

# Association Analysis

---

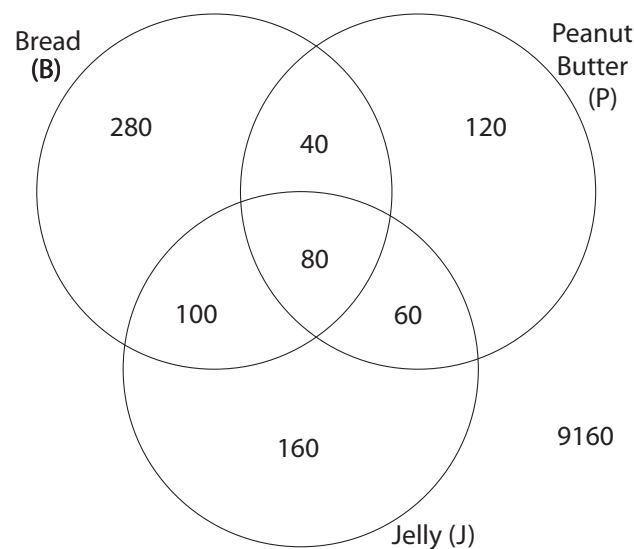
## Exercises

- Using the following Venn Diagram, determine the **confidence** of the following rules. (note: there are 10,000 transactions total in this data set. 9,160 transactions did not involve any of the 3 items of interest.)

$$\text{Bread} \implies \text{Jelly}$$

$$\text{PeanutButter} \implies \text{Jelly}$$

Also, determine the **lift** of the first rule,  $\text{Bread} \implies \text{Jelly}$ , and explain the result in a sentence.



- Give an example of an association rule that might satisfy the given conditions. For visual folks, it might help to draw a Venn diagram with two circles representing the antecedent and consequent of the rule. Consider how the relative size of each circle and the overlap would be representative of support and confidence. Comment on whether such a rule might be subjectively interesting.
  - A rule with high support and high confidence.
  - A rule with low support and low confidence.
  - A rule with low support and high confidence.
- Suppose you collected some data for association analysis, and you found the support of the rule

$$\text{itemA} \implies \text{itemB}$$

was 100%. What can you say about the confidence, expected confidence and lift of this rule? What can you take away from this about rules with very high support? Are they likely to be interesting? Discuss briefly.

4. In an effort to learn more about visitors to news websites, you've collected data on whether or not individuals have visited certain sites in the past week. These are binary variables indicating if the website has been visited ("1") or not ("0"). You'd like to perform an association analysis to determine rules indicating whether an individual is likely to visit an alternative site, given their browsing history. Using the following dataset, answer the questions below.

Name	NPR	CBS	MSNBC	ABC	FOX	POST	WSJ
Tito	1	1	0	0	0	0	0
Jojo	0	1	0	0	0	1	1
Loki	0	1	1	1	1	1	1
Niko	0	0	1	1	1	1	1
Yaya	1	0	0	1	0	1	1
Kujo	1	1	1	0	0	0	1
Arod	0	0	0	0	1	0	0
Park	1	1	1	1	0	1	1
Yorp	0	1	0	1	0	0	0
Hoki	0	0	1	1	0	1	1

- What is the support of the rule  $CBS \implies MSNBC$ ?
- What is the confidence of the rule  $NPR \implies WSJ$ ?
- What is the confidence of the rule  $WSJ \implies FOX$ ?
- What is the lift of the rule  $WSJ \implies POST$ ?
- What is the lift of the rule  $POST \implies WSJ$ ?
- Suppose you put this binary data into a matrix,  $\mathbf{X}$ . Let  $\mathbf{S} = \mathbf{X}\mathbf{X}^T$  and explain the meaning behind  $\mathbf{S}_{ij}$ .  
Hint: You should actually perform the multiplication of  $\mathbf{X}\mathbf{X}^T$  and see what happens to answer this question!

# List of Key Terms

support

confidence

lift

association rule

transaction dataset

antecedent

consequent