

Questions Q&A 5 Collaborative Filtering, Cross-Selling, Upselling – Tuesday Nov 30th

Questions Week 5 – Collab Filt, Cross and UP

- [Konstantinos];
 - (1) W5.3 / Slide 28: Regarding, item-based approach, I cannot understand how we derive (by eyeballing) these values for Predator and Notting Hill. Shouldn't these guesses, based on the correlations, be reversed?
 - (2) W5.3/ Slide 32: How shall we split the dataset in training/validation/testing? You mention 80/20 for calibration/validation, this means something like 80/10/10 (training/validation/testing)? Cannot we use K-CV instead and training/testing split?
- [Noa];
 - (1) Clip 5.3 - around min 15-16:

Based on the item-based collaborative filtering we argue that Notting Hill is positively correlated to Pretty Woman and that Predator is positively correlated to Erin Brockovich and Total Recall.

However, when really making the prediction regarding Adam's ratings for Predator and Notting Hill you base your recommendations for Predator on Pretty Woman and for Notting Hill on Total Recall and Erin Brockovich. It should be the other way around, right? Instead of predicting Adam's rating to be 2 for Predator it should be around 4 or 5 (based on TR and EB) whereas for Notting Hill the rating should be around 2 (based on PW). Or am I mistaken?

Below you find the screenshot of the video

Item-based collaborative filtering

	Pretty Woman	Total Recall	Erin Brockovich	Terminator 2	Predator	Notting Hill
Adam	2	5	4	2	? 2	? 4.5
Ben	5	1	2		1	
Cindy	5	5	5	5	5	5
Dave	2	5		3		
Emily	5	3	5	3		5
Fred	1	5				1
George	2		5		5	

Notting Hill's ratings are positively correlated to those of Pretty Woman