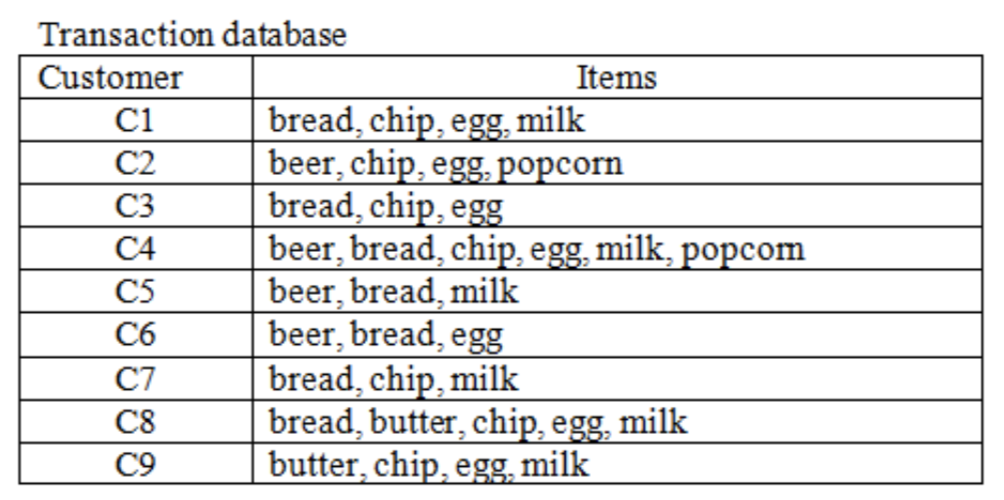
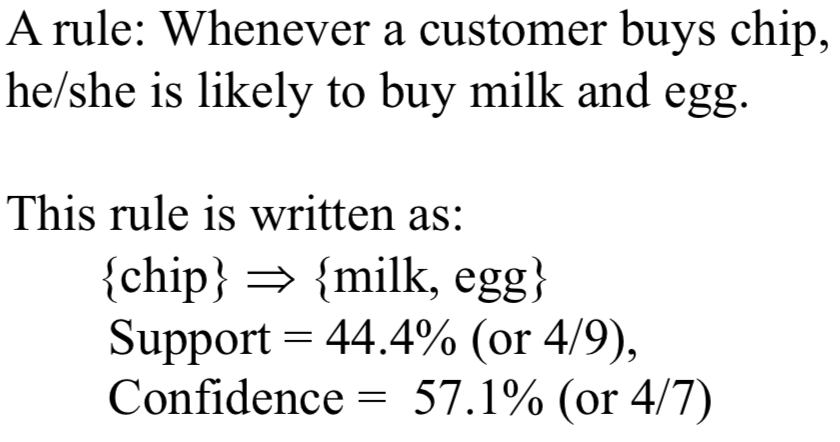
**Module 1**

**Association and Correlation: Frequent patterns**





**Module 2**

**Positively skewed**: slope is on left of range

**Negatively skewed**: slope is on right of range

*Opposite of what I imagined*

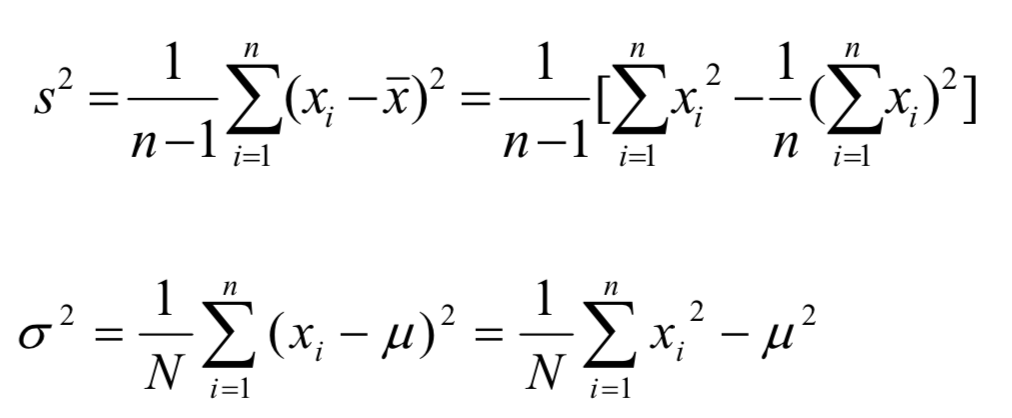
**Inter Quartile Range** (Q3-Q1)

**Outliers**:

Less than Q1 – 1.5\*IQR

Greater than Q3 + 1.5\*IQR

**Standard deviation & Variance**:



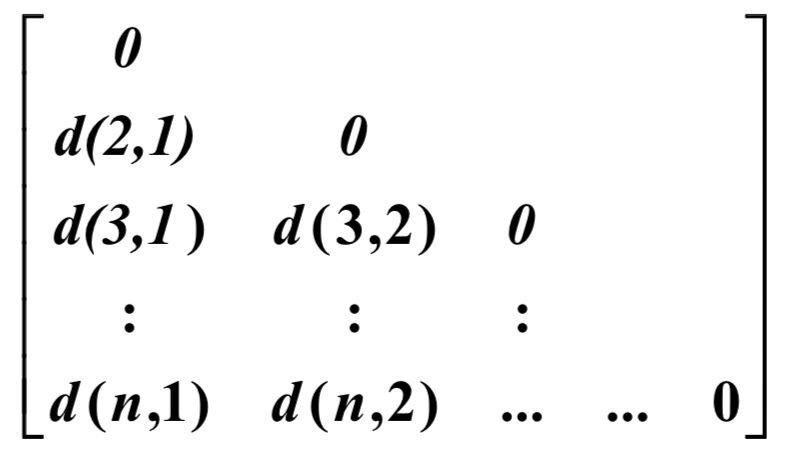
**Properties of Normal Distribution Curve**:

μ–σ to μ+σ: contains about 68%

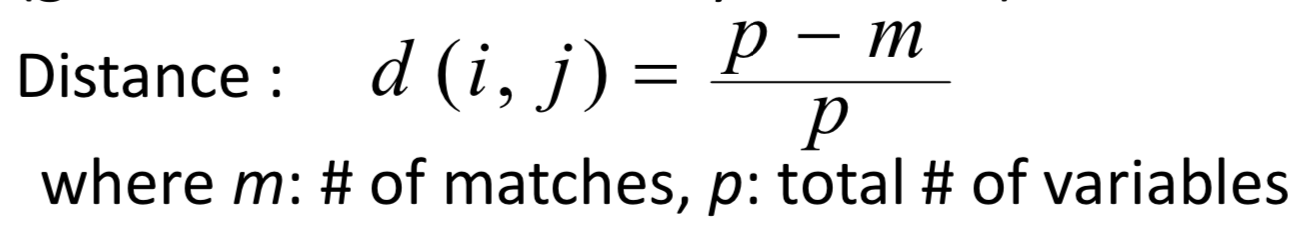
μ–2σ to μ+2σ: contains about 95%

μ–3σ to μ+3σ: contains about 99.7%

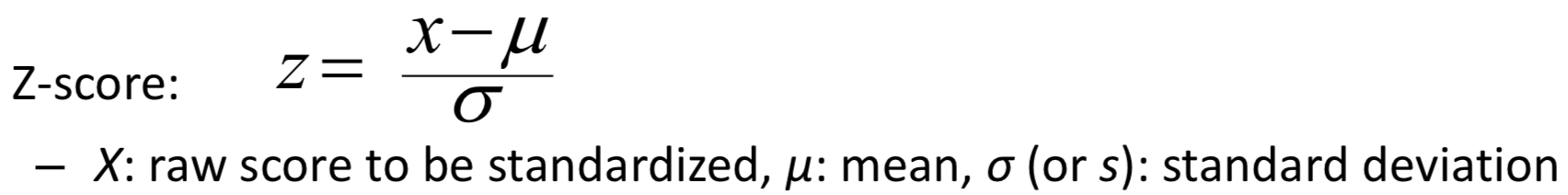
**Dissimilarity matrix**:



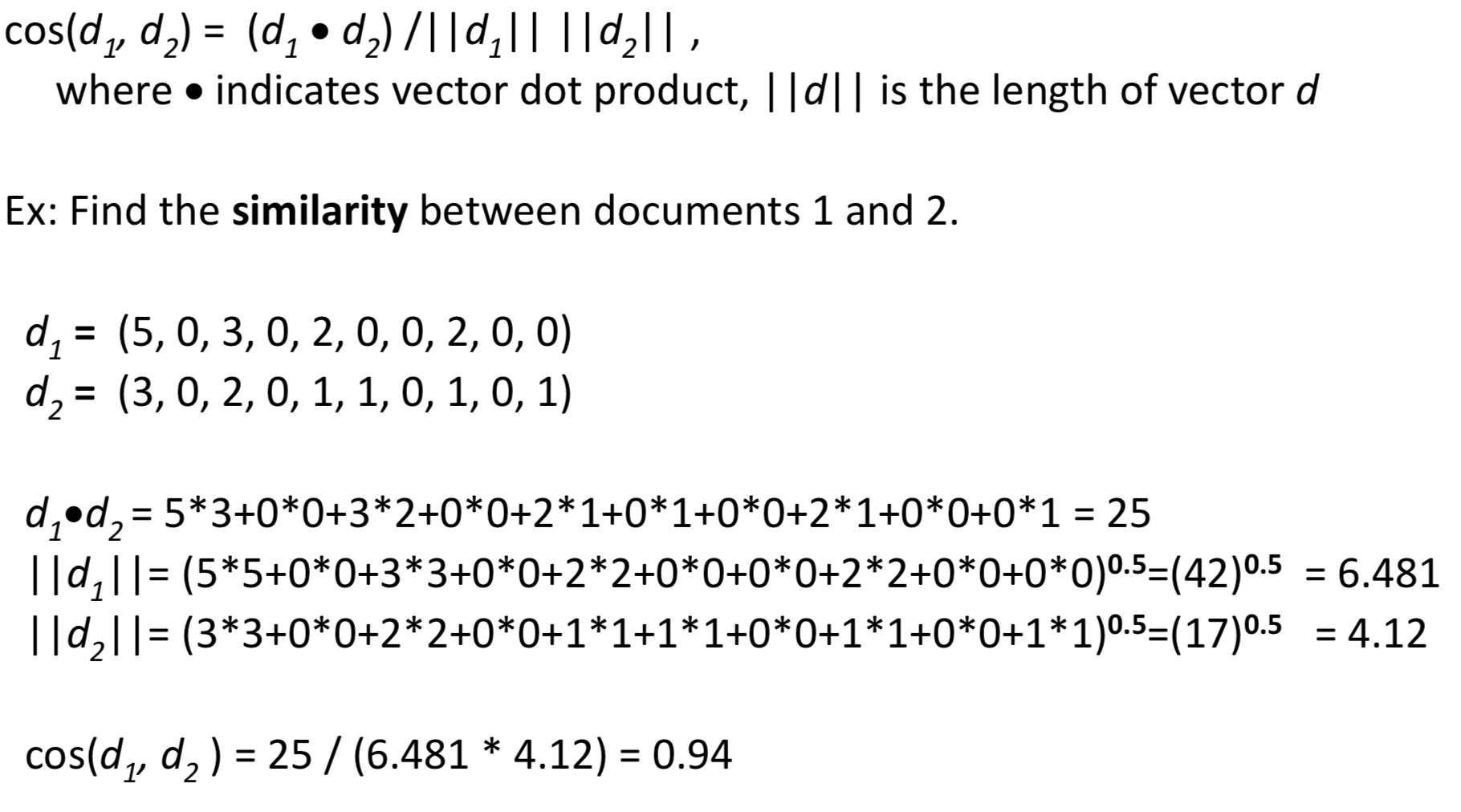
**Nominal attributes**:



**Z-score standardization**:



**Cosine similarity**:



**Module 3**

**Binning:**

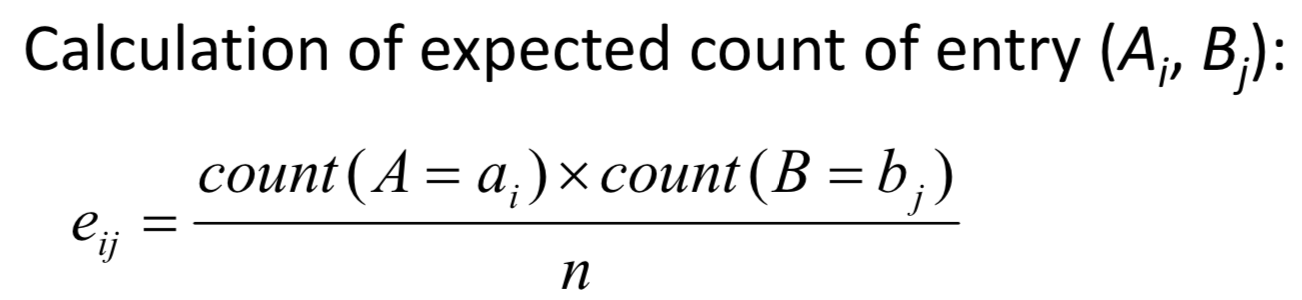
**Equal width:** *W* = (*B* –*A*)/*N*

**Equal depth:** approx. equal amounts in each bin

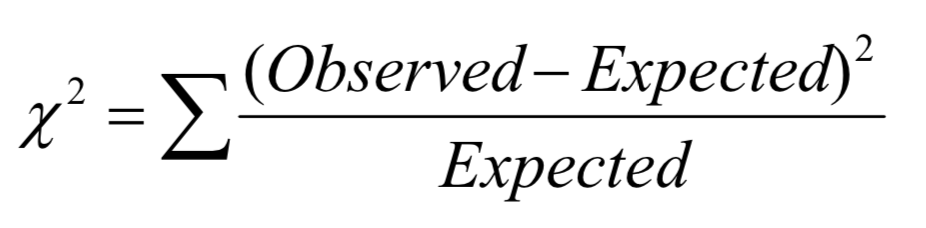
**Bin means:** mean of each bin

**Bin boundaries:** which bin boundary each element is closest to

**Compute Expected Values**:

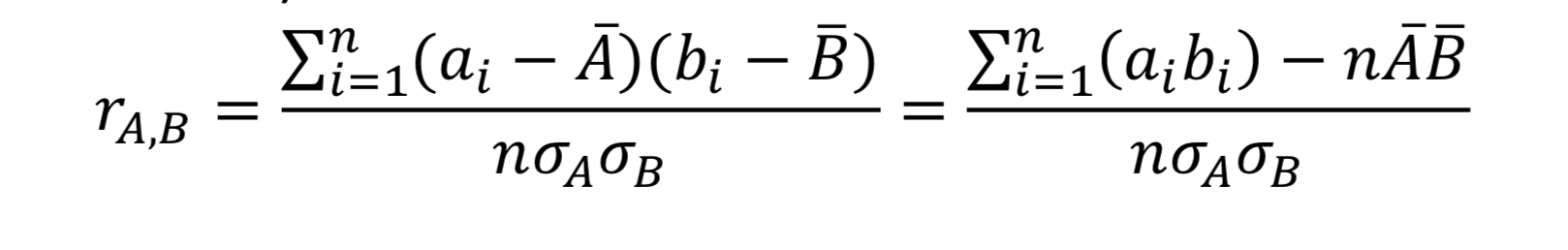


**Chi-squared Test:**

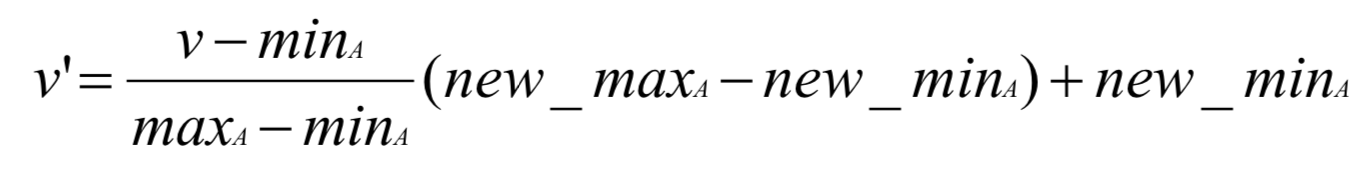
****

The larger the *Χ*2 value, the more likely the variables are related.

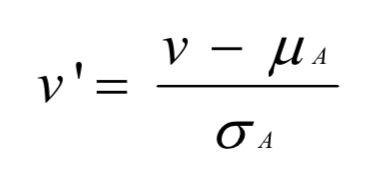
**Correlation Coefficient:**

****

**Min-max normalization:**

****

**Z-score normalization:**

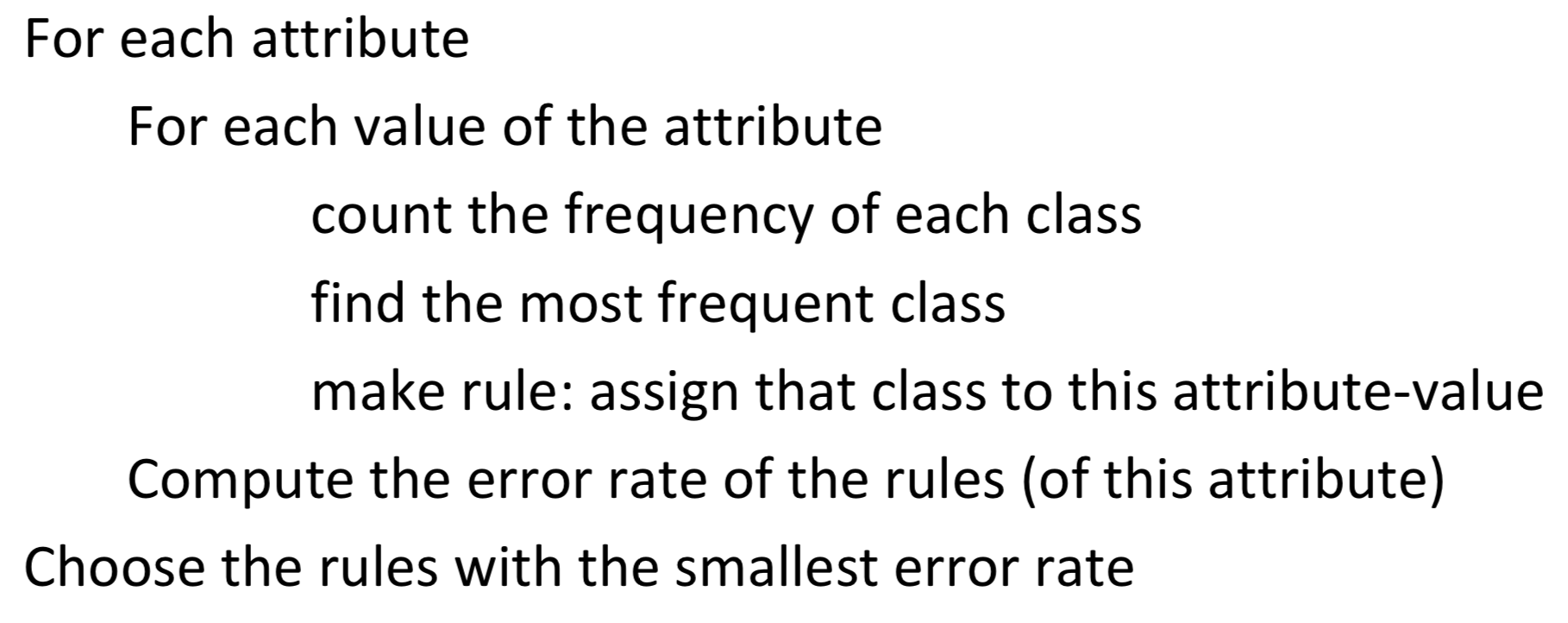
****

**Normalization by decimal scaling:**

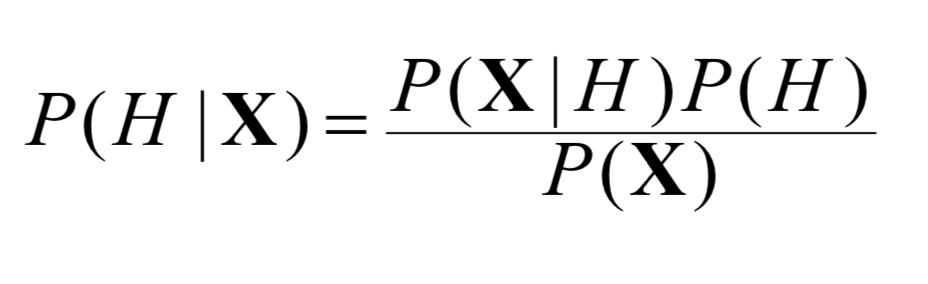


**Module 4**

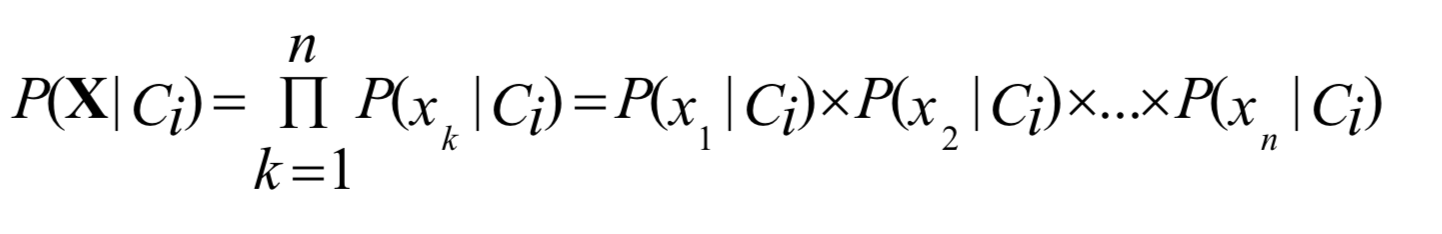
**1R Classifier:**

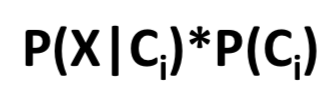


**Bayes Theorem:**



**Naïve Bayes Classifier:**





**Decision Tree Classifier:**

Pick test attribute: A1, separate

For each node, pick test attribute: A2, separate

…

Evaluate purity of attribute selection