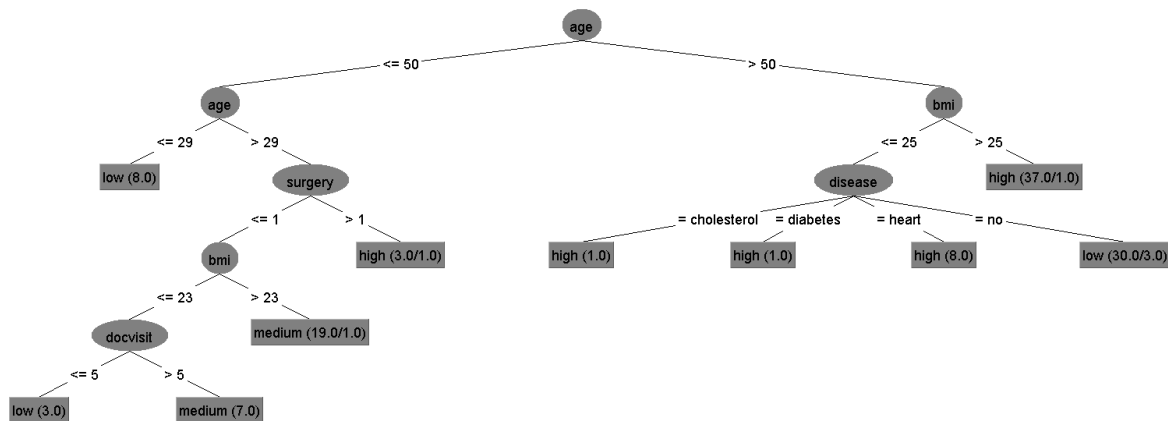


# Assignment: Case-Based Reasoning

A health insurance company developed a knowledge-based system for deciding whether an application for a health insurance can be accepted. From this data about previous applications they learned rules to determine the risk value of the applicant. This decision tree shows the rules:



Eligible candidates with low risk score low are accepted, applicants with risk score high are rejected. Applicants with medium risks may either pay a surcharge on the premium to compensate the risk or some diseases are excluded from the insurance policy.

In order to support the decision making for these medium risk cases, the company builds a case base with the information about the diseases and the decision that was made. It was determined that the following attributes are relevant for making the decision how to deal with medium-risk cases:

- Age: a number between (18 and 70)
- Gender: one of the value male (m) or female (f)
- Disease: There can be diseases of *organs, bones or ligaments*. Affected organs can be *kidney, heart or liver*. Heart diseases are *heart attack or heart rhythm disturbance*. We also know *fractures and amputation* of bones as well as *torn ligament and stretched ligament*.

The solution part consists of two attributes

- Surcharge: the additional insurance premium
- Policy adaptation: adaptation on the insurance policy like exclusion of diseases or medications.

When a new applicant is classified as medium risk, the underwriter looks in the case base to find the most similar cases in order to determine whether a surcharge should be added to the premium and/or some changes on the policy should be made

## Exercise 1

1. Define local similarity measures for the attributes age, gender and disease. The value of the similarity measures must be between 0 and 1. The following conditions must hold:
  - The gender of a person is either male (m) or female (f).
  - the similarity function for age should be symmetric
  - The local similarity measure of diseases is based on a taxonomy. Draw the taxonomy of diseases and add sample similarity values.
2. What would be a similarity measure, if the age would be one of the values young, adult, senior?
3. Define a formula for the global similarity such that the following conditions hold:
  - The attribute disease is as important as the other two attributes together.
  - The weight of the attribute gender is 0.1

## Exercise 2

Assume that the system contains documents to the following three cases:

Case C1:

Gender: m  
Age: 33  
Disease: fracture of the tibia  
Policy adaption: exclusion of ligament diseases

Case C2:

Gender: m  
Age: 43  
Disease: heart attack  
Surcharge: surcharge of € 80

Case C3:

Gender: f  
Age: 69  
Disease: heart attack  
Surcharge: surcharge of € 120  
Policy adaptation: percentage share of 10% for heard diseases

There is a new application with the following data:

Application A1

Gender: f  
Age: 49  
Disease: heart rhythm disturbance

1. Compute the local similarity values for each of the three attributes of each case with this application.
2. Compute the global similarity of A1 with each of the three cases. To compute the global similarity, the similarity values of each attribute can be weighted. For the global similarity in this example the following conditions must hold:
  - The attribute disease is as important as the other two attributes together.
  - The weight of the attribute gender is 0.1



## Solutions

### Exercise 1:

#### 1. Local similarity measures

Gender:

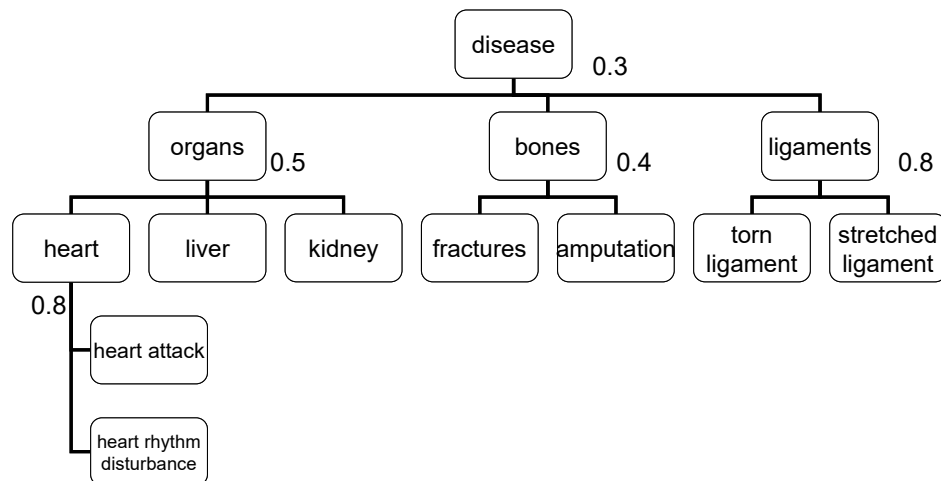
$$\text{Sim}_{\text{gender}}(x,y)$$

	m	f
m	1	0
f	0	1

Age:

$$1 - |x-y|/52$$

Diseases:



*It is important that in the taxonomy the similarity values of the more special categories are less than those of the more general categories.*

*The given similarity values for the inner nodes are examples that satisfy this condition; they do not follow from the task description.*

#### 2. Alternative for age:

*Regard the symbolic values as ordered, assign a number to each value and use a similarity function for numeric values, e.g.*

*Numeric values: young = 1, adult = 2, senior = 3*

*Formula:  $\text{sim}_{\text{age}}(x,y) = 1 - |x-y|/2$*

#### 3. Global similarity: $\text{sim}(x,y) = 0.4 * \text{sim}_{\text{age}}(x,y) + 0.1 * \text{sim}_{\text{gender}}(x,y) + 0.5 * \text{sim}_{\text{disease}}(x,y)$



## Exercise 2

Local similarities

$$C1: \text{sim}_{age} = 0.82; \text{sim}_{gender} = 0; \text{sim}_{disease} = 0.3$$

$$C2: \text{sim}_{age} = 0.93; \text{sim}_{gender} = 0; \text{sim}_{disease} = 0.8$$

$$C3: \text{sim}_{age} = 0.78; \text{sim}_{gender} = 1; \text{sim}_{disease} = 0.8$$

Global similarity:  $0.4 * \text{sim}_{age} + 0.1 * \text{sim}_{gender} + 0.5 * \text{sim}_{disease}$

$$G(C1) = 0.4 * 0.82 + 0.1 * 0 + 0.5 * 0.3 = \mathbf{0.478}$$

$$G(C2) = 0.4 * 0.93 + 0.1 * 0 + 0.5 * 0.8 = \mathbf{0.772}$$

$$G(C3) = 0.4 * 0.78 + 0.1 * 1 + 0.5 * 0.8 = \mathbf{0.812}$$

