Recommender Systems - Solutions

Q1

Q. The user-item matrix below shows the purchasing history of 5 users with respect to 9 different books in a user-based CF system.
Who will be *U3*'s nearest neighbour in the data?
Calculate similarities with the Jaccard Index.

	Book1	Book2	Book3	Book4	Book5	Book6	Book7	Book8	Book9
U1		1		1		1	1		
U2			1		1			1	
U3	1	1		1			1		1
U4	1	1				1		1	
U5			1	1	1				

Jaccard Index

$$sim(p,q) = \frac{|B_p \cap B_q|}{|B_p \cup B_q|}$$

B_p = Books purchased by p
B_q = Books purchased by q

Q1

	Book1	Book2	Book3	Book4	Book5	Book6	Book7	Book8	Book9
U1		1		1		1	1		
U2			1		1			1	
U3	1	1		1			1		1
U4	1	1				1		1	
U5			1	1	1				

Target	Other	Intersection	Union	Jaccard
U3	U1	3	6	**0.50**
U3	U2	0	8	0.00
U3	U3	5	5	1.00
U3	U4	2	7	0.29
U3	U5	1	7	0.14

$$sim(p,q) = \frac{|B_p \cap B_q|}{|B_p \cup B_q|}$$

B_p = Books purchased by p

B_q = Books purchased by q

Based on past book purchases, U3's nearest neighbour is U1

Q2

Q. The user-item matrix below shows the purchasing history of 6 users for 10 different products in a user-based CF system.
Who will be *U1*'s nearest neighbour in the data?
Calculate similarities with the Jaccard Index.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
U1		1		1		1		1		1
U2	1		1				1			
U3				1				1		
U4		1		1						
U5				1		1		1		1
U6	1	1	1							

Q2

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
U1		1		1		1		1		1
U2	1		1				1			
U3				1				1		
U4		1		1						
U5				1		1		1		1
U6	1	1	1							

$$sim(p,q) = \frac{|B_p \cap B_q|}{|B_p \cup B_q|}$$

Target	Other	Intersection	Union	Jaccard
U1	U1	5	5	1.00
U1	U2	0	8	0.00
U1	U3	2	5	0.40
U1	U4	2	5	0.40
U1	U5	4	5	**0.80**
U1	U6	1	7	0.14

Based on past purchase history, U1's nearest neighbour is U5

Q3

- The table below was generated as part of the evaluation of a system to predict star ratings (1-5) for movies. The predicted and true ratings for 7 test examples are reported.
- Calculate the (a) MAE, (b) RMSE

Movie	True Rating (Stars)	Predicted Rating
Brooklyn	4	3.2
Toy Story 3	5	4.7
Batman v Superman	2	1.8
Angry Birds	1	3.2
The Shallows	2	2.0
Spectre	3	3.9
The Martian	4	4.1

Q3

$$MAE = \frac{1}{|Q|} \sum_{(u_f, i_g) \in Q} |r_{fg} - \hat{r}_{fg}|$$

$$RMSE = \sqrt{\frac{1}{|Q|} \sum_{(u_f, i_g) \in Q} (r_{fg} - \hat{r}_{fg})^2}$$

	True Rating (Stars)	Predicted Rating	Difference	Absolute Difference	Squared Difference
Brooklyn	**4**	**3.2**	0.8		
Toy Story 3	**5**	**4.7**	0.3		
Batman v Superman	**2**	**1.8**	0.2		
Angry Birds	**1**	**3.2**	-2.2		
The Shallows	**2**	**2.0**	0.0		
Spectre	**3**	**3.9**	-0.9		
The Martian	**4**	**4.1**	-0.1		
Sum					

Q3

$$MAE = \frac{1}{|Q|} \sum_{(u_f, i_g) \in Q} |r_{fg} - \hat{r}_{fg}|$$

$$RMSE = \sqrt{\frac{1}{|Q|} \sum_{(u_f, i_g) \in Q} (r_{fg} - \hat{r}_{fg})^2}$$

	True Rating (Stars)	Predicted Rating	Difference	Absolute Difference	Squared Difference
Brooklyn	**4**	**3.2**	0.8	0.8	0.64
Toy Story 3	**5**	**4.7**	0.3	0.3	0.09
Batman v Superman	**2**	**1.8**	0.2	0.2	0.04
Angry Birds	**1**	**3.2**	-2.2	2.2	4.84
The Shallows	**2**	**2.0**	0.0	0.0	0.00
Spectre	**3**	**3.9**	-0.9	0.9	0.81
The Martian	**4**	**4.1**	-0.1	0.1	0.01
Sum				4.5	6.43

$$MAE = \frac{4.5}{7} = 0.6429$$

$$RMSE = \sqrt{\frac{6.43}{7}} = 0.9584$$