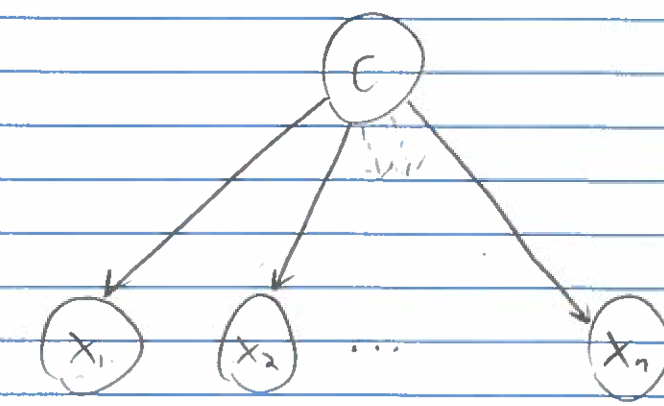
3) Expectation Maximization in a Naive Bayes Model



Class variable C, discrete evidence variables $X_1, \dots, X_n$

CPDs parametrized: $P(C=c) = \Theta_c$, $P(X_i = x \mid C=c) = \Theta_{x_i \mid c}$ for $i = 1, \dots, n$ and for all assignments $x_i \in Val(X_i)$, classes $c \in Val(C)$

Given dataset $D = \{x[1], \dots, x[M]\}$ each $x[m]$ assignment to $X_1, \dots, X_n$
C not observed

E-Step 1: $X[m] = (X_1 = x_1, X_2 = x_2, \dots, X_n = x_n, C = c_1)$
$(\quad \text{"} \quad , C = c_2)$
$\vdots$
$(\quad \text{"} \quad , C = c_{|Val(C)|})$

Expand $x[m]$ into $|Val(C)|$ datapoints, one for each $c \in C$

$$w = \frac{1}{Z} \Theta_c^{\circ} \prod_{i=1}^{n} \Theta_{x_i \mid c}^{\circ} = \frac{1}{Z} \cdot \frac{1}{|Val(C)|} \cdot \prod_{i=1}^{n} \frac{1}{|Val(X_i)|}$$

$$Z = \sum_{j=1}^{|Val(C)|} \frac{1}{|Val(C)|} \cdot \prod_{i=1}^{n} \frac{1}{|Val(X_i)|} = \prod_{i=1}^{n} \frac{1}{|Val(X_i)|}$$

$$w = \frac{\frac{1}{|Val(C)|} \cdot \prod_{i=1}^{n} \frac{1}{|Val(X_i)|}}{\prod_{i=1}^{n} \frac{1}{|Val(X_i)|}} = \frac{1}{|Val(C)|}$$

Weight w is same for all datapoints

M-step: $$\Theta'_{c_1} = \frac{\sum weight(c_1)}{\# \text{data}} = \frac{M \cdot \frac{1}{|Val(C)|}}{M \cdot |Val(C)|} = \frac{1}{|Val(C)|^2}$$

$$\Theta'_{X_i = x_i \mid C = c_1} = \frac{n \leq M \cdot \frac{1}{|Val(C)|}}{M \cdot \frac{1}{|Val(C)|}} = \frac{n}{M}, \quad n \leq M \rightarrow n \rightarrow t$$

E-Step 2: $$w = \frac{1}{Z} \cdot \frac{1}{|Val(C)|^2} \cdot \prod_{i=1}^{n} \frac{t_i}{M} = \frac{1}{Z} \cdot \frac{1}{|Val(C)|^3} \cdot \frac{1}{M} \cdot \prod_{i=1}^{n} t_i$$

$$Z = |Val(C)| \cdot \frac{1}{|Val(C)|^2} \cdot \frac{1}{M} \cdot \prod_{i=1}^{n} t_i = \frac{1}{|Val(C)|} \cdot \frac{1}{M} \cdot \prod_{i=1}^{n} t$$

$$\rightarrow w = \frac{\frac{1}{|Val(C)|^2} \cdot \frac{1}{M} \cdot \prod_{i=1}^{n} t_i}{\frac{1}{|Val(C)|} \cdot \frac{1}{M} \cdot \prod_{i=1}^{n} t_i} = \frac{|Val(C)|}{|Val(C)|^2} = \frac{1}{|Val(C)|}$$