CS229 Problem Set 3

Tianyu Du

Saturday 10th August, 2019

1 Problem 4: Semi-supervised EM

1.1 (a) Convergence

Proof.

$$
\begin{aligned}
\ell_{\text{semi-sup}}(\theta^{(t+1)}) &= \ell_{\text{unsup}}(\theta^{(t+1)}) + \alpha \ell_{\text{sup}}(\theta^{(t+1)}) && (1.1)\\
&= \sum_{i=1}^{n} \log \left(\sum_{z^{(i)}} Q_i^{(t)} \frac{p(x^{(i)}, z^{(i)}; \theta^{(t+1)})}{Q_i^{(t)}} \right) + \alpha \ell_{\text{sup}}(\theta^{(t+1)}) && (1.2)\\
&= \sum_{i=1}^{n} \log \mathbb{E}_{z^{(i)} \sim Q_i^{(t)}} \left[\frac{p(x^{(i)}, z^{(i)}; \theta^{(t+1)})}{Q_i^{(t)}} \right] + \alpha \ell_{\text{sup}}(\theta^{(t+1)}) && (1.3)\\
&\geq \sum_{i=1}^{n} \mathbb{E}_{z^{(i)} \sim Q_i^{(t)}} \left[\log \frac{p(x^{(i)}, z^{(i)}; \theta^{(t+1)})}{Q_i^{(t)}} \right] + \alpha \ell_{\text{sup}}(\theta^{(t+1)}) \because \text{Jensen's Inequality} && (1.4)\\
&= \sum_{i=1}^{n} \sum_{z^{(i)}} Q_i^{(t)} \log \left(\frac{p(x^{(i)}, z^{(i)}; \theta^{(t+1)})}{Q_i^{(t)}} \right) + \alpha \ell_{\text{sup}}(\theta^{(t+1)}) && (1.5)\\
&\geq \sum_{i=1}^{n} \sum_{z^{(i)}} Q_i^{(t)} \log \left(\frac{p(x^{(i)}, z^{(i)}; \theta^{(t)})}{Q_i^{(t)}} \right) + \alpha \ell_{\text{sup}}(\theta^{(t)}) \because \text{M-step is maximizing w.r.t. } \theta && (1.6)\\
&= \sum_{i=1}^{n} \log \left(\sum_{z^{(i)}} Q_i^{(t)} \frac{p(x^{(i)}, z^{(i)}; \theta^{(t)})}{Q_i^{(t)}} \right) + \alpha \ell_{\text{sup}}(\theta^{(t)}) && (1.7)\\
&= \ell_{\text{semi-sup}}(\theta^{(t)}) && (1.8)
\end{aligned}
$$

The last two steps were derive from the fact that $Q(\cdot)$ was specifically chosen in the E-step so that $\ell_{\text{semi-sup}}(\theta^{(t)})$ was equal to it's ELBO. ■

1.2 (b) Semi-supervised E-step

Answer $z^{(i)}$ for all unlabelled examples should be estimated. Specifically, posterior $p(z^{(i)}|x^{(i)};\mu,\Sigma,\phi)$ for all $i \in \{1,\cdots,n\}$ are estimated. Define $w_j^{(i)} = p(z^{(i)}=j|x^{(i)};\mu,\Sigma,\phi)$.

Proof.

$$
\begin{aligned}
w_j^{(i)} &:= p(z^{(i)}=j|x^{(i)};\mu,\Sigma,\phi) &(1.9)\\
&= \frac{p(x^{(i)}|z^{(i)}=j;\mu,\Sigma,\phi)p(z^{(i)}=j;\mu,\Sigma,\phi)}{p(x^{(i)};\mu,\Sigma,\phi)} &(1.10)\\
&= \frac{p(x^{(i)}|z^{(i)}=j;\mu,\Sigma,\phi)p(z^{(i)}=j;\mu,\Sigma,\phi)}{\sum_{\ell=1}^k \left\{p(x^{(i)}|z^{(i)}=\ell;\mu,\Sigma,\phi)p(z^{(i)}=\ell;\mu,\Sigma,\phi)\right\}} &(1.11)\\
&= \frac{\frac{1}{(2\pi)^{d/2}|\Sigma_j|^{1/2}}\exp\left[-\frac{1}{2}(x^{(i)}-\mu_j)^T\Sigma_j^{-1}(x^{(i)}-\mu_j)\right]\phi_j}{\sum_{\ell=1}^k \left\{\frac{1}{(2\pi)^{d/2}|\Sigma_\ell|^{1/2}}\exp\left[-\frac{1}{2}(x^{(i)}-\mu_\ell)^T\Sigma_\ell^{-1}(x^{(i)}-\mu_\ell)\right]\phi_\ell\right\}} &(1.12)
\end{aligned}
$$

■

1.3 (c) Semi-supervised M-step

1.3.1 Choosing $\mu_\ell^{(t+1)}$

Answer Let $\Theta := \{\mu_\ell, \Sigma_\ell, \phi_\ell\}_{\ell=1}^k$. The first order condition is for optimal μ_ℓ is:

Proof.

$$\begin{aligned}
\nabla\mu_\ell \; \ell_{\text{semi-sup}}(\Theta) &= \nabla\mu_\ell \; \sum_{i=1}^{n}\sum_{j=1}^{k} w_j^{(i)} \log\left(\frac{p(x^{(i)}, z^{(i)};\Theta)}{w_j^{(i)}}\right) + \alpha\sum_{i=1}^{\tilde{n}} \log\Big(p(\tilde{x}^{(i)}, \tilde{z}^{(i)};\Theta)\Big) && (1.13)\\
&= \nabla_{\mu_\ell} \; \sum_{i=1}^{n} w_\ell^{(i)} \log\left(\frac{p(x^{(i)}, z^{(i)};\Theta)}{w_\ell^{(i)}}\right) + \alpha\sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\} \log\Big(p(\tilde{x}^{(i)}, \tilde{z}^{(i)};\Theta)\Big) && (1.14)\\
&= \nabla_{\mu_\ell} \; \sum_{i=1}^{n} w_\ell^{(i)} \log\Big(p(x^{(i)}, z^{(i)};\Theta)\Big) + \alpha\sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\} \log\Big(p(\tilde{x}^{(i)}, \tilde{z}^{(i)};\Theta)\Big) && (1.15)\\
&= \nabla_{\mu_\ell} \; \sum_{i=1}^{n} w_\ell^{(i)} \log\Big(p(x^{(i)}|z^{(i)};\Theta)p(z^{(i)};\Theta)\Big) + \alpha\sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\} \log\Big(p(\tilde{x}^{(i)}|\tilde{z}^{(i)};\Theta)p(\tilde{z}^{(i)};\Theta)\Big) && (1.16)\\
&= \nabla_{\mu_\ell} \; \sum_{i=1}^{n} w_\ell^{(i)} \Big\{\log\Big(p(x^{(i)}|z^{(i)};\Theta)\Big)\Big\} + \alpha\sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\} \log\Big(p(\tilde{x}^{(i)}|\tilde{z}^{(i)};\Theta)\Big) && (1.17)\\
&= \nabla_{\mu_\ell} \; \sum_{i=1}^{n} w_\ell^{(i)} \left\{-\frac{1}{2}(x^{(i)} - \mu_\ell)^T\Sigma_\ell^{-1}(x^{(i)} - \mu_\ell)\right\} && (1.18)\\
&+ \alpha\sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\}\left\{-\frac{1}{2}(\tilde{x}^{(i)} - \mu_\ell)^T\Sigma_\ell^{-1}(\tilde{x}^{(i)} - \mu_\ell)\right\} && (1.19)\\
&= \sum_{i=1}^{n} w_\ell^{(i)} \left\{x^{(i)T}\Sigma_\ell^{-1} - \mu_\ell^T\Sigma_\ell^{-1}\right\} + \alpha\sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\}\left\{\tilde{x}^{(i)T}\Sigma_\ell^{-1} - \mu_\ell^T\Sigma_\ell^{-1}\right\} && (1.20)\\
&= 0 && (1.21)
\end{aligned}$$

By right multiplying Σ_ℓ^{-1},

$$\sum_{i=1}^{n} w_\ell^{(i)}\left\{x^{(i)} - \mu_\ell\right\} + \alpha\sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\}\left\{\tilde{x}^{(i)} - \mu_\ell\right\} = 0 \quad (1.22)$$

$$\implies \mu_\ell = \frac{\sum_{i=1}^{n} w_\ell^{(i)} x^{(i)} + \alpha\sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\}\tilde{x}^{(i)}}{\sum_{i=1}^{n} w_\ell^{(i)} + \alpha\sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\}} \quad (1.23)$$

■

1.4 Choosing Σ_ℓ

Proof. Let $S_\ell := \Sigma_\ell^{-1}$. From lecture we know that the first order condition of optimizing Σ_ℓ is the same as finding the first order condition for S_ℓ.

$$\nabla_{S_\ell}\ \ell_{\text{semi-sup}}(\Theta) = \nabla_{S_\ell}\ \sum_{i=1}^{n}\sum_{j=1}^{k} w_j^{(i)} \log\left(\frac{p(x^{(i)}, z^{(i)}; \Theta)}{w_j^{(i)}}\right) + \alpha \sum_{i=1}^{\tilde{n}} \log\left(p(\tilde{x}^{(i)}, \tilde{z}^{(i)}; \Theta)\right) \tag{1.24}$$

$$= \nabla_{S_\ell}\ \sum_{i=1}^{n} w_\ell^{(i)} \log\left(\frac{p(x^{(i)}, z^{(i)}; \Theta)}{w_\ell^{(i)}}\right) + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\} \log\left(p(\tilde{x}^{(i)}, \tilde{z}^{(i)}; \Theta)\right) \tag{1.25}$$

$$= \nabla_{S_\ell} \sum_{i=1}^{n} w_\ell^{(i)} \left(\log(|\Sigma_\ell|^{-1}) - \frac{1}{2}(x^{(i)} - \mu_\ell)^T S_\ell (x^{(i)} - \mu_\ell)\right) \tag{1.26}$$

$$+ \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\} \left(\log(|\Sigma_\ell|^{-1}) - \frac{1}{2}(\tilde{x}^{(i)} - \mu_\ell)^T S_\ell (\tilde{x}^{(i)} - \mu_\ell)\right) \tag{1.27}$$

$$= \nabla_{S_\ell} \sum_{i=1}^{n} w_\ell^{(i)} \left(\log(|S_\ell|) - \frac{1}{2}(x^{(i)} - \mu_\ell)^T S_\ell (x^{(i)} - \mu_\ell)\right) \tag{1.28}$$

$$+ \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\} \left(\log(|S_\ell|) - \frac{1}{2}(\tilde{x}^{(i)} - \mu_\ell)^T S_\ell (\tilde{x}^{(i)} - \mu_\ell)\right) \tag{1.29}$$

$$= \sum_{i=1}^{n} w_\ell^{(i)} S_\ell^{-T} - w_\ell^{(i)} (x^{(i)} - \mu_\ell)(x^{(i)} - \mu_\ell)^T + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\} \left(S_\ell^{-T} - (\tilde{x}^{(i)} - u_\ell)(\tilde{x}^{(i)} - u_\ell)^T\right) \tag{1.30}$$

$$= 0 \tag{1.31}$$

Since Σ_ℓ is symmetric, so $S_\ell^{-T} = \Sigma_\ell$. Above first order condition implies

$$\Sigma_\ell \left(\sum_{i=1}^{n} w_\ell^{(i)} + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\}\right) = \sum_{i=1}^{n} w_\ell^{(i)} (x^{(i)} - \mu_\ell)(x^{(i)} - \mu_\ell)^T + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\} (\tilde{x}^{(i)} - u_\ell)(\tilde{x}^{(i)} - u_\ell)^T \tag{1.32}$$

$$\implies \Sigma_\ell = \frac{\sum_{i=1}^{n} w_\ell^{(i)} (x^{(i)} - \mu_\ell)(x^{(i)} - \mu_\ell)^T + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\} (\tilde{x}^{(i)} - u_\ell)(\tilde{x}^{(i)} - u_\ell)^T}{\sum_{i=1}^{n} w_\ell^{(i)} + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\}} \tag{1.33}$$

■

1.5 Choosing ϕ_ℓ

Proof.

$$\nabla_{\phi_\ell} \ell_{\text{semi-sup}}(\Theta) = \nabla_{\phi_\ell} \sum_{i=1}^{n} \sum_{\ell=1}^{k} w_\ell^{(i)} \log \left(\frac{p(x^{(i)}|z^{(i)} = \ell; \Theta) p(z^{(i)} = \ell; \Theta)}{p(x^{(i)}|z^{(i)} = \ell; \Theta)} \right) + \alpha \sum_{i=1}^{\tilde{n}} \log \left(p(\tilde{x}^{(i)}|\tilde{z}^{(i)}; \Theta) p(\tilde{z}; \Theta) \right) \tag{1.34}$$

$$= \nabla_{\phi_\ell} \sum_{i=1}^{n} \sum_{\ell=1}^{k} w_\ell^{(i)} \log \left(p(z^{(i)} = \ell; \Theta) \right) + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\} \log(\phi_\ell) \tag{1.35}$$

$$= \nabla_{\phi_\ell} \sum_{i=1}^{n} \sum_{\ell=1}^{k} w_\ell^{(i)} \log(\phi_\ell) + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\} \log(\phi_\ell) \tag{1.36}$$

$$= \nabla_{\phi_\ell} \sum_{i=1}^{n} w_\ell^{(i)} \log(\phi_\ell) + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\} \log(\phi_\ell) \tag{1.37}$$

Using the constraint that $\sum_{\ell=1}^{k} \phi_\ell = 1$, the Lagrangian can be constructed as

$$\mathcal{L}(\cdot) = \ell_{\text{semi-sup}}(\phi, \cdot) + \lambda \left(1 - \sum_{\ell=1}^{k} \phi_\ell \right) \tag{1.38}$$

Solving the stationary point for $\mathcal{L}(\cdot)$ gives

$$\frac{\partial \mathcal{L}(\phi_\ell, \cdot)}{\partial \phi_\ell} = \nabla_{\phi_\ell} \ell_{\text{semi-sup}} - \lambda \tag{1.39}$$

$$= \frac{1}{\phi_\ell} \left(\sum_{i=1}^{n} w_\ell^{(i)} + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\} \right) - \lambda = 0 \tag{1.40}$$

$$\implies \frac{1}{\lambda} \left(\sum_{i=1}^{n} w_\ell^{(i)} + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\} \right) = \phi_\ell \tag{1.41}$$

Then, $\sum_{\ell=1}^{k} \phi_\ell = 1$ implies

$$\sum_{\ell=1}^{k} \frac{1}{\lambda} \left(\sum_{i=1}^{n} w_\ell^{(i)} + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\} \right) = 1 \tag{1.42}$$

$$\implies \sum_{\ell=1}^{k} \left(\sum_{i=1}^{n} w_\ell^{(i)} + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\} \right) = \lambda \tag{1.43}$$

Therefore,

$$\phi_\ell = \frac{\sum_{i=1}^{n} w_\ell^{(i)} + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\}}{\sum_{j=1}^{k} \left(\sum_{i=1}^{n} w_j^{(i)} + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = j\} \right)} \tag{1.44}$$

$$= \frac{\sum_{i=1}^{n} w_\ell^{(i)} + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\}}{\sum_{j=1}^{k} \sum_{i=1}^{n} w_j^{(i)} + \alpha \sum_{j=1}^{k} \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = j\}} \tag{1.45}$$

$$= \frac{\sum_{i=1}^{n} w_\ell^{(i)} + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\}}{\sum_{i=1}^{n} \sum_{j=1}^{k} w_j^{(i)} + \alpha \sum_{i=1}^{\tilde{n}} \sum_{j=1}^{k} \mathbb{1}\{\tilde{z}^{(i)} = j\}} \tag{1.46}$$

$$= \frac{\sum_{i=1}^{n} w_\ell^{(i)} + \alpha \sum_{i=1}^{\tilde{n}} \mathbb{1}\{\tilde{z}^{(i)} = \ell\}}{n + \alpha \tilde{n}} \tag{1.47}$$

$\blacksquare$