

REVIEW EXERCISE 08

Question 1. Roger Federer is one of the greatest tennis players since tennis have been invented. We want to learn a little about what makes Federer win or lose a match. To do so, we gathered data from games played by R. Federer, shown in the below table.

ID	Time	Match Type	Court Surface	Best Effort	Outcome
1	Morning	Master	Grass	True	Win
2	Afternoon	Grand slam	Clay	True	Win
3	Night	Friendly	Hard	False	Win
4	Afternoon	Friendly	Mixed	False	Lose
5	Afternoon	Master	Clay	True	Lose
6	Afternoon	Grand slam	Grass	True	Win
7	Afternoon	Grand slam	Hard	True	Win
8	Afternoon	Grand slam	Hard	True	Win
9	Morning	Master	Grass	True	Win
10	Afternoon	Grand slam	Clay	True	Lose
11	Night	Friendly	Hard	False	Win
12	Night	Master	Mixed	True	Lose
13	Afternoon	Master	Clay	True	Lose
14	Afternoon	Master	Grass	True	Win
15	Afternoon	Grand slam	Hard	True	Win
16	Afternoon	Grand slam	Clay	True	Win

- a. Build a classification model using ID3 decision tree from the gathered data.

The entropy of the whole dataset

- $H(\text{Dataset}) =$

The information gain of the attribute Time

- $H(\text{Time} = \text{Morning}) =$
- $H(\text{Time} = \text{Afternoon}) =$
- $H(\text{Time} = \text{Night}) =$
- $AE(\text{Time}) =$
- $IG(\text{Time}) =$

The information gain of the attribute Match Type

- $H(\text{Match Type} = \text{Master}) =$
- $H(\text{Match Type} = \text{Grand slam}) =$

- $H(\text{Match Type} = \text{Friendly}) = \dots\dots\dots$
- $AE(\text{Match Type}) = \dots\dots\dots$
- $IG(\text{Match Type}) = \dots\dots\dots$

The information gain of the attribute Court Surface

- $H(\text{Court Surface} = \text{Grass}) = \dots\dots\dots$
- $H(\text{Court Surface} = \text{Hard}) = \dots\dots\dots$
- $H(\text{Court Surface} = \text{Clay}) = \dots\dots\dots$
- $H(\text{Court Surface} = \text{Mixed}) = \dots\dots\dots$
- $AE(\text{Court Surface}) = \dots\dots\dots$
- $IG(\text{Court Surface}) = \dots\dots\dots$

The information gain of the attribute Best Effort

- $H(\text{Best Effort} = \text{True}) = \dots\dots\dots$
- $H(\text{Best Effort} = \text{False}) = \dots\dots\dots$
- $AE(\text{Best Effort}) = \dots\dots\dots$
- $IG(\text{Best Effort}) = \dots\dots\dots$

The root attribute will be $\dots\dots\dots$

Repeat for every branch of the root attribute that has examples not fully classified into a single class

- b. Knowing the conditions in which a tennis match takes place, we would like to predict whether R. Federer will win or lose the match, using ID3 decision tree in a.

Time	Match type	Court Surface	Best Effort	Outcome
Morning	Grand Slam	Grass	True	?
Afternoon	Friendly	Clay	False	?

Question 2. Repeat the question above by using naïve Bayes (with or without Laplacian correction).