\[ \text{(Returning to derivation for } P(c|\ldots, x_n)) \]
\[
= \arg \max_{c \in C} p(c)^n \prod_{i=1}^{n} p(x_i|c) \quad \text{- small numbers! estimated from data} \]
\[
\left[ \arg \max_x \ell(x) = \arg \max_x \log{\ell(x)} \right] \]
\[
= \arg \max_{c \in C} \log{p(c)} + \sum_{i=1}^{n} \log{p(x_i|c)} \quad \text{(for implementation reasons)} \]

Example: \( x_i \): presence or absence of a word from a news story
\( C \): news story categories

\[
p(c) = \frac{\left| \{(x, y) \in D \mid y = c \} \right|}{|D|} \]

\[
p(x_i|c) = \frac{\left| \{(x, y) \in D \mid y = c \text{ and } i \text{th attribute is } x_i \} \right|}{\left| \{(x, y) \in D \mid y = c \} \right|} \]

Problem: numerator often = 0

Laplace smoothing

If \( x_i \in \{0, 1, 3\} \), given no info \( p(x_i) = \frac{1}{2} \)

\[
p(x_i|c) = \frac{\left| \{(x, y) \in D \mid y = c \text{ and } i \text{th attribute is } x_i \} \right| + \alpha}{\left| \{(x, y) \in D \mid y = c \} \right| + \alpha \beta} \]

\[
p(x_i|c) = \frac{\left| \{(x, y) \in D \mid y = c \} \right| + \alpha}{\left| \{(x, y) \in D \mid y = c \} \right| + \alpha \beta} \]

Unsupervised learning

Clustering

- Given $D = \tilde{x}^1, \ldots, \tilde{x}^n$
  
  No $\mathcal{C}(\tilde{x})$!

- Assign each $\tilde{x}^i$ to one of $c_1, \ldots, c_k$ s.t.
  if $\tilde{x}^i$ and $\tilde{x}^j$ are in same $c_l$, $\tilde{x}^i$ and $\tilde{x}^j$ are "close"
  if $\tilde{x}^i$ and $\tilde{x}^j$ are in diff. $c_l$, $\tilde{x}^i$ and $\tilde{x}^j$ are "far"

$c_1, \ldots, c_k$ are called "clusters"

K-means clustering

- Assume $c_1, \ldots, c_k$ are disjoint
  
  $k$ is given

K-means $(D,k)$

Randomly select $\tilde{u}^1, \ldots, \tilde{u}^k$ from $D$

Repeat
  
  For $i=1$ to $k$
    
    $S_i = \{ \tilde{x} \in D | i = \text{argmin}_j \sum_{i=1}^n (\tilde{x} - \tilde{u}_j)^2 \}$
    
    $\tilde{u}_i = \frac{\sum_{\tilde{x} \in S_i} \tilde{x}}{|S_i|}$

Until the $S_i$ don't change