

Tutorial Business Analytics

Tutorial 6

Exercise 6.1

Calculate:

- $\text{entropy}(0.1, 0.9)$
- $\text{entropy}(0.8, 0.2)$
- $\text{entropy}(0.3, 0.7)$
- $\text{entropy}(0.5, 0.5)$
- $\text{entropy}(0.8, 0.1, 0.1)$

Exercise 6.2

Calculate:

- $\text{info}([2, 3])$
- $\text{info}([5, 4])$
- $\text{info}([2, 3], [5, 4])$
- $\text{info}([2, 3], [9, 0])$

Exercise 6.3

Construct a tree:

Temperature	Visibility	Snow depth	Sport
<(-5)	Clear	>=50	Skiing
<(-5)	Fog	>=50	Swimming
<(-5)	Fog	<50	Swimming
<(-5)	Rain	>=50	Skiing
<(-5)	Rain	<50	Swimming
>=(-5)	Clear	>=50	Skiing
>=(-5)	Clear	<50	Skiing
>=(-5)	Fog	<50	Swimming
>=(-5)	Rain	>=50	Skiing

Exercise 6.4

Construct a tree:

The data record has an ID attribute now

ID	Temperature	Visibility	Snow depth	Sport
A	<(-5)	Clear	>=50	Skiing
B	<(-5)	Fog	>=50	Swimming
C	<(-5)	Fog	<50	Swimming
D	<(-5)	Rain	>=50	Skiing
E	<(-5)	Rain	<50	Swimming
F	>=(-5)	Clear	>=50	Skiing
G	>=(-5)	Clear	<50	Skiing
H	>=(-5)	Fog	<50	Swimming
I	>=(-5)	Rain	>=50	Skiing

Exercise 6.5

Construct the tree from exercise 6.4 a second time using gain ratio:

ID	Temperature	Visibility	Snow depth	Sport
A	<(-5)	Clear	>=50	Skiing
B	<(-5)	Fog	>=50	Swimming
C	<(-5)	Fog	<50	Swimming
D	<(-5)	Rain	>=50	Skiing
E	<(-5)	Rain	<50	Swimming
F	>=(-5)	Clear	>=50	Skiing
G	>=(-5)	Clear	<50	Skiing
H	>=(-5)	Fog	<50	Swimming
I	>=(-5)	Rain	>=50	Skiing

Exercise 6.6

Find the optimal binary splits.

a)

60	60	120	120	180	180	180
F	F	T	F	F	T	T

b)

5	5	7	7	7	8	9	9
T	T	T	T	F	T	F	F