CSE 250A HW 4

Jiping Lin A15058075

1.(a) $\mathcal{L} = \log P(data) = \log \prod_{d=1}^{2D} P_d^{C_d} = \sum_{d=1}^{2D} C_d \log P_d$.

(b) max. $\mathcal{L}$, s.t. $\sum P_d = 1$, $P_d \geq 0$.

$Lag(P_1, \cdots, P_{2D}, \lambda) = \sum_{d=1}^{2D} C_d \log P_d + \lambda(\sum_{d=1}^{2D} P_d - 1)$

$\frac{\partial Lag}{\partial P_i} = \frac{C_i}{P_i} + \lambda = 0$, $\frac{\partial Lag}{\partial \lambda} = \sum_{d=1}^{2D} P_d - 1 = 0$.

↳ implies $\frac{C_1}{P_1} = \cdots = \frac{C_{2D}}{P_{2D}}$,

$\sum P_d = 1 \Rightarrow \hat{P}_i = \frac{C_i}{\sum_{i=1}^{2D} C_i} \geq 0$

(c) $P(X_{even}) = P_2 + P_4 + \cdots + P_{2D}$, $P(X_{odd}) = P_1 + P_3 + \cdots + P_{2D-1}$

$P(X_{even}) = P(X_{odd}) \Leftrightarrow -P_1 + P_2 - P_3 + \cdots - P_{2D+1} + P_{2D} = 0$

$\Leftrightarrow \sum_{d=1}^{2D} (-1)^d P_d = 0$

(d) We have one more constraint, so

$Lag(P_1, \cdots, P_{2D}, \lambda, \mu) = \sum_{d=1}^{2D} C_d \log P_d + \lambda(\sum_{d=1}^{2D} P_d - 1) + \mu(\sum_{d=1}^{2D} (-1)^d P_d)$

$\frac{\partial Lag}{\partial P_i} = \frac{C_i}{P_i} + \lambda + (-1)^i \mu = 0$, $\frac{\partial Lag}{\partial \lambda} = \sum P_d - 1 = 0$, $\frac{\partial Lag}{\partial \mu} = \sum (-1)^d P_d = 0$.

For odd i, $\frac{C_i}{P_i} + \lambda - \mu = 0$; for even i, $\frac{C_i}{P_i} + \lambda + \mu = 0$.

then $\frac{C_1}{P_1} = \frac{C_3}{P_3} = \cdots = \frac{C_{2D-1}}{P_{2D-1}}$, $\frac{C_2}{P_2} = \frac{C_4}{P_4} = \cdots = \frac{C_{2D}}{P_{2D}}$.

Since $\sum_{i\,odd} P_i = \sum_{i\,even} P_i$, we have $P_i = \begin{cases} \frac{C_i}{2C_{odd}}, & \text{if } i \text{ odd} \\ \frac{C_i}{2C_{even}}, & \text{if } i \text{ even} \end{cases}$

then $\sum_{i\,odd} P_i = \sum_{i\,even} P_i = \frac{1}{2}$.

2.(a) $P_{ML}(X_1 = x) = \frac{Count_1(X)}{T}$, $i \geq 1$: $P_{ML}(X_{i+1} = x' | X_i = x) = \frac{count_i(X, X')}{count_i(X)}$

(b) $P_{ML}(X_n = x) = \frac{count_n(X)}{T}$, $i < n$: $P_{ML}(X_i = x | X_{i+1} = x') = \frac{count_i(X, X')}{count_{i+1}(X')}$

(c) $P_{G1}(X_1 = x_1, \cdots, X_n = x_n) = P(X_1 = x_1) P(X_2 = x_2 | X_1 = x_1) \cdots P(X_n = x_n | X_{n-1} = x_{n-1})$

$= \frac{Count_1(X_1)}{T} \prod_{i=1}^{n-1} \frac{count_i(X_i, X_{i+1})}{count_i(X_i)} = \frac{\cancel{Count_1(X_1)}}{T} \frac{Count_1(X_1, X_2)}{\cancel{Count_1(X_1)}} \cdots \frac{Count_{n-1}(X_{n-1}, X_n)}{Count_{n-1}(X_{n-1})}$

$P_{G2}(X_1 = x_1, \cdots, X_n = x_n) = P(X_n = x_n) P(X_{n-1} = x_{n-1} | X_n = x_n) \cdots P(X_1 = x_1 | X_2 = x_2)$

$= \frac{Count_n(X_n)}{T} \prod_{i=1}^{n-1} \frac{count_i(X_i, X_{i+1})}{count_{i+1}(X_{i+1})} = \frac{\cancel{Count_n(X_n)}}{T} \frac{Count_1(X_1, X_2)}{Count_2(X_2)} \cdots \frac{count_{n-1}(X_{n-1}, X_n)}{\cancel{Count_n(X_n)}}$

So P_{G1}, P_{G2} have the same joint distribution

(d) Since $P(X_{n-1} | X_{n-2}) \neq P(X_{n-1} | X_{n-2}, X_{n-3})$ 「d.sep(3) fails」, then we can not expand the joint probability as conditional only on one previous node.