CSE 250A HW6

Jiping Lin A15058075

6.2 (a) $P(a,b|c,d) = \dfrac{P(a,b,c,d)}{P(c,d)}$

$= \dfrac{P(a)P(b|a)P(c|a,b)P(d|a,b,c)}{P(c,d)}$

$d \to a | b,c$

$= \dfrac{P(a)P(b|a)P(c|a,b)P(d|b,c)}{\sum_{a'}\sum_{b'} P(a=a', b=b', c, d)}$ 「d.sep (1)」

$= \dfrac{P(a)P(b|a)P(c|a,b)P(d|b,c)}{\sum_{a'}\sum_{b'} P(a=a')P(b=b'|a=a')P(c|a=a', b=b')P(d|b=b', c)}$

(b) $P(a|c,d) = \sum_{b'} P(a, b=b'|c,d)$ (part (a))

$P(b|c,d) = \sum_{a'} P(a=a', b|c,d)$ (part (a))

(c) $\mathcal{L} = \sum_t \log \sum_{a'}\sum_{b'} P(A=a', B=b', C=c_t, D=d_t)$

$= \sum_t \log \sum_{a'}\sum_{b'} P(A=a')P(B=b'|A=a')P(C=c_t|A=a', B=b')P(D=d_t|B=b', C=c_t)$

↳ d.sep (1) like pt. (a).

(d) $P(A=a) \leftarrow \frac{1}{T}\sum_t P(A=a|C=c_t, D=d_t)$

$P(B=b|A=a) \leftarrow \dfrac{\sum_t P(B=b, A=a|C=c_t, D=d_t)}{\sum_t P(A=a|C=c_t, D=d_t)}$

$P(C=c|A=a, B=b) \leftarrow \dfrac{\sum_t I(c, c_t) P(A=a, B=b|C=c_t, D=d_t)}{\sum_t P(A=a, B=b|C=c_t, D=d_t)}$

$P(D=d|B=b, C=c) \leftarrow \dfrac{\sum_t I(c, c_t) I(d, d_t) P(B=b|C=c_t, D=d_t)}{\sum_t I(c, c_t) P(B=b|C=c_t, D=d_t)}$

6.3 (a). We need to show in second BN,

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

Summing over all possible $(z_1, \dots, z_n) = z \in \{0,1\}^n$, we have

$P(Y=1|X) = \sum_{z \in \{0,1\}^n} P(Y=1, z|X) = \sum_z P(z|X)P(Y=1|z, X) = \sum_z P(z|X)P(Y=1|z)$ (*)

「Marginalization」 「d.sep(1)」

For any z, $P(Y=1|z) = \begin{cases} 1, & \text{if } z_i=1 \text{ for some } i \\ 0, & \text{if } z_i=0 \ \forall i \end{cases}$

then $P(Y=1|z)$ only is 0 when all z_i's are 0.

$X=0 \Rightarrow \prod = 1$

$X=1 \Rightarrow \prod = 1-p_i$

So $(*) = \underbrace{\sum_z P(z|X)}_{\text{All situation}} - \underbrace{P(\text{All } z_i=0|X)}_{\text{only case}} = 1 - \prod_{i=1}^{n} P(z_i=0|X) = 1 - \prod_{i=1}^{n} (1-p_i)^{x_i}$

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

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

d.sep(1)

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

Z_i only relates to X_i

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

$$= \begin{cases} 0, & \text{if } x_i=0 \\ p_i, & \text{if } x_i=1 \end{cases} \Leftrightarrow p_i I(X_i=1)$$

since $P(Y=1 \mid Z_i=1)=1$ so y must be 1

$$= \frac{I(X_i,1)\,I(y,1)\left(p_i I(x_i,1)\right)}{P(Y=1 \mid X)}$$

$$= \frac{x_i y p_i}{1-\prod_j (1-p_j)^{x_j}}$$

$\rightarrow$ part (a)

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

$$= \frac{\sum_t P(Z_i=1, X_i=1 \mid X=x^{(t)}, Y=y^{(t)})}{\sum_t I(x_i^{(t)},1)}$$

$\sum_t I(x_i^{(t)},1) \rightarrow = T_i$ by def.

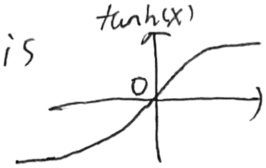
$$= \frac{1}{T_i}\sum_t P(Z_i=1, X_i=1 \mid X=x^{(t)}, Y=y^{(t)}).$$

(d)

iter	M	$\mathcal{L}$
0	175	-0.95809
1	56	-0.49592
2	43	-0.40822
4	42	-0.36461
8	44	-0.34750
16	40	-0.33462
32	37	-0.32258
64	37	-0.31482
128	36	-0.31116
256	36	-0.31016

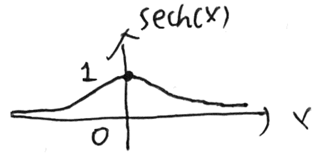
6.4 (a) $f'(x) = \frac{\sinh(x)}{\cosh(x)} = \tanh(x)$

The graph of $\tanh(x)$ is



so $f'(0)=0$, and f' has positive slope at $x=0$, so f has minimum at $x=0$

(b). $f''(x) = (\tanh(x))' = \operatorname{sech}^2(x)$. where the graph of $\operatorname{sech}(x)$ is



This implies $\operatorname{sech}^2(x) \le 1$ since $0 \le \operatorname{sech}(x) \le 1$, $\forall x$.

so $f''(x) \le 1 \ \forall x$.

(c) See code

(d) $Q(x,x) = f(x) + 0 + 0 = f(x)$.

$$f(x) = f(y) + \int_y^x du\left[f'(y) + \int_y^u dv\, f''(v)\right]$$

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

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

$$\le f(y) + f'(y)(x-y) + \int_y^x du \int_y^u dv \qquad \ulcorner f''(v) \le 1 \text{ by part (b)} \lrcorner$$

$$= f(y) + f'(y)(x-y) + \int_y^x du\,(u-y)$$

$$= f(y) + f'(y)(x-y) + \left.\frac{u^2 - y^2}{2}\right|_y^{u=x}$$

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

(e) $Q(x, x_n) = f(x_n) + f'(x_n)(x - x_n) + \frac{1}{2}(x - x_n)^2$

$$\frac{\partial Q(x, x_n)}{\partial x} = 0 \Rightarrow f'(x_n) + (x - x_n) = 0 \qquad \frac{\partial^2 Q(x, x_n)}{\partial x^2} = 1 > 0 \Rightarrow \min.$$

so $x = x_n - f'(x_n) \Rightarrow x_{n+1} = x_n - f'(x_n) = x_n - \tanh(x)$.

(f) See code

(g) $x_{n+1} = x_n - f'(x_n)/f''(x_n)$

$$= x_n - \frac{\tanh(x_n)}{\operatorname{sech}^2(x_n)} = x_n - \tanh(x_n)\cosh^2(x_n)$$

$$= x_n - \sinh(x_n)\cosh(x_n)$$

Update rule diverges to infinity (see code).
since $\sinh(x_n)\cosh(x_n)$ can be very large.
$|x_1| < |x_0| \iff |x_0 - \sinh(x_0)\cosh(x_0)| < |x_0|$.
this implies $|x_0| \leq 1.06$ 「Inequality Solver」

(h). The minimum in this case is NOT 0.

(i) It suffices to show that $g''(x) \leq 1$, since $Q(x,y)$ is aux. fn. for all f s.t. $f''(x) \leq 1$ $\forall x$.

$$g'(x) = \frac{1}{10}\sum_{k=1}^{10} \tanh\left(x + \frac{2}{\sqrt{k}}\right)$$

$$g''(x) = \frac{1}{10}\sum_{k=1}^{10} \operatorname{sech}^2\left(x + \frac{2}{\sqrt{k}}\right)$$

$\operatorname{sech}^2(x)$ has maxima $= 1$ at $x = 0$, so $g''(x) \leq g''(0)$, $\forall x$

$$\Rightarrow g''(0) = \frac{1}{10}\sum_{k=1}^{10} \operatorname{sech}^2\left(\frac{2}{\sqrt{k}}\right) \leq \frac{1}{10}\sum_{k=1}^{10} \operatorname{sech}^2(0) \leq \frac{1}{10} \times 10 = 1$$

$\Rightarrow g''(x) \leq 1$, $\forall x$.

(j) $\frac{\partial R(x, x_n)}{\partial R} = 0 \Rightarrow g'(x_n) + (x - x_n) = 0$

$$\Rightarrow x_{n+1} = x_n - g'(x_n)$$

$$= x_n - \frac{1}{10}\sum_{k=1}^{10} \tanh\left(x_n + \frac{2}{\sqrt{k}}\right).$$

(k) -0.9800