CSE 150 Homework 5

Pedro Sousa Meireles
A15677282

Fall 2018

1 Maximum Likelihood Estimation

(a) **Complete data**

$$P(B=b|A=a) = \frac{\sum_{t=1}^{T} I(a,a_t)\cdot I(b,b_t)}{\sum_{t=1}^{T} I(a,a_t)}$$

$$P(C=c|A=a,B=b) = \frac{\sum_{t=1}^{T} I(a,a_t)\cdot I(b,b_t)\cdot I(c,c_t)}{\sum_{t=1}^{T} I(a,a_t)\cdot I(b,b_t)}$$

$$P(D=d|A=a,C=c) = \frac{\sum_{t=1}^{T} I(a,a_t)\cdot I(c,c_t)\cdot I(d,d_t)}{\sum_{t=1}^{T} I(a,a_t)\cdot I(d,d_t)}$$

(b) **Posterior probability**

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

(c) **Posterior probability**

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

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

(d) **Log-likelihood**

$$\begin{aligned} L &= \sum_t \log P(B=b_T, D=d_t) \\ &= \sum_t \log \sum_a \sum_c P(A=a,B=b_T,C=c,D=d_t) \\ &= \sum_t \log \sum_a \sum_c P(A=a)\cdot P(B=b_T|A=a)\cdot P(C=c|A=a,B=b_t)\cdot P(D=d_t|A=a,C=c) \end{aligned}$$

(e) **EM algorithm**

$$\begin{aligned} P(A=a) &\longleftarrow \frac{\widehat{count}(A=a)}{T} \\ &\longleftarrow \frac{\sum_{t=1}^{T} P(A=a|B=b_t,D=d_t)}{T} \\ &\longleftarrow \frac{\sum_{t=1}^{T}\sum_c P(A=a,C=c|B=b_t,D=d_t)}{T} \end{aligned}$$

$$
\begin{aligned}
P(B=b|A=a) &\longleftarrow \frac{\widehat{count}(A=a,B=b)}{\widehat{count}(A=a)} \\
&\longleftarrow \frac{\sum_{t=1}^{T} P(A=a,B=b|B=b_t,D=d_t)}{\widehat{count}(A=a)} \\
&\longleftarrow \frac{\sum_{t=1}^{T} I(b,b_t)\cdot P(A=a|B=b_t,D=d_t)}{\widehat{count}(A=a)} \\
&\longleftarrow \frac{\sum_{t=1}^{T}\sum_c I(b,b_t)\cdot P(A=a,C=c|B=b_t,D=d_t)}{\widehat{count}(A=a)} \\
&\longleftarrow \frac{\sum_{t=1}^{T}\sum_c I(b,b_t)\cdot P(A=a,C=c|B=b_t,D=d_t)}{\sum_{t=1}^{T}\sum_c P(A=a,C=c|B=b_t,D=d_t)}
\end{aligned}
$$

$$
\begin{aligned}
P(C=c|A=a,B=b) &\longleftarrow \frac{\widehat{count}(A=a,B=b,C=c)}{\widehat{count}(A=a,B=b)} \\
&\longleftarrow \frac{\sum_{t=1}^{T} P(A=a,B=b,C=c|B=b_t,D=d_t)}{\widehat{count}(A=a,B=b)} \\
&\longleftarrow \frac{\sum_{t=1}^{T} I(b,b_t)\cdot P(A=a,C=c|B=b_t,D=d_t)}{\widehat{count}(A=a,B=b)} \\
&\longleftarrow \frac{\sum_{t=1}^{T} I(b,b_t)\cdot P(A=a,C=c|B=b_t,D=d_t)}{\sum_{t=1}^{T}\sum_{c'} I(b,b_t)\cdot P(A=a,C=c'|B=b_t,D=d_t)}
\end{aligned}
$$

$$
\begin{aligned}
P(D=d|A=a,C=c) &\longleftarrow \frac{\widehat{count}(A=a,C=c,D=d)}{\widehat{count}(A=a,C=c)} \\
&\longleftarrow \frac{\sum_{t=1}^{T} P(A=a,C=c,D=d|B=b_t,D=d_t)}{\widehat{count}(A=a,C=c)} \\
&\longleftarrow \frac{\sum_{t=1}^{T} I(d,d_t)\cdot P(A=a,C=c|B=b_t,D=d_t)}{\widehat{count}(A=a,C=c)} \\
&\longleftarrow \frac{\sum_{t=1}^{T} I(d,d_t)\cdot P(A=a,C=c|B=b_t,D=d_t)}{\sum_{t=1}^{T} P(A=a,C=c|B=b_t,D=d_t)}
\end{aligned}
$$

2 EM algorithm for noisy-OR

(a) **Equivalence of models**

(i) Marginalization over $\overrightarrow{z}$

(ii) Product/chain rule was applied along with conditional independence between every z_i since all paths between different z_i are d-separated due to rule #3.

(iii) $P_B(Y=0|\overrightarrow{Z}=\overrightarrow{z})=0$ for any $z_i=1$, so the only remaining term is for $\overrightarrow{Z}=0$.

(iv) $P_B(Z_i=0|X_i=1)=1-p_i$ and $P_B(Z_i=0|X_i=0)=1$. Then $P_B(Z_i=0|X_i=x_i)=(1-p_i)^{x_i}$

(b) **EM Implementation: Per-iteration statistics**

(c) **EM Implementation: Estimated values for p_i**

(d) **EM Implementation: Source code**

3 EM algorithm for binary matrix completion

(a) **Sanity check**

(b) **Likelihood**

(c) **E-step**

(d) **M-step**

(e) **Implementation**

(f) **Personal categorization**

(g) **Personal movie recommendations**

(h) **Source code**