Summary

The final model is selected by selecting the xgboost model hyperparameters and the utilized features.

Main questions:

1) Should x, y and accuracy be included? **Maybe – one submission each**
2) Model hyperparameters **Relatively modest trees**
3) What is the constant for the top M selection prediction? **Try both very low and relatively moderate settings in final submission**
4) Is better accuracy achieved by dropping some prediction columns? **Not obviously, seems to alywas decrease after pruning => use all base features**

1) Should x, y and accuracy be included?

Experiment 1: leaderboard with and without x, y and accuracy (submissions 22 and 23)

Random test batches 1-27

- With x, y and accuracy achieves better results MAP@3 (5.857>5.855) – extrapolated diff of 0.02%

Experiment 2: Local validation batches 21-30

- With x, y and accuracy achieves similar results MAP@3 (60.48 ~= 60.47)

2) Model hyperparameters Stage 1

Phase one July 4, 2016

100 rounds for train batches 131-135 and row sampling of 0.5; no x, y or accuracy

Validated on batches 21:30

Hyperset	ETA constant	Colsample	Tree depth	MAP@3
1	12	0.6	8	60.39
2	8	0.6	8	60.46
3	12	1	8	60.36
4	8	1	8	60.46
5	8	0.6	6	60.45
6	8	0.6	10	60.4

Conclusions

Importance of ETA > Importance of colsample

- **ETAC at 8 for now**
- **Colsample at 0.6 for now (lower is faster for model building)**
- **Treedepth at 7 for now**

More rounds in final models!

More checks needed for finetuning of hyperparameters after feature selection (Stage 2 hyperparameter selection)

3) What is the constant for the top M selection prediction?

200 rounds, etaC at 8, colsample at 0.6; depth of 7

Train batches second level learner: 131:140

Test batches 1-27

Top M constant	MAP@3 Public LB
1e-6	5.857
1	5.857
0.5	5.857

The top M constant is proportional to the number of train batches

Oh yes, I checked, the submissions **ARE** different!

In final submission:

- Keep at 1e-6
- Try 1/10 of #train batches

4) Feature selection

200 rounds, etaC at 8, colsample at 0.6; depth of 7

Train batches second level learner: 131:140

Validation batches 21:30

# Features	MAP@3
103	60.50
75	60.48
98	60.48
121 (3+18 covar)	60.52

- Use all **121** features (if they don't help they probably don't hurt)

Accuracy has rank 64, x has rank 99 and y has rank 101

5) Model hyperparameters Stage 2

Phase two July 5, 2016

Train batches 131-135 and row sampling of 0.5; 121 features; Colsample of 0.6 and depth of 7.

Validated on batches 21:30

Hyperset	ETA constant	# rounds	MAP@3
1	8	100	60.49
2	6	100	60.5
3	10	100	60.48
4	4	100	60.47
5	8	200	60.52

6	8	400	60.52

Conclusions

- **ETA not too important => pick a relatively high setting (8)**
- **More rounds do not lead to drastic improvements => keep them modest at 200**

Final two models:

Train batches 131:160

121 features

200 rounds

etaC of 8

rowSample of 0.5

colsample of 0.6

depth of 7

top M prediction constants of 1e-6 and 1/10*nbTrainBatches. The 1/10 comes from a manual estimation of the optimal setting given limited time before the deadline. Manual estimation proportional to the base model counts and the average model validation performance.