Overview

In the first part of this exercise session, you will build a recommender system for predicting movies using different techniques. In the second part, you will explore an important concept in time series analysis: dynamic time warping.
Have fun!

1 Recommending movies using collaborative filtering

1.1 Similarity-based

A). First, you will apply the similarity-based collaborative filtering algorithm presented in class by hand to make a prediction for an unrated movie. The following table illustrates the ratings.

	User A	User B	User C
Movie 1	1		3
Movie 2	2		3
Movie 3	3		3
Movie 4	3	4	
Movie 5	5	4	
Movie 6	2	4	**?**
Movie 7		2	
Movie 8		4	2
Movie 9		4	5
Movie 10			4

Use the collaborative filtering technique with Pearson correlation to predict the rating of user C for movie 6.
Next, the formulas that should be used to solve this exercise:

$$\hat{R}_{u,i} = \bar{R}_u + \frac{1}{\alpha} \sum_{v:i \in I_v} w(u,v)(R_{v,i} - \bar{R}_v) \tag{1}$$

$$w(u,v) = \frac{\sum_{j \in I_u \cap I_v} (R_{u,j} - \bar{R}_u)(R_{v,j} - \bar{R}_v)}{\sqrt{\sum_{j \in I_u \cap I_v} (R_{u,j} - \bar{R}_u)^2 \sum_{j \in I_u \cap I_v} (R_{v,j} - \bar{R}_v)^2}} \tag{2}$$

$$\alpha = \sum_{v:i \in I_v} |w(u,v)| \tag{3}$$

$$\bar{R}_u = \frac{1}{|I_u|} \sum_{j \in I_u} R_{u,j} \tag{4}$$

Remember that the symbols used in the previous formulas have the following meaning:

- $\hat{R}_{u,i}$ is the prediction for user u for unrated item i;
- I_u is the set of all items rated by user u;
- $w(u,v)$ is the Pearson correlation between user u and user v. Note that the sums run over the items, $i \in I_u \cap I_v$, that are rated by both user u AND user v;
- $\alpha_{u,i}$ is a normalization constant;
- $\bar{R}_u$ is the average rating for user u. The average is taken over *all* items (I_u) that user u has rated.

B). Now download from Toledo the `collaborative-filtering.ipynb` notebook and the data `u.data` which contains 100,000 movie ratings by different users. The goal is to build a recommendation engine that implements the functions above.

1.2 Model-based

Another approach to collaborative filtering is to use a model-based method such as singular value decomposition or non-negative matrix factorization. In the `collaborative-filtering.ipynb` notebook, you will use an implementation of non-negative matrix factorization to predict ratings.

2 Dynamic Time Warping

For the time series $s_1 = [1, 1, 1, 3, 2, 0]$ and $s_2 = [2, 4, 3, 2, 1, 1]$, do the following:

1. Compute the Euclidean distance.

2. Compute the DTW distance.

	1	**1**	**1**	**3**	**2**	**0**
2						
4						
3						
2						
1						
1						

3. Draw a plot of s_1 and s_2 and show visually which points of s_1 are compared to which points of s_2, for both the Euclidean distance and the DTW distance.

4. Compute the DTW distance, but now use a warping constraint of 1.

	1	**1**	**1**	**3**	**2**	**0**
2						
4						
3						
2						
1						
1						

Let's now apply these distance measures to real-world data in `dtw.ipynb` which you can find on Toledo.