**1.** (2 credits) What is the confidence for the rules A 🡪 ∅ ?  Herein, ∅ indicates an empty set.

Ans:

Confidence: 100%

Confidence = σ (A, ∅) / σ(A). Since, σ (A, ∅) = σ(A), σ(A)/ σ(A) = 1

**2.** (2 credits) Let c1, c2, and c3 be the confidence values of the rules {p} 🡪{q}, {p} 🡪{q, r}, and {p, r} 🡪{q}, respectively. If we assume that c1, c2, and c3 have different values, which rule has the lowest confidence? Explain the reason.

Ans: C2

C1 = , C2 = , C3 =

According to Anti-monotone Property, σ (p, q) ≥ σ (p, q, r) and σ (p) ≥ σ (p, r). The lowest one means the numerator is the smallest, and denominator is the largest. Therefore, C2 is the lowest, because σ (p, q, r) is the smallest and σ (p) is the largest.

**3.** (3 credits) Consider the following traffic accident data. To discover the associate patterns, we need to binarize each categorical attribute.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | weather\_good | weather\_bad | driver\_sober | driver\_alcohol | violation\_none | violation\_speed | violation\_stop | violation\_traffic | belt\_no | belt\_yes | severity\_major | severity\_minor |
| 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |
| 2 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 3 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| 4 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 |
| 5 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 |
| 6 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| 7 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| 8 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| 9 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| 10 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 |
| 11 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 |
| 12 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |



Ans:

After binarized attributes, the table looks like.

My minimal support is 2. Get 1 frequent itemsets.

|  |  |
| --- | --- |
| ID | Items |
| 1 | weather\_bad, driver\_sober, violation\_traffic, belt\_no, severity\_major |

**4.** (4 credits) Based on the previous results, write an Apriori algorithm to generate the association rules with high confidences. Please set a proper confidence threshold to get 1-50 rules.

Ans:

Setting confidence as 80%, we get 25 rules.

|  |  |
| --- | --- |
| Rule ID | Association rules |
| 1 | ( driver\_sober, violation\_traffic, belt\_no, severity\_major ) 🡺 ( weather\_bad ) |
| 2 | ( weather\_bad, violation\_traffic, belt\_no, severity\_major ) 🡺 ( driver\_sober ) |
| 3 | ( weather\_bad, driver\_sober, belt\_no, severity\_major ) 🡺 ( violation\_traffic ) |
| 4 | ( weather\_bad, driver\_sober, violation\_traffic, severity\_major ) 🡺 ( belt\_no ) |
| 5 | ( weather\_bad, driver\_sober, violation\_traffic, belt\_no ) 🡺 ( severity\_major ) |
| 6 | ( violation\_traffic, belt\_no, severity\_major ) 🡺 ( weather\_bad, driver\_sober ) |
| 7 | ( driver\_sober, belt\_no, severity\_major ) 🡺 ( weather\_bad, violation\_traffic ) |
| 8 | ( driver\_sober, violation\_traffic, severity\_major ) 🡺 ( weather\_bad, belt\_no ) |
| 9 | ( driver\_sober, violation\_traffic, belt\_no ) 🡺 ( weather\_bad, severity\_major ) |
| 10 | ( weather\_bad, belt\_no, severity\_major ) 🡺 ( driver\_sober, violation\_traffic ) |
| 11 | ( weather\_bad, violation\_traffic, severity\_major ) 🡺 ( driver\_sober, belt\_no ) |
| 12 | ( weather\_bad, violation\_traffic, belt\_no ) 🡺 ( driver\_sober, severity\_major ) |
| 13 | ( weather\_bad, driver\_sober, severity\_major ) 🡺 ( violation\_traffic, belt\_no ) |
| 14 | ( weather\_bad, driver\_sober, belt\_no ) 🡺 ( violation\_traffic, severity\_major ) |
| 15 | ( weather\_bad, driver\_sober, violation\_traffic ) 🡺 ( belt\_no, severity\_major ) |
| 16 | ( belt\_no, severity\_major ) 🡺 ( weather\_bad, driver\_sober, violation\_traffic ) |
| 17 | ( violation\_traffic, severity\_major ) 🡺 ( weather\_bad, driver\_sober, belt\_no ) |
| 18 | ( violation\_traffic, belt\_no ) 🡺 ( weather\_bad, driver\_sober, severity\_major ) |
| 19 | ( driver\_sober, severity\_major ) 🡺 ( weather\_bad, violation\_traffic, belt\_no ) |
| 20 | ( driver\_sober, belt\_no ) 🡺 ( weather\_bad, violation\_traffic, severity\_major ) |
| 21 | ( driver\_sober, violation\_traffic ) 🡺 ( weather\_bad, belt\_no, severity\_major ) |
| 22 | ( weather\_bad, severity\_major ) 🡺 ( driver\_sober, violation\_traffic, belt\_no ) |
| 23 | ( weather\_bad, belt\_no ) 🡺 ( driver\_sober, violation\_traffic, severity\_major ) |
| 24 | ( weather\_bad, violation\_traffic ) 🡺 ( driver\_sober, belt\_no, severity\_major ) |
| 25 | ( weather\_bad, driver\_sober ) 🡺 ( violation\_traffic, belt\_no, severity\_major ) |