**Data Mining – Assignment 2**

**1930026143 薛劭杰**

Q1.

For {e}:

, where is the number of transactions.

For {b, d}

, where is the number of transactions.

For {b, d, e}

, where is the number of transactions.

(i).

|  |  |
| --- | --- |
| Item | Support |
|  | 5 |
|  | 3 |
|  | 4 |
|  | 5 |
|  | 4 |

For : , so it should be pruned.

|  |  |
| --- | --- |
| Item | Support |
|  | 4 |
|  | 5 |
|  | 4 |
|  | 4 |
|  | 3 |
|  | 4 |

For : , so it should be pruned.

|  |  |
| --- | --- |
| Item | Support |
|  | 4 |
|  | 4 |
|  | 3 |

For : , so it should be pruned.

|  |  |
| --- | --- |
| Item | Support |
|  |  |

(ii).

|  |  |
| --- | --- |
| Item | Count |
| {a} | 5 |
| {c} | 4 |
| {d} | 4 |
| {e} | 4 |

|  |  |
| --- | --- |
| Item | Count |
| {a, c} | 4 |
| {a, d} | 5 |
| {a, e} | 4 |
| {c, d} | 4 |
| {d, e} | 4 |

|  |  |
| --- | --- |
| Item | Count |
| {a, c, d} | 4 |
| {a, d, e} | 4 |

|  |  |
| --- | --- |
| Item | Count |
|  |  |

(c).

1

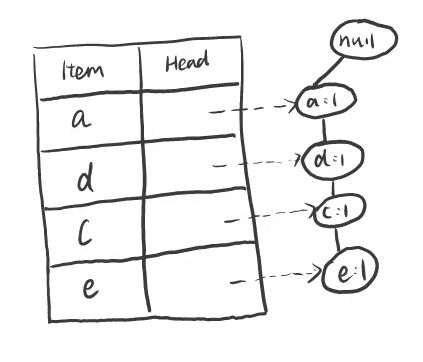
the list of strong rules are:

(d).

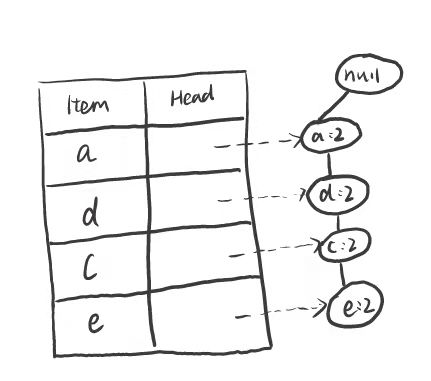
At the beginning, we should modify the database by removing the infrequent items. According to the (b) in this question, is infrequent so it should be removed. Then sort the remaining items in every transaction according to the count order.

|  |  |
| --- | --- |
| Transaction ID | Items Bought |
| T1 |  |
| T2 |  |
| T3 |  |
| T4 |  |
| T5 |  |

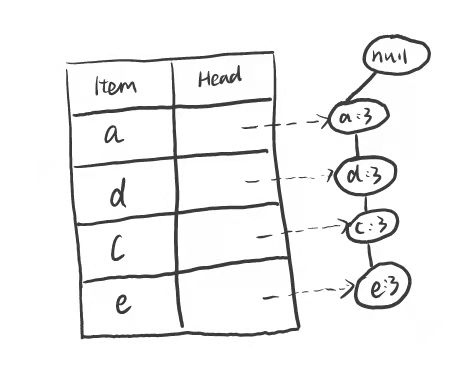
Create a branch from the root and insert T1:



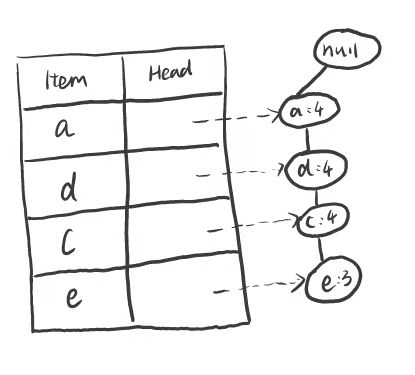
Insert T2:



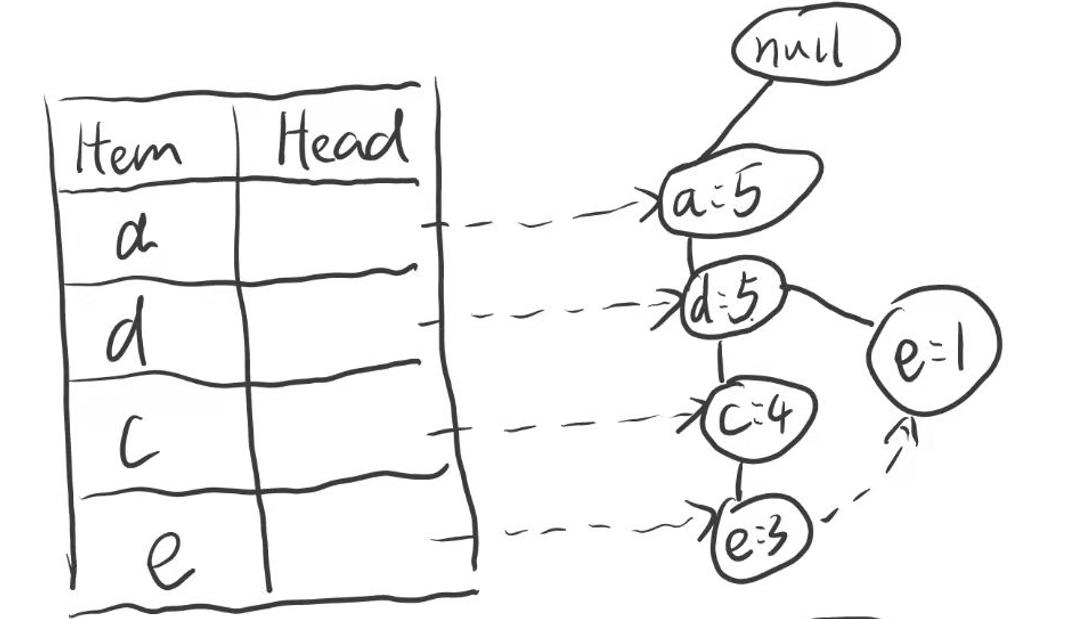
Insert T3:



Insert T4:

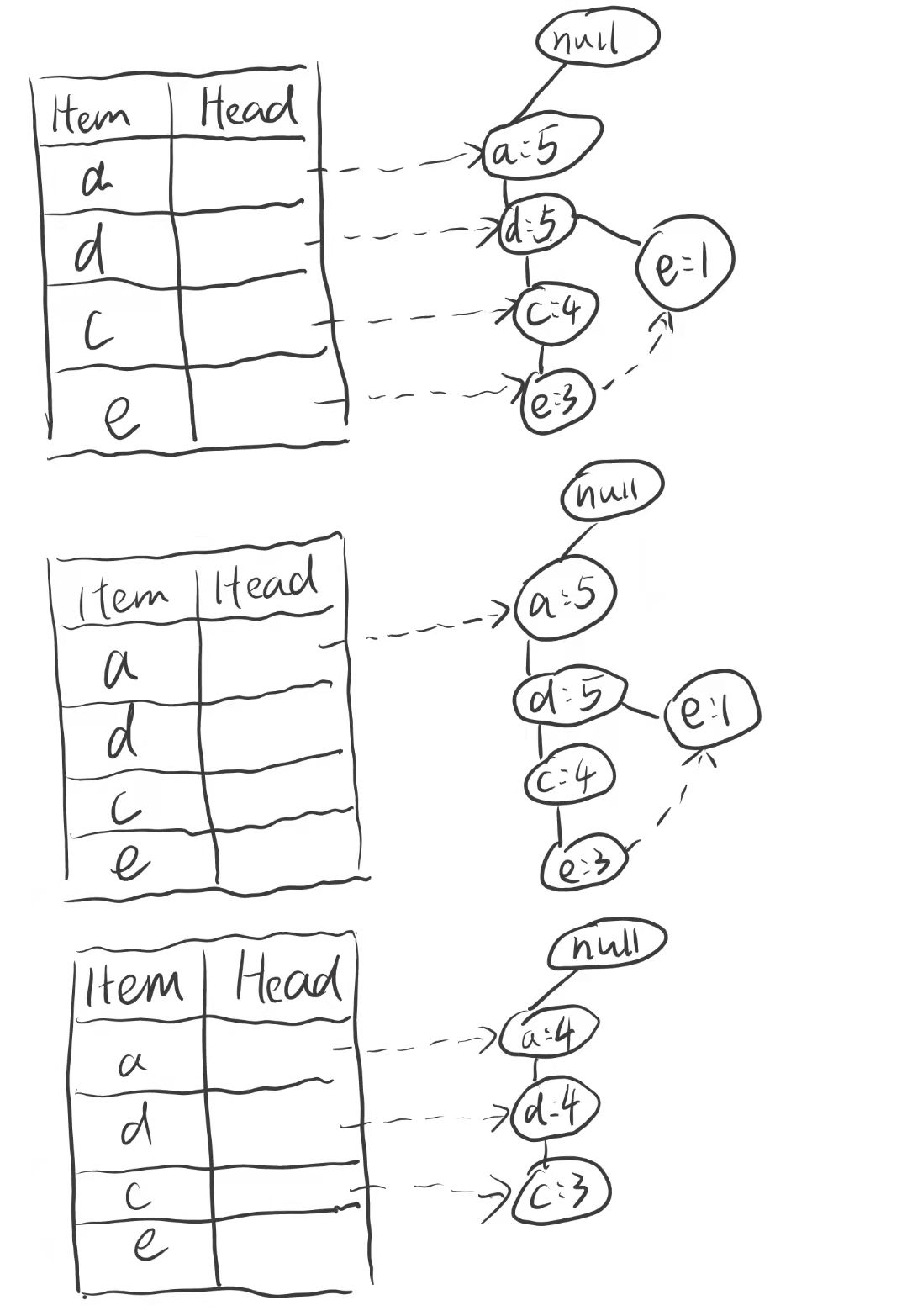


Insert T5:

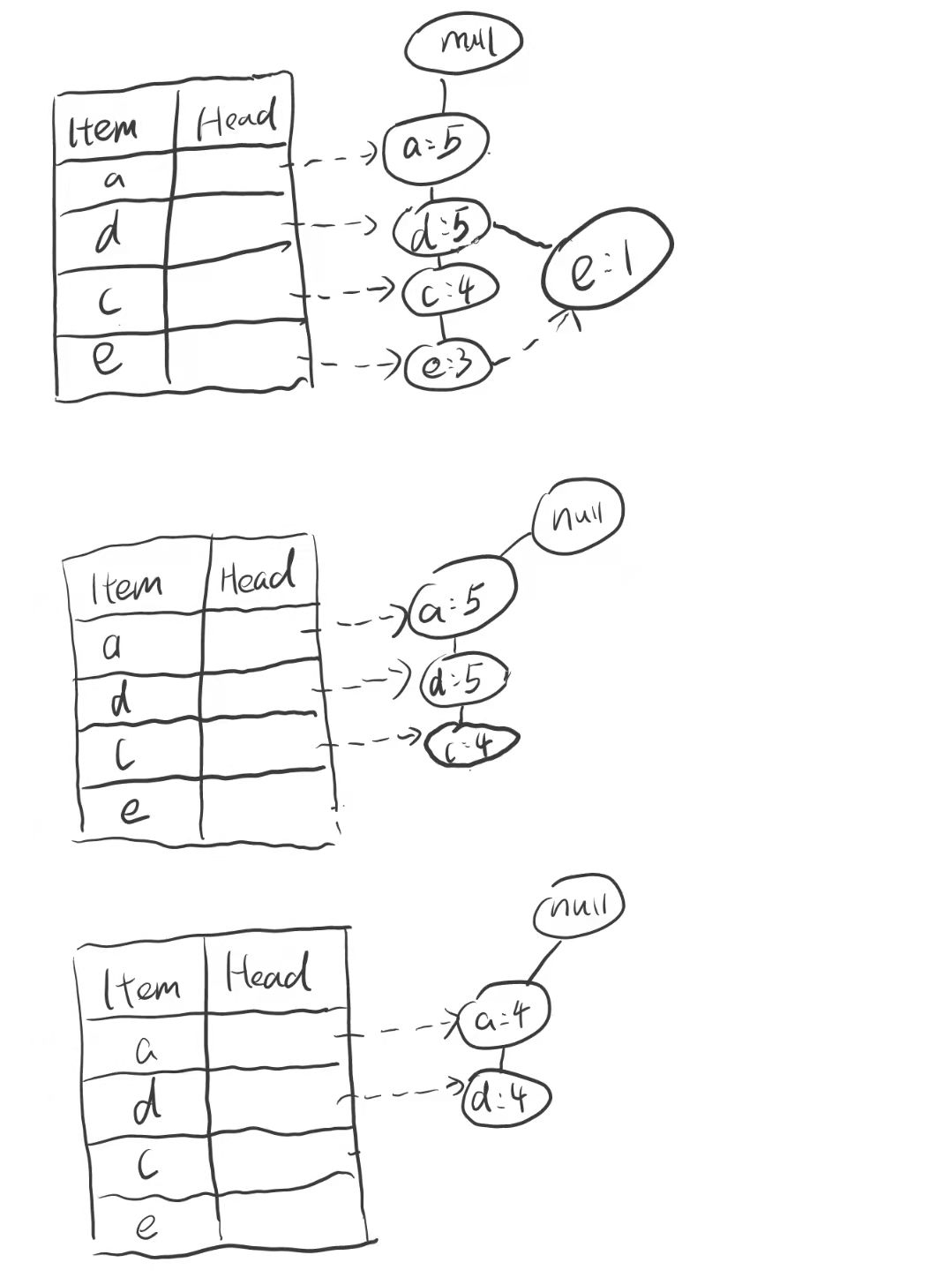




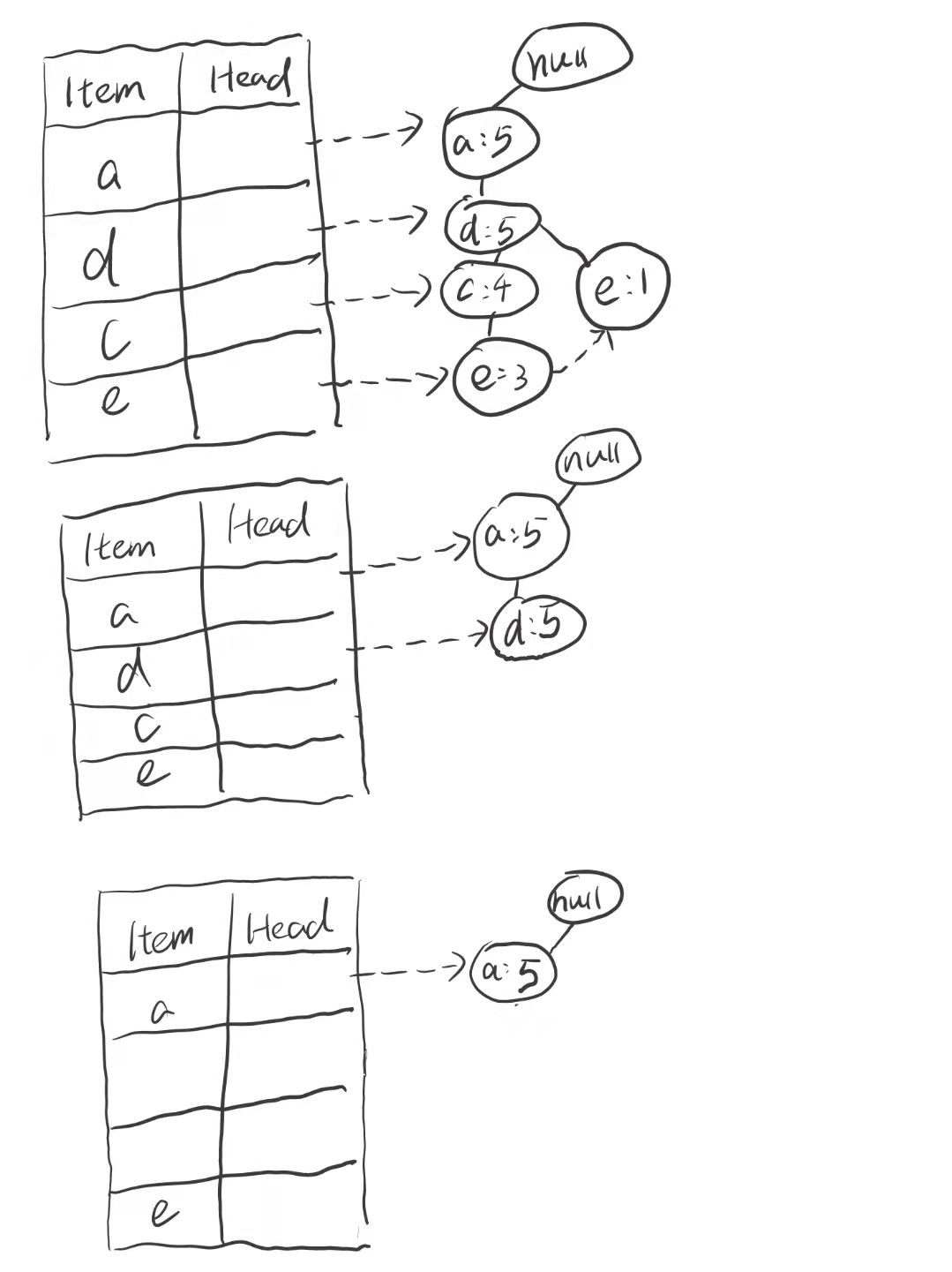
Generate pattern {e}:



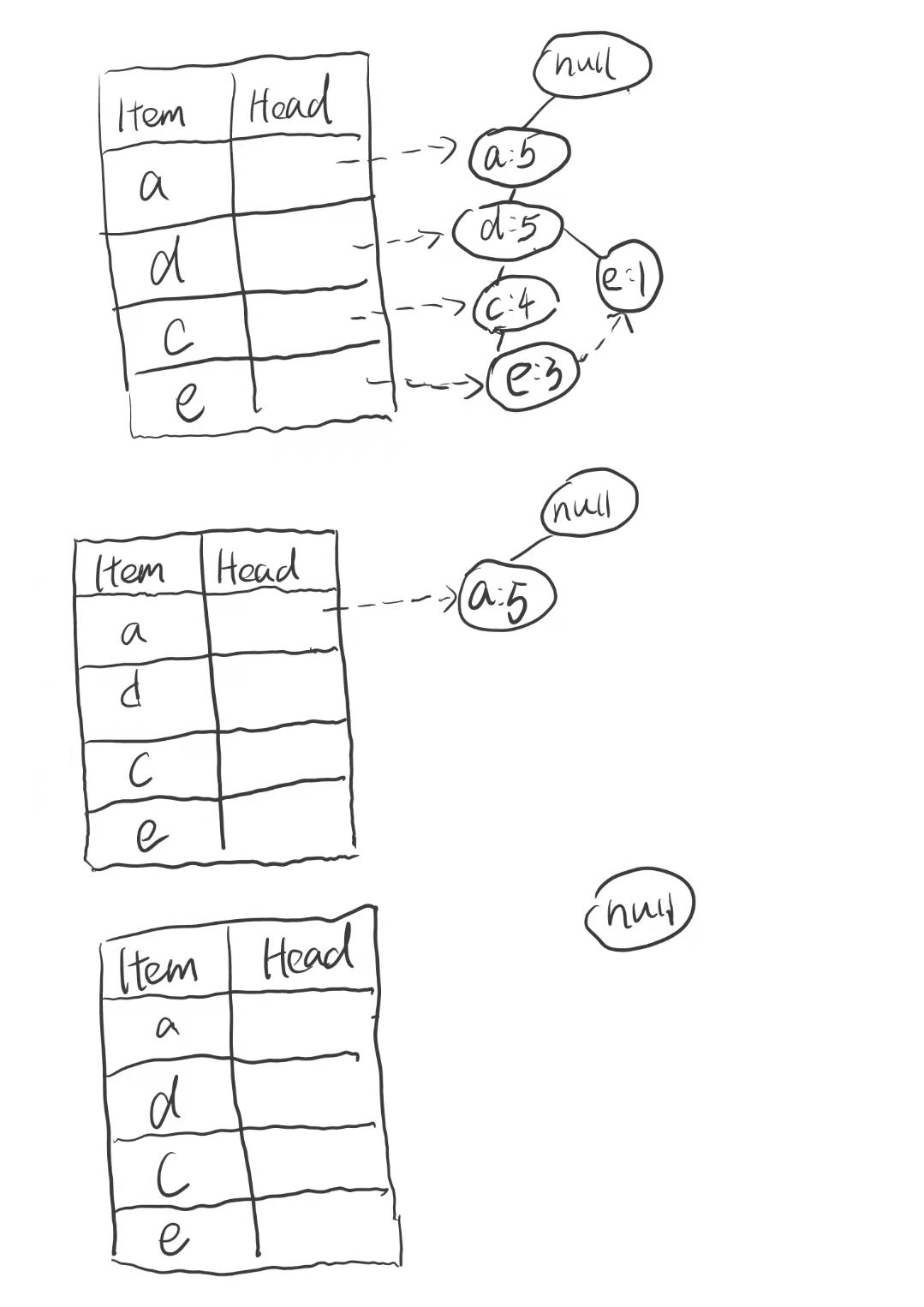
Generate pattern {c}:



Generate pattern {d}:



Generate pattern {a}:



Q2.

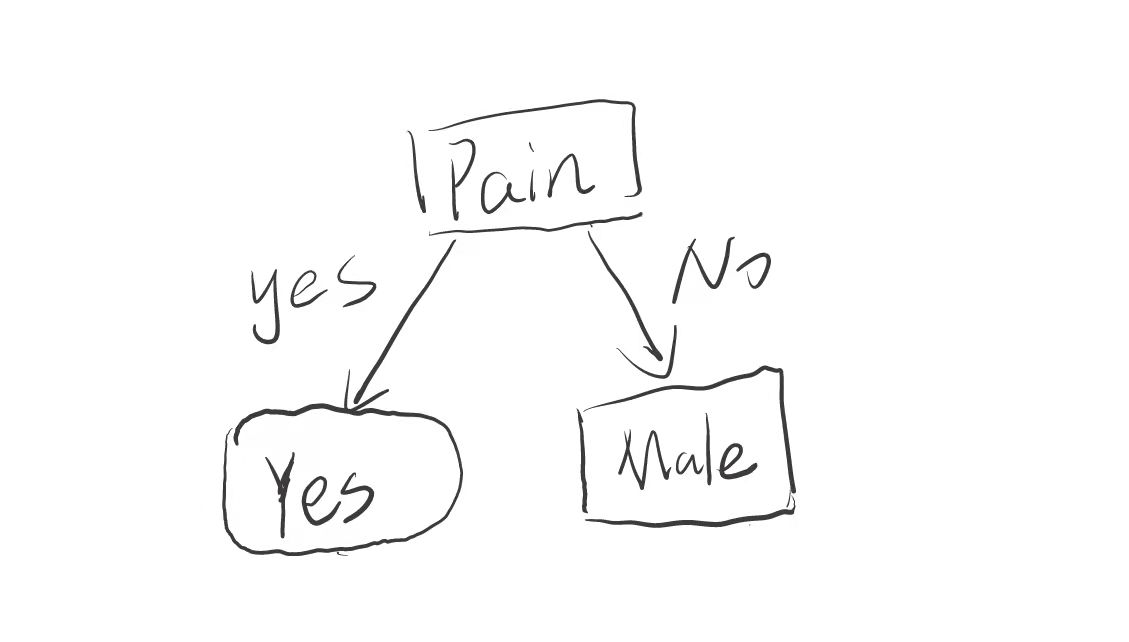
For chest pain:

For male:

For smoke:

For exercise:

Split using attribute chest pain.



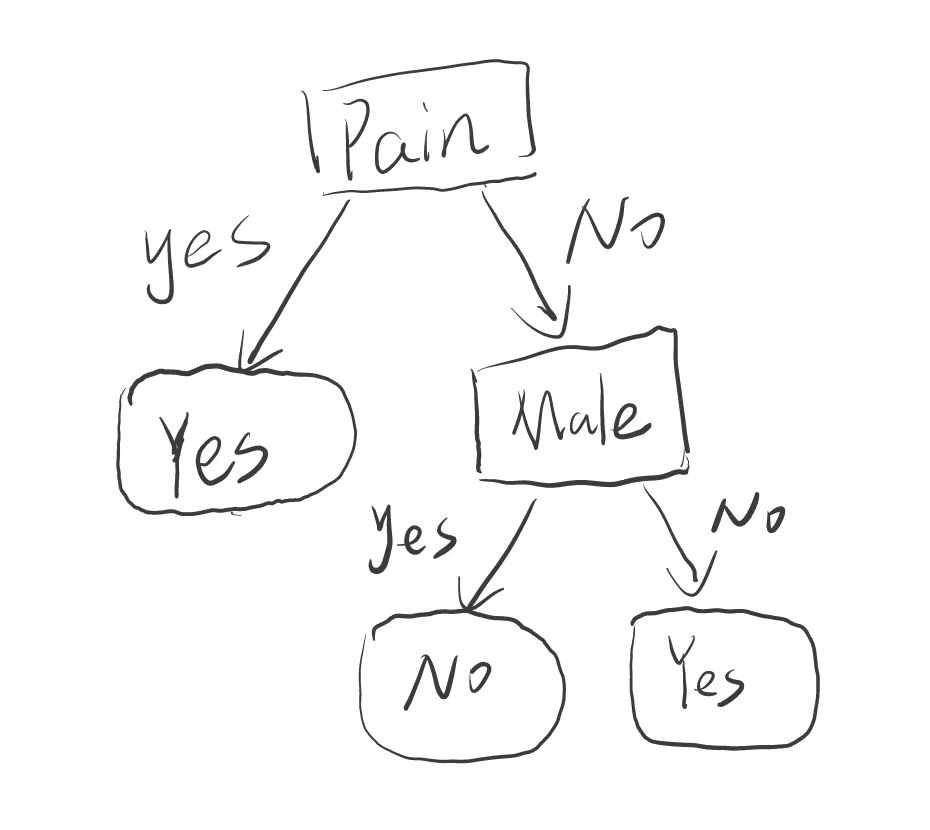
Then for the rest of item:

For male:

For smoke:

For exercise:

Split using attribute male.



(b).

According to the (a) decision,

|  |  |
| --- | --- |
| Predicted | Actual |
| Yes | No |
| Yes | Yes |
| Yes | Yes |
| No | No |
| No | Yes |
| Yes | No |

|  |  |  |
| --- | --- | --- |
| Classes | C1 | C2 |
| C1 | 2 (TP) | 1 (FN) |
| C2 | 2 (FP) | 1 (TN) |

-

Q3.

For Patient 7:

Because , the patient 7 may have heart attack.

For Patient 8:

Because , the patient 8 may have heart attack.