(Returning to derivation for \( P \in \{ x_1, \ldots, x_n \} \))

\[
\text{argmax}_{c \in C} \ p(c) \prod_{i=1}^{n} p(x_i | c) \quad \text{small numbers! (estimated from data)}
\]

\[
\text{argmax}_{x} \ E(x) = \text{argmax}_{x} \ \log \ (E(x))
\]

\[
\text{argmax}_{c \in C} \ \log \ (p(c)) + \sum_{i=1}^{n} \ \log \ (p(x_i | c))
\]

(for implementation reasons)

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

\( C \): news story categories

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

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

Problem: numerator often \( = 0 \)

Laplace smoothing
If \( x_i \in \{0, 1\} \) given \( n \) in \( D \) \( p(x_i) = \frac{1}{2} \)

\[
p(x_i | c) = \frac{\left| \{ (x, y) \in D | y = c \text{ and } \text{ith attribute is } x \} \right| + q}{\left| \{ (x, y) \in D | y = c \} \right| + 2q}
\]
Unsupervised learning

Clustering
- Given \( D = \bar{X}_1, \bar{X}_2, \ldots, \bar{X}_n \)
  No \( F(\bar{x}) \)
- Place each \( \bar{x}_i \) in a group (a cluster) (disjoint)
  such that \( \bar{x}_i \) and \( \bar{x}_j \) are close if
  they're in the same group and
  far if not

k-mean clustering
- Assume each \( \bar{x}_i \) is from 1 of \( k \) disjoint
  clusters that are not known
- Takes \( k \) as input

k-means \((D, k)\)

\[
\text{Repeat } S_1, \ldots, S_k \leftarrow \{\}
\]
Randomly select \( \bar{U}_1, \ldots, \bar{U}_k \) from \( D \)
Repeat
For \( i = 1 \) to \( m \)
  \( j \leftarrow \arg \min_{1 \leq q \leq k} \sum_{1 \leq \ell \leq n} (\bar{x}_i - \bar{U}_q)^2 \)
  \( S_j \leftarrow S \)
Classification learning: Classifying data

Given \( D = (x_1', y_1'), \ldots, (x_m', y_m') \quad y_i = f(x_i') \)

\( x_i' \in \{0, 1\}^n \) Boolean attributes

\( y_i \in C \) Set of categories

\( (x_1, \ldots, x_n) \) New example

Predict \( f(x_1, \ldots, x_n) \)

Bayesian learning:

Predict: \( \arg \max_{c \in C} p(c | x_1, \ldots, x_n) \)

Bayes rule: \( p(a|b) = \frac{p(b|a) \cdot p(a)}{p(b)} \)

\( p(c | x_1, \ldots, x_n) = \frac{p(x_1, \ldots, x_n | c) \cdot p(c)}{p(x_1, \ldots, x_n)} \)

\( \arg \max_{c \in C} p(c | x_1, \ldots, x_n) \)

\( = \arg \max_{c \in C} \frac{p(x_1, \ldots, x_n | c) \cdot p(c)}{p(x_1, \ldots, x_n)} \) \hspace{1cm} \text{Doesn't depend on } c \)

\( = \arg \max_{c \in C} p(x_1, \ldots, x_n | c) \cdot p(c) \)

"Naive Bayes"

Conditional independence

\( p(a_1, a_2, \ldots, a_n | b) \)

\( = p(a_1 | b) \cdot p(a_2 | b) \cdot \ldots \cdot p(a_n | b) \)

\( = \prod_{i=1}^{n} p(a_i | b) \)