6.1) CSE 250A - Homework 6

a) $P(a,b|c,d) = \frac{P(a,b,c,d)}{P(c,d)}$ Baye's Rule

$P(a,b,c,d) =$

$\underbrace{p(d|a,b,c)}_{\text{causal chain}} \cdot P(c|a,b) \cdot P(b|a) \cdot P(a) =$ Chain Rule

$P(d|b,c) \cdot P(c|a,b) \cdot P(b|a) \cdot P(a)$

$P(c,d) = \sum_a \sum_b P(a,b,c,d) =$

$\sum_a \sum_b \underbrace{P(d|a,b,c)}_{\text{causal chain}} \cdot P(c|a,b) \cdot P(b|a) \cdot P(a) =$ Chain Rule

$\sum_a \sum_b P(d|b,c) \cdot P(c|a,b) \cdot P(b|a) \cdot P(a)$

$$\boxed{P(a,b|c,d) = \frac{P(d|b,c) \cdot P(c|a,b) \cdot P(b|a) \cdot P(a)}{\sum_a \sum_b P(d|b,c) \cdot P(c|a,b) \cdot P(b|a) \cdot P(a)}}$$

b) $P(a|c,d) = \sum_b P(a,b|c,d) =$

$$\boxed{\frac{\sum_b P(d|b,c) \cdot P(c|a,b) \cdot P(b|a) \cdot P(a)}{\sum_a \sum_{b'} P(d|b',c) \cdot P(c|a,b') \cdot P(b'|a) \cdot P(a)}}$$

$P(b|c,d) = \sum_a P(a,b|c,d) =$

$$\boxed{\frac{\sum_a P(d|b,c) \cdot P(c|a,b) \cdot P(b|a) \cdot P(a)}{\sum_{a'} \sum_b P(d|b,c) \cdot P(c|a',b) \cdot P(b|a') \cdot P(a')}}$$

c) $P(C=c_t, D=d_t) =$

$\sum_a \sum_b P(a,b,C=c_t,D=d_t) =$ From part a

$\sum_a \sum_b P(D=d_t|b,C=c_t) \cdot P(C=c_t|a,b) \cdot P(b|a) \cdot P(a)$

$\downarrow$

$\mathcal{L} = \sum_t \log(P(C=c_t, D=d_t))$

$$\boxed{\mathcal{L} = \sum_t \log\left(\sum_a \sum_b P(D=d_t|b,C=c_t) \cdot P(C=c_t|a,b) \cdot P(b|a) \cdot P(a)\right)}$$

$$\frac{\dfrac{P(Y=y,\ X=x,\ Z_i=1)}{P(X=x,\ Z_i=1)} \cdot \dfrac{P(X_i=1,\ Z_i=1)}{P(Z_i=1)}}{\dfrac{P(Y=y,\ X=x)}{P(X=x)}} =$$

$$\frac{P(Y=y,\ X=x,\ Z_i=1) \cdot P(X_i=1,\ Z_i=1) \cdot P(X=x)}{P(X=x,\ Z_i=1) \cdot P(Y=y,\ X=x) \cdot P(Z_i=1)} =$$

6.2)

a) $P(Y=1|X) = \sum_{Z=\{0,1\}^n} P(Y=1, Z|X)$

Compute $P(Y=1, Z|X) = P(Y=1|Z) \cdot P(Z|X)$

Note that

$$P(Y=1|Z) = \begin{cases} 1 & \text{if } Z_i = 1 \text{ for any } i \\ 0 & \text{if } Z_i = 0 \text{ for all } i \end{cases}$$

$$P(Y=1|Z) \cdot P(Z|X) = \prod_{i=1}^{n} P(Z_i|X_i)$$

Substitute this back into $P(Y=1|X)$

$P(Y=1|X) =$

$\sum_{Z=\{0,1\}^n} P(Y=1, Z|X) =$

$\sum_{Z=\{0,1\}^n} \left(\prod_{i=1}^{n} P(Z_i|X_i)\right) =$

Note that

$P(Z_i=1|X_i=0) = 0$

$P(Z_i=1|X_i=1) = P_i$

This also implies that

$P(Z_i=0|X_i=0) = 1$

$P(Z_i=1|X_i=1) = 1-P_i$

$$\sum_{Z=\{0,1\}^n} \prod_{i=1}^{n} P(Z_i=1|X_i)^{X_i} \cdot P(Z_i=0|X_i)^{1-X_i} =$$

$$\sum_{Z=\{0,1\}^n} \prod_{i=1}^{n} \left(P_i^{X_i} \cdot (1-P_i)^{1-X_i}\right)$$

Note: for the cases where Z has at least one $Z_i=1$, the product term for each i such that $Z_i=0$ becomes 0, since $(1-P_i)^{1-X_i}$ is 0 when $X_i=0$.

Thus, for the sum, we can simplify it to only consider cases where Z has at least one $z_i = 1$.

$P(Y=1|X) =$

$$\sum_{\substack{z \in \{0,1\}^n \\ \text{at least one } z_i = 1}} \prod_{i=1}^{n} \left((p_i^{x_i}) \cdot (1-p_i)^{1-x_i} \right) =$$

This is effectively a noisy-OR implementation.

$$P(Y=1|X) = 1 - \prod_{i=1}^{n} (1 - p_i^{x_i})$$

b) $P(Z_i = 1, X_i = 1 | X = x, Y = y) =$ Baye's rule

$$\frac{P(X=x, Y=y | Z_i = 1, X_i = 1) \cdot P(Z_i = 1 | X_i = 1)}{P(X=x, Y=y)}$$

Compute $P(X=x, Y=y | Z_i = 1, X_i = 1)$
Recall from the extended belief network that Z_i is the intermediary between X_i and Y. Also recall that $P(Z_i = 1 | X_i = 1) = p_i$. Thus, $Z_i = 1$ implies that $X_i = 1$, so we can simplify as follows:
$P(X=x, Y=y | Z_i = 1, X_i = 1) =$
$P(X=x, Y=y | X_i = 1) =$
$P(Y=y | X=x) \leftarrow$ noisy-OR

For $P(Z_i = 1 | X_i = 1)$, we are given that $P(Z_i = 1 | X_i = 1) = p_i$

Regarding the denominator (next page),

$P(X=x, Y=y) = \sum_{z \in \{0,1\}^n} P(X=x, Y=y \mid z) \cdot P(z)$

Recall from the extended belief network's structure and from part a that

$P(X=x, Y=y \mid z) =$

$\prod_{i=1}^{n} P(z_i \mid X_i)$

and

$P(z) =$

$\prod_{i=1}^{n} (p_i^{X_i} \cdot (1-p_i)^{1-X_i})$

Thus,

$P(X=x, Y=y) =$

$\sum_{z \in \{0,1\}^n} P(X=x\ Y=y \mid z) \cdot P(z) =$

$\sum_{z \in \{0,1\}^n} \prod_{i=1}^{n} P(z_i \mid X_i) \cdot \prod_{i=1}^{n} (p_i^{X_i} \cdot (1-p_i)^{1-X_i}) =$ from part a

$P(Y=1 \mid X) =$

$1 - \prod_{i=1}^{n} (1 - p_i^{X_i})$

Putting these all together, we get

$P(Z_i=1, X_i=1 \mid X=x, Y=y) =$

$\frac{P(X=x, Y=y \mid Z_i=1, X_i=1) \cdot P(Z_i=1 \mid X_i=1)}{P(X=x, Y=y)} =$

$\frac{P(Y=y \mid X=x) \cdot p_i}{1 - \prod_{i=1}^{n} (1-p_i^{X_i})}$

Regarding $P(Y=y \mid X=x)$, note that if $X_i=1$, it implies that $Y=1$ and $X_i=1$ based on the definitions given in this problem. Thus, the event that $X=x$ and $Y=y$ can only occur when $y \cdot x_i = 1$.

This shows that

$$P(Z_i=1, X_i=1 \mid X=x, Y=y) = \frac{y \cdot x_i \cdot p_i}{1-\prod_{i=1}^{n}(1-p_i)^{x_i}}$$

c) From part a, recall that $Z_i=1$ implies that $X_i=1$. From the problem statement, $P(Z_i=1 \mid X_i=1)=p_i$, but since $Z_i=1$ implies that $X_i=1$, we can simply compute that $P(X_i=1)=p_i$

The special case of the general formula from lecture is

$$P(X_i=x)=\frac{1}{T_i}\sum_t P(X_i=x \mid V_t=v_t)$$

Our data set is $\{\vec{x}^{(t)}, \vec{y}^{(t)}\}_{t=1}^{T}$.
Thus, we have

$$p_i \leftarrow \frac{1}{T_i}\sum_t P(X_i=x \mid X=x^{(t)}, Y=y^{(t)})$$

6.3)

a) $f'(x) = \frac{d}{dx} \log(\cosh(x)) =$

$\frac{1}{\cosh(x)} \cdot \frac{d}{dx}(\cosh(x)) =$

$\frac{\sinh(x)}{\cosh(x)}$

Find critical points of $f'(x)$

$\frac{\sinh(x)}{\cosh(x)} = 0$ when $x = 0$ because

$\frac{\sinh(x)}{\cosh(x)}\Big|_{x=0} =$

$\frac{\sinh(0)}{\cosh(0)} =$

$\frac{0}{1} =$

0

$f''(x) = \frac{d}{dx}\left(\frac{\sinh(x)}{\cosh(x)}\right) =$

$\frac{\cosh(x) \cdot \cosh(x) - \sinh(x) \cdot \sinh(x)}{\cosh^2(x)} =$

$\frac{\cosh^2(x) - \sinh^2(x)}{\cosh^2(x)} =$

$\frac{1}{\cosh^2(x)}$

This implies that $f''(x) > 0$ for all values of x.

By the second derivative $f''(x) > 0$ and $f'(x) = 0$ when $x = 0$, this implies that $x = 0$ is the minimum.

b) $f''(x) = \dfrac{1}{\cosh^2(x)}$ from part a.

Note that $\cosh(x) \geq 1$ for all x. This means that $\cosh^2(x) \geq 1$ for all x too.

Because $\cosh^2(x)$ is in the denominator, this means that $f''(x) \leq 1$ for all x. Thus, $f''(x) \leq 1$ for all x.

c) $Q = (x, y) = \log(\cosh(y)) + \dfrac{\sinh(y)}{\cosh(y)} \cdot (x-y) + \dfrac{1}{2} \cdot (x-y)^2$

d)

i) $Q(x, x) = f(x)$

$$Q(x,x) = f(x) + f'(x)\cdot(x-x) + \tfrac{1}{2}\cdot(x-x)^2 =$$
$$f(x) + f'(x)\cdot 0 + \tfrac{1}{2}\cdot 0^2 =$$
$$f(x) + 0 + 0 =$$
$$f(x)$$

Thus, $Q(x,x) = f(x)$.

ii) $Q(x,y) \geq f(x)$

First, let's solve parts of the integral:

$$f(x) = f(y) + \int_y^x du\cdot\left[f'(y) + \int_y^u dv\cdot f''(v)\right] =$$
$$f(y) + \int_y^x du\cdot f'(y) + \int_y^x du \int_y^u dv \cdot f''(v) =$$
$$f(y) + \left[f'(y)\cdot u\right]_{u=y}^{u=x} + \int_y^x du \int_y^u f''(v)\cdot dv =$$
$$\underbrace{f(y) + (x-y)\cdot f'(y)} + \int_y^x du \int_y^u f''(v)\cdot dv$$

We'll come back to this in a bit.

Recall that

$$Q(x,y) = \underbrace{f(y) + f'(y)\cdot(x-y)} + \tfrac{1}{2}\cdot(x-y)^2$$

Notice that this is equal to the first 2 parts of the solution for $f(x)$

$$f(x) = f(y) + f'(y)\cdot(x-y) + \int_y^x du \int_y^u f''(v)\cdot dv \rightarrow$$
$$f(x) - \int_y^x \int_y^u f''(v)\cdot dv = f(y) + f'(y)\cdot(x-y)$$

Consider $Q(x,y) \geq f(x)$. Rewrite as

$Q(x,y) - f(x) \geq 0$

Solve for $Q(x,y) - f(x)$ on the next page.

$Q(x,y) - f(x) =$

$f(y) + f'(y)\cdot(x-y) + \frac{1}{2}\cdot(x-y)^2 - f(x) =$

$f(x) - \int_y^x du \int_y^u f''(v)\cdot dv + \frac{1}{2}\cdot(x-y)^2 - f(x) =$

$\frac{1}{2}\cdot(x-y)^2 - \underbrace{\int_y^x du \int_y^u f''(v)\cdot dv}$

Let's analyze this.

From part b, we know that $f''(x) \leq 1$ for all x. This implies that

$\int_y^u f''(v)\cdot dv \leq \int_y^u 1\cdot dv = u - y \rightarrow$

$\int_y^x \int_y^u f''(v)\cdot dv \leq \int_y^x du \int_y^u 1\cdot dv =$

$\int_y^x (u-y)\cdot du =$

$\frac{1}{2}\cdot u^2 - u\cdot y \Big|_{u=y}^{u=x} =$

$\frac{1}{2}\cdot x^2 - x\cdot y - (\frac{1}{2}\cdot y^2 - y^2) =$

$\frac{1}{2}\cdot x^2 - x\cdot y - (-\frac{1}{2}\cdot y^2) =$

$\frac{1}{2}\cdot x^2 - x\cdot y + \frac{1}{2}\cdot y^2 =$

$\frac{1}{2}\cdot(x-y)^2$

So, plugging this back into $Q(x,y) - f(x)$

$Q(x,y) - f(x) =$ Because this is $\leq \frac{1}{2}\cdot(x-y)^2$

$\frac{1}{2}\cdot(x-y)^2 - \int_y^x du \int_y^u f''(v)\cdot dv \geq$

$\frac{1}{2}\cdot(x-y)^2 - \frac{1}{2}\cdot(x-y)^2 =$

0

So, this implies that $Q(x,y) - f(x) \geq 0$ and that

$Q(x,y) \geq f(x)$