**opics of exercises**

* Metrics, dissimilarity matrices

**Exercise 3.1**

Consider the euclidean distance in R^m.

1. \* Verify that this distance really forms a metrics(Hint: make use of the Cauchy's-Schwartz inequality).
2. Define the euclidean distance via the appropriate function of 4 variables in R.
3. Define the following open balls in their corresponding metrics and compare the form of the newly obtained balls with the euclidean ball if:

* **K(1,1/3)** in the metrical space (N,*d*) with *d(m, n)* = |*1/m* - *1/n*|, for natural *m, n*,
* **K(2,3)** in the metrical space (R,*d*), where *d(x,y)* = |*x*-*y*|,
* **K((1,1),3)** in the metrical space (R^2, *d*) with *d(x,y)* = max{|x1-y1|, |x2-y2|}.

**Exercise 3.2**

Consider two vectors: v1 = c(1,2,3,4,5,6), v2 = c(-4,-5,-6,-6,6,7). Compute the distance between them using:

1. Euclidean metric
2. Canberra metric
3. Minkowski metrics (p = 2, p = 3)
4. Manhattan metric
5. Maximum metric

Compare the results. Explain the similarity between distances in Euclidean metrics and the Minkowski one for p = 2, and the Manhattan metric and the Minkowski one for p = 1. Why R-computation is feasible even if the length of vectors is different?

**Exercise 3.3**

Suppose *f* and *g* are metrics. Check if the following functions are also metrics:

1. *f* + *g*
2. max{*f*, *g*}
3. min{*f*, *g*}
4. 1/3 *f* + 2/3 *g*
5. *f\*g*

**Exercise 3.4**

Let the following ordinal attributes be given:

* **cloud.cover**, levels: clear, scattered, broken, overcast
* **precipitation**, levels: none, light, moderate, heavy, extreme
* **wind**, levels: calm, light (breeze), moderate (breeze), strong (breeze), gale, storm
* **temperature**, levels: cold, chilly, normal, warm, heat

In all cases the levels are given from the smallest to the largest.

The following objects contains information about weather conditions in Krakow (8 days of November).

| **cloud.cover** | **precipitation** | **wind** | **temperature** |
| --- | --- | --- | --- |
| overcast | light | strong | chilly |
| overcast | heavy | moderate | chilly |
| scattered | light | light | cold |
| broken | none | light | cold |
| broken | light | light | cold |
| overcast | moderate | light | chilly |
| scattered | none | calm | chilly |
| overcast | light | calm | normal |

1. Define 4 factors for attributes: **cloud.cover**, **precipitation**, **wind**, and **temperature**.
2. Define **weather** data frame that contains data from the above table.
3. Determine the dissimilarity matrix for these objects using Euclidean metric (see slide 31 part 2).

**Exercise 3.5**

1. Define **weather2** data frame. Copy two first columns from **weather** data frame and include the following 3 columns.
   * **wind** (m/s): 24, 14, 6, 5, 5, 6, 2, 3
   * **temperature** (Celsius): 9, 9, 5, 5, 7, 12, 8, 10
   * **humidity** (%): 87, 81, 73, 76, 80, 85, 81, 88
2. Determine the dissimilarity matrix for these objects (see slide 34, part 2).

**Exercise 3.6**

1. Transform the data from Exercise 3.2 into a data frame **weather.binary** with asymmetric binary attributes (see slide 30 part 2). How many attributes do you need?
2. Determine the dissimilarity matrix for the objects from **weather.binary** data frame.

**Exercise 3.7**

Let the following categorical attributes be given:

* **chocolate**, levels: milk, white, dark
* **nuts**, levels: almonds, hazelnuts, pistachios, walnuts
* **filling**, levels: cream, strawberry, coffee, marzipan
* **shape**, levels: heart, leaf, snowflake, ball

The following objects contains information about chocolates with some nuts and filling.

| **chocolate** | **nuts** | **filling** | **shape** |
| --- | --- | --- | --- |
| milk | pistachios | cream | heart |
| milk | pistachios | coffee | heart |
| milk | walnuts | cream | leaf |
| milk | hazelnuts | cream | snowflake |
| milk | almonds | marzipan | ball |
| white | pistachios | cream | ball |
| white | pistachios | coffee | ball |
| white | hazelnuts | strawberry | heart |
| white | almonds | marzipan | leaf |
| white | almonds | cream | snowflake |
| dark | walnuts | cream | snowflake |
| dark | almonds | marzipan | leaf |
| dark | almonds | cream | leaf |
| dark | hazelnuts | coffee | heart |
| dark | hazelnuts | strawberry | heart |

1. Define 4 factors for attributes: **chocolate**, **nuts**, **filling**, and **shape**.
2. Define **sweets** data frame that contains data from the above table.
3. Determine the dissimilarity matrix for these objects.

**Exercise 3.8**

Load **Orange** data set. Analyse the contents of the entire data set. How can you compare these 5 objects? Determine the dissimilarity matrix for these objects using Euclidean metric.

data("Orange")

**Exercise 3.9**

Consider the following table briefly describing 3 Polish cities in the 'Pomeranian Voivodeship'. Compute both the symmetric and antisymmetric dissimilarities between them. Find the most simmilar cities among them.

| **City\Property** | **>250.000 people** | **German location** | **History > 1000 year** | **Part of 'Trójmiasto'** |
| --- | --- | --- | --- | --- |
| **Gdańsk** | 1 | 1 | 1 | 1 |
| **Gdynia** | 1 | 0 | 0 | 1 |
| **Sopot** | 0 | 1 | 0 | 1 |

**Exercise 3.10**

Check, which of the following functions are metrics in their domains:  
- *d(x, y)* = | arctg(*x*)- arctg(*y*)|, for real *x, y*,  
- *d(x, y)* = |sin(*x - y*)/2|cos(*x + y*)/2 in [-pi/2, pi/2],  
- *d(x, y)* = 1 - cos(*x-y*), for *x, y* in [0, pi],  
- *d(x, y)* = *|x^2 - y^2|*, for real *x, y*,  
- *d(n, m)* = |*1/n* - *1/m*|, for natural *n, m*.