

Information Mining - winter semester 2020/21**Exercise sheet 2**

Exercise 1: Understanding

List all types of attributes and describe them with your own words. Which operations can be applied on which type?

nominal : a name \rightarrow ($=$ / \neq)
ordinal : names with an order. ($>$, $<$, $=$, \neq)
interval : numeric, but no zero point. ($+$, $-$, \dots)
ratio : numeric, have a zero point. ($*$, $/$, $+$, $-$, \dots)

Exercise 2: More understanding

Give for each of the attribute type from above an example and explain why it fits to that type.

nominal : color
ordinal : hot, warm, cold
interval : Celsius temperature
ratio : length

Exercise 3: More understanding

There are situations where there are some attributes for which not all values are available. For instance, you might have a dataset with persons and each person is recorded with several attributes such as earnings, gender, etc. It might happen that the attribute values for earnings is not always available. This is also called missing values. How would you deal with such missing values? Also how would you deal with missing values for the gender attribute?

replace by mean value,
for some special cases, regard "missing" as a value.

Exercise 4: More understanding

What do you understand under data normalisation? When do you use data normalisation?

different features may have very different scales, we need normalization, to improve learner's performance.
e.g. Number of red blood cells, number of white blood cells

Exercise 5: RapidMiner

Furthermore, for practical exercises we will use *RapidMiner*, a data mining tool.

- (a) Please install *RapidMiner* and get familiar with the application.
The free version can be downloaded at *rapidminer*¹. There're also installation guide and documentation on the site.
- (b) Which attribute types are supported by *RapidMiner*?

¹Application: <https://rapidminer.com/download-rapidminer/>