

Knowledge Discovery in Databases II
 SS 2018

Exercise 7: Data Stream Classification

Exercise 7-1 Hoeffding trees

Predict the risk class of a car driver based on the following attributes:

- Time since getting the driving license (1 – 2 years, 2 – 7 years, > 7 years)
- Gender (male, female)
- Residential area (urban, rural)

These are the first 8 examples.

Person	Time since license	Gender	Area	Risk class
1	1 – 2	m	urban	low
2	2 – 7	m	rural	high
3	> 7	f	rural	low
4	1 – 2	f	rural	high
5	> 7	m	rural	high
6	1 – 2	m	rural	high
7	2 – 7	f	urban	low
8	2 – 7	m	urban	low

- Incrementally construct a Hoeffding tree for this example.
 Use information gain and $\delta = 0.2$ and $N_{\min} = 2$.
- Compute the value of δ at which the tree would still consist of the leaf only.

Exercise 7-2 Cohen's Kappa

Gegeben seien die folgenden Konfusionsmatrizen zu den Zeitpunkten $t = 1, 2, 3$:

	$t = 1$			$t = 2$			$t = 3$	
	positiv	negativ		positiv	negativ		positiv	negativ
positiv	37	14	positiv	65	8	positiv	90	4
negativ	17	32	negativ	7	20	negativ	5	1

Berechnen Sie Accuracy und Cohen's Kappa, und vergleichen Sie die Ergebnisse.