a better network

Final Prediction

- Our final prediction is based on the previous neural network which generates around 2000 customers
- We change the threshold a bit lower, which enlarges our final prediction to around 3000.
- Finally, we use pairwise connections to eliminate around 300 customers (eliminate people who have had high market values in wealth purchases)

Thank You!