2 Task 2

2.1 Task2.a

We need to extract binary features from the Rat data and save the data in transaction form. From Exercise 3 Task 2, there are some features are already extracted,but remaining features should extract by ourselves. And after extraction of features, we convert the original rat data into transaction form for searching association rules with Kingfisher later.

The rules of extracting data features (some are original from *description.txt* in E3T2) are shown below:

1. Ratid (not needed in rule discovery, but doesn't harm, helps to identify rats if needed; note that each id is now a binary attribute)

2. **Summer** if day 116-300 and **winter** if day 1-115 or 301-365.

3. **Freezer** if day=0

4. **weightlow** if weight≤162 and **weightnormal** otherwise (these probably correspond to puppies/young and adult rats)

5. Gender. **female, male**

6. Femstate. Accoring to 1-4, the features are **pregnant, nursing, pregnant+nursing, femstate_neither**.

7. **Liversmall** if liverind≤0.037 and **liverlarge** if liverind>0.064, in between are **livernormal**.

8. **Batlow** if batind≤0.00067 and **bathigh** if batind>0.00184, in between are **batnormal**.

9. **tailshort** if tailind ≤ 0.74 and **taillong** if tailind > 0.85,in between are **tailnormal**.

10. **wild** if place=1-3, and **lab** otherwise.

11. **adrenalsmall** if ADWBind≤0.21 and **adrenalarge** if ADWBind>0.48, in between are **adrenalnormal**.

12. **BMIsmall** if BMI ≤ 0.47 and **BMIlarge** if BMI > 0.75, in between are **BMInormal**.

13. **heartsmall** if heartind ≤0.0036 **heartlarge** if heartind >0.0046, in between are **heartnormal**.

14. **appsmall** if appind ≤0.0104, **applarge** if appind >0.0167, in between are **appnormal**.

15. **gonsmall** if gonfatind ≤0.0018, **gonfat** if gonfatind >0.0103.in between are **gonnormal**.

16. sulcer. **sulcermild** ≤ 4, **sulcerserious** > 4.

17. kmethod. **According to 1-5, one attribute for each method**.

18. blength. **blengthshort** ≤ 19.5 , **blengthlong** > 22.5, in between are **blengthnormal**.

So after extraction of features, we obtain [**summer, winter, freezer, weightlow, weightnormal, female, male, pregnant, nursing, pregnant+nursing, femstateneither, liversmall, liverlarge, batlow, batnormal, bathigh, tailshort, tailnormal, taillong, wild, lab, adrenalsmall, andadrenalarge, adrenalnormal, BMIsmall, BMIlarge, BMInormal, heartsmall, heartlarge, heartnormal, appsmall, applarge, appnormal, gonsmall, gonfat, gonnormal, sulcermild, sulcerserious, kmethod1, kmethod2, kmethod3, kmethod4, kmethod5, blengthshort, blengthnormal, blengthlong**], totally 46 features.

2.2   Task2.b

The threshold of Kingfisher search in rat data are $In(P_F)$ is set to -50 with option -M. The number of rules is not defined, so it will output by default max 100 rules having $ln(P_F) \leq -50$. It searches for both positive and negative dependencies. An upperbound for the number of attributes is set to 700 with option -k. From all rules generated in program, we have chosen to keep the rules of practical significance in the following Table.3.

Table 3: Selected rules in rat data

Rule	fr_X	Cf	γ	δ	$In(P_F)$
1 weightlow $\rightarrow$ blengthshort	99	1.0	3.484	0.126	-153.7
2 lab $\rightarrow$ kmethod4	34	1.0	16.5	0.057	-125.6
3 femstate_neither gonfat heartsmall blengthnormal $\rightarrow$ lab	33	0.846	13.962	0.055	-104.4
4 femstate_neither weightnormal gonfat heartsmall $\rightarrow$ lab	33	0.846	13.962	0.055	-104.4
6 BMIsmall $\rightarrow$ weightlow	63	0.808	4.577	0.088	-96.9
7 female liversmall gonfat blengthnormal $\rightarrow$ lab	29	0.967	15.95	0.048	-95.6
8 female liversmall gonfat blengthnormal $\rightarrow$ kmethod4	29	0.967	15.95	0.048	-95.6
9 female liversmall gonfat heartsmall $\rightarrow$ lab	29	0.935	15.543	0.048	-92.9
10 female sulcermild $\rightarrow$ femstate_neither	155	0.442	1.598	0.103	-89.6
11 sulcermild weightnormal $\rightarrow$ blengthnormal	291	0.634	1.222	0.094	-86.9
12 blengthnormal $\rightarrow$ weightnormal	291	1.0	1.241	0.092	-83.9
13 heartlarge blengthshort $\rightarrow$ BMIsmall	53	0.639	4.593	0.074	-75.7

2.3   Task2.c

Now from the rules in the table above, we can summarize some interesting findings. The features in these rules 1, 7, 13, 14 all have some linear relationship with each other. It is natural to think that after body length becomes larger, the weight must also become larger, so the BMI will also change with it. These characteristics will change together in general. From rule 2, we can speculate that the rats produced in the experimental environment are killed by specific way. This is because rats can be killed more easily in a laboratory environment,

whereas rats in a wild environment may not be treated according to laboratory methods because of bacteria and other problems. From rules 2,3,8,9, we can observe that rats in a laboratory environment, even though their weight and height were normal, were more likely to have smaller hearts and to have more gonadal fat suggesting their excessive food intake. My guess is that they do not have enough space to move around in the lab and are fed regularly every day. So they became lazy and inactive and their hearts did not get exercise. Compared to the rats in the wild environment, they had a better food intake situation every day. From rule 13, we can observe that female rats with mild stomach ulcer are likely not in any specific female state, indicating Rats who are pregnant or breastfeeding are more likely to develop stomach ulcers.