**Association Rules: Apriori Algorithm. Prof. Carolina Ruiz, WPI.**

Consider the following subset of the weather.nominal dataset, with 4 attributes and 10 instances:

|  |  |  |  |
| --- | --- | --- | --- |
| **Outlook** | **Temperature** | **Humidy** | **Windy** |
| sunny | hot | high | true |
| sunny | mild | high | false |
| sunny | cool | normal | false |
| sunny | mild | normal | true |
| overcast | cool | normal | true |
| overcast | mild | high | true |
| overcast | hot | normal | false |
| rainy | mild | high | false |
| rainy | cool | normal | true |
| rainy | mild | high | true |

Each of these instances can be seen as a transaction consisting of attribute-value pairs. For example, the first data instance above represents the transaction:

{outlook=sunny, temperature=hot, humidity=high, windy=true}.

To simplify the notation in this handout, instead of writing an itemset as a set of attribute-value pairs, we will use positional notation as follows: the itemset {temperature=cool, windy=false} will be represented as:

|  |  |  |  |
| --- | --- | --- | --- |
| **Outlook** | **Temperature** | **Humidy** | **Windy** |
|  | cool |  | false |

1. Let minimum support = 20% (that is, support count = 2). Using the Apriori algorithm, construct all frequent itemsets level by level. Remember to use the join (= merge) and the subset conditions.

**Level 1:** Frequent 1-itemsets. Below is the list of all 1-itemsets with support count ≥ 2 (each row represents a frequent itemset).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outlook** | **Temperature** | **Humidy** | **Windy** | Support Count |
| sunny |  |  |  | 4 |
| overcast |  |  |  | 3 |
| rainy |  |  |  | 3 |
|  | hot |  |  | 2 |
|  | mild |  |  | 5 |
|  | cool |  |  | 3 |
|  |  | high  |  | 5 |
|  |  | normal |  | 5 |
|  |  |  | true | 6 |
|  |  |  | false | 4 |

**Level 2:** Frequent 2-itemsets. Below is the list of all 2-itemsets with support count ≥ 2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outlook** | **Temperature** | **Humidy** | **Windy** | Support Count |
| sunny | mild |  |  | 2 |
| sunny |  | high |  | 2 |
| sunny |  | normal |  | 2 |
| sunny |  |  | true | 2 |
| sunny |  |  | false | 2 |
| overcast |  | normal |  | 2 |
| overcast |  |   | true | 2 |
| rainy | mild |  |  | 2 |
| rainy |  | high |  | 2 |
| rainy |  |  | true | 2 |
|  | mild | high |  | 4 |
|  | mild |  | true | 3 |
|  | mild |  | false | 2 |
|  | cool | normal |  | 3 |
|  | cool |  | true | 2 |
|  |  | high | true | 3 |
|  |  | high | false | 2 |
|  |  | normal | true | 3 |
|  |  | normal | false | 2 |

**Level 3:** Frequent 3-itemsets: Below, only one frequent 3-itemset is given. Find the remaining frequent 3-itemsets. Remember to use the join (= merge) and the subset conditions. Show your work.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outlook** | **Temperature** | **Humidy** | **Windy** | Support Count |
| rainy | mild | high |  | 2 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Level 4:** Frequent 4-itemsets. Find all frequent 4-itemsets. Remember to use the join (= merge) and the subset conditions. Show your work.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outlook** | **Temperature** | **Humidy** | **Windy** | Support Count |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

1. Rules: Generate all association rules with confidence ≥ 90% from the 3-itemset: {rainy, mild, high}. Show your work.