Process Mining

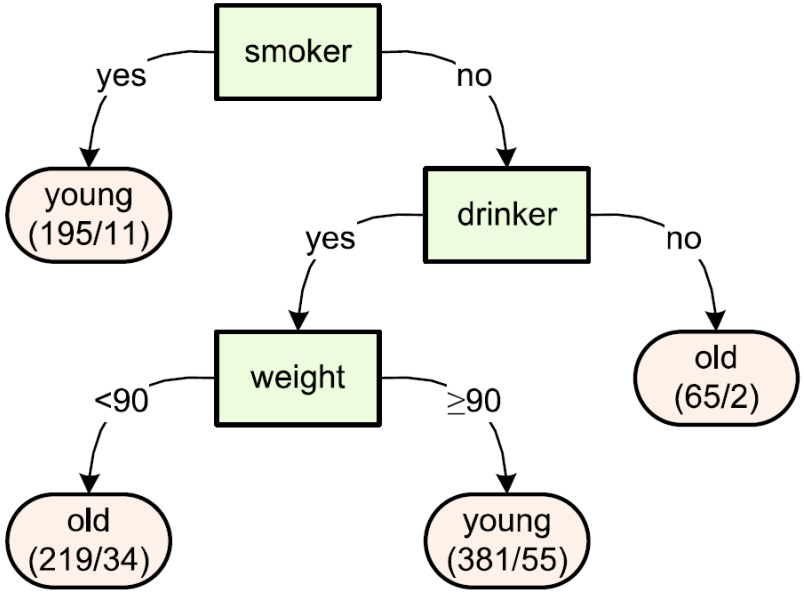
Lesson 1 – Data Mining Techniques

Doc. Cognini Riccardo

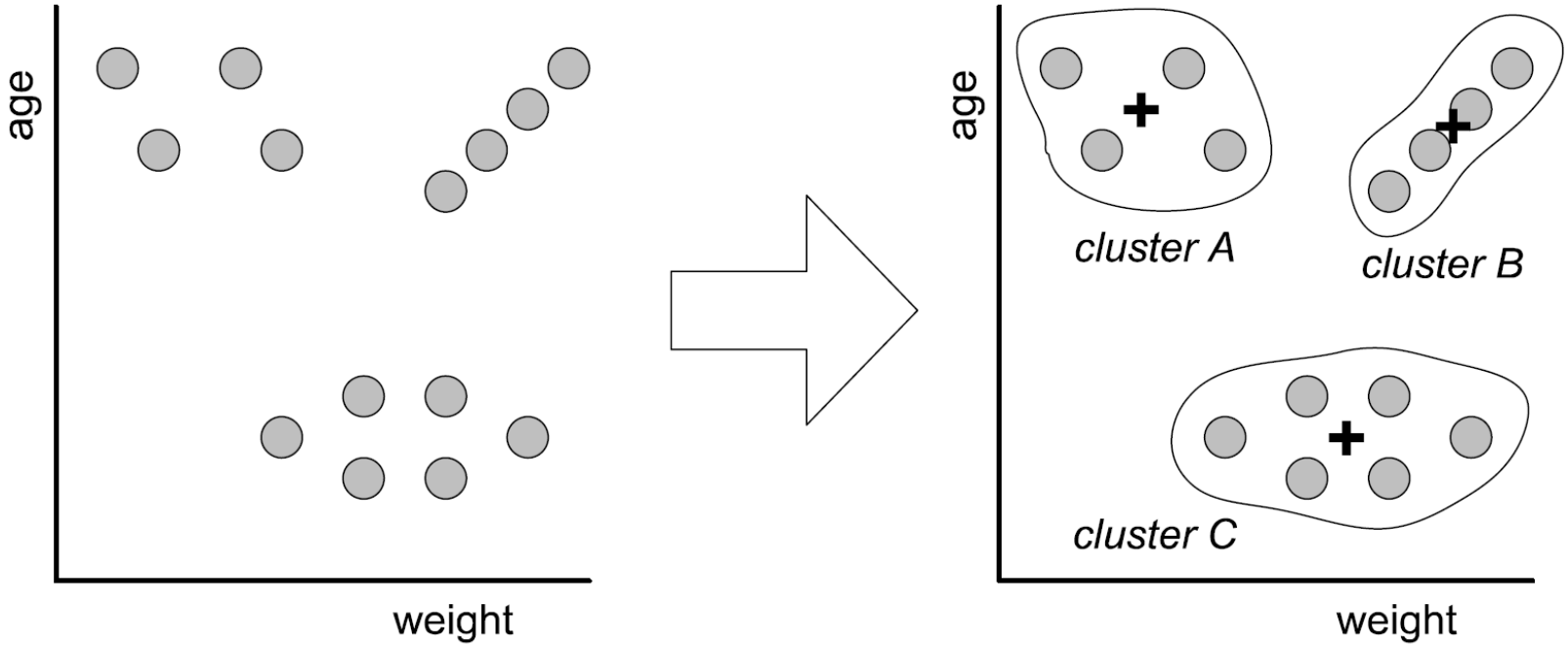
Supervised (Decision Tree)

Table 3.1 Data set 1: Data about 860 recently deceased persons to study the effects of drinking, smoking, and body weight on the life expectancy

Drinker	Smoker	Weight	Age
Yes	Yes	120	44
No	No	70	96
Yes	No	72	88
Yes	Yes	55	52
No	Yes	94	56
No	No	62	93
...	...	...	



Unsupervised (Clustering)



Supervised (Sequence Mining)

Table 3.4 A fragment of a data set used for sequence mining: each line corresponds to an order

Customer	Seq. number	Timestamp	Items
Wil	1	02-01-2011:09.02	{*cappuccino*}
	2	03-01-2011:10.06	{*espresso*, *muffin*}
	3	05-01-2011:15.12	{*americano*, *cappuccino*}
	4	06-01-2011:11.18	{*espresso*, *muffin*}
	5	07-01-2011:14.24	{*cappuccino*}
	6	07-01-2011:14.24	{*americano*, *cappuccino*}
Mary	1	30-12-2010:11.32	{*tea*}
	2	30-12-2010:12.12	{*cappuccino*}
	3	30-12-2010:14.16	{*espresso*, *muffin*}
	4	05-01-2011:11.22	{*bagel*, *tea*}
Bill	1	30-12-2010:14.32	{*cappuccino*}
	2	30-12-2010:15.06	{*cappuccino*}
	3	30-12-2010:16.34	{*bagel*, *espresso*, *muffin*}
	4	06-01-2011:09.18	{*ristretto*}
	5	06-01-2011:12.18	{*cappuccino*}
...	...	...	...

QUESTIONS??