Project Assessment

To assess your project work, you should be able to answer the following questions:

1. For query classification:
    1. How many unique categories did you see in your rolled up training data when you set the minimum number of queries per category to 100? To 1000?
        i. Threshold 100: 720 categories
        ii. Threshold 1000: 163 categories
    2. What values did you achieve for P@1, R@3, and R@5? You should have tried at least a few different models, varying the minimum number of queries per category as well as trying different fastText parameters or query normalization. Report at least 3 of your runs.

Round	Params	Results

1	`Min query / category: 100`
	
	`Fasttext:`
	`-epoch 25 \`
	`-lr .1 \`
	`-loss hs`
		N 56365
		P@1 0.293
		R@1 0.293
		
		N 56365
		P@3 0.148
		R@3 0.444
		
		N 56365
		P@5 0.101
		R@5 0.507

2	`Min query / category: 1000`
	
	`Fasttext:`
	`-epoch 25 \`
	`-lr .1 \`
	`-loss hs`
		N 56366
		P@1 0.347
		R@1 0.347
		
		N 56366
		P@3 0.172
		R@3 0.517
		
		N 56366
		P@5 0.116
		R@5 0.582

3	`Min query / category: 1000`
	
	`Fasttext:`
	`-epoch 25 \`
		N 56366
		P@1 0.332
		R@1 0.332
		
		N 56366
		P@3 0.166
		R@3 0.499

	`-lr .05 \`
	`-loss hs \`
	`-wordNgrams 2`
		N 56366
		P@5 0.113
		R@5 0.563

2. For integrating query classification with search:
    1. Give 2 or 3 examples of queries where you saw a dramatic positive change in the results because of filtering. Make sure to include the classifier output for those queries.

flashdrive Analysis: Good first prediction, should probably limit the prediction to 1 or increase the score threshold; The search without query classification produces no results	[X] label: abcat0504010, score: 0.12476956844329834, name: USB Flash Drives [X] label: pcmcat242800050021, score: 0.11017191410064697, name: Health, Fitness & Sports [X] label: pcmcat248700050021, score: 0.10797383636236191, name: Home
ear piec Analysis: The classifier seems works well with misspelled words	[X] label: abcat0515038, score: 0.19110167026519775, name: Speakers & Headsets [X] label: pcmcat143000050011, score: 0.1645003855228424, name: Over-Ear & On-Ear Headphones [X] label: abcat0802000, score: 0.14901256561279297, name: Telephones & Communication
beat head phones Analysis: Returning much better results comparing to the default	[X] label: pcmcat143000050011, score: 0.3178328573703766, name: Over-Ear & On-Ear Headphones [X] label: pcmcat144700050004, score: 0.3021498918533325, name: All Headphones [X] label: pcmcat143000050007, score: 0.1499129831790924, name: Earbud Headphones
internal harddrive Analysis: The search without query classification produces no	[X] label: abcat0504001, score: 0.625261664390564, name: Hard Drives [X] label: pcmcat247400050000, score: 0.11751869320869446, name: PC Laptops

results	

2. Given 2 or 3 examples of queries where filtering hurt the results, either because the classifier was wrong or for some other reason. Again, include the classifier output for those queries.

Resident evil 4 HD Analysis: Wrong predictions, probably it has HD in it	[X] label: pcmcat247400050000, score: 0.27900516986846924, name: PC Laptops [X] label: abcat0101001, score: 0.115447998046875, name: All Flat-Panel TVs
i cook	[X] label: abcat0515028, score: 0.18950538337230682, name: Laptop Bags & Cases
guitar hero	[X] label: pcmcat245100050028, score: 0.13957132399082184, name: Office