Microsoft: DAT203x Data Science and Machine Learning Essentials

KEY POINTS

- Recommender systems are often based on a matrix of ratings that users give to items, as shown in the following image:

	Aliens	Bug's Life	Cars	Dark Knight
Carmen	5	4	1	1
Joseph	5	4	2	
Leonore	1		3	3
Esmerelda	5		1	

- User-based collaborative filtering predicts unknown ratings based on ratings of other users who gave similar ratings for other items. For example, in the following image, the Joseph's rating for *The Dark Knight* can be predicted based on the ratings given for that movie by other users who gave similar ratings to Joseph for other movies. In this case, Carmen gave similar ratings for *Aliens* and *A Bug's Life* to the ratings that Joseph gave for those movies, so it's reasonable to predict that Joseph will give *The Dark Knight* a similar rating to Carmen's.

	Aliens	Bug's Life	Cars	Dark Knight
Carmen	5	4	1	1
Joseph	5	4	2	?
Leonore	1		3	3
Esmerelda	5		1	

- Item-based collaborative filtering predicts unknown ratings based on ratings for other items that received similar ratings from other users. For example, in the following image, *Cars* and *The Dark Knight* have received similar ratings from each user that has rated both. It is therefore reasonable to predict that Joseph will give *The Dark Knight* a similar rating to the one he gave to *Cars*.

	Aliens	Bug's Life	Cars	Dark Knight
Carmen	5	4	1	1
Joseph	5	4	2	?
Leonore	1		3	3
Esmerelda	5		1	

- Matrix Factorization attempts to determine latent features for items and users, and correlate them to predict ratings. For example, features for "scariness" and "kiddiness" could be determined for movies (so *Aliens* is a scary movie, *Cars* is a kid's movie, and *The Dark Knight* has an even degree of "scariness" and "kiddiness"). Based on her known ratings, Esmerelda has a high "scariness" feature (she likes scary movies) and a low "kiddiness" feature (she dislikes kid's movies). By applying these derived latent features, a predicted rating that Esmerelda would give to *The Dark Knight* can be calculated by multiplying the "scariness" and "kiddiness" features for her and the movie, and adding the results together.

	Scary	Kiddy
Esmerelda	5	1

	Scary	Kiddy
Aliens	1	0
Cars	0	1
Dark Knight	½	½

	Aliens	Bug's Life	Cars	Dark Knight	
Carmen	5	4	1	1	
Joseph	5	4	2		
Leonore	1		3	3	
Esmerelda	5		1	?	(5 x ½) + (1 x ½) = 3