

Tutorial 11

1. A local retailer has a database that stores 10,000 transactions of last summer. After analyzing the data, a data science team has identified the following statistics:
  - {battery} appears in 6000 transactions
  - {sunscreen} appears in 5000 transactions
  - {sandals} appears in 4000 transactions
  - {bowls} appears in 2000 transactions
  - {battery, sunscreen} appears in 1500 transactions
  - {battery, sandals} appears in 1000 transactions
  - {battery, bowls} appears in 250 transactions
  - {battery, sunscreen, sandals} appears in 600 transactions
  - (a) What are the support values of the preceding itemsets?
  - (b) Assuming the minimum support is 0.05, which itemsets are considered frequent?
  - (c) What are the confidence values of  $\{\text{battery}\} \rightarrow \{\text{sunscreen}\}$  and  $\{\text{battery, sunscreen}\} \rightarrow \{\text{sandals}\}$ ? Which of these two rules is more interesting, i.e. has higher values of confidence?
2. Suppose for three products  $A$ ,  $B$  and  $C$ ,  $\text{support}(\{A\}) = 0.6$ ,  $\text{support}(\{B\}) = 0.6$ ,  $\text{confidence}(\{B\} \rightarrow \{A\}) = 0.9$  and  $\text{confidence}(\{C\} \rightarrow \{A, B\}) = 0.5$ . Compute the following quantities.
  - (a)  $\text{Lift}(\{A\} \rightarrow \{B\})$
  - (b)  $\text{Leverage}(\{A\} \rightarrow \{B\})$
  - (c)  $\text{Confidence}(\{A\} \rightarrow \{B\})$
  - (d)  $\text{Lift}(\{A, B\} \rightarrow \{C\})$