

Exercise 1 - Classification – alternative methods (11 points)

Given the training dataset below, predict the class of the below new test data by using k-Nearest Neighbor for k=3. For similarity measure use a simple match of attribute values: Similarity(A,B) that is computed by the following formula

$$\sum_{i=1}^4 w_i * \partial(a_i, b_i) / 4$$

where $\partial(a_i, b_i)$ is 1 if a_i equals b_i and 0 otherwise. The only exception is the attribute Income, for which we have that $\partial(\text{Low}, \text{Medium}) = \partial(\text{Medium}, \text{Low}) = \partial(\text{High}, \text{Medium}) = \partial(\text{Medium}, \text{High}) = 0.5$, while for all the other cases the same rule apply as above, i.e. $\partial(a_{\text{income}}, b_{\text{income}})$ is 1 if a_{income} equals b_{income} and 0 otherwise. Weights are all 0.3, except for Age, which is 0.1.

Training Data

Income	Student	Age	Sex	Credit
Low	yes	Young	F	No
High	yes	Young	M	Yes
Low	no	Young	M	No
High	yes	Old	F	Yes
Medium	yes	Young	M	Yes
Low	no	Old	F	No
Medium	no	Young	M	No
High	yes	Old	M	No

Test Data

Income	Student	Age	Sex	Credit
High	yes	Young	F	
High	no	Old	F	

Exercise 2 - Sequential patterns (11 points)

Given the following input sequence

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

A) 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 max-gap = 2 (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>.

B) list all the subsequences that contain the event D, and satisfy both min-gap=3 (i.e. all gaps must be >3) and max-span=4.

	Occurrences	Occurrences with max-gap=2
ex.: $\langle \{D\}\{B\} \rangle$	$\langle 1,3 \rangle \langle 1,6 \rangle \langle 1,7 \rangle$	$\langle 1,3 \rangle$
$w_1 = \langle \{A\} \{C\} \rangle$		
$w_2 = \langle \{E\} \{E\} \rangle$		
$w_3 = \langle \{A\} \{E\} \{B\} \rangle$		

Exercise 3 - Time series / Distances (10 points)

Given the following dataset of time series (on the left):

ID	Time Series
W	$\langle 13, 9, 2, 15, 13 \rangle$
X	$\langle 10, 17, 1, 12, 14, 17 \rangle$
Y	$\langle 9, 14, 13, 20 \rangle$

	W	X	Y
W			
X			
Y			

compute the matrix of distances among all pairs of time series (on the right) adopting a Dynamic Time Warping distance, and computing the distances between single points as $d(x,y) = |x - y|$. For each pair of time series compared also show the matrix used to compute the final result.

W - X

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	
[1,]		3	7	19	20	21	25
[2,]		4	11	15	18	23	29
[3,]		12	19	12	22	30	38
[4,]		17	14	26	15	16	18
[5,]		20	18	26	16	16	20

W - Y

	[,1]	[,2]	[,3]	[,4]	
[1,]		4	5	5	12
[2,]		4	9	9	16
[3,]		11	16	20	27
[4,]		17	12	14	19
[5,]		21	13	12	19

X - Y

	[,1]	[,2]	[,3]	[,4]	
[1,]		1	5	8	18
[2,]		9	4	8	11
[3,]		17	17	16	27
[4,]		20	19	17	24
[5,]		25	19	18	23
[6,]		33	22	22	21