CS-E4650 - Assignment 3

Arttu Häkkinen - 596077

November 9, 2021

Task 2

a)

Firstly I used the exact same binarization as was used in exercise 3 task 2. I did not include the ratIDs and kept motherhood and gender separate.

Ex. 3 task 2 attribute descriptions and how they were extracted:

- summer if day 116...300 and winter if day 1...115 or 301...365
- freezer if day=0
- weightlow if weight$\leq$162 and weightnormal otherwise (these probably correspond to puppies/young and adult rats)
- non-mother-female, mother-female and male derived from features gender and femstate (you could keep gender and motherhood separate (I did!))
- liversmall if liverind$\leq$0.037 and liverlarge if liverind$>$0.064
- pregnant and nursing from femstate
- batlow if batind$\leq$0.00067 and bathigh if batind$>$0.00184
- tailshort if tailind$\leq$0.74 and taillong if tailind$>$0.85
- place1,..., place9: one binary attribute for each place
- wild if place=1-3, and lab otherwise
- adralsmall if ADWBind$\leq$0.21 and adrenalarge if ADWBind$>$0.48
- BMIsmall if BMI$\leq$0.47 and BMIlarge if BMI$>$0.75

The attributes that I invented myself and how they were extracted:

- bigheart if heartind $>$ 0.006, smallheart if heartind$\leq$0.003
- high-appind if appind$>$0.025, normal-appind otherwise

- high-food-intake if gonfatind > 0.015, normal-food-intake otherwise
- has-sulcer if sulcer > 1, no-sulcer otherwise
- kmethod1,..., kmethod4: one binary attribute for each place
- adult if blength$\geq$15, otherwise child
- extra-melatonin if place=7 or place=8, no-extra-melatonin otherwise
- late-decade if year > 5, early-decade otherwise

So in total, I had 44 attributes.

b)

10 most interesting rules are in the table below. I selected them by trying to find rules with the best significance and so that they would make some sense without making too much sense and base only on the description of the data. Basically, I tried to avoid reporting any rules that were obvious after the binarization made in the part a. Many of the listed rules had similar combinations, but I chose the most interesting conditioning set and consequent combinations in order to have meaningful interpretations about certain things in part c. The non-default command parameters when using kingfisher from the command line were -M-10 and -k100: So I used goodness measure threshold -10 and looked at the top 100 associations/rules found by the program.

	rule	fr_X	cf.	lift γ	lev. δ	$ln(p_F))$
1	late-decade $\rightarrow$ place3	349	1.000	1.607	0.235	-3.686e+02
2	wild kmethod1 $\rightarrow$ late-decade	332	0.965	1.551	0.210	-2.614e+02
3	wild early-decade $\rightarrow$ kmethod3	120	0.909	3.273	0.156	-2.013e+02
4	winter kmethod3 early-decade $\rightarrow$ place2	62	0.969	7.246	0.095	-1.520e+02
5	early-decade summer kmethod1 $\rightarrow$ place9	46	0.836	10.200	0.074	-1.337e+02
6	kmethod4 $\rightarrow$ freezer	34	1.000	16.500	0.057	-1.256e+02
7	female high-food-intake lab $\rightarrow$ kmethod4	33	0.767	12.663	0.054	-9.798e+01
8	high-food-intake, lab, smallheart $\rightarrow$ kmethod4	30	1.000	16.500	0.050	-1.037e+02
9	liversmall smallheart $\rightarrow$ kmethod4	28	1.000	16.500	0.047	-9.455e+01
10	extra-melatonin $\rightarrow$ kmethod4	26	1.000	16.500	0.044	-8.601e+01

c)

From the table in last section one can see that many of the rules tell us something about the methods the rat were killed. Firstly, from these and the other non-reported associations, it was clear to see that only 3 different killing methods were common, and had reasonable associations related to them. Even after the analysis, killing method 1 remains as the biggest mystery. We know that if the rat was wild and died with method 1, it was most likely a late-decade

rat. On the other hand, if the rat was a early decade rat born in summer, and it died with killing method 1 it was most likely from place 9 which was a laboratory. Based on these facts is hard to say anything else than the fact that this killing method did as likely happen to lab or wild rat, but for a wild rat in the late-decades and for the lab rat in the early-decade.

More interesting findings were made when it comes to killing method 3. based on our evidence, which are the rules, we can see that when late-decade wild rats were likely to die with method1, the early-decade wild rats were likely to die with method 3. In fact, killing method 3 didn't relate to any lab rats in the top 100 rules returned by the kingfisher program. It gets even more interesting when we look at the rule number 4. It clearly states that rats that died with method 3 in the early decades, in the winter, were rats from the place2 which is the fur farm. This is concerning information since it tells that many rats died in the winter because of bad circumstances. For animal activists this information would be useful to know. Also it is quite natural since it is clear that rats in the fur farm don't have the protecting fur against the cold so they were more likely to die to coldness. So, basically with this information we could assume that the killing method 3 is freezing to death or something similar.

Killing method 4 was clearly related to lab rats and it didn't have many associations with the wild ones. Also if a rat died with method 4, it was a freezer rat with high confidence. The following four attributes were clearly associated with the killing method 4: Gender was female, high food intake/high gonadal fat index, small heart index, small liver and extra-melatonin received. Based on these facts one could assume that the killing method 4 is some disease, e.g. a cardiovascular disease which would like cause death if the gonadal fat index was high and liver and the heart were small. The fact that female rats and rats that received extra melatonin died with this method often, is quite interesting finding in my opinion. It is hard to say it directly, an would lead to further investigations.

References