

**Data Mining II**

July 17th, 2015

**Exercise 1 - Classification – alternative methods (11 points)**

Given the training dataset below, predict the class of the below new test data by using k-Nearest Neighbor for k=3. For similarity measure use a simple match of attribute values: Similarity(A,B) that is computed by the following formula

$$\sum_{i=1}^4 w_i * \partial(a_i, b_i) / 4$$

where  $\partial(a, b)$  is 1 if  $a_i$  equals  $b_i$  and 0 otherwise.  $a_i$  and  $b_i$  are either age, income, student or sex. Weights are all 1 except for income it is 2.

**Training Data**

| Income | Student | Age   | Sex | Credit |
|--------|---------|-------|-----|--------|
| Low    | yes     | Young | F   | No     |
| High   | yes     | Young | M   | Yes    |
| Low    | no      | Old   | M   | No     |
| High   | yes     | Old   | F   | No     |
| Medium | yes     | Young | M   | Yes    |
| Low    | no      | Old   | F   | No     |
| Medium | no      | Young | M   | No     |
| High   | yes     | Old   | M   | Yes    |

**Test Data**

| Income | Student | Age   | Sex | Credit    |
|--------|---------|-------|-----|-----------|
| Medium | yes     | Old   | F   | <b>NO</b> |
| Low    | no      | Young | M   | <b>NO</b> |

**Solution:**

**Sim(T1,1)= 0.5 (NO)**

Sim(T1,2)= 0.25

Sim(T1,3)= 0.25

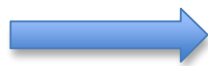
**Sim(T1,4)= 0.75 (NO)**

**Sim(T1,5)= 0.75 (YES)**

Sim(T1,6)= 0.5

Sim(T1,7)= 0.5

Sim(T1,8)= 0.5



NO

**Sim(T2,1)= 0.75 (NO)**

Sim(T2,2)= 0.5

**Sim(T2,3)= 1 (NO)**

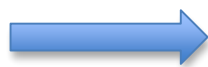
Sim(T2,4)= 0

Sim(T2,5)= 0.5

**Sim(T2,6)= 0.75 (NO)**

Sim(T2,7)= 0.75

Sim(T2,8)= 0.25



NO