Converting to a Different Scale Type

some ML algorithms can use only data of a particular scale type. The good news is that it is possible to convert data from a qualitative scale to a quantitative scale

Data set related with to work colleagues, showing their favourite food, age, how far from us they live and if they are good or bad company.

Food preferences of our colleagues.

Food	Age	Distance	Company
Chinese	51	Close	Good
Italian	43	Very close	Good
Italian	82	Close	Good
Burgers	23	Far	Bad

Converting Nominal to Relative

The nominal scale does not assume an order between its values, to keep this information, nominal values should be converted to relative or binary values.

The most common conversion is called "1-of-n", also known as canonical or one-attribute- per-value conversion, which transforms *n* values of a nominal attribute into *n* binary attributes. A binary attribute has only two values, 0 or 1

Conversion from nominal scale torelative scale.

Nominal	Relative
Green	001
Yellow	010
Blue	100

Conversion from the nominal scale to binary values.

Original data				Converted data					
Food	Age	Distance	Company	F1	F2	F3	Age	Distance	Company
Chinese	51	Close	Good	0	0	1	51	2	1
Italian	43	Very close	Good	0	1	0	43	1	1
Italian	82	Close	Good	0	1	0	82	2	1
Burgers	23	Far	Bad	1	0	0	23	3	0

Converting Ordinal to Relative or Absolute

we can convert to natural numbers, starting with the value 0 for the smallest value and, for each subsequent value, adding 1 to the previous value.

some algorithms may work only with binary values.

If we want to convert ordinal values to binary values, we can use the gray code, which keeps the distance between two consecutive values as a different value in one of the binary values.

Another binary code, called the thermometer code, starts with a binary vector with only 0 values and substitutes one 0 value by 1, from right to left, as the ordinal value increases.

Nominal	Natural number	Gray code	Thermometer code
Small	0	00	000
Medium	1	01	001
Large	2	11	011
Very large	3	10	111

Conversion from the ordinal scale to the relative or absolute scale.

Quantitative values can be converted to nominal or ordinal values. This process is called "discretization" and, depending whether we want to keep the order between the values, will be referred to as "nominal" or "ordinal" discretization.

Discretization has two steps

The first step is the definition of the number of qualitative values, which is usually defined by the data analyst. This number of qualitative values is called the number of "bins"

The Second Step is define the interval of values to be associated with each bin. This asso ciation is usually done with an algorithm.

There are two alternatives for the association: by width or by frequency.

Conversion from the ordinal scale to the relative scale.

Quantiative		Conversion by width	Conversion by frequency
2	A	A	
3	A	A	
5	A	A	
7	A	B	
10	B	B	
15	B	B	
16	C	C	
19	C	C	
20	C	C	

Converting to a Different Scale

Converting data in a scale to another scale of the same type is necessary, in several situations, such as when using distance measures.

This kind of conversion is typically done in order to have different attributes expressed on the same scale; a process known as "normalization".

The most similar friends are Bernhard and James, while the most dissimilar are Bernhard and Gwyneth. Let us do the same calculation measuring the ages in decades: 4.3, 3.8 and 4.2 for Bernhard, Gwyneth and James respectively

Euclidean distances of ages expressed in years.

Age in years	B–G	B–J	G–J
Euclidean distance	5.46	2.33	4.00

Table 4.13 Euclidean distance with age expressed in decades.

Age in decades	B–G	B–J	G–J
Euclidean distance	2.26	2.10	0.41

Table 4.14 Normalization using min–max rescaling.

Friend	Age	Education	Rescaled age	Rescaled education
Bernhard	43	2.0	1.0	0.0
Gwyneth	38	4.2	0.0	1.0
James	42	4.0	0.8	0.91

Table 4.15 Normalization using standardization.

Friend	Age	Education	Rescaled age	Rescaled education
Bernhard	43	2.0	0.76	−1.15
Gwyneth	38	4.2	−1.13	0.66
James	42	4.0	0.38	0.49

Data Transformation

Data transformation is the process of converting data from one format, such as a database file, XML document or Excel spreadsheet, into another.
Transformations typically involve converting a raw data source into a cleansed, validated and ready-to-use format.

- *Apply a logarithmic function to the values of a predictive attribute:* This is usually performed for skewed distributions, when some of the values are much larger (or much smaller) than the others. The logarithm makes the dis-tribution less skewed. Thus, log transformations make the interpretation of highly skewed data easier.
- *Conversion to absolute values:* For some predictive attributes, the value’s magnitude is more important than its sign, if the value is positive or negative.