COMP9321 Sample Exam Solution

Apr 17,2019

Part One

1. D
2. D
3. D
4. C
5. C
6. A
7. B
8. D
9. B
10. B
11. B
12. C
13. C
14. D
15. C
16. B
17. C
18. D
19. A
20. A

Part Two

1. Main difference between POST and GET method is that GET carries request parameter appended in URL string while POST carries request parameter in message body which makes it more secure way of transferring data from client to server in http protocol. e.g., log in

2. Data quality problems which contains incomplete data,noisy data,inconsistent (and duplicate) data will affect the analyze result. Data cleansing is used to deal with that.

3. Transformation, Standardization... (More in lecture 2.1)

4. Use out-of-bag error as the estimate of the generalization error. Measuring variable importance through permutation.

5. Classification is the process of classifying the data with the help of class labels. Clustering is similar to classification but there are no predefined class labels.Classification is geared with supervised learning. As against, clustering is also known as unsupervised learning.

Part Three

Association Rule Mining

1.
minimal support is $\frac{2}{9} * 9 = 2$ which means the support for a frequent itemset must larger than or equal to 2. so the $C_1 = \{M1, M2, M3, M4, M5\}$ the support is:

M1	M2	M3	M4	M5
6	7	6	2	2

So $L_1 = \{M1, M2, M3, M4, M5\}$
C_2 is generated from L_1 by enumerating all pairs as: $\{(M1, M2), (M1, M3), (M1, M4), (M1, M5), (M2, M3), (M2, M4), (M2, M5), (M3, M4), (M3, M5), (M4, M5)\}$ scan the db to get the support:

M1,M2	M1,M3	M1,M4	M1,M5	M2,M3	M2,M4	M2,M5	M3,M4	M3,M5	M4,M5
4	4	1	2	4	2	2	0	1	0

Therefore the $L_2 = \{(M1, M2), (M1, M3), (M1, M5), (M2, M3), (M2, M4), (M2, M5)\}$
C_3 is generated from L_2 by enumerating all pairs as:$\{(M1, M2, M3), (M1, M2, M5), (M1, M3, M5), (M2, M3, M4), (M2, M3, M5), (M2, M4, M5)\}$, scan to db and do the pruning we get:

M1,M2,M3	M1,M2,M5
2	2

We stop here, if we go further to C_4 which is $\{M1, M2, M3, M5\}$, you will find the support is 1 which is less than the minimal support.
2.
We list the frequent itemsets related to $M1, M2, M5$:

M1	M2	M5	(M1,M2)	(M1,M5)	(M2,M5)	(M1,M2,M5)
6	7	2	4	2	2	2

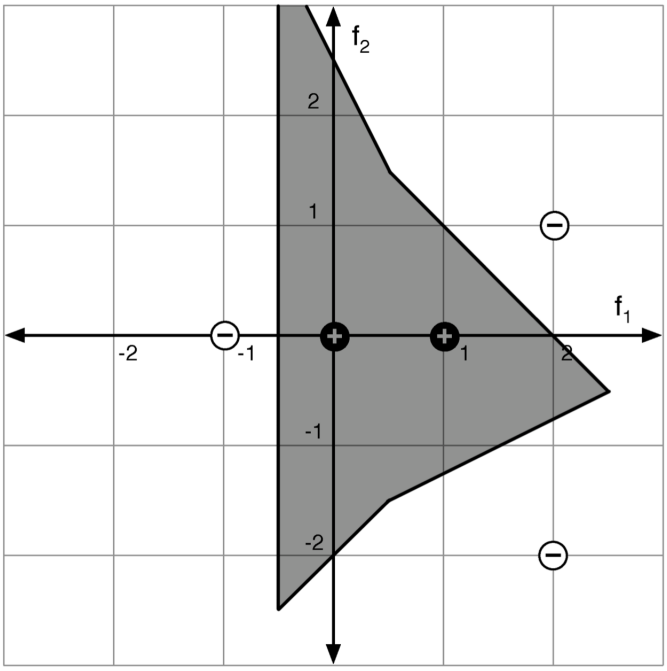
So what we will do is try all the permutation like: $M1 \to M2, M1 \to M5, M2 \to M5$.... Consider about the case $M1 \to M2$, as the formula to calculate the confidence is: $\frac{\text{supp}(M1, M2)}{\text{supp}(M1)} = \frac{4}{6}$ which less than the minimal confident. It's easy to find that the $M2$ has highest support which means that any rule in that form: $M2 \to ...$ will not meet the minimal confident. As well as the $M1 \to ....$ Therefore, we can repeat the above procedure and find the following rules:

$$M5 \to (M1, M2)$$
$$(M1, M5) \to M2$$
$$(M2, M5) \to M1$$

Which all the rules have confident with $\frac{2}{2} = 1$ which larger than $\frac{7}{9}$

kNN

1. Your solution may vary. If you can successfully divide it into two different classes, will be full mark



2. Positive (+) since this is the class of the closest data point (1,0).
3. Positive (+) since it is the majority class of the three closest data points (0,0), (1,0) and (2,-2).