CSE 250A: Assignment 5

Jiaxu Zhu A53094655

November 10, 2015

5.1 EM algorithm

(a) Posterior probability

$$\begin{aligned} P(a,c|b,d) &= \frac{P(a,c,b,d)}{P(b,d)} \\ &= \frac{P(a,c,b,d)}{\sum_{a',c'} P(a',b,c',d)} \\ &= \frac{P(a)P(b|a)P(c|a,b)P(d|b,c)}{\sum_{a',c'} P(a')P(b|a')P(c'|a',b)P(d|b,c')} \end{aligned}$$

(b) Posterior probability

$$\begin{aligned} P(a|b,d) &= \sum_{c'} P(a,c'|b,d) \\ P(c|b,d) &= \sum_{a'} P(a',c|b,d) \end{aligned}$$

(c) Log-likelihood

$$\begin{aligned} \mathcal{L} &= \sum_t \log P(B=b_t, D=d_t) \\ &= \sum_t \log \sum_{a',c'} P(A=a', B=b_t, C=c', D=d_t) \\ &= \sum_t \log \sum_{a',c'} P(A=a')P(B=b_t|A=a')P(C=c'|A=a', B=b_t)P(D=d_t|B=b_t, C=c') \end{aligned}$$

(d) EM algorithm

E-step

$$\begin{aligned} P(b,a|b_t,d_t) &= I(b,b_t)P(a|b_t,d_t) \\ P(c,a,b|b_t,d_t) &= I(b,b_t)P(a,c|b_t,d_t) \\ P(d,b,c|b_t,d_t) &= I(b,b_t)I(d,d_t)P(c|b_t,d_t) \end{aligned}$$

M-step

$$
\begin{aligned}
P(b|a) &= \frac{\sum_t P(b,a|b_t,d_t)}{\sum_t P(a|b_t,d_t)} \\
&= \frac{\sum_t I(b,b_t)P(a|b_t,d_t)}{\sum_t P(a|b_t,d_t)} \\
P(c,a,b|b_t,d_t) &= \frac{\sum_t P(c,b,a|b_t,d_t)}{\sum_t P(a,b|b_t,d_t)} \\
&= \frac{\sum_t I(b,b_t)P(a,c|b_t,d_t)}{\sum_t I(b,b_t)P(a|b_t,d_t)} \\
P(d,b,c|b_t,d_t) &= \frac{\sum_t P(d,c,b|b_t,d_t)}{\sum_t P(b,c|b_t,d_t)} \\
&= \frac{\sum_t I(b,b_t)I(d,d_t)P(c|b_t,d_t)}{\sum_t I(b,b_t)P(c|b_t,d_t)}
\end{aligned}
$$

5.2 EM algorithm for noisy-OR

(a)

$$
\begin{aligned}
P(Y=1|X) &= \sum_{Z\in\{0,1\}^n} P(Y=1,Z|X) \\
&= \sum_{Z\in\{0,1\}^n} P(Y=1|Z,X)P(Z|X) \\
&= \sum_{Z\in\{0,1\}^n} P(Y=1|Z)P(Z|X) \quad (d-separation \quad I) \\
&= \sum_{Z\in\{0,1\}^n} P(Y=1|Z)\prod_{i=1}^{n} P(Z_i|X_i) \quad (d-separation \quad III) \\
&= 1-\sum_{Z\in\{0,1\}^n} P(Y=0|Z)\prod_{i=1}^{n} P(Z_i|X_i) \quad (d-separation \quad III) \\
&= 1-P(Y=0|Z=\vec{0})\prod_{i=1}^{n} P(Z_i=0|X_i) \\
&= 1-\prod_{i=1}^{n}(1-p_i)^{X_i}
\end{aligned}
$$

(b)

$$
\begin{aligned}
P(Z_i=1,X_i=1|X=x,Y=y) &= P(x_i=1)P(Z_i=1|X=x,Y=y) \\
&= x_i\frac{P(Y=y|X=x,Z_i=1)P(Z_i=1|X=x)}{P(Y=y|X=x)} \\
&= \frac{yx_ip_i}{1-\prod_{i=1}^{n}(1-p_i)^{X_i}}
\end{aligned}
$$

(c)

$$
\begin{aligned}
p_i &= \frac{\sum_t P(Z_i=1,X_i=1|X=x^{(t)},Y=y^{(t)})}{\sum_t P(X_i=1|X=x^{(t)},Y=y^{(t)})} \\
&= \frac{\sum_t P(Z_i=1,X_i=1|X=x^{(t)},Y=y^{(t)})}{\sum_t P(X_i=1)} \\
&= \frac{1}{T_i}\sum_t P(Z_i=1,X_i=1|X=x^{(t)},Y=y^{(t)})
\end{aligned}
$$

#	0	1	2	4	8	16	32	64
$\mathcal{L}$	-2.5709	-0.6938	-0.5442	-0.5385	-0.5372	-0.5365	-0.5363	-0.5365

Table 1: the log-likelihood

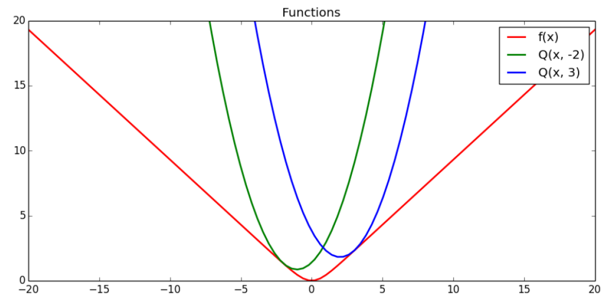

Figure 1: 5.3 (c)

where T_i is the number of examples in which $X_i = 1$.

(d) The completed version of table is shown in Tab. 1.

5.3 EM algorithm for noisy-OR

(a)

$$
\begin{aligned}
\cosh x &= \frac{e^x + e^{-x}}{2} \\
f(x) &= \log \frac{e^x + e^{-x}}{2} \\
f'(x) &= \frac{2}{e^x + e^{-x}} \frac{e^x - e^{-x}}{2} \\
&= \frac{e^{2x} - 1}{e^{2x} + 1} \\
f''(x) &= \frac{2e^{2x}(e^{2x} + 1) - 2e^{2x}(e^{2x} - 1)}{(e^{2x} + 1)^2} \\
&= \frac{4e^{2x}}{(e^{2x} + 1)^2}
\end{aligned}
$$

Let $f'(x) = 0$, we get $x = 0$. And since $f''(x) \geq 0$, we can tell that minimum occurs at $x = 0$.

(b)

$$
\begin{aligned}
f''(x) &= \frac{4e^{2x}}{(e^{2x} + 1)^2} \\
&= \left(\frac{2e^x}{e^{2x} + 1} \right)^2 \\
&= \left(\frac{1}{coshx} \right)^2 \\
&\leq 1
\end{aligned}
$$

(c) The plot is shown in Fig.

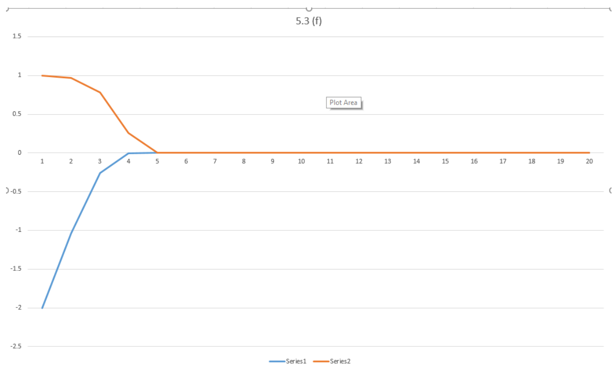

Figure 2: 5.3 (f)

(d) Compute the partial derivative of y:

$$
\begin{aligned}
\frac{\partial Q(x,y)}{\partial y} &= f'(y) + xf''(y) - f'(y) - yf''(y) - (x-y) \\
&= (f''(y) - 1)(x-y)
\end{aligned}
$$

Let this partial derivative equal zero, we get $y = x, y = 1$. And according to $\frac{\partial^2 Q(x,y)}{\partial^2 y}$,$y = x$ is the minimum point, which means $Q(x,y) \geq Q(x,x) = f(x)$.

(e)

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

It's a quatratic equation. So the minimum is $x_{n+1} = x_n - f'(x_n)$.

$$
x_{n+1} = x_n - \frac{e^{2x_n} - 1}{e^{2x_n} + 1}
$$

(f) The convergences are shown in Fig. .

(g) After several iterations, the update results changes to NAN. The reason is that $''(x)$ decrease too fast (quadratically). And the upper bound on $|x_0|$ so that Newtons method converges is 1.1.