Heinrich Heine Universität Düsseldorf

HEIDERLAB

Exercise 6 on Machine Learning WS 2023/24
Prof. Dr. Dominik Heider
M. Sc. Arsam (Mohammad) Tajabadi
Submission: Until 06.12.2023, 23:59 on Ilias

---

## Task 1. K-nearest neighbor (10 points)

**Note:** It is recommended to do this exercise by hand.

a) Given the following data set: (7 P.)

| Day | Outlook | Temperature | Humidity | Wind | PlayTennis |
|---|---|---|---|---|---|
| D1 | Sunny | 26 | High | | No |
| D2 | Sunny | 28 | High | Strong | No |
| D3 | Overcast | 29 | High | Weak | Yes |
| D4 | Rain | 23 | High | Weak | Yes |
| D5 | Rain | | Normal | Weak | Yes |
| D6 | Rain | 12 | Normal | Strong | No |
| D7 | Overcast | 8 | | Strong | Yes |
| D8 | Sunny | 25 | High | Weak | No |
| D9 | Sunny | 18 | Normal | Weak | Yes |
| D10 | Rain | 20 | Normal | Weak | Yes |
| D11 | Sunny | 20 | Normal | Strong | |
| D12 | Overcast | 21 | High | Strong | Yes |
| D13 | | 26 | Normal | Weak | Yes |
| D14 | Rain | 24 | High | Strong | No |
| D15 | Sunny | 23 | Normal | Weak | No |
| D16 | Sunny | 21 | Normal | Weak | Yes |

Use the Nearest Neighbors method to determine missing values. Choose k = 3. Normalize the attributes to [0, 1]. Use the Manhattan metric for distance or the 0/1 distance for nominal attributes.

$$d_A(v_1, v_2) = \begin{cases} 0 & \text{if } v_1 = v_2 \\ 1 & \text{if } v_1 \neq v_2 \end{cases} \tag{1}$$

b) Do the classification labels (PlayTennis) have to be included? Why or why not? (1 P.)

c) Classify the new sample D17 for k=1. (1 P.)
D17: Outlook=Sunny, Temperature=23, Humidity=High, Wind=Strong

d) Test different values of k. At what value of k does the assignment change compared to k=1? (1 P.)