Archit Jain
AT
Crib sheet

Co Association $(X \to Y)$ (2 step)

Apriori algorithm {Diaper → Beer} → (Co-occurrence) (not Casuality)
(Inventory mangment)

(Frequent Itemset)

Before → rules

1) Itemset - Collection of items {Milk, Bread}
2) Support Count → Count → σ {Milk, Bread} = 2
3) Support → fraction of $\frac{\{\text{Milk bread}\}}{\text{Total}} = 2/5$
4) Frequent - Itset → Support is greater or min Support

☆ Support - fraction of transaction ← (Contain X & Y)

☆ Confidence Mes, How often items in X appear in transaction X

Support ≥ minsup threshold
Confidence

☆ Prune → delete rules. (Computational → Brute force)

Two step approach: 1) frequent itemset generation (Support ≥ min sup)
(Computationally expensive) 2) Rule generation (Confidence ≥ min Conf)

$\sim O(NMw)$ — (width)
Transactions → frequent itmset set

$$\left[Conf_{in} = \frac{\sigma(Y)}{\sigma(X)}\right]$$

Minsupport issue → too high → miss itemset with rare items
to low → Computationally expensive.

Drawback Confidence.

hides negative Correlation

	Coffee	
Tea	15	20
Tea		
	90	

Statistical Independence $P(S \cap B) = P(S) \times P(B)$ = Independent
$P(S \cap B) > P(S) \times P(B)$ = Positive ; Correlation
$P(S \cap B) < P(S) \times P(B)$ = Negative

$P(Coffee) = .75$

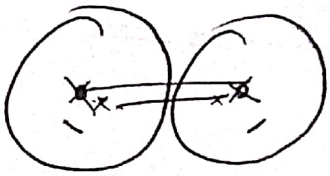
$P(Coffee | Tea) = \frac{15}{P\,Tea} = \frac{15}{20}$

$$Lift = \frac{Conf(X \to Y)}{Sup(Y)} = \frac{P(X,Y)}{P(X).P(Y)} = \begin{matrix} 1 \text{ independent} \\ 1 > \text{positive} \\ 1 < \text{-ve} \end{matrix}$$

Interest = when positive

(k-means)

fix number of iterations



Stopping criteria - iteration
Same Centroid.

Han Initial Centroid random,

6 points →

Euclidaan distance:
Manhattan → Summation of absolute value
Cosine Simi :

Convergence

[k means → Important to initial Centroids.
doesn't give global, just local solution.

Min (SSE = Sum of Squared error)

$$SSE = \sum_{i=1}^{K} \sum_{x \in C} dist^2(m_i, x)$$

minimize → distance from Center
choose low SSE
min(SSE) increase → (k)
good cluster with smaller k have smaller (SSE)