Problem Set 3 Solutions

Calvin Walker

1.

(a) Consider the M-projection with the factored approximation $Q(X,Y) = Q(X)Q(Y)$,

$$\begin{aligned}
D(P||Q) &= \sum_{X,Y} P(X,Y) \log \frac{P(X,Y)}{Q(X)Q(Y)} \\
&= \mathbb{E}_P[\log P(X,Y) - \log Q(X)Q(Y)] \\
&= \mathbb{E}_P[\log P(X,Y)] - \Big(\mathbb{E}_p[\log Q(X)] + \mathbb{E}_p[\log Q(Y)]\Big) \pm \log(P(X)P(Y)) \\
&= \mathbb{E}_p\left[\log \frac{P(X,Y)}{P(X)P(Y)}\right] + \mathbb{E}_p\left[\log \frac{P(X)}{Q(X)}\right] + \mathbb{E}_p\left[\log \frac{P(Y)}{Q(Y)}\right]
\end{aligned}$$

Letting $Q_M^* = P(X)P(Y)$,

$$D(P||Q) = D(P||Q_M^*) + \mathbb{E}_p\left[\log \frac{P(X)}{Q(X)}\right] + \mathbb{E}_p\left[\log \frac{P(Y)}{Q(Y)}\right]$$

Thus, $D(P||Q) \geq D(P||Q_M^*)$, and $Q_M^* = P(X)P(Y)$ minimizes the M-projection.

(b)

$$\begin{aligned}
\theta^* &= \arg\max_\theta \prod_{i=1}^{M} Q(X^{(i)};\theta) \\
&= \arg\min_\theta \sum_{i=1}^{M} -\log Q(X^{(i)};\theta) \\
&= \arg\min_\theta \sum_{i=1}^{M} -\log Q(X^{(i)};\theta) + P(X^{(i)})
\end{aligned}$$

So if the sample size M is significantly large,

$$\begin{aligned}
\theta^* &= \arg\min_\theta \ \mathbb{E}_P\left[\log \frac{P(X)}{Q(X;\theta)}\right] \\
&= \arg\min_\theta \ D(P||Q(X;\theta))
\end{aligned}$$

Therefore, the MLE solution θ^* minimizes the KL-Divergence $D(P||Q(X;\theta))$, and is equivalent to solving for the M-projection $D(P||Q)$.

(c)

2.

(a) $\mathcal{M} \subseteq Local[\mathcal{U}]$, since any valid distribution P over $\mathcal{X}$ will satisfy the constraints of the local-consistency polytope. For a clique tree $\mathcal{T}$, under the constraints of the local-consistencey polytope, the pseudo marginals must be locally consistent,

$$\mu_{i,j}(S_{i,j}) = \sum_{C_i - S_{i,j}} \beta_i(C_i) = \sum_{C_j - S_{i,j}} \beta_i(C_j)$$

which implies that the clique tree is calibrated. So we have the clique tree invariant

$$\tilde{P}_\Phi(\mathcal{X}) = \frac{\prod_i \beta_i(C_i)}{\prod_{i,j} \mu_{i,j}(S_{i,j})}$$

where $\beta_i(C_i) \propto \tilde{P}_\Phi(C_i)$ and $\mu_{i,j}(S_{i,j}) \propto \tilde{P}_\Phi(S_{i,j})$. Therefore, the clique and sepset marginals define a valid distribution over $\mathcal{X}$, and $Local[\mathcal{U}] \subseteq \mathcal{M}$. So we have that $Local[\mathcal{U}]$ is equivalent to $\mathcal{M}$ for a clique tree $\mathcal{T}$.

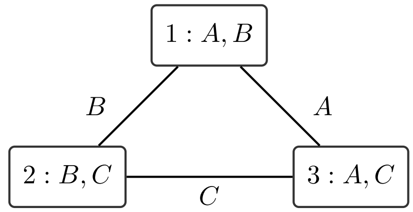

Figure 1: Example Cluster Graph

(b)

Consider the above Cluster Graph for a pairwise MRF over three binary variables A, B, and C. We can satisfy the local consistency constraints with the following clique beliefs:

$$\beta_1(A,B) = \begin{bmatrix} 0.45 & 0.05 \\ 0.05 & 0.45 \end{bmatrix} \; \beta_2(B,C) = \begin{bmatrix} 0.45 & 0.05 \\ 0.05 & 0.45 \end{bmatrix} \; \beta_3(A,C) = \begin{bmatrix} 0.05 & 0.45 \\ 0.45 & 0.05 \end{bmatrix}$$

Such that the sepset marginals are given by: $\mu_{1,3}(A) = \mu_{1,2}(B) = \mu_{2,3}(C) = (0.5, 0.5)$. If we assume that these are marginals for a valid probability distribution,

$$\begin{aligned} P(A=0,B=0) &= \beta_1(A=0,B=0) = 0.45 = P(0,0,1) + P(0,0,0) \\ P(A=0,C=0) &= \beta_3(A=0,C=0) = 0.05 = P(0,1,0) + P(0,0,0) \\ P(B=0,C=1) &= \beta_2(B=0,C=1) = 0.05 = P(1,0,1) + P(0,0,1) \end{aligned}$$

So $P(0,0,1) \leq 0.05$ and $P(0,0,0) \leq 0.05$, but $P(0,0,1) + P(0,0,0) = 0.45$. So the parameterization does not correspond to a valid probability distribution, and we can see for a cluster graph that is not a tree, the marginal polytope is strictly contained by the local consistency polytope.

3.

(a) (i) We can see that the pseudo-marginal distributions satisfy the marginalization condition, since:

$$\begin{aligned} \mu_{1,3}(X_1) = (0.5, 0.5) &= \sum_{X_2} \beta_1(X_1, X_2) = \sum_{X_3} \beta_3(X_1, X_3) \\ \mu_{1,2}(X_2) = (0.5, 0.5) &= \sum_{X_1} \beta_1(X_1, X_2) = \sum_{X_3} \beta_2(X_2, X_3) \\ \mu_{2,3}(X_3) = (0.5, 0.5) &= \sum_{X_2} \beta_2(X_2, X_3) = \sum_{X_1} \beta_3(X_1, X_3) \end{aligned}$$

And that the normalization conditions hold:

$$\sum_{X_1,X_2} \beta_1(X_1,X_2) = \sum_{X_2,X_3} \beta_2(X_2,X_3) = \sum_{X_1,X_3} \beta_3(X_1,X_3) = 1$$

and $\beta_i(c_i) \geq 0 \; \forall i$. Therefore, they are calibrated and locally consistent.

(ii) Assume that there is a valid distribution $P(X_1, X_2, X_3)$ with the beliefs as its marginals. Then,

$$\begin{aligned} P(X_1=0,X_2=0) &= \beta_1(X_1=0,X_2=0) = 0.4 = P(0,0,1) + P(0,0,0) \\ P(X_1=0,X_3=0) &= \beta_3(X_1=0,X_3=0) = 0.1 = P(0,1,0) + P(0,0,0) \\ P(X_2=0,X_3=1) &= \beta_2(X_2=0,X_3=1) = 0.1 = P(1,0,1) + P(0,0,1) \end{aligned}$$

So $P(0,0,1) \leq 0.1$ and $P(0,0,0) \leq 0.1$, but $P(0,0,1) + P(0,0,0) = 0.4$, a contradiction. So the pseudo-marginals can't constitute a vaild distribution.

(b)

$$\begin{aligned} P_\Phi(A,B) - \beta_1(A,B) &= \sum_{C,D} P_\Phi(A,B,C,D) - \sum_{C,D} P_\Phi(A,B,C,D) r(A,D) \\ &= P_\Phi(A,B) - P_\Phi(A,B) \sum_D r(A,D) \end{aligned}$$

So $\beta_1(A,B) = P_\Phi(A,B) \sum_D r(A,D)$, and we have a bound on the error for the estimated marginal $\beta_1(A,B)$

4.

(a)

Table 1: GMM Mean Vectors (Foreground)

1	2	3	4	5
36.52	84.58	28.38	54.43	54.73
-0.13	17.19	10.62	22.38	-1.72
-46.72	16.56	-23.34	4.80	-27.85

Table 2: GMM Covariance Traces (Foreground)

1	2	3	4	5
7.05	35.98	44.74	131.42	108.44
10.16	10.80	15.04	78.22	5.41
27.44	12.82	99.20	75.32	70.76

Table 3: GMM Mean Vectors (Background)

1	2	3	4	5
87.89	67.91	44.28	97.85	75.99
1.36	18.52	5.19	0.98	-2.30
-3.51	13.05	-9.32	0.89	-11.15

Table 4: GMM Covariance Traces (Background)

1	2	3	4	5
19.86	6.27	25.90	0.18	67.36
3.90	1.66	12.03	0.45	1.74
10.53	4.70	18.18	1.49	6.58

5.

(a)