National University of Singapore
School of Computing

BT2101 Decision Making Methods and Tools

Tutorial 1:
Decision Tree

Read Chapter 3, 4 and 5 of "Principles of Data Mining" before answering the following questions.

RMS Titanic sank in the Atlantic Ocean during its maiden voyage from the UK to New York City after colliding with an iceberg in 1912. Frightened people scrambling to escape a sinking ship in chaos. With an inadequate number of lifeboats available only a fraction of the passengers survived. The disaster is famous for saving "women and children first".

In the dataset, there is survival information of each passenger. We also collected additional attributes (features) (i.e., age, passenger class, gender, number of siblings or spouses on board, and number of parents or children on board), to predict whether a certain passenger will survive.

ID	Age Class	Passenger Class	Gender	No of Siblings or Spouses on Board	No of Parents or Children on Board	Survived
1	Young Adulthood	Third	Male	0	0	No
2	Young Adulthood	Second	Male	0	0	No
3	Childhood	Third	Female	>=2	>=2	Yes
4	Childhood	Third	Male	0	0	No
5	Middle Adulthood	Second	Female	0	0	No
6	Elder Citizen	First	Male	0	0	No
7	Middle Adulthood	First	Female	1	0	Yes
8	Childhood	Third	Male	1	1	No
9	Young Adulthood	Third	Male	0	0	No
10	Childhood	Second	Male	0	0	No
11	Young Adulthood	Third	Male	0	0	No
12	Young Adulthood	Second	Female	1	0	Yes
13	Childhood	Third	Female	1	>=2	Yes

14	Young Adulthood	First	Male	1	0	Yes
15	Young Adulthood	Third	Male	>=2	0	No
16	Elder Citizen	Third	Male	0	0	No
17	Childhood	Third	Female	0	0	Yes
18	Young Adulthood	Third	Male	1	>=2	No
19	Young Adulthood	Third	Male	1	0	Yes
20	Childhood	Second	Female	0	1	Yes
21	Young Adulthood	Second	Female	0	0	No
22	Young Adulthood	Second	Male	>=2	1	No
23	Young Adulthood	Third	Male	0	0	No
24	Elder Citizen	Third	Male	0	0	No
25	Childhood	Third	Male	>=2	>=2	No
26	Young Adulthood	First	Female	0	0	Yes
27	Middle Adulthood	First	Female	0	0	Yes
28	Young Adulthood	Third	Male	0	0	No
29	Young Adulthood	First	Female	1	0	Yes
30	Young Adulthood	Second	Female	1	0	Yes

Q1: Calculate the information gain for all five attributes (Age Class, Passenger Class, Gender, No of Siblings or Spouses on Board, No of Parents or Children on Board).

Q2. Based on your answer to the question above, what root attribute would ID3 select to split upon for the Titanic data?

Q3: Suppose you use Gini index instead, what root attribute would decision tree model select to split upon for the Titanic data? Why?

Submission Instructions:

Upload your answers to IVLE Tutorial1_Paper folder and name it: matric#_t1_paper i.e. AXXXXX_t1_paper
You can submit either pdf or excel

Submit by Aug 28 before Tutorial 1 (by midnight)