Problem Set #4: Unsupervised Learning and Reinforcement Learning

Trung H. Nguyen

1. **EM for Supervised Learning**

   (a) We have the log-likelihood of the parameters as following:

$$\begin{aligned}
\ell(\phi, \theta_0, \theta_1) &= \log \prod_{i=1}^{m} p(y^{(i)}|x^{(i)}; \phi, \theta_0, \theta_1) \\
&= \log \prod_{i=1}^{m} p(y^{(i)}|x^{(i)}, z^{(i)}; \theta_0, \theta_1) p(z^{(i)}|x^{(i)}; \phi) \\
&= \sum_{i=1}^{m} \left(\log p(y^{(i)}|x^{(i)}, z^{(i)} = 0; \theta_0) + \log p(z^{(i)} = 0|x^{(i)}; \phi) \right) + \\
&\quad \sum_{i=1}^{m} \left(\log p(y^{(i)}|x^{(i)}, z^{(i)} = 1; \theta_1) + \log p(z^{(i)} = 1|x^{(i)}; \phi) \right)
\end{aligned}$$

Taking the gradient of the log-likelihood w.r.t θ_0 and setting it equal to zero, we have:

$$\begin{aligned}
0 &= \nabla_{\theta_0} \ell(\phi, \theta_0, \theta_1) \\
&= \nabla_{\theta_0} \sum_{i=1}^{m} \log p(y^{(i)}|x^{(i)}, z^{(i)} = 0; \theta_0) \\
&= \nabla_{\theta_0} \sum_{i=1}^{m} \log \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y^{(i)} - \theta_0^T x^{(i)})^2}{2\sigma^2} \right) \\
&= \nabla_{\theta_0} \sum_{i=1, z^{(i)}=0}^{m} (y^{(i)} - \theta_0^T x^{(i)})^2
\end{aligned}$$

We can easily see that this is just the least squares problem on a subset of the data. Then, we can obtain the maximum likelihood estimate of θ_0:

$$\theta_0 = (X_{z=0}^T X_{z=0})^{-1} X_{z=0}^T \vec{y}_{z=0}$$

where $X_{z=0}$ and $\vec{y}_{z=0}$ are design matrix and vector created by examples where $z^{(i)} = 0$. Similarly, the maximum likelihood estimate of θ_1 is:

$$\theta_1 = (X_{z=1}^T X_{z=1})^{-1} X_{z=1}^T \vec{y}_{z=1}$$

And for the case of ϕ, we have that:

$$\begin{aligned}
\nabla_\phi \ell(\phi, \theta_0, \theta_1) &= \nabla_\phi \sum_{i=1}^{m} \log p(z^{(i)} = 0|x^{(i)}; \phi) + \sum_{i=1}^{m} \log p(z^{(i)} = 1|x^{(i)}; \phi) \\
&= \nabla_\phi \sum_{i=1}^{m} (1 - z^{(i)})(1 - g(\phi^T x^{(i)})) + \sum_{i=1}^{m} z^{(i)} g(\phi^T x^{(i)})
\end{aligned}$$

which is the standard form of logistic regression problem. Thus, we can obtain the gradient and Hessian:

$$\begin{aligned}
\nabla_\phi \ell(\phi, \theta_0, \theta_1) &= X^T(\vec{z} - \vec{h}), \text{ where } \vec{h}_i = h(x^{(i)}) = g(\phi^T x^{(i)}) \\
H &= \sum_{i=1}^{m} -h(x^{(i)}) \left(1 - h(x^{(i)})\right) x^{(i)} {x^{(i)}}^T
\end{aligned}$$

(b) When z is a latent variable, the log-likelihood is:

$$\begin{aligned}
\ell(\phi, \theta_0, \theta_1) &= \log \prod_{i=1}^{m} p(y^{(i)}|x^{(i)}; \phi, \theta_0, \theta_1) \\
&= \log \prod_{i=1}^{m} \sum_{z^{(i)}} p(y^{(i)}|x^{(i)}, z^{(i)}; \theta_0, \theta_1) p(z^{(i)}|x^{(i)}; \phi) \\
&= \sum_{i=1}^{m} \log \sum_{z^{(i)}} p(y^{(i)}|x^{(i)}, z^{(i)}; \theta_0, \theta_1) p(z^{(i)}|x^{(i)}; \phi) \\
&= \sum_{i=1}^{m} \log \left((1 - g(\phi^T x^{(i)})) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y^{(i)} - \theta_0^T x^{(i)})^2}{2\sigma^2} \right) + \right. \\
&\quad \left. g(\phi^T x^{(i)}) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y^{(i)} - \theta_1^T x^{(i)})^2}{2\sigma^2} \right) \right)
\end{aligned}$$

For each E-step, set:

$$\begin{aligned}
Q_i(z^{(i)}) &= p(z^{(i)}|x^{(i)}, y^{(i)}; \phi, \theta_0, \theta_1) \\
&= \frac{p(y^{(i)}|x^{(i)}, z^{(i)}; \theta_0, \theta_1) p(z^{(i)}|x^{(i)}; \phi)}{\sum_z p(y^{(i)}|x^{(i)}, z; \theta_0, \theta_1) p(z|x^{(i)}; \phi)}
\end{aligned}$$

In the M-step, we define:

$$w_j^{(i)} = p(z^{(i)} = j|x^{(i)}, y^{(i)}; \phi, \theta_j), \quad j = 0, 1$$

Taking the gradient of the lower bound $\sum_i \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(y^{(i)}|x^{(i)}, z^{(i)}; \phi, \theta_0, \theta_1) p(z^{(i)}|x^{(i)}; \phi)}{Q_i(z^{(i)})}$ of the log-likelihood w.r.t θ_0, we have:

$$\begin{aligned}
&\nabla_{\theta_0} \sum_i \sum_{j=0}^{1} w_j^{(i)} \log \frac{p(y^{(i)}|x^{(i)}, z^{(i)} = j; \phi, \theta_0, \theta_1) p(z^{(i)} = j|x^{(i)}; \phi)}{w_j^{(i)}} \\
&= \nabla_{\theta_0} \sum_i \left(w_0^{(i)} \log \frac{1}{w_0^{(i)}} (1 - g(\phi^T x^{(i)})) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y^{(i)} - \theta_0^T x^{(i)})^2}{2\sigma^2} \right) \right) \\
&= \frac{1}{2\sigma^2} \nabla_{\theta_0} \sum_i -w_0^{(i)} (y^{(i)} - \theta_0^T x^{(i)})^2
\end{aligned}$$

Setting this gradient equal to zero, we simply have:

$$\nabla_{\theta_0} \sum_i -w_0^{(i)} (y^{(i)} - \theta_0^T x^{(i)})^2 = 0$$

which is the weighted linear regression problem. And thus, it has the solution:

$$\theta_0 = (X_{z=0}^T W X_{z=0})^{-1} X_{z=0}^T W \vec{y}_{z=0},$$

where $W_{ii} = w_0^{(i)}$ Likewise, we have that:

$$\theta_1 = (X_{z=1}^T W X_{z=1})^{-1} X_{z=1}^T W \vec{y}_{z=1}$$

For the case of ϕ, we have:

$$\begin{aligned}
&\nabla_\phi \sum_i \sum_{j=0}^{1} w_j^{(i)} \log \frac{p(y^{(i)}|x^{(i)}, z^{(i)} = j; \phi, \theta_0, \theta_1) p(z^{(i)} = j|x^{(i)}; \phi)}{w_j^{(i)}} \\
&= \nabla_\phi \sum_i \left(w_0^{(i)} \log(1 - g(\phi^T x^{(i)})) + w_1^{(i)} \log g(\phi^T x^{(i)}) \right)
\end{aligned}$$

which is in the form of a logistic regression problem (since $w_0^{(i)} = 1 - w_1^{(i)}$). Therefore, the gradient and Hessian are obtained by:

$$\nabla_\phi \sum_i \sum_{j=0}^{1} w_j^{(i)} \log \frac{p(y^{(i)}|x^{(i)}, z^{(i)} = j; \phi, \theta_0, \theta_1) p(z^{(i)} = j|x^{(i)}; \phi)}{w_j^{(i)}} = X^T (\vec{w} - \vec{h})$$

$$H = \sum_{i=1}^{m} -h(x^{(i)})\left(1 - h(x^{(i)})\right) x^{(i)} {x^{(i)}}^T,$$

where $\vec{h}_i = h(x^{(i)}) = g(\phi^T x^{(i)})$

2. **Factor Analysis and PCA**

   (a) We have that:

$$\begin{aligned} z &\sim \mathcal{N}(0,\, I) \\ x|z &\sim \mathcal{N}(Uz,\, \sigma^2 I) \end{aligned}$$

is equivalent to:

$$\begin{aligned} z &\sim \mathcal{N}(0,\, I) \\ \epsilon &\sim \mathcal{N}(0,\, \sigma^2 I) \\ x &= Uz + \epsilon \end{aligned}$$

where $z,\ \epsilon$ are independent. And thus, we can easily obtain:

$$\mathbb{E}x = \mathbb{E}(Uz + \epsilon) = U\mathbb{E}z + \mathbb{E}\epsilon = 0$$

The joint distribution over (x, z) can be defined as:

$$\begin{aligned} \begin{bmatrix} z \\ x \end{bmatrix} &\sim \mathcal{N}(\mu,\, \Sigma), \text{ where:} \\ \mu &= \begin{bmatrix} \mu_z \\ \mu_x \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \Sigma &= \begin{bmatrix} \Sigma_{zz} & \Sigma_{zx} \\ \Sigma_{xz} & \Sigma_{xx} \end{bmatrix} \end{aligned}$$

Using the fact about the independence between z, ϵ, we have that:

$$\begin{aligned} \Sigma_{zz} &= \mathbb{E}(z - \mu_z)(z - \mu_z)^T \\ &= \mathbb{E}zz^T = \mathrm{Cov}(z) = I \\ \Sigma_{zx} &= \mathbb{E}(z - \mu_z)(x - \mu_x)^T \\ &= \mathbb{E}\left(z(Uz + \epsilon)^T\right) \\ &= \mathbb{E}(zz^T U^T + z\epsilon^T) \\ &= \mathbb{E}(zz^T U^T) + \mathbb{E}z\mathbb{E}\epsilon^T \\ &= U^T \\ \Sigma_{xz} &= \Sigma_{xz}^T = U \\ \Sigma_{xx} &= \mathbb{E}(x - \mu_x)(x - \mu_x)^T \\ &= \mathbb{E}(Uz + \epsilon)(Uz + \epsilon)^T \\ &= \mathbb{E}(Uzz^T U^T + Uz\epsilon^T + \epsilon z^T U^T + \epsilon\epsilon^T) \\ &= U\mathbb{E}(zz^T)U^T + U\mathbb{E}z\mathbb{E}\epsilon^T + \mathbb{E}\epsilon\mathbb{E}z^T U^T + \mathbb{E}(\epsilon\epsilon^T) \\ &= UU^T + \sigma^2 I \end{aligned}$$

Putting everything together, we therefore have that:

$$\begin{bmatrix} z \\ x \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I & U^T \\ U & UU^T + \sigma^2 I \end{bmatrix}\right)$$

And also, $z|x \sim \mathcal{N}(\mu_{z|x}, \Sigma_{z|x})$, where:

$$\begin{aligned}
\mu_{z|x} &= \mu_z + \Sigma_{zx}\Sigma_{xx}^{-1}(x - \mu_x) \\
&= U^T(UU^T + \sigma^2 I)^{-1}x \\
&= \frac{U^T x}{U^T U + \sigma^2} \\
\Sigma_{z|x} &= \Sigma_{zz} - \Sigma_{zx}\Sigma_{xx}^{-1}\Sigma_{xz} \\
&= I - U^T(UU^T + \sigma^2 I)^{-1}U \\
&= 1 - \frac{U^T U}{U^T U + \sigma^2}
\end{aligned}$$

(b) For the E-step, we have that:

$$\begin{aligned}
Q_i(z^{(i)}) &= p(z^{(i)}|x^{(i)}; U) \\
&= \frac{1}{(2\pi)^{k/2}|\Sigma_{z^{(i)}|x^{(i)}}|^{1/2}} \exp\left(-\frac{1}{2}(z^{(i)} - \mu_{z^{(i)}|x^{(i)}})^T \Sigma_{z^{(i)}|x^{(i)}}^{-1}(z^{(i)} - \mu_{z^{(i)}|x^{(i)}})\right)
\end{aligned}$$

where $\mu_{z^{(i)}|x^{(i)}}$ and $\Sigma_{z^{(i)}|x^{(i)}}$ can be computed as above. For the M-step, we need to maximize:

$$\begin{aligned}
&\sum_{i=1}^{m} \int_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; U)}{Q_i(z^{(i)})} dz^{(i)} \\
&= \sum_{i=1}^{m} \mathbb{E}_{z^{(i)} \sim Q_i}\left[\log p(x^{(i)}|z^{(i)}; U) + \log p(z^{(i)}) - \log Q_i(z^{(i)})\right]
\end{aligned}$$

Dropping terms that do not depend on the parameter U and taking the gradient w.r.t U, we have that:

$$\begin{aligned}
&\nabla_U \sum_{i=1}^{m} \mathbb{E}_{z^{(i)} \sim Q_i}\left[\log p(x^{(i)}|z^{(i)}; U)\right] \\
&= \nabla_U \sum_{i=1}^{m} \mathbb{E}\left[\log \frac{1}{(2\pi)^{n/2}\sigma} - \frac{1}{2\sigma^2}(x^{(i)} - Uz^{(i)})^T(x^{(i)} - Uz^{(i)})\right] \\
&= \frac{1}{2\sigma^2}\nabla_U \sum_{i=1}^{m} \mathbb{E}\left({x^{(i)}}^T U z^{(i)} + {z^{(i)}}^T U^T x^{(i)} - {z^{(i)}}^T U^T U z^{(i)}\right) \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{m} \mathbb{E}\left(x^{(i)}{z^{(i)}}^T - Uz^{(i)}{z^{(i)}}^T\right) \\
&= \frac{1}{\sigma^2}\left[\sum_{i=1}^{m} x^{(i)}\mathbb{E}\left[{z^{(i)}}^T\right] - U\sum_{i=1}^{m}\mathbb{E}\left[z^{(i)}{z^{(i)}}^T\right]\right]
\end{aligned}$$

Setting this equal to zero and solving for U, we have that:

$$U = \left(\sum_{i=1}^{m} x^{(i)}\mathbb{E}_{z^{(i)} \sim Q_i}\left[{z^{(i)}}^T\right]\right)\left(\sum_{i=1}^{m}\mathbb{E}_{z^{(i)} \sim Q_i}\left[z^{(i)}{z^{(i)}}^T\right]\right)^{-1} \tag{1}$$

From our definition of Q_i being Gaussian with mean $\mu_{z^{(i)}|x^{(i)}}$ and covariance $\Sigma_{z^{(i)}|x^{(i)}}$, we easily find:

$$\begin{aligned}
\mathbb{E}_{z^{(i)} \sim Q_i}\left[{z^{(i)}}^T\right] &= \mu_{z^{(i)}|x^{(i)}} \\
\mathbb{E}_{z^{(i)} \sim Q_i}\left[z^{(i)}{z^{(i)}}^T\right] &= \mu_{z^{(i)}|x^{(i)}}^T \mu_{z^{(i)}|x^{(i)}} + \Sigma_{z^{(i)}|x^{(i)}}
\end{aligned}$$

Substituting this back into equation (1), we get the M-step update for U:

$$U = \left(\sum_{i=1}^{m} x^{(i)}\mu_{z^{(i)}|x^{(i)}}\right)\left(\sum_{i=1}^{m}\mu_{z^{(i)}|x^{(i)}}\mu_{z^{(i)}|x^{(i)}}^T + \Sigma_{z^{(i)}|x^{(i)}}\right)^{-1}$$

(c) When $\sigma^2 \to 0$, we have that:

$$\mu_{z|x} \to \frac{U^T x}{U^T U}$$
$$\Sigma_{z|x} \to 0$$

From the definition of w that $w_i = \mu_{z^{(i)}|x^{(i)}}$ and results obtained from part (a), we have:

$$w = \frac{XU}{U^T U}$$

From what we have got from part (b), we have that:

$$\begin{aligned} U &= \left(\sum_{i=1}^m x^{(i)} \mu_{z^{(i)}|x^{(i)}}\right)\left(\sum_{i=1}^m \mu_{z^{(i)}|x^{(i)}} \mu_{z^{(i)}|x^{(i)}}^T + \Sigma_{z^{(i)}|x^{(i)}}\right)^{-1} \\ &= \left(\sum_{i=1}^m x^{(i)} w_i\right)\left(\sum_{i=1}^m w_i w_i\right)^{-1} \\ &= \frac{X^T w}{w^T w} \end{aligned}$$

If U remains unchanged after the update, it must satisfy:

$$\begin{aligned} U &= \frac{X^T \frac{XU}{U^T U}}{\frac{U^T X^T}{U^T U} \frac{XU}{U^T U}} \\ &= X^T X U \frac{U^T U}{U^T X^T X U} \\ &= \Sigma U \frac{m^2 U^T U}{U^T X^T X U} \\ &= \Sigma U \frac{1}{\lambda} \end{aligned}$$

3. **PCA and ICA for Natural Images**

4. **Convergence of Policy Iteration**

$$\begin{aligned} V'(s) &= B(V)(s) \\ &= R(s) + \gamma \sum_{s' \in \mathcal{S}} P_{s\pi(s)}(s') V(s') \end{aligned}$$

(a) Since $V_1(s) < V_2(s), \forall s \in \mathcal{S}$, we have:

$$\begin{aligned} B(V_1)(s) &= R(s) + \gamma \sum_{s' \in \mathcal{S}} P_{s\pi(s)}(s') V_1(s') \\ &\leq R(s) + \gamma \sum_{s' \in \mathcal{S}} P_{s\pi(s)}(s') V_2(s') \\ &= B(V_2)(s) \end{aligned}$$

(b) For any V, we have:

$$\begin{aligned} ||B^\pi(V) - V^\pi||_\infty &= \max_{s \in \mathcal{S}} \left| R(s) + \gamma \sum_{s' \in \mathcal{S}} P_{s\pi(s)}(s') V(s') - R(s) - \gamma \sum_{s' \in \mathcal{S}} P_{s\pi(s)}(s') V^\pi(s') \right| \\ &= \max_{s' \in \mathcal{S}} \gamma \left| \sum_{s' \in \mathcal{S}} P_{s\pi(s)}(s')(V(s') - V^\pi(s')) \right| \\ &\leq \max_{s' \in \mathcal{S}} \gamma \left| V(s') - V^\pi(s') \right| \\ &= ||V - V^\pi||_\infty \end{aligned}$$

In the third step, we use the fact that $\sum_{s' \in \mathcal{S}} P_{s\pi(s)} = 1$ and $\sum_i \alpha_i x_i \leq \max_i x_i$ if $\sum_i \alpha_i = 1, \alpha_i \geq 0$

(c) Since $\pi'(s) = \arg\max_{a \in \mathcal{A}} \sum_{s' \in \mathcal{S}} P_{sa}(s')V^{\pi}(s')$, we have:

$$\begin{aligned} V^{\pi}(s) &= R(s) + \gamma \sum_{s' \in \mathcal{S}} P_{s\pi(s)}(s')V^{\pi}(s') \\ &\leq R(s) + \gamma \sum_{s' \in \mathcal{S}} P_{s\pi'(s)}(s')V^{\pi}(s') \\ &= B^{\pi'}(V^{\pi})(s) \end{aligned}$$

By part (a), applying $B^{\pi'}$ on both sides of the inequality, we have:

$$V^{\pi}(s) \leq B^{\pi'}(V^{\pi})(s) \Rightarrow B^{\pi'}(V^{\pi})(s) \leq B^{\pi'}(B^{\pi'}(V^{\pi}))(s)$$

Using the result obtained from part (b), we have:

$$\begin{aligned} V^{\pi}(s) &\leq B^{\pi'}(V^{\pi})(s) \\ &\leq B^{\pi'}(B^{\pi'}(V^{\pi}))(s) \\ &\leq B^{\pi'}(B^{\pi'}(...B^{\pi'}(V^{\pi})...))(s) \\ &= V^{\pi'}(s) \end{aligned}$$

(d) As result obtained from part (c), the policies are monotonically improving. On the other hand, there are a finite number of possible policies. Hence, at some point, the policy's gonna stop updating, or in other words, we will produce a $\pi' = \pi$. If $\pi' = \pi$, we have:

$$V^{\pi} = V^{\pi'} = R(s) + \gamma \max_{a \in \mathcal{A}} \sum_{s' \in \mathcal{S}} P_{sa}(s')V^{\pi}(s')$$

or $V^{\pi} = V^{*}$, and thus, $\pi = \pi^{*}$

5. **Reinforcement Learning: The Mountain Car**