

dataset

170k recipes

25k user

700k interactions

Additional informations for recipes and users

Cosine Signature Approximation

-	Index 1	Index 2	Index 3
Plane 1	-1	1	1
Plane 2	-1	-1	1

-	User 1	User 2	User 3
Recipe 1	1	3	5
Recipe 2	2	4	4
Recipe 3	1	5	5
Recipe 4	3	2	3
Recipe 5	1	4	4

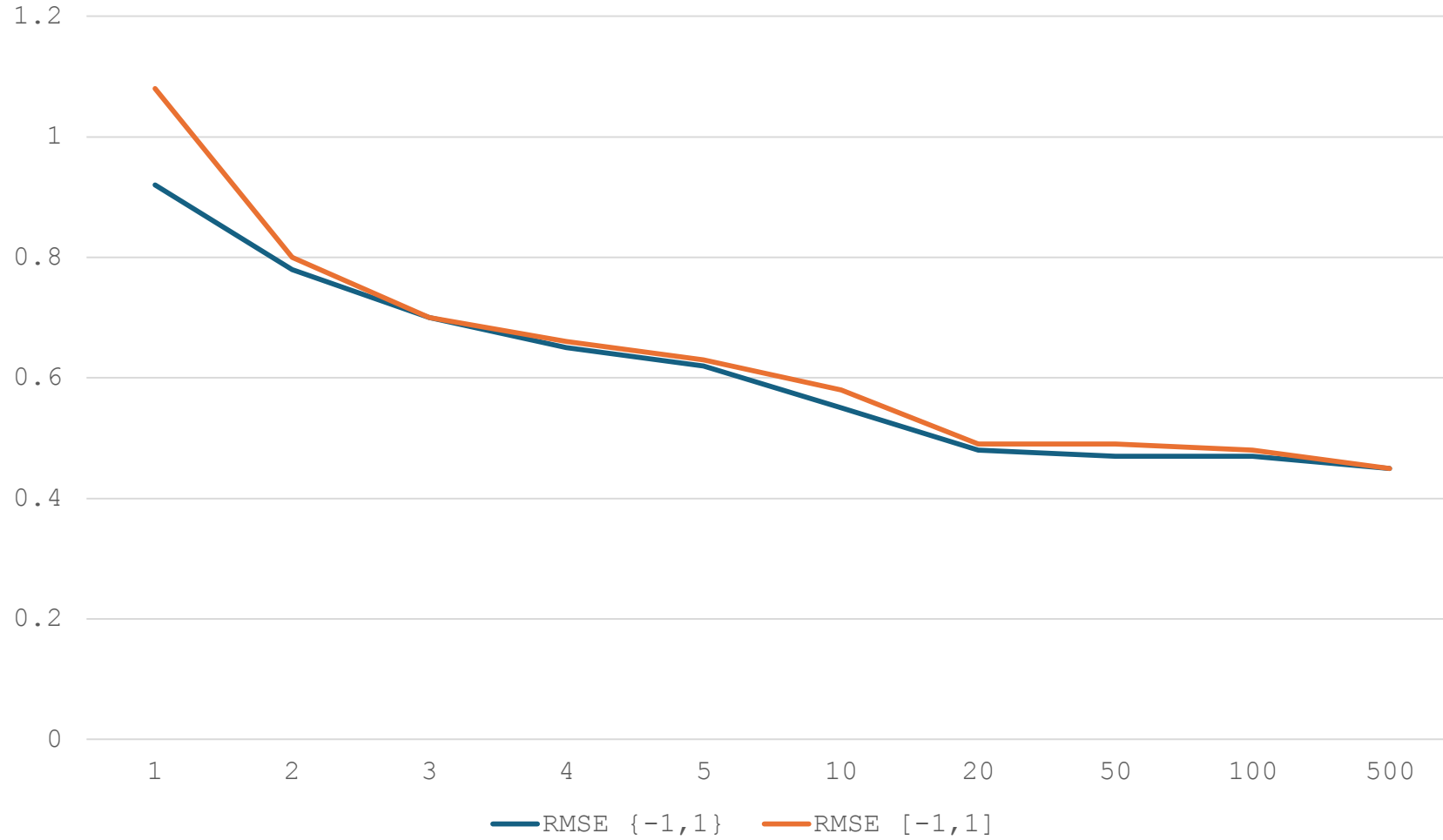
`sign(recipe * plane)`



-	Plane 1	Plane 2
Recipe 1	1	1
Recipe 2	1	-1
Recipe 3	1	-1
Recipe 4	1	-1
Recipe 5	1	-1

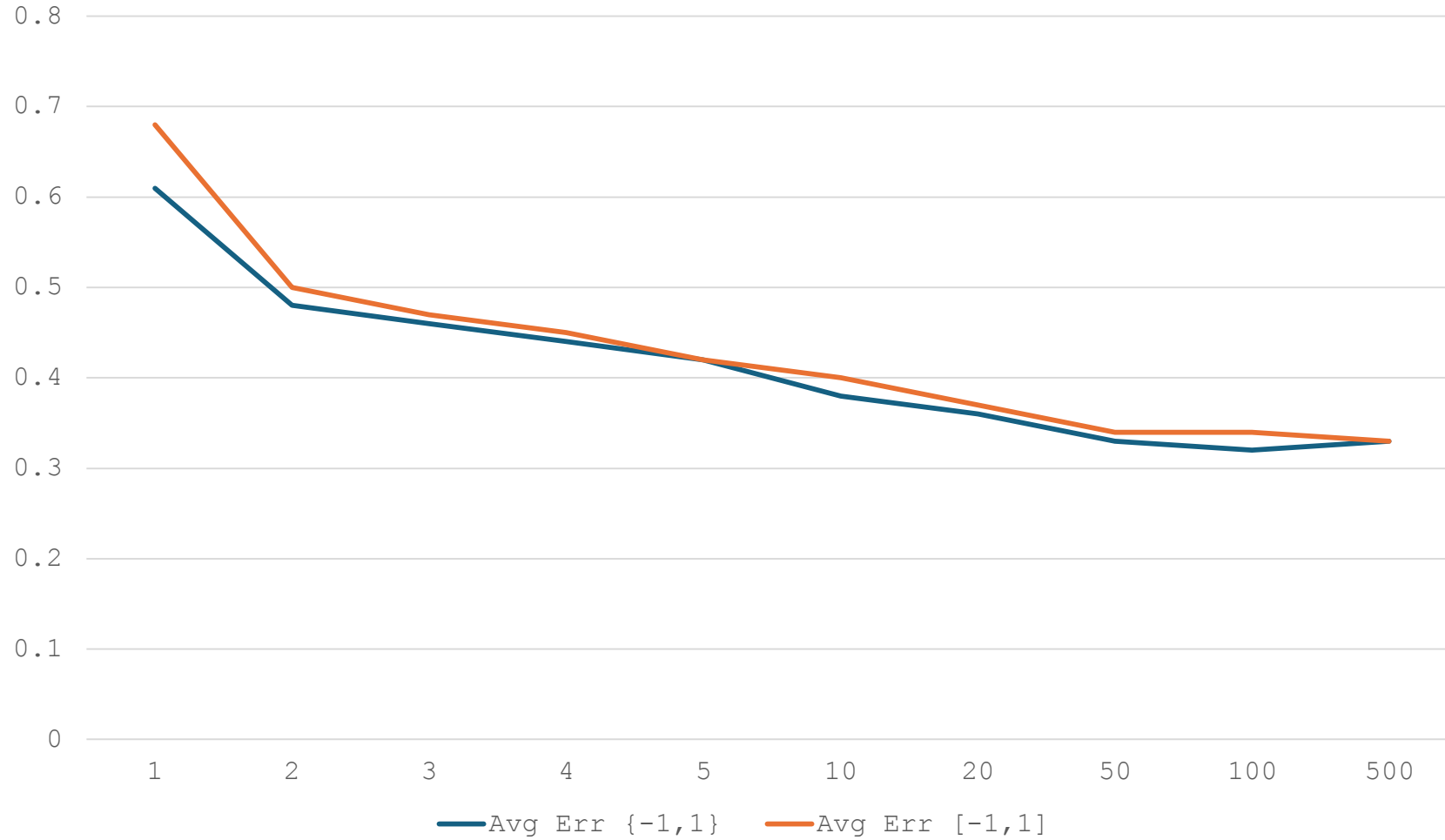
Approximation Evaluation

RMSE between real cosine dist and signature
matrix cosine dist for 10k samples



Approximation Evaluation

Average Error between real cosine dist and
signature matrix cosine dist for 10k samples

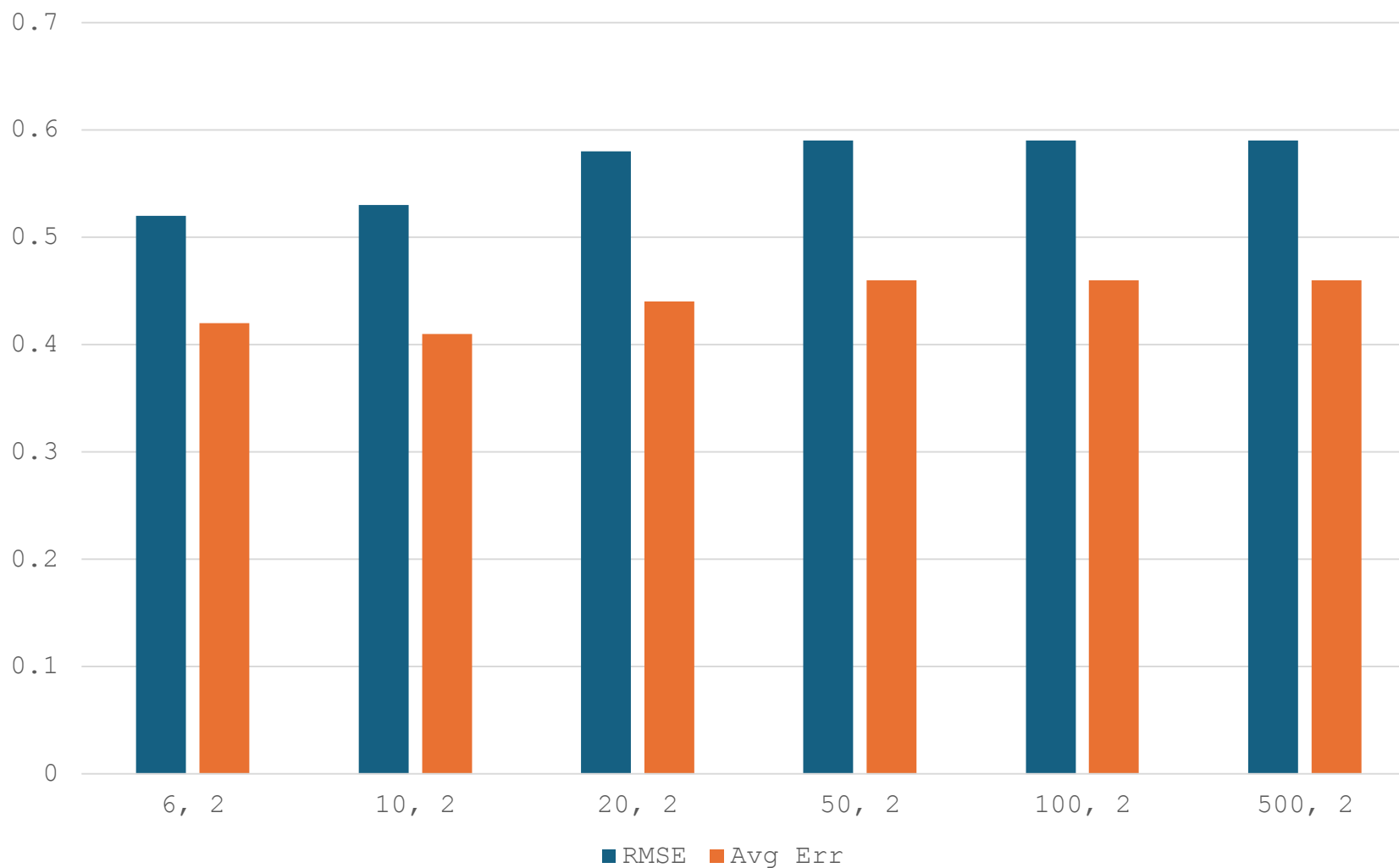


CF Algorithm

	1	2	3	4	5	6	7	8	9	10	11	12	sim(1,m)
1	1		3		?	5			5		4		1.00
2			5	4			4			2	1	3	-0.18
<u>3</u>	2	4		1	2		3		4	3	5		<u>0.41</u>
4		2	4		5			4			2		-0.10
5			4	3	4	2					2	5	-0.31
<u>6</u>	1		3		3			2			4		<u>0.59</u>

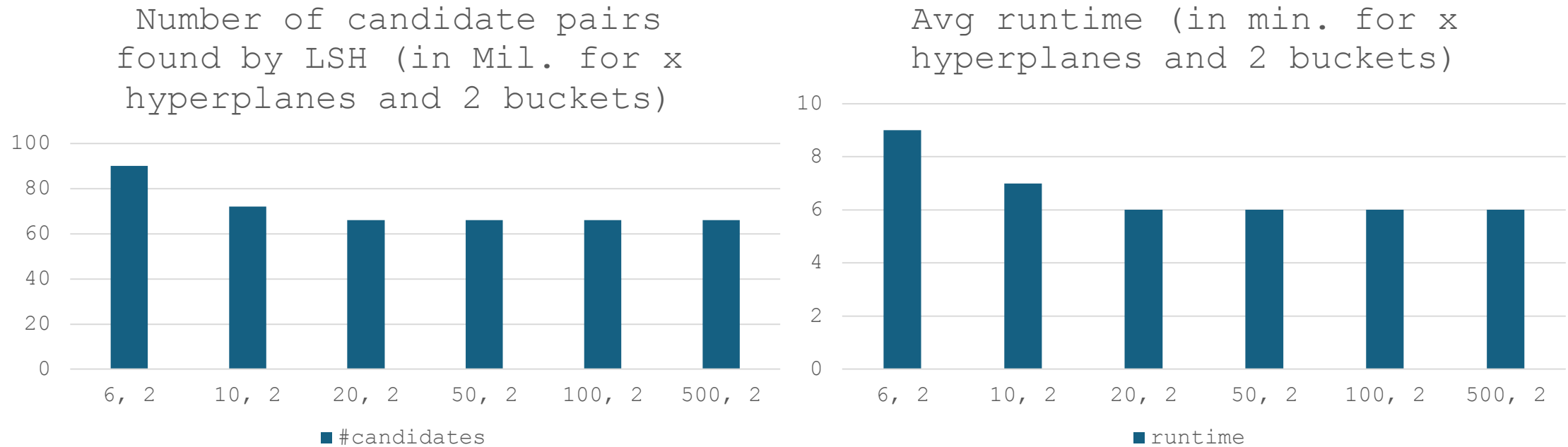
CF Evaluation

RMSE, Avg Err for x hyperplanes, 2 buckets



Goes up because of less samples
2k samples

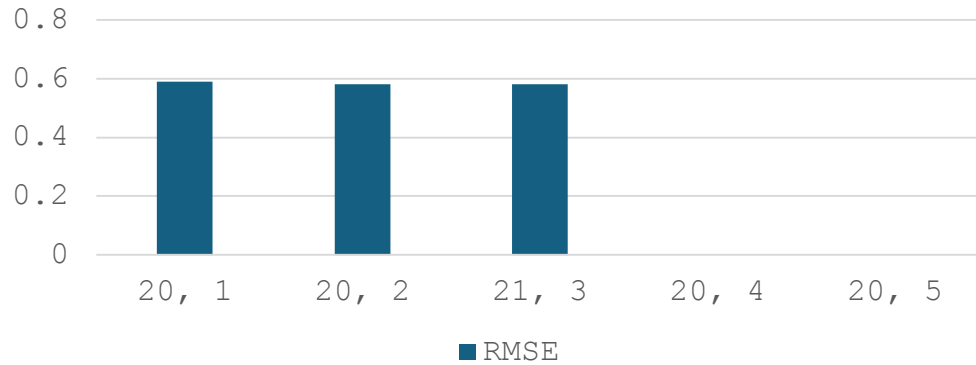
CF Evaluation



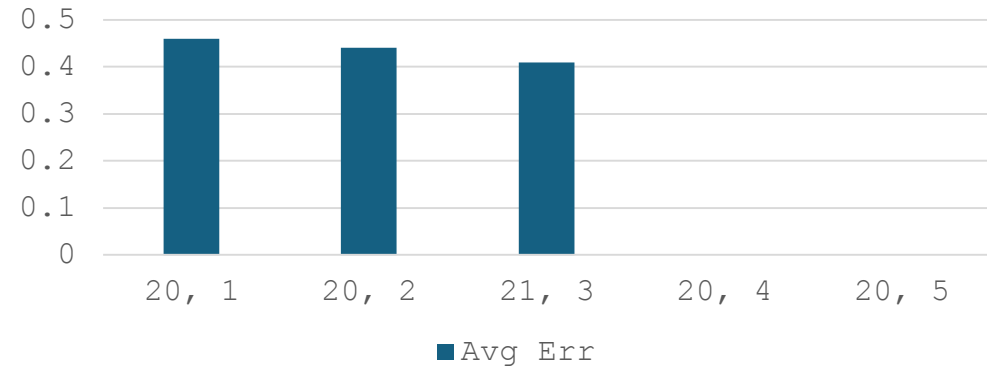
170k recipes => 28 Billion pairs, LSH filters down to 66 Million
Flattens out at 20 because the distance approximation also flattens out

CF Evaluation

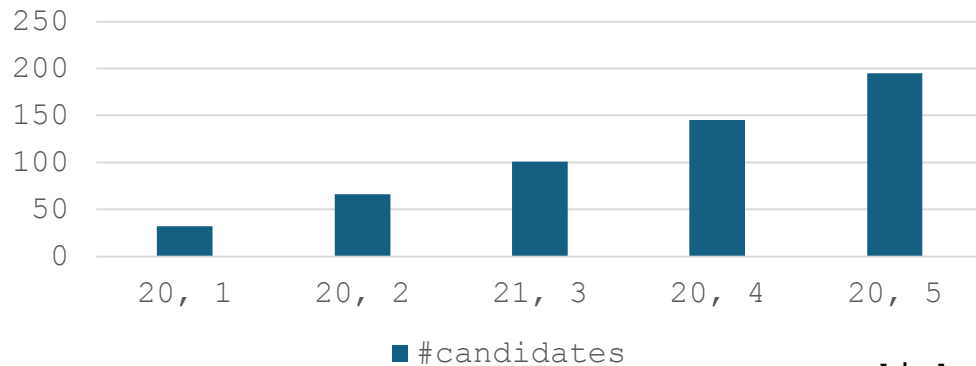
RMSE for 20 hyperplanes, x buckets



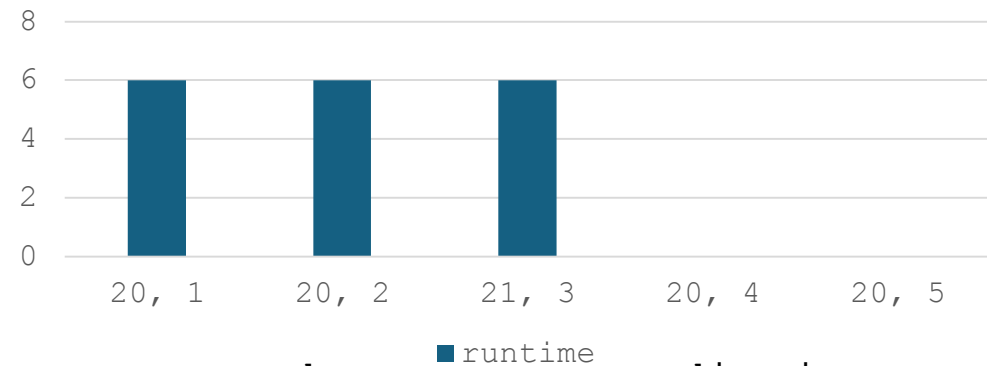
Avg Err for 20 hyperplanes, x buckets



#candidates for 20 hyperplanes, x buckets



Runtime for 20 hyperplanes, x buckets



Too many candidates = program aborts, no prediction
21, 3 because we want to have clean division
More candidates, same time, idk what happens here

Evaluation

- Rmse, avg err so low because most users have low variance in the rating
- More planes => less candidates => worse approximation
- centralized rmse/me, 20,1: 0.59 / 0.45
- Uncentralized rmse/me, 20,1: 0.054 / 0.61

Improved CF: Item profile from da

Just for the memes, predict rating only based on the date the recipe was published. we use the hyperplane approach, but this does not matter since all dates are $\gg 0$ and therefore have the same signature. So this is basically a compare all approach.

short: prediction = summary of rat
that were uploaded in the same wee

rating	date
5	27/03/2019
3	01/01/2011

RMSE: 1.4

ME: 0.9

Runtime: FASSSSSST (12 sec)

Improved CF: Item profile from da

Prediction based in nutrition

RMSE: 0.77

ME: 0.62

Runtime: FASSSSSST (15 sec)

rating	Calorie	Fat	Suggar	Sodium	Protein	Sat. fat	idk
4	269	22	32	48	39	27	5
3	315	0	202	9	6	0	21