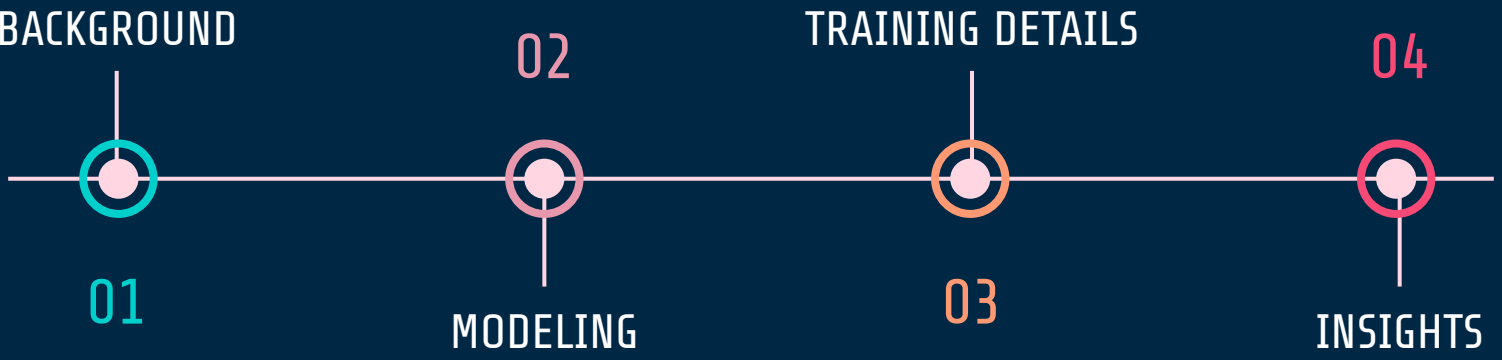
Tuning Deep Neural Networks on real-life Business Data

AGENDA



BACKGROUND

- DATA:

  Online Retail Industry: One line in the data represents a product selling on the company's e-commerce website (5000 rows)

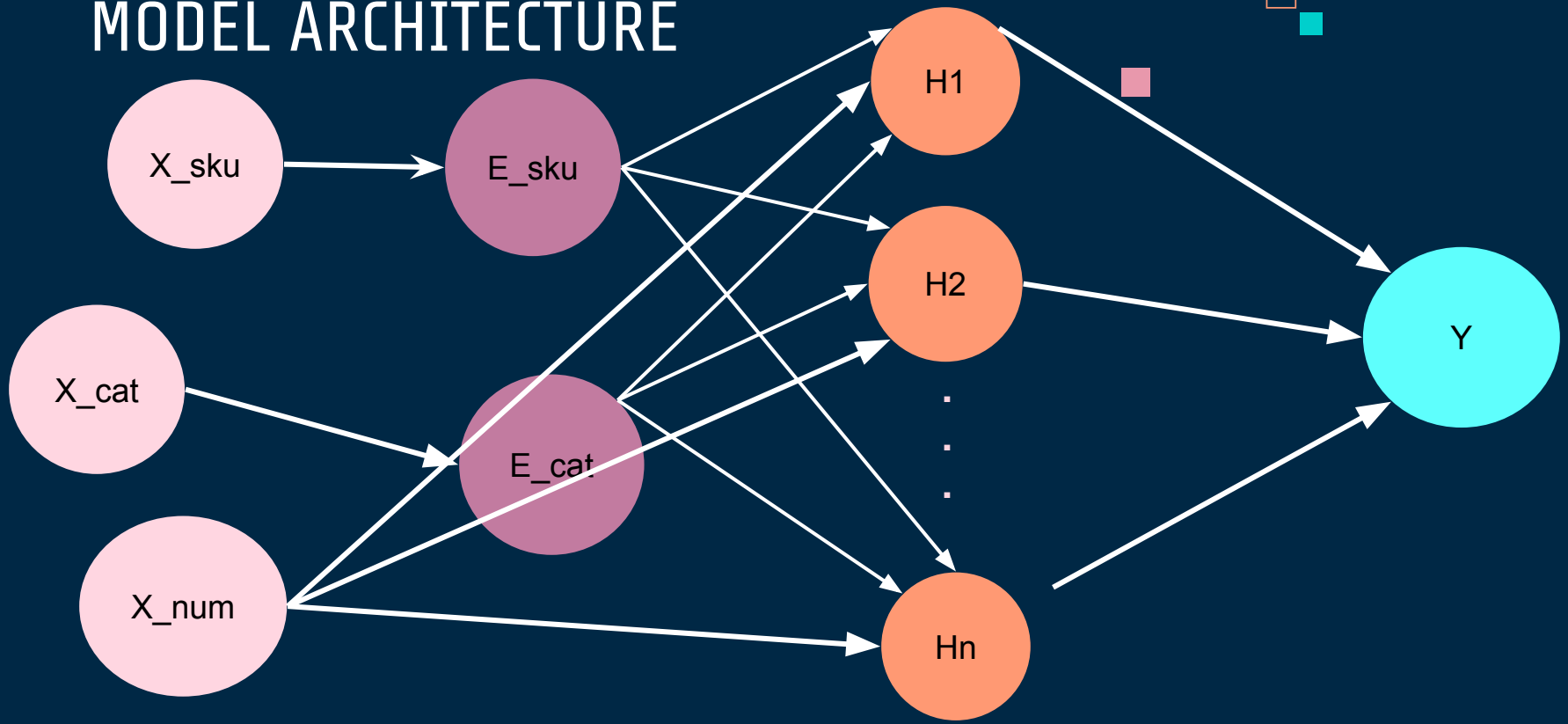
- GOAL:

  Predict quantity sold of a given product as accurately as possible by tuning the learning procedure

- TEAMWORK:



MODEL ARCHITECTURE



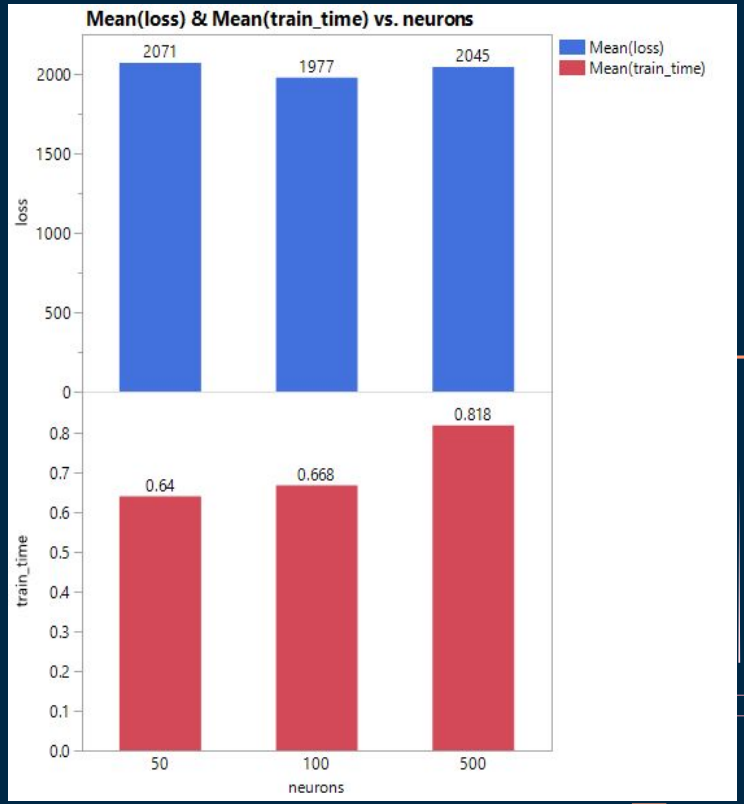
Combinations

Neurons	50, 100, 500
Activation	sigmoid, tanh, relu
Optimizer	SGD, RMSprop, Adam
Learning Rate	0.01, 0.05, 0.1
Batch Size	50, 100, 200
Epochs	5, 10, 20

Insights--Neurons

5000 dataset

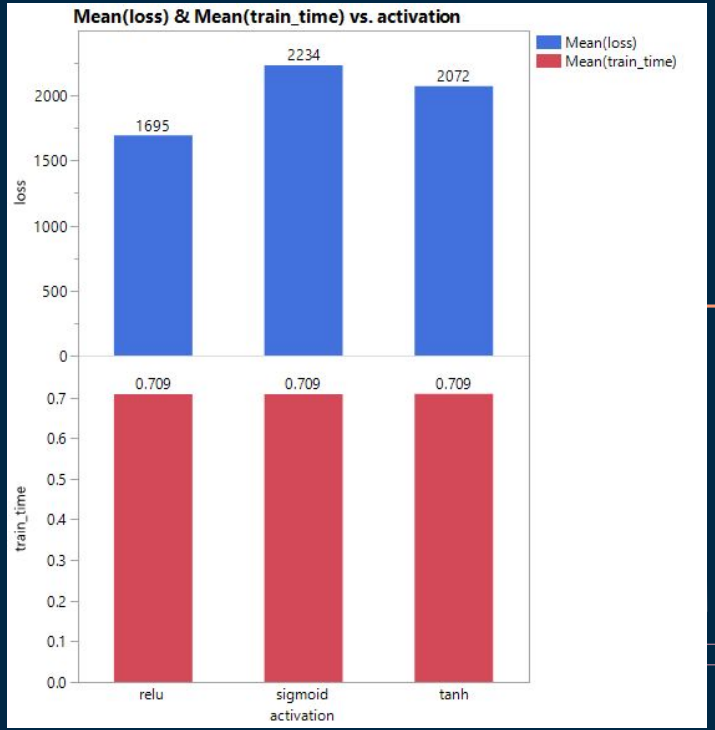
- 100 neurons has the smallest loss
- The training time increase with the increase of neuron



Insights--Activation function

5000 dataset

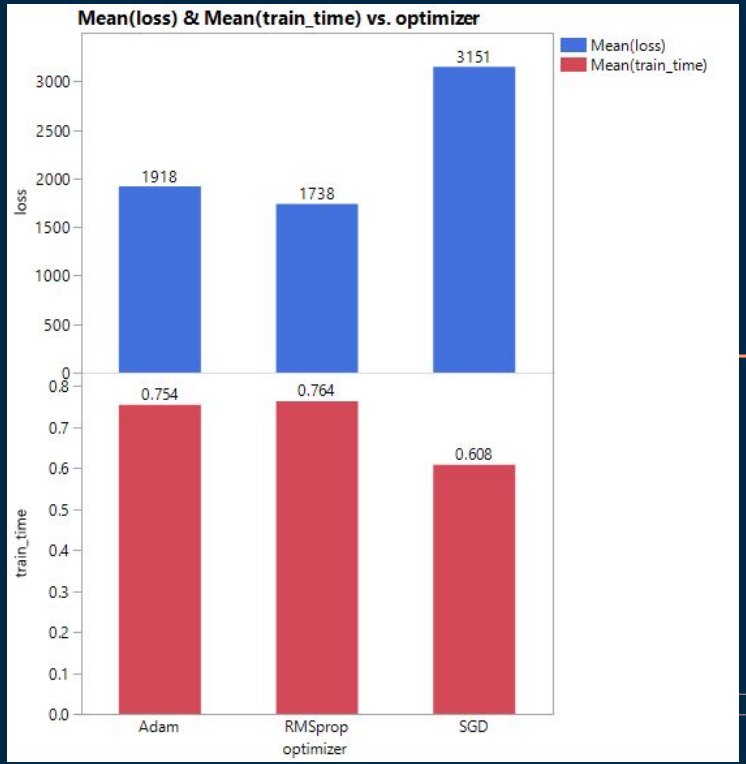
- Relu has the smallest loss
- 3 activation functions has the same training time



Insights--Optimizer

5000 dataset

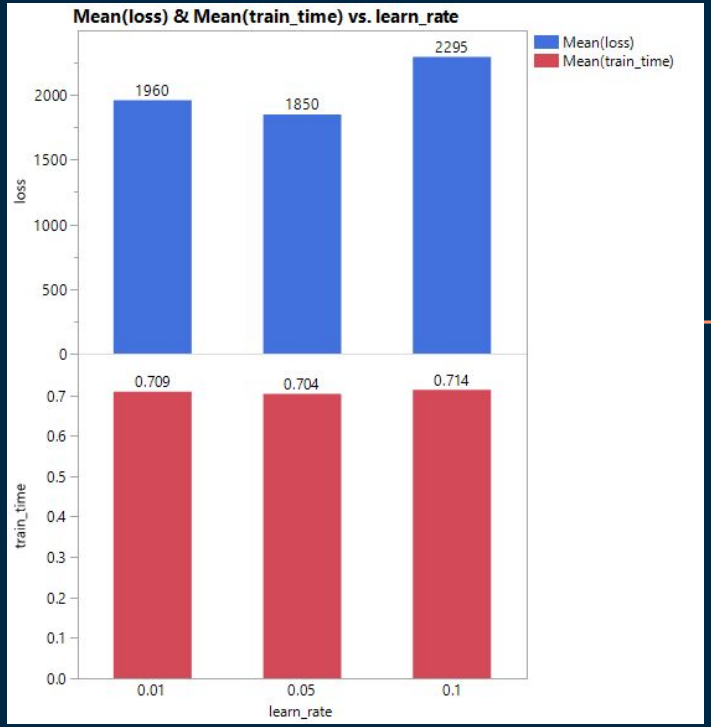
- RMSprop has the smallest loss
- SGD has the shortest training time



Insights-Learning Rate

5000 dataset

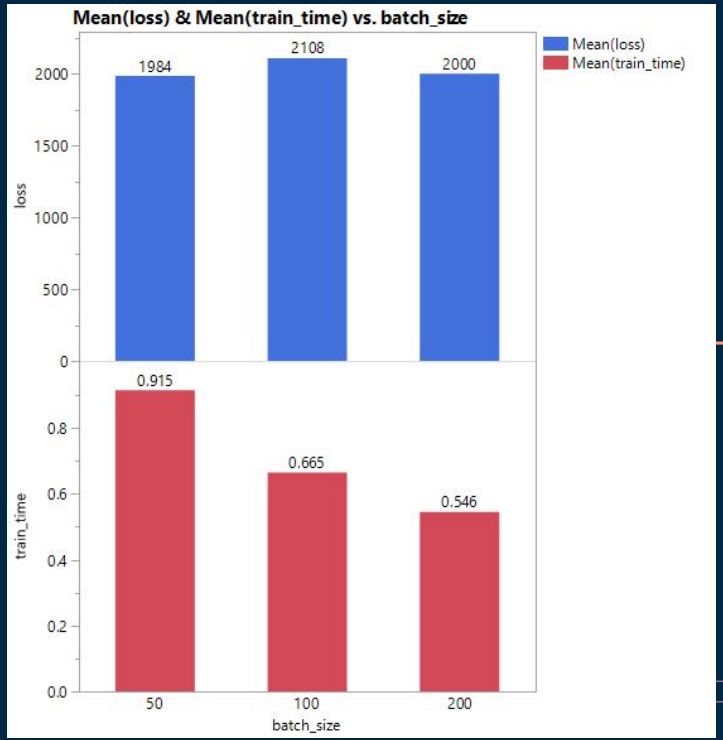
- 0.05 learning rate has the smallest loss and train time



Insights--Batch Size

5000 dataset

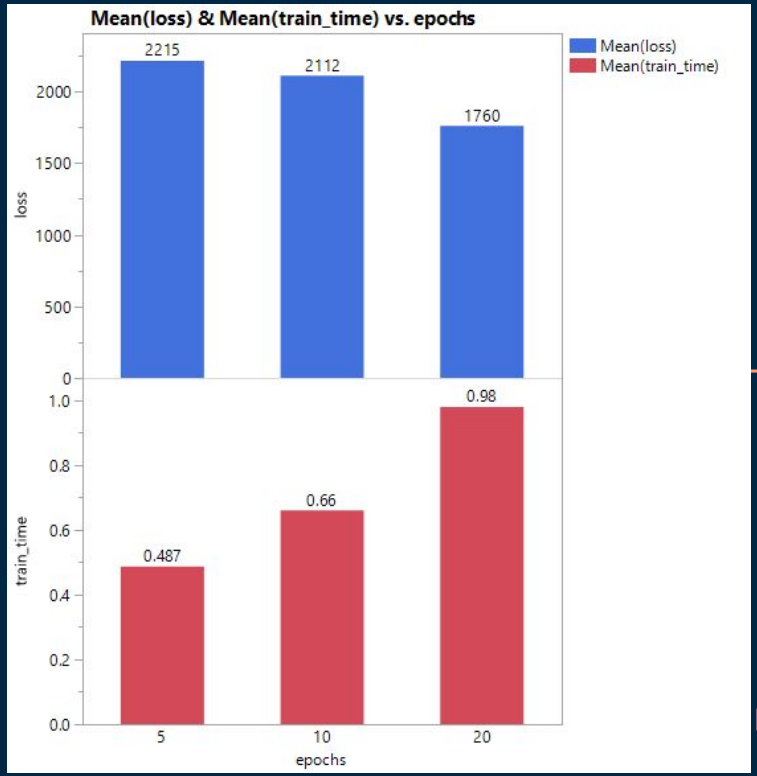
- 50 batch size has the smallest loss
- The larger the batch size the shorter the training time



Insights--Epochs

5000 dataset

- The more epochs the smaller the loss
- More epochs will significantly increase the training time.



Summary

Sensitivity:

- Training Time — Epochs, Batch Size
- Loss — Optimizer, Activation

Optimal Parameters Combination:

- Neurons = 100
- Activation function = Relu learning
- Optimizer = Adam
- Learning rate = 0.05
- Batch size = 200
- Epochs = 20

Questions?

Appendix

20 combination with lowest loss

	neurons	activation	optimizer	learn_rate	batch_size	epochs	loss	train_time
1	100	relu	Adam	0.05	200	20	1095.890503	0.643144846
2	500	relu	Adam	0.05	200	20	1103.684082	0.837501287
3	50	relu	Adam	0.05	200	20	1129.258667	0.688154697
4	100	relu	Adam	0.01	50	20	1151.962402	1.272286177
5	500	relu	Adam	0.01	200	20	1154.644409	0.824185133
6	500	relu	Adam	0.01	100	20	1155.107056	1.077242136
7	500	relu	Adam	0.01	50	20	1168.930908	1.684378386
8	500	tanh	Adam	0.01	200	20	1171.074829	0.828186035
9	100	relu	Adam	0.01	200	20	1198.17041	0.653147697
10	500	tanh	Adam	0.01	100	20	1200.280273	1.191267967
11	500	sigmoid	Adam	0.01	100	20	1201.819702	1.082243919
12	50	relu	Adam	0.01	50	20	1204.539429	1.209270954
13	100	relu	Adam	0.1	200	20	1208.450195	0.658148766
14	500	sigmoid	Adam	0.01	200	20	1209.071411	0.823184967
15	100	relu	Adam	0.01	100	20	1212.878296	0.85419178
16	500	relu	Adam	0.05	50	20	1214.769165	1.868419647
17	50	relu	Adam	0.01	100	20	1220.747437	0.819183826
18	50	relu	Adam	0.05	100	20	1220.971436	0.827187061
19	100	relu	Adam	0.05	50	20	1227.879883	1.257283211

Appendix

Sort it with training time

	neurons	activation	optimizer	learn_rate	batch_size	epochs	loss	train_time
1	100	relu	Adam	0.05	200	20	1095.890503	0.643144846
2	100	relu	Adam	0.01	200	20	1198.17041	0.653147697
3	100	relu	Adam	0.1	200	20	1208.450195	0.658148766
4	50	relu	Adam	0.05	200	20	1129.258667	0.688154697
5	50	relu	Adam	0.01	100	20	1220.747437	0.819183826
6	500	sigmoid	Adam	0.01	200	20	1209.071411	0.823184967
7	500	relu	Adam	0.01	200	20	1154.644409	0.824185133
8	50	relu	Adam	0.05	100	20	1220.971436	0.827187061
9	500	tanh	Adam	0.01	200	20	1171.074829	0.828186035
10	500	relu	Adam	0.05	200	20	1103.684082	0.837501287
11	100	relu	Adam	0.01	100	20	1212.878296	0.85419178
12	500	relu	Adam	0.01	50	10	1229.105835	1.065240622
13	500	relu	Adam	0.01	100	20	1155.107056	1.077242136
14	500	sigmoid	Adam	0.01	100	20	1201.819702	1.082243919
15	500	tanh	Adam	0.01	100	20	1200.280273	1.191267967
16	50	relu	Adam	0.01	50	20	1204.539429	1.209270954
17	100	relu	Adam	0.05	50	20	1227.879883	1.257283211
18	100	relu	Adam	0.01	50	20	1151.962402	1.272286177
19	500	relu	Adam	0.01	50	20	1168.930908	1.684378386
20	500	relu	Adam	0.05	50	20	1214.769165	1.868419647