CSE 150A/250A. Assignment 5 (Solutions)

5.1 Survey *(2 pts)*

5.2 EM algorithm *(19 pts)*

(a) **Complete data** *(3 pts)*

$$\boxed{\begin{aligned} P(B\!=\!b|A\!=\!a) &= \frac{\sum_t I(a,a_t)\, I(b,b_t)}{\sum_t I(a,a_t)} \\ P(C\!=\!c|A\!=\!a,B\!=\!b) &= \frac{\sum_t I(a,a_t)\, I(b,b_t)\, I(c,c_t)}{\sum_t I(a,a_t)\, I(b,b_t)} \\ P(D\!=\!d|A\!=\!a,C\!=\!c) &= \frac{\sum_t I(a,a_t)\, I(c,c_t)\, I(d,d_t)}{\sum_t I(a,a_t)\, I(c,c_t)} \end{aligned}}$$

(b) **Posterior probability** *(3 pts)*

$$\begin{aligned} P(a,b|c,d) &= \frac{P(a,b,c,d)}{P(c,d)} \quad \boxed{\textbf{product rule}} \\ &= \frac{P(a)\,P(b|a)\,P(c|a,b)\,P(d|b,c)}{P(c,d)} \quad \boxed{\textbf{product rule}} \\ &= \frac{P(a)\,P(b|a)\,P(c|a,b)\,P(d|b,c)}{\sum_{a',b'} P(a',b',c,d)} \quad \boxed{\textbf{marginalization}} \\ &= \frac{P(a)\,P(b|a)\,P(c|a,b)\,P(d|b,c)}{\sum_{a',b'} P(a')\,P(b'|a')\,P(c|a',b')\,P(d|b',c)} \quad \boxed{\textbf{from numerator}} \end{aligned}$$

(c) **Posterior probability** *(2 pts)*)

$$\begin{aligned} P(a|c,d) &= \sum_b P(a,b|c,d) \quad \boxed{\textbf{marginalization}} \\ P(b|c,d) &= \sum_a P(a,b|c,d) \quad \boxed{\textbf{marginalization}} \end{aligned}$$

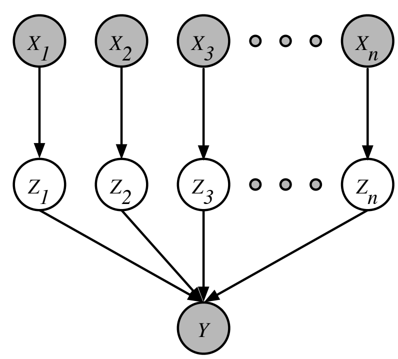
(d) **Log-likelihood** *(3 pts)*

$$\begin{aligned}
\mathcal{L} &= \sum_t \log P(C\!=\!c_t, D\!=\!d_t) \\
&= \sum_t \log \sum_{a,b} P(A\!=\!a, B\!=\!b, C\!=\!c_t, D\!=\!d_t) \quad \boxed{\textbf{marginalization}} \\
&= \sum_t \log \sum_{a,b} P(a)\, P(b|a)\, P(c_t|a,b)\, P(d_t|b,c_t) \quad \boxed{\textbf{product rule}}
\end{aligned}$$

(e) **EM algorithm** *(8 pts)*

$$\begin{aligned}
P(A\!=\!a) &\leftarrow \frac{\sum_t P(a|c_t,d_t)}{T} \\
P(B\!=\!b|A\!=\!a) &\leftarrow \frac{\sum_t P(a,b|c_t,d_t)}{\sum_t P(a|c_t,d_t)} \\
P(C\!=\!c|A\!=\!a,B\!=\!b) &\leftarrow \frac{\sum_t P(a,b|c_t,d_t)\, I(c,c_t)}{\sum_t P(a,b|c_t,d_t)} \\
P(D\!=\!d|B\!=\!b,C\!=\!c) &\leftarrow \frac{\sum_t P(b|c_t,d_t)\, I(c,c_t)\, I(d,d_t)}{\sum_t P(b|c_t,d_t)\, I(c,c_t)}
\end{aligned}$$

5.3 EM algorithm for noisy-OR *(24 pts)*



(a) **Noisy-OR** *(4 pts)*

$$
\begin{aligned}
P(Y\!=\!1|X) &= \sum_{Z\in\{0,1\}^n} P(Y\!=\!1, Z|X) \\
&= \sum_{Z\in\{0,1\}^n} P(Z|X)\, P(Y\!=\!1|Z,X) \quad \boxed{\textbf{product rule}} \\
&= \sum_{Z\in\{0,1\}^n} P(Z|X)\, P(Y\!=\!1|Z) \quad \boxed{\textbf{conditional independence}} \\
&= \sum_{Z\in\{0,1\}^n} P(Z|X) \left[1 - I(Z, \{0\}^n)\right] \quad \boxed{\textbf{logical OR}} \\
&= \left[\sum_{Z\in\{0,1\}^n} P(Z|X)\right] \; - \; P(Z = \{0\}^n|X) \quad \boxed{\textbf{separating terms}} \\
&= 1 \; - \; \prod_{i=1}^{n} P(Z_i\!=\!0|x_i) \quad \boxed{\textbf{conditional independence}} \\
&= 1 \; - \; \prod_{i=1}^{n} (1\!-\!p_i)^{x_i} (1)^{1-x_i} \quad \boxed{\textbf{noisy copy}} \\
&= 1 \; - \; \prod_{i=1}^{n} (1\!-\!p_i)^{x_i}
\end{aligned}
$$

(b) **E-step** *(4 pts)*

$$\begin{aligned}
P(Z_i\!=\!1, X_i\!=\!1|x,y) &= I(x_i,1)\,P(Z_i\!=\!1|x,y) \quad \boxed{\textbf{logical consistency}} \\
&= I(x_i,1)\left[\frac{P(y|x,Z_i\!=\!1)\,P(Z_i\!=\!1|x)}{P(y|x)}\right] \quad \boxed{\textbf{Bayes rule}} \\
&= I(x_i,1)\left[\frac{I(y,1)\,P(Z_i\!=\!1|x)}{P(y|x)}\right] \quad \boxed{\textbf{logical OR}} \\
&= I(x_i,1)\left[\frac{I(y,1)\,P(Z_i\!=\!1|x_i)}{P(y|x)}\right] \quad \boxed{\textbf{conditional independence}} \\
&= I(x_i,1)\left[\frac{I(y,1)\,p_i I(x_i,1)}{P(y|x)}\right] \quad \boxed{\textbf{noisy copy}} \\
&= \frac{y\,x_i\,p_i}{P(y|x)} \quad \boxed{\textbf{notation}} \\
&= \frac{y\,x_i\,p_i}{P(y\!=\!1|x)} \quad \boxed{\textbf{because numerator selects } y=1} \\
&= \frac{y\,x_i\,p_i}{1-\prod_j (1-p_j)^{x_j}}
\end{aligned}$$

(c) **M-step** *(3 pts)*

The general formula for the EM update gives:

$$P(Z_i\!=\!1|X_i\!=\!1) \;=\!\longleftarrow\; \frac{\sum_t P\left(X_i\!=\!1, Z_i\!=\!1|X\!=\!x^{(t)}, Y\!=\!y^{(t)}\right)}{\sum_t P\left(X_i\!=\!1|X\!=\!x^{(t)}, Y\!=\!y^{(t)}\right)}.$$

The denominator simplifies to $\sum_t I(x_{it},1) = \sum_t x_{it}$, or the number of inputs with $x_i\!=\!1$. Call this number T_i, and let p_i denote the conditional probability $P(Z_i\!=\!1|X_i\!=\!1)$. Then finally we obtain:

$$p_i \;\longleftarrow\; \frac{1}{T_i}\sum_t P\left(Z_i\!=\!1, X_i\!=\!1|X\!=\!x^{(t)}, Y\!=\!y^{(t)}\right).$$

(d) **Log-likelihood and error rates** *(5 pts)*

$$\begin{aligned}
\mathcal{L} &= \frac{1}{T}\sum_{t=1}^{T}\log P(y_t|\vec{x}_t) \\
&= \frac{1}{T}\sum_{t=1}^{T}\left[(1\!-\!y_t)\log P(y=0|\vec{x}_t) + y_t \log P(y\!=\!1|\vec{x}_t)\right] \\
&= \frac{1}{T}\sum_{t=1}^{T}\left[(1\!-\!y_t)\sum_{i=1}^{n} x_{it}\log(1\!-\!p_i) + y_t\log\left(1-\prod_{i=1}^{n}(1\!-\!p_i)^{x_{it}}\right)\right] \quad \boxed{\textbf{noisy-OR}}
\end{aligned}$$

Convergence

iteration	# mistakes M	log-likelihood $\mathcal{L}$
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	-0.31483
128	36	-0.31116
256	36	-0.31016

(e) **Source code** *(8 pts)*

5.4 Identifying and Formulating a Problem for Hidden Markov Modeling *(7 pts)*

(a) **Identify a different real-world or plausible scenario** Modeling the average health of servers in a cluster based on percentage of requests the cluster eventually failed to respond to.

(b) **Briefly explain *why* your chosen scenario is suitable for HMM modeling** We will model the health of the server cluster as the states, and the (buketed) percentage of failed requests as the observable.

The percentage of failed requests depends on the health of the server cluster at that moment. However, we assume that due to some system limitations, the number of failed servers itself is not visible at any moment (Hidden).

The health of the server cluster is assumed to depend only on its health at a preceeding moment (eg. the previous hour) - Markov property.

(c) **Formulate the problem for your chosen scenario using the components of a Hidden Markov Model.**

i **Hidden States (S_t)**: Health of a server cluster at hour t. $High$ if % of healthy servers is $> 90\%$. $Medium$ if % of servers is between 50 and 90 % and Low if % of healthy servers is $< 50\%$. $S_t \in \{High, Medium, Low\}$

ii **Observations (O_t)**: Percentage of failed requests in hour t. $High$ if % of Failed servers is $> 50\%$. $Medium$ if % of failed requests is between 10 and 50 % and Low if % of failed requests is $< 10\%$. $O_t \in \{High, Medium, Low\}$

iii **Initial State Distribution (π)**: We may start a server cluster at full health (i.e. all servers running well). $\pi(High) = 1, \pi(Medium) = \pi(Low) = 0$

iv **Transition Probabilities ($P(S_{t+1}|S_t)$)**: The number of healthy servers (servers ready to respond to requests) in the next hour is assumed to depend only on the number of healthy servers in the current hour. $P(S_{t+1} = High|S_t = High)$ is large. Maintainance activities (failure recovery processes) may determine $P(S_{t+1}|S_t = Low)$

v **Emission Probabilities ($P(O_t|S_t)$)**: If the average heath of the server cluster is high, the number of failed requests will be low, and vice versa. However, the number/percentage of failed requests won't tell us exactly how many servers have failed, so it is non-deterministic (even in the bucketed case). $P(O_t = Low|S_t = High) \approx 1$, and the percentage of failed requests won't depend on the health of the server at any other time.

5.5 Conditional independence *(12 pts)*

Consider the hidden Markov model (HMM) shown below, with hidden states S_t and observations O_t for times $t \in \{1, 2, \dots, T\}$. State whether the following statements of conditional independence are true or false.

False	$P(S_t \mid S_{t-1})$	=	$P(S_t \mid S_{t-1}, S_{t+1})$
True	$P(S_t \mid S_{t-1})$	=	$P(S_t \mid S_{t-1}, O_{t-1})$
False	$P(S_t \mid S_{t-1})$	=	$P(S_t \mid S_{t-1}, O_t)$
False	$P(S_t \mid O_{t-1})$	=	$P(S_t \mid O_1, O_2, \dots, O_{t-1})$
True	$P(O_t \mid S_{t-1})$	=	$P(O_t \mid S_{t-1}, O_{t-1})$
False	$P(O_t \mid O_{t-1})$	=	$P(O_t \mid O_1, O_2, \dots, O_{t-1})$
True	$P(S_2, S_3, \dots, S_T \mid S_1)$	=	$\prod_{t=2}^{T} P(S_t \mid S_{t-1})$
True	$P(S_1, S_2, \dots, S_{T-1} \mid S_T)$	=	$\prod_{t=1}^{T-1} P(S_t \mid S_{t+1})$
False	$P(S_1, S_2, \dots, S_T \mid O_1, O_2, \dots, O_T)$	=	$\prod_{t=1}^{T} P(S_t \mid O_t)$
False	$P(S_1, S_2, \dots, S_T, O_1, O_2, \dots, O_T)$	=	$\prod_{t=1}^{T} P(S_t, O_t)$
True	$P(O_1, O_2, \dots, O_T \mid S_1, S_2, \dots, S_T)$	=	$\prod_{t=1}^{T} P(O_t \mid S_t)$
True	$P(O_1, O_2, \dots, O_T)$	=	$\prod_{t=1}^{T} P(O_t \mid O_1, \dots, O_{t-1})$