Big Data Management and Analytics
Assignment 10

Assignment 10-2

- Find the CUR-decomposition of the matrix, when we pick two „random" rows and columns. The columns we pick are Alien and StarWars and the rows are the ones of Jack and Jill.

- From the lecture (Ch.7, Sl. 46) we know that the scaled column for Alien is: $[1.54, 4.63, 6.17, 7.72, 0, 0, 0]^T$. The second column for Star Wars is the same. We thus define C as follows:

$$C = \begin{pmatrix} 1.54 & 1.54 \\ 4.63 & 4.63 \\ 6.17 & 6.17 \\ 7.72 & 7.72 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Assignment 10-2

- The unscaled rows for R are:

$$R_{unscaled} = \begin{pmatrix} 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \end{pmatrix}$$

- The probability p_i with which we select row i is given by:

$$p_i = \sum_j {m_{i,j}^2} \Big/ {\|M\|_F^2}$$

- The **square of the Frobenius norm for M** is $\|M\|_F^2 = 243$
- The **square of the Frobenius norm for Jack** is: $row_{jack} = \sum_j m_{3,j}^2 = 5^2 + 5^2 + 5^2 = 75$
- The **square of the Frobenius norm for Jill** is: $row_{jill} = \sum_j m_{4,j}^2 = 4^2 + 4^2 = 32$
- The **probability for selecting Jack** is: $p_{jack} = 75/243 = 0.309$
- The **probability for selecting Jill** is: $p_{jill} = 32/243 = 0.132$

Assignment 10-2

- The unscaled rows for R are:

$$R_{unscaled} = \begin{pmatrix} 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \end{pmatrix}$$

- Scaling the row for Jack, we divide all its row entries by:

$$\sqrt{r * p_{jack}} = \sqrt{2 * 0.309} = 0.786$$

- Scaling the row for Jill, we divide all its row entries by:

$$\sqrt{r * p_{jill}} = \sqrt{2 * 0.132} = 0.514$$

- This yields the scaled matrix R:

$$\mathrm{R} = \begin{pmatrix} 6.36 & 6.36 & 6.36 & 0 & 0 \\ 0 & 0 & 0 & 7.78 & 7.78 \end{pmatrix}$$

Assignment 10-2

- Now that we have C and R, we construct the middle matrix U:

- First construct a matrix W from the intersection of the selected rows from R and the columns from C:

$$W = \begin{pmatrix} 5 & 5 \\ 0 & 0 \end{pmatrix}$$

	Matrix	Alien	Star Wars	Casablanca	Titanic
Joe	1	1	1	0	0
Jim	3	3	3	0	0
John	4	4	4	0	0
Jack	5	5	5	0	0
Jill	0	0	0	4	4
Jenny	0	0	0	5	5
Jane	0	0	0	2	2

Assignment 10-2

- $W = \begin{pmatrix} 5 & 5 \\ 0 & 0 \end{pmatrix}$
- Take the SVD from W:

$$W = X\Sigma Y^T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{50} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$$

- Taking the Moore-Penrose pseudoinverse of Σ leads to:

$$\Sigma^+ = \begin{pmatrix} 1/\sqrt{50} & 0 \\ 0 & 0 \end{pmatrix}$$

Assignment 10-2

- Now we can compute $U = Y(\Sigma^+)^2 X^T$

$$U = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \left(\begin{pmatrix} 1/\sqrt{50} & 0 \\ 0 & 0 \end{pmatrix} \right)^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/50 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1/50\sqrt{2} & 0 \\ 1/50\sqrt{2} & 0 \end{pmatrix}$$