# Recommendation Systems

**Question 1**: Here is a table of 1-5 star ratings for five movies (M, N, P. Q. R) by three raters (A, B, C).

| | M | N | P | Q | R |
|---|---|---|---|---|---|
| A | 1 | 2 | 3 | 4 | 5 |
| B | 2 | 3 | 2 | 5 | 3 |
| C | 5 | 5 | 5 | 3 | 2 |

Normalize the ratings by subtracting the average for each row and then subtracting the average for each column in the resulting table. Then, identify the largest element and entry of (C,P) about the normalized table.

avg of A=15/5=3, avg of B=15/5=3, avg of C=20/5=4

By subtracting avg from each row, we get-

| | M | N | P | Q | R |
|---|---|---|---|---|---|
| A | -2 | -1 | 0 | 1 | 2 |
| B | -1 | 0 | -1 | 2 | 0 |
| C | 1 | 1 | 1 | -1 | -2 |

avg of M=-⅔, avg of N=0, avg of P=0, avg of Q=⅔, avg of R=0

Subtracting avg foreach column, we get-

| | M | N | P | Q | R |
|---|---|---|---|---|---|
| A | -4/3 | -1 | 0 | 1/3 | 2 |
| B | -1/3 | 0 | -1 | 4/3 | 0 |
| C | 5/3 | 1 | 1 | -5/3 | -2 |

Largest element is 2 and entry of (C,P) is 1

**Question 2**: Below is a table giving the profile of three items.

| | | | | | | |
|---|---|---|---|---|---|---|
| **A** | 1 | 0 | 1 | 0 | 1 | 2 |
| **B** | 1 | 1 | 0 | 0 | 1 | 6 |
| **C** | 0 | 1 | 0 | 1 | 0 | 2 |

The first five attributes are Boolean, and the last is an integer "rating." Assume that the scale factor for the rating is α. Compute, as a function of α, the cosine distances between each pair of profiles. For each of α = 0, 0.5, 1, and 2, determine the cosine of the angle between each pair of vectors.

② $\cos(A,B) = \dfrac{AB}{|A||B|} = \dfrac{2+12\alpha^2}{\sqrt{9+120\alpha^2+144\alpha^4}}$

$\cos(B,C) = \dfrac{BC}{|B||C|} = \dfrac{1+12\alpha^2}{\sqrt{6+84\alpha^2+144\alpha^4}}$

$\cos(C,A) = \dfrac{CA}{|C||A|} = \dfrac{4\alpha^2}{\sqrt{6+20\alpha^2+144\alpha^4}}$

if

$\alpha = 0 \Rightarrow \cos(A,B) = 0.66$, $\cos(B,C) = 0.408$, $\cos(C,A) = 0$

$\alpha = 0.5 \Rightarrow \cos(A,B) = 0.721$, $\cos(B,C) = 0.666$, $\cos(C,A) = 0.288$

$\alpha = 1 \Rightarrow \cos(A,B) = 0.847$, $\cos(B,C) = 0.849$, $\cos(C,A) = 0.617$

$\alpha = 2 \Rightarrow \cos(A,B) = 0.946$, $\cos(B,C) = 0.992$, $\cos(C,A) = 0.865$

**Question 3**: Below is a utility matrix representing ratings by users A, B, and C for items *a* through h.

| | a | b | c | d | e | f | g | h |
|---|---|---|---|---|---|---|---|---|
| **A** | 4 | 5 | | 5 | 1 | | 3 | 2 |
| **B** | | 3 | 4 | 3 | 1 | 2 | 1 | |
| **C** | 2 | | 1 | 3 | | 4 | 5 | 3 |

Treat ratings of 3, 4, and 5 as 1 and 1, 2, and blank as 0. Compute the Jaccard distance between each pair of items. Then, cluster the items hierarchically into four clusters, using the Jaccard distance. When a cluster consists of more than one item, take the distance between clusters to be the minimum over all pairs of items, one from each cluster, of the Jaccard distance between those items. Break ties lexicographically. That is, sort the items that would be merged alphabetically, and merge those clusters whose resulting set would be first alphabetically.

Note: if you are not familiar with hierarchical clustering, read Sect. 7.2 of the MMDS book.

Update the matrix by replacing 3,4 and 5 with 1 and 1,2 and blanks with 0-

| | a | b | c | d | e | f | g | h |
|---|---|---|---|---|---|---|---|---|
| A | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| B | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| C | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |

- Jaccard dis(a,b)=1-½=1/2
- Jaccard dis(a,c)=1-0=1
- Jaccard dis(a,d)=1-⅓=1/3
- Jaccard dis(a,e)=1-0=1
- Jaccard dis(a,f)=1-0=1
- Jaccard dis(a.g)=1-½=1/2
- Jaccard dis(a,h)=1-0=1
- Jaccard dis(b,c)=1-½=1/2
- Jaccard dis(b,d)=1-⅔=⅓
- Jaccard dis(b,e)=1-0=1
- Jaccard dis(b,f)=1-0=1
- Jaccard dis(b,g)=1-⅓=2/3
- Jaccard dis(b,h)=1-0=1
- Jaccard dis(c,d)=1-⅓=2/3
- Jaccard dis(c,e)=1-0=1
- Jaccard dis(c,f)=1-0=1
- Jaccard dis(c,g)=1-0=1
- Jaccard dis(c,h)=1-0=1

- Jaccard dis(d,e)=1-0=1
- Jaccard dis(d,f)=1-⅓=2/3
- Jaccard dis(d,g)=1-⅔=1/3
- Jaccard dis(d,h)=1-⅓=2/3
- Jaccard dis(e,f)=1-0=1
- Jaccard dis(e,g)=1-0=1
- Jaccard dis(e,h)=1-0=1
- Jaccard dis(f,g)=1-½=1/2
- Jaccard dis(f,h)=1-0=1
- Jaccard dis(g,h)=1=½=1/2

  The final clusters are- {f,h},(b,d,g,a},{c},{e}

**Question 4**: We want to do an approximate UV-decomposition of the matrix M =

| | | |
|---|---|---|
| 1 | 2 | 3 |
| 4 | 5 | 6 |
| 7 | 8 | 9 |

We shall use only a single column for U and a single row for V, so the goal is to make the product UV as close as possible to M. Initially, we shall set V to [5,5,5] and make the entries of U unknown. Then in the first step, we choose the values of x, y, and z that minimize the root-mean-square error (RMSE) between the product

| |
|---|
| x |
| y |
| z |

| | | |
|---|---|---|
| 5 | 5 | 5 |

and the matrix M.

Find the values of x, y, and z that minimize the RMSE

Product of U and V- 5x 5x 5x

5y 5y 5y

5z 5z 5z

Let the eqn be-
$(5x-1)^2+(5x-2)^2+(5x-3)^2+(5y-4)^2+(5y-5)^2+(5y-6)^2+(5z-7)^2+(5z-8)^2+(5z-9)^2=0$

Differentiation with x gives, $150x-60=0 \Rightarrow x=\frac{2}{5}$

Differentiation with y gives, $150y-150=0 \Rightarrow y=1$

Differentiation with z gives, $150z-240=0 \Rightarrow z=8/5$