BIRZEIT UNIVERSITY

Electrical and Computer Engineering Department

SP.TOP: Machine Learning - ENCS5396

Assignment #1

Submission deadline: 6.5.2022

1. **Binary Decision Trees [15]**

   Consider the following set of training examples

Outlook	Temperature	Wind	Enjoy Sport
Sunny	High	Weak	-
Sunny	High	Strong	-
Overcast	High	Weak	-
Rain	Low	Weak	+
Rain	Low	Weak	+
Rain	Low	Strong	+
Overcast	Low	Strong	+
Sunny	High	Strong	+
Sunny	Low	Weak	+
Overcast	High	Strong	-
Rain	High	Strong	-

   Construct a binary decision tree using the information gain for attribute selection. Write down all your calculation steps as well as the final result tree.

2. **K-NN [15]**

   Consider the `data.csv` file, a modified dataset from the UCI machine learning repository. It has 9 attributes and a binary target attribute for the prediction value.

   - The first line contains a sequence of attribute type declarations of the form:

     $$a_1: T_1 , a_2: T_2 , \ldots, a_9: T_9 , a_{10}: T_{10}$$

     where a_i is the id of attribute I and T_i is its type ($1 \leq i \leq 10$). All attributes have one of the following types:
     - "n": numerical
     - "c": categorial with at least three attribute values
     - "b" and "t" (target): binary with attribute values "yes" and "no"

     In the `data.csv` file, the first line looks like this:

a:n, b:c, c:c, d:n, e:b, f:b, g:c, h:c, i:b, j:t

The last attribute "j" is the target attribute and is binary.

- Each subsequent line describes a labelled example, e.g.,

  24, bb, cc, 3, yes, no, gb, hc, yes, yes
  33, bd, cc, 5, yes, no, gb, hd, yes, no

Tasks:

(a) Design a distance function for this dataset.
(b) Select 5 instances at random and calculate for each one the k nearest neighbors for k = 1, 2, 3
(c) Select parameter k $\in$ {1, 2, 3, 4, 5} by using 10-fold cross validation.