Università di Pisa	A.A. 2014- 2015

Data Mining II April 13th, 2015

Exercise 1 - Classification – alternative methods **(11 points)**

Exercise 2 - Sequential patterns **(10 points)**

Given the following input sequence

<	{A}	{B,C,D}	{C,D}	{A,D,E}	{A,B}	{F}	{A,B}	{B,D} >
	t=0	t=1	t=2	t=3	t=4	t=5	t=6	t=7

show all the occurrences (there can be more than one or none, in general) of each of the following subsequences in the input sequence above. Repeat the exercise twice: the first time considering no temporal constraints (left column): the second time considering min-gap = 1 (right column). Each occurrence should be represented by its corresponding list of time stamps, e.g..: <0,2,3> = <t=0, t=2, t=3>.

	Occurrences	*Occurrences with min-gap=1*
ex.: <{A}{D}>	**<0,2> <0,3> <0,7> <3,7> <4,7> <6,7>**	**<0,2> <0,3> <0,7> <3,7> <4,7>**
w_1 = <{A} {C} {D} >	**<0,1,2> <0,1,3> <0,1,7> <0,2,3> <0,2,7>**	**<0,2,7>**
w_2 = <{B}{A}>	**<1,3> <1,4> <1,6> <4,6>**	**<1,3> <1,4> <1,6> <4,6>**
w_3 = <{A}{E} {B}>	**<0,3,4> <0,3,6><0,3,7>**	**<0,3,6><0,3,7>**
w_4 = <{A}{D}{E}>	**<0,1,3><0,2,3>**	**none**

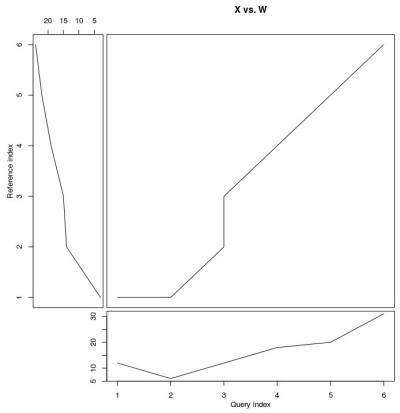
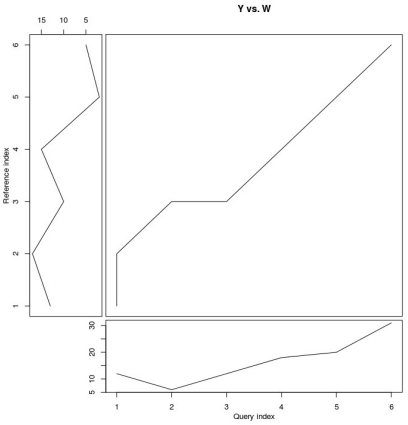
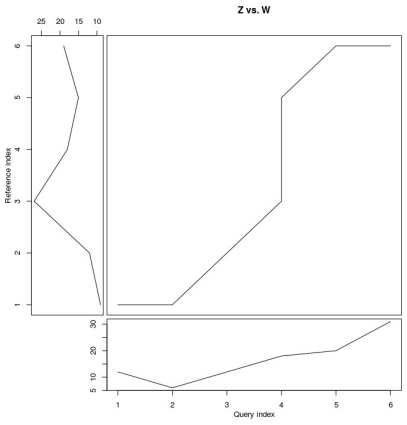
Exercise 3 - Time series / Distances **(11 points)**

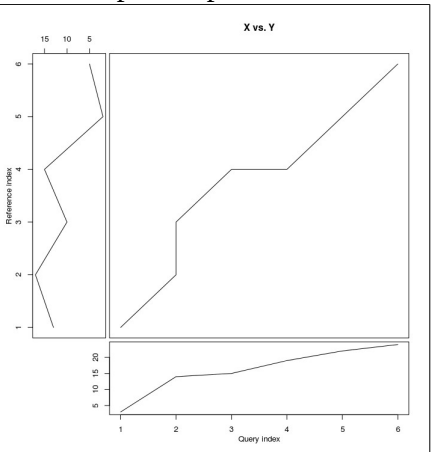
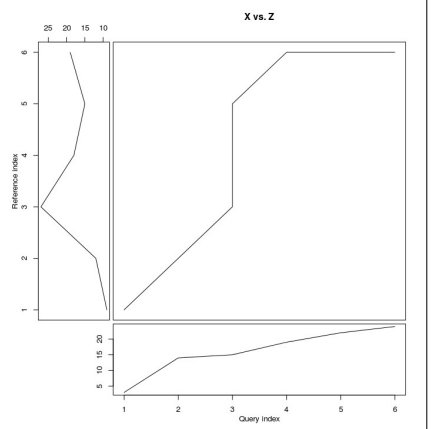
Given the following dataset of time series:

ID	Time series
W	< 12, 6, 12, 18, 20, 31 >
X	< 3, 14, 15, 19, 22, 24 >
Y	< 13, 17, 10, 15, 2, 5 >
Z	< 9, 12, 27, 18, 15, 19 >

compute the matrix of distances among all pairs of time series adopting a Dynamic Time Warping distance.

	W	X	Y	Z
W		27	59	31
X			60	31
Y				48
Z				

	Point-to-point costs	Cost matrix	Optimal path
	[,1] [,2] [,3] [,4] [,5] [,6] [1,] 9 2 3 7 10 12 [2,] 3 8 9 13 16 18 [3,] 9 2 3 7 10 12 [4,] 15 4 3 1 4 6 [5,] 17 6 5 1 2 4 [6,] 28 17 16 12 9 7	[,1] [,2] [,3] [,4] [,5] [,6] [1,] 9 11 14 21 31 43 [2,] 12 17 20 27 37 49 [3,] 21 14 17 24 34 46 [4,] 36 18 17 18 22 28 [5,] 53 24 22 18 20 24 [6,] 81 41 38 30 27 27	
	[,1] [,2] [,3] [,4] [,5] [,6] [1,] 1 5 2 3 10 7 [2,] 7 11 4 9 4 1 [3,] 1 5 2 3 10 7 [4,] 5 1 8 3 16 13 [5,] 7 3 10 5 18 15 [6,] 18 14 21 16 29 26	[,1] [,2] [,3] [,4] [,5] [,6] [1,] 1 6 8 11 21 28 [2,] 8 12 10 17 15 16 [3,] 9 13 12 13 23 22 [4,] 14 10 18 15 29 35 [5,] 21 13 20 20 33 44 [6,] 39 27 34 36 49 59	
	[,1] [,2] [,3] [,4] [,5] [,6] [1,] 3 0 15 6 3 7 [2,] 3 6 21 12 9 13 [3,] 3 0 15 6 3 7 [4,] 9 6 9 0 3 1 [5,] 11 8 7 2 5 1 [6,] 22 19 4 13 16 12	[,1] [,2] [,3] [,4] [,5] [,6] [1,] 3 3 18 24 27 34 [2,] 6 9 24 30 33 40 [3,] 9 6 21 27 30 37 [4,] 18 12 15 15 18 19 [5,] 29 20 19 17 20 19 [6,] 51 39 23 30 33 31	

	Point-to-point costs	Cost matrix	Optimal path
	[,1] [,2] [,3] [,4] [,5] [,6] [1,] 10 14 7 12 1 2 [2,] 1 3 4 1 12 9 [3,] 2 2 5 0 13 10 [4,] 6 2 9 4 17 14 [5,] 9 5 12 7 20 17 [6,] 11 7 14 9 22 19	[,1] [,2] [,3] [,4] [,5] [,6] [1,] 10 24 31 43 44 46 [2,] 11 13 17 18 30 39 [3,] 13 13 18 17 30 40 [4,] 19 15 22 21 34 44 [5,] 28 20 27 28 41 51 [6,] 39 27 34 36 50 60	
	[,1] [,2] [,3] [,4] [,5] [,6] [1,] 6 9 24 15 12 16 [2,] 5 2 13 4 1 5 [3,] 6 3 12 3 0 4 [4,] 10 7 8 1 4 0 [5,] 13 10 5 4 7 3 [6,] 15 12 3 6 9 5	[,1] [,2] [,3] [,4] [,5] [,6] [1,] 6 15 39 54 66 82 [2,] 11 8 21 25 26 31 [3,] 17 11 20 23 23 27 [4,] 27 18 19 20 24 23 [5,] 40 28 23 23 27 26 [6,] 55 40 26 29 32 31	
	[,1] [,2] [,3] [,4] [,5] [,6] [1,] 4 1 14 5 2 6 [2,] 8 5 10 1 2 2 [3,] 1 2 17 8 5 9 [4,] 6 3 12 3 0 4 [5,] 7 10 25 16 13 17 [6,] 4 7 22 13 10 14	[,1] [,2] [,3] [,4] [,5] [,6] [1,] 4 5 19 24 26 32 [2,] 12 9 15 16 18 20 [3,] 13 11 26 23 21 27 [4,] 19 14 23 26 21 25 [5,] 26 24 39 39 34 38 [6,] 30 31 46 52 44 48