April 25

(Returning to derivation for $P(c \mid x_1, \ldots, x_n)$)

$$= \underset{c \in C}{\operatorname{argmax}} \; p(c) \times \prod_{i=1}^{n} p(x_i \mid c)$$ - small numbers!

estimated from data

$$\left[\underset{x}{\operatorname{argmax}} \; f(x) = \underset{x}{\operatorname{argmax}} \; \log(f(x)) \right]$$

$$= \underset{c \in C}{\operatorname{argmax}} \; \log(p(c)) + \sum_{i=1}^{n} \log(p(x_i \mid c))$$

(for implementation reasons)

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

C: news story categories

$$p(c) = \frac{|\{(\bar{x}, y) \in D \mid y = c\}|}{|D|}$$

$$p(x_i \mid c) = \frac{|\{(\bar{x}, y) \in D \mid y = c \text{ and } i\text{th attribute is } x_i\}|}{|\{(\bar{x}, y) \in D \mid y = c\}|}$$

problem: numerator often $= 0$

Laplace smoothing

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

$$p(x_i \mid c) = \frac{|\{(\bar{x}, y) \in D \mid y = c \text{ and } i\text{th attribute is } x_i\}| + k}{|\{(\bar{x}, y) \in D \mid y = c\}| + 2k}$$

Unsupervised learning

Clustering

- Given $D = \bar{x}^1, \dots, \bar{x}^m$
  No $f(\bar{x})$!
- Assign each x^i to one of $c_1, \dots, c_k$ s.t.
  if $\bar{x}^i$ and $\bar{x}^j$ are in same c_ℓ $\bar{x}^i$ and $\bar{x}^j$ are "close"
  if $\bar{x}^i$ and x^j are in diff. c_ℓ $\bar{x}^i$ and $\bar{x}^j$ are "far"

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

k-means clustering

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

k-means (D, k)

Randomly select $\bar{u}^1, \dots, \bar{u}^k$ from D

Repeat

For $i = 1$ to k

$$S_i = \left\{ \bar{x} \in D \mid i = \underset{\ell}{\operatorname{argmin}} \sum_{j=1}^{n} (x_j - u_j^\ell)^2 \right\}$$

$$\bar{u}^i = \frac{\sum_{\bar{x}^j \in S^i} \bar{x}^j}{|S^i|}$$

Until the S^i don't change