ps3-4-solve

June 3, 2021

1 (a)

$$
\begin{aligned}
\ell_{\text{semi-sup}}(\theta^{(t+1)}) &= \ell_{\text{unsup}}(\theta^{(t+1)}) + \alpha\ell_{\text{sup}}(\theta^{(t+1)}) \\
&= \sum_{i=1}^{m} \log \sum_{z^{(i)}} p(x^{(i)}, z^{(i)}; \theta^{(t+1)}) + \alpha \sum_{i=1}^{\tilde{m}} \log p(\tilde{x}^{(i)}, \tilde{z}^{(i)}; \theta^{(t+1)}) \\
&= \sum_{i=1}^{m} \log \sum_{z^{(i)}} Q_i^{(t)}(z^{(i)}) \frac{p(x^{(i)}, z^{(i)}; \theta^{(t+1)})}{Q_i^{(t)}(z^{(i)})} + \alpha \sum_{i=1}^{\tilde{m}} \log p(\tilde{x}^{(i)}, \tilde{z}^{(i)}; \theta^{(t+1)}) \\
&= \sum_{i=1}^{m} \log \mathbb{E}\big(\frac{p(x^{(i)}, z^{(i)}; \theta^{(t+1)})}{Q_i^{(t)}(z^{(i)})}\big) + \alpha \sum_{i=1}^{\tilde{m}} \log p(\tilde{x}^{(i)}, \tilde{z}^{(i)}; \theta^{(t+1)}) \\
&\geq \sum_{i=1}^{m} \mathbb{E}\big(\log \frac{p(x^{(i)}, z^{(i)}; \theta^{(t+1)})}{Q_i^{(t)}(z^{(i)})}\big) + \alpha \sum_{i=1}^{\tilde{m}} \log p(\tilde{x}^{(i)}, \tilde{z}^{(i)}; \theta^{(t+1)}) \\
&= \sum_{i=1}^{m} \sum_{z^{(i)}} Q_i^{(t)}(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \theta^{(t+1)})}{Q_i^{(t)}(z^{(i)})} + \alpha \sum_{i=1}^{\tilde{m}} \log p(\tilde{x}^{(i)}, \tilde{z}^{(i)}; \theta^{(t+1)}) \\
&\geq \sum_{i=1}^{m} \sum_{z^{(i)}} Q_i^{(t)}(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \theta^{(t)})}{Q_i^{(t)}(z^{(i)})} + \alpha \sum_{i=1}^{\tilde{m}} \log p(\tilde{x}^{(i)}, \tilde{z}^{(i)}; \theta^{(t)}) \\
&= \ell_{\text{unsup}}(\theta^{(t)}) + \alpha\ell_{\text{sup}}(\theta^{(t)}) \\
&= \ell_{\text{semi-sup}}(\theta^{(t)})
\end{aligned}
$$

2 (b)

we have:

$$
p(x^{(i)}|z^{(i)} = j; \mu_j, \Sigma_j) = \big((2\pi)^{\frac{n}{2}} |\Sigma_j|^{\frac{1}{2}}\big)^{-1} \exp\big(-\frac{1}{2}(x^{(i)} - \mu_j)^T \Sigma_j^{-1} (x^{(i)} - \mu_j)\big) p(z^{(i)} = j; \phi) = \phi_j
$$

for $Q_i^{(t)}(z^{(i)} = j)$:

$$
\begin{aligned}
Q_i^{(t)}(z^{(i)}=j) &:= p(z^{(i)}=j|x^{(i)};\phi,\mu_j,\Sigma_j) \\
&= \frac{p(x^{(i)}|z^{(i)}=j;\mu_j,\Sigma_j)p(z^{(i)}=j;\phi)}{\sum_j p(x^{(i)}|z^{(i)}=j;\mu_j,\Sigma_j)p(z^{(i)}=j;\phi)} \\
&= \frac{\big((2\pi)^{\frac{n}{2}}|\Sigma_j|^{\frac{1}{2}}\big)^{-1}\exp\big(-\frac{1}{2}(x^{(i)}-\mu_j)^T\Sigma_j^{-1}(x^{(i)}-\mu_j)\big)\phi_j}{\sum_j\big((2\pi)^{\frac{n}{2}}|\Sigma_j|^{\frac{1}{2}}\big)^{-1}\exp\big(-\frac{1}{2}(x^{(i)}-\mu_j)^T\Sigma_j^{-1}(x^{(i)}-\mu_j)\big)\phi_j} \\
&= \frac{\big(|\Sigma_j|^{\frac{1}{2}}\big)^{-1}\exp\big(-\frac{1}{2}(x^{(i)}-\mu_j)^T\Sigma_j^{-1}(x^{(i)}-\mu_j)\big)\phi_j}{\sum_j\big(|\Sigma_j|^{\frac{1}{2}}\big)^{-1}\exp\big(-\frac{1}{2}(x^{(i)}-\mu_j)^T\Sigma_j^{-1}(x^{(i)}-\mu_j)\big)\phi_j}
\end{aligned}
$$

3 (c)

let

$$
w_j^{(i)} = Q_i^{(t)}(z^{(i)}=j)
$$

for ℓ_{unsup}:

$$
\begin{aligned}
\ell_{\text{unsup}} = & \sum_{i=1}^{m}\sum_j w_j^{(i)}\log\frac{p(x^{(i)},z^{(i)}=j;\mu_j^{(t)},\Sigma_j^{(t)},\phi^{(t)})}{w_j^{(i)}} \\
= & \sum_{i=1}^{m}\sum_j w_j^{(i)}\log\frac{p(x^{(i)}|z^{(i)}=j;\mu_j^{(t)},\Sigma_j^{(t)})p(z^{(i)}=j;\phi^{(t)})}{w_j^{(i)}} \\
= & \sum_{i=1}^{m}\sum_j w_j^{(i)}\log\frac{\big((2\pi)^{\frac{n}{2}}|\Sigma_j^{(t)}|^{\frac{1}{2}}\big)^{-1}\exp\big(-\frac{1}{2}(x^{(i)}-\mu_j^{(t)})^T(\Sigma_j^{(t)})^{-1}(x^{(i)}-\mu_j^{(t)})\big)\phi_j^{(t)}}{w_j^{(i)}}
\end{aligned}
$$

for ℓ_{sup}:

$$
\begin{aligned}
\ell_{\text{sup}} = &\sum_{i=1}^{\tilde{m}} \log p(\tilde{x}^{(i)}, \tilde{z}^{(i)}; \mu_j^{(t)}, \Sigma_j^{(t)}, \phi^{(t)}) \\
= &\sum_{i=1}^{\tilde{m}} \log \big(\sum_j p(\tilde{x}^{(i)} | \tilde{z}^{(i)} = j; \mu_j^{(t)}, \Sigma_j^{(t)}) p(\tilde{z}^{(i)} = j; \phi^{(t)}) \big) \\
= &\sum_{i=1}^{\tilde{m}} \log \big(\sum_j 1\{\tilde{z}^{(i)} = j\} p(\tilde{x}^{(i)} | \tilde{z}^{(i)} = j; \mu_j^{(t)}, \Sigma_j^{(t)}) p(\tilde{z}^{(i)} = j; \phi^{(t)}) \big) \\
= &\sum_{i=1}^{\tilde{m}} \sum_j 1\{\tilde{z}^{(i)} = j\} \log \big(p(\tilde{x}^{(i)} | \tilde{z}^{(i)} = j; \mu_j^{(t)}, \Sigma_j^{(t)}) p(\tilde{z}^{(i)} = j; \phi^{(t)}) \big) \\
= &\sum_{i=1}^{\tilde{m}} \sum_j 1\{\tilde{z}^{(i)} = j\} \log \Big(\big((2\pi)^{\frac{n}{2}} |\Sigma_j^{(t)}|^{\frac{1}{2}} \big)^{-1} \exp \big(-\frac{1}{2} (\tilde{x}^{(i)} - \mu_j^{(t)})^T (\Sigma_j^{(t)})^{-1} (\tilde{x}^{(i)} - \mu_j^{(t)}) \big) \phi_j^{(t)} \Big)
\end{aligned}
$$

for $\bigtriangledown_{\mu_l} \ell_{\text{unsup}}$:

$$
\begin{aligned}
\bigtriangledown_{\mu_l} \ell_{\text{unsup}} = &\bigtriangledown_{\mu_l} \sum_{i=1}^{m} \sum_j w_j^{(i)} (-\frac{1}{2}) (x^{(i)} - \mu_j^{(t)})^T (\Sigma_j^{(t)})^{-1} (x^{(i)} - \mu_j^{(t)}) \\
= &\sum_{i=1}^{m} w_l^{(i)} (\Sigma_l^{(t)})^{-1} (x^{(i)} - \mu_l^{(t)})
\end{aligned}
$$

for $\bigtriangledown_{\mu_l} \ell_{\text{sup}}$:

$$
\begin{aligned}
\bigtriangledown_{\mu_l} \ell_{\text{sup}} = &\bigtriangledown_{\mu_l} \sum_{i=1}^{\tilde{m}} \sum_j 1\{\tilde{z}^{(i)} = j\} (-\frac{1}{2}) (\tilde{x}^{(i)} - \mu_j^{(t)})^T (\Sigma_j^{(t)})^{-1} (\tilde{x}^{(i)} - \mu_j^{(t)}) \\
= &\sum_{i=1}^{\tilde{m}} 1\{\tilde{z}^{(i)} = l\} (\Sigma_l^{(t)})^{-1} (\tilde{x}^{(i)} - \mu_l^{(t)})
\end{aligned}
$$

let $\bigtriangledown_{\mu_l} \ell_{\text{semi-sup}} = 0$:

$$
\begin{aligned}
\nabla_{\mu_l}\ell_{\text{semi-sup}} &= 0\\
0 &= \nabla_{\mu_l}\ell_{\text{unsup}} + \alpha\,\nabla_{\mu_l}\ell_{\text{sup}}\\
0 &= \sum_{i=1}^{m} w_l^{(i)}(\Sigma_l^{(t)})^{-1}(x^{(i)}-\mu_l^{(t)}) + \alpha\sum_{i=1}^{\tilde{m}} 1\{\tilde{z}^{(i)}=l\}(\Sigma_l^{(t)})^{-1}(\tilde{x}^{(i)}-\mu_l^{(t)})\\
\mu_l^{(t+1)} &= \frac{\sum_{i=1}^{m} w_l^{(i)}x^{(i)} + \alpha\sum_{i=1}^{\tilde{m}} 1\{\tilde{z}^{(i)}=l\}\tilde{x}^{(i)}}{\sum_{i=1}^{m} w_l^{(i)} + \alpha\sum_{i=1}^{\tilde{m}} 1\{\tilde{z}^{(i)}=l\}}
\end{aligned}
$$

for $\nabla_{\Sigma_l}\ell_{\text{unsup}}$:

$$
\begin{aligned}
\nabla_{\Sigma_l}\ell_{\text{unsup}} &= \nabla_{\Sigma_l}\sum_{i=1}^{m}\sum_{j} w_j^{(i)}\big(-\frac{1}{2}\log|\Sigma_j^{(t)}| - \frac{1}{2}(x^{(i)}-\mu_j^{(t)})^T(\Sigma_j^{(t)})^{-1}(x^{(i)}-\mu_j^{(t)})\big)\\
&= \sum_{i=1}^{m} w_l^{(i)}\big(-\frac{1}{2}(\Sigma_l^{(t)})^{-1} - \frac{1}{2}\nabla_{\Sigma_l}\operatorname{tr}(x^{(i)}-\mu_l^{(t)})^T(\Sigma_l^{(t)})^{-1}(x^{(i)}-\mu_l^{(t)})\big)\\
&= \sum_{i=1}^{m} w_l^{(i)}\big(-\frac{1}{2}(\Sigma_l^{(t)})^{-1} - \frac{1}{2}\nabla_{\Sigma_l}\operatorname{tr}(x^{(i)}-\mu_l^{(t)})(x^{(i)}-\mu_l^{(t)})^T(\Sigma_l^{(t)})^{-1}\big)\\
&= \sum_{i=1}^{m} w_l^{(i)}\big(-\frac{1}{2}(\Sigma_l^{(t)})^{-1} + \frac{1}{2}(x^{(i)}-\mu_l^{(t)})(x^{(i)}-\mu_l^{(t)})^T(\Sigma_l^{(t)})^{-2}\big)
\end{aligned}
$$

for $\nabla_{\Sigma_l}\ell_{\text{sup}}$:

$$
\begin{aligned}
\nabla_{\Sigma_l}\ell_{\text{sup}} &= \nabla_{\Sigma_l}\sum_{i=1}^{\tilde{m}}\sum_{j} 1\{\tilde{z}^{(i)}=j\}\big(-\frac{1}{2}\log|\Sigma_j^{(t)}| - \frac{1}{2}(\tilde{x}^{(i)}-\mu_j^{(t)})^T(\Sigma_j^{(t)})^{-1}(\tilde{x}^{(i)}-\mu_j^{(t)})\big)\\
&= \sum_{i=1}^{\tilde{m}} 1\{\tilde{z}^{(i)}=l\}\big(-\frac{1}{2}(\Sigma_l^{(t)})^{-1} + \frac{1}{2}(\tilde{x}^{(i)}-\mu_l^{(t)})(\tilde{x}^{(i)}-\mu_l^{(t)})^T(\Sigma_l^{(t)})^{-2}\big)
\end{aligned}
$$

let $\nabla_{\Sigma_l}\ell_{\text{semi-sup}} = 0$:

$$
\begin{aligned}
\triangledown_{\Sigma_l}\ell_{\text{semi-sup}} &= 0 \\
0 &= \triangledown_{\Sigma_l}\ell_{\text{unsup}} + \alpha \triangledown_{\Sigma_l}\ell_{\text{sup}} \\
0 &= \sum_{i=1}^{m} w_l^{(i)}\big(-\frac{1}{2}(\Sigma_l^{(t)})^{-1} + \frac{1}{2}(x^{(i)} - \mu_l^{(t)})(x^{(i)} - \mu_l^{(t)})^T(\Sigma_l^{(t)})^{-2}\big) \\
&\quad + \alpha\sum_{i=1}^{\tilde{m}} 1\{\tilde{z}^{(i)} = l\}\big(-\frac{1}{2}(\Sigma_l^{(t)})^{-1} + \frac{1}{2}(\tilde{x}^{(i)} - \mu_l^{(t)})(\tilde{x}^{(i)} - \mu_l^{(t)})^T(\Sigma_l^{(t)})^{-2}\big) \\
\Sigma_l^{(t+1)} &= \frac{\sum_{i=1}^{m} w_l^{(i)}(x^{(i)} - \mu_l^{(t)})(x^{(i)} - \mu_l^{(t)})^T + \alpha\sum_{i=1}^{\tilde{m}} 1\{\tilde{z}^{(i)} = l\}(\tilde{x}^{(i)} - \mu_l^{(t)})(\tilde{x}^{(i)} - \mu_l^{(t)})^T}{\sum_{i=1}^{m} w_l^{(i)} + \alpha\sum_{i=1}^{\tilde{m}} 1\{\tilde{z}^{(i)} = l\}}
\end{aligned}
$$

for $\triangledown_{\phi_l}\ell_{\text{unsup}}$:

$$
\begin{aligned}
\triangledown_{\phi_l}\ell_{\text{unsup}} &= \triangledown_{\phi_l}\sum_{i=1}^{m}\sum_{j} w_j^{(i)}\log\phi_j \\
&= \sum_{i=1}^{m} w_l^{(i)}\phi_l^{-1}
\end{aligned}
$$

for $\triangledown_{\phi_l}\ell_{\text{sup}}$:

$$
\begin{aligned}
\triangledown_{\phi_l}\ell_{\text{sup}} &= \triangledown_{\phi_l}\sum_{i=1}^{\tilde{m}}\sum_{j} 1\{\tilde{z}^{(i)} = j\}\log\phi_j \\
&= \sum_{i=1}^{m} 1\{\tilde{z}^{(i)} = l\}\phi_l^{-1}
\end{aligned}
$$

create $\mathcal{L}(\phi)$:

$$
\mathcal{L}(\phi) = \ell_{\text{unsup}} + \alpha\ell_{\text{sup}} - \beta(\sum_{j}\phi_j - 1)\triangledown_{\phi_l}\mathcal{L}(\phi) = \sum_{i=1}^{m} w_l^{(i)}\phi_l^{-1} + \alpha\sum_{i=1}^{m} 1\{\tilde{z}^{(i)} = l\}\phi_l^{-1} - \beta\triangledown_{\beta}\mathcal{L}(\phi) = 1 - \sum_{j}\phi_j
$$

let $\triangledown_{\phi_l}\mathcal{L}(\phi) = 0$:

$$
\begin{aligned}
\bigtriangledown_{\phi_l}\mathcal{L}(\phi) &= 0 \\
\beta\phi_l &= \sum_{i=1}^{m} w_l^{(i)} + \alpha\sum_{i=1}^{m} 1\{\tilde{z}^{(i)} = l\} \\
\phi_l &= \big(\sum_{i=1}^{m} w_l^{(i)} + \alpha\sum_{i=1}^{m} 1\{\tilde{z}^{(i)} = l\}\big)\beta^{-1}
\end{aligned}
$$

let $\bigtriangledown_{\beta}\mathcal{L}(\phi) = 0$:

$$
\begin{aligned}
\bigtriangledown_{\beta}\mathcal{L}(\phi) &= 0 \\
1 &= \sum_j \phi_j \\
1 &= \sum_j \big(\sum_{i=1}^{m} w_j^{(i)} + \alpha\sum_{i=1}^{m} 1\{\tilde{z}^{(i)} = j\}\big)\beta^{-1} \\
\beta &= \sum_j \big(\sum_{i=1}^{m} w_j^{(i)} + \alpha\sum_{i=1}^{m} 1\{\tilde{z}^{(i)} = j\}\big)
\end{aligned}
$$

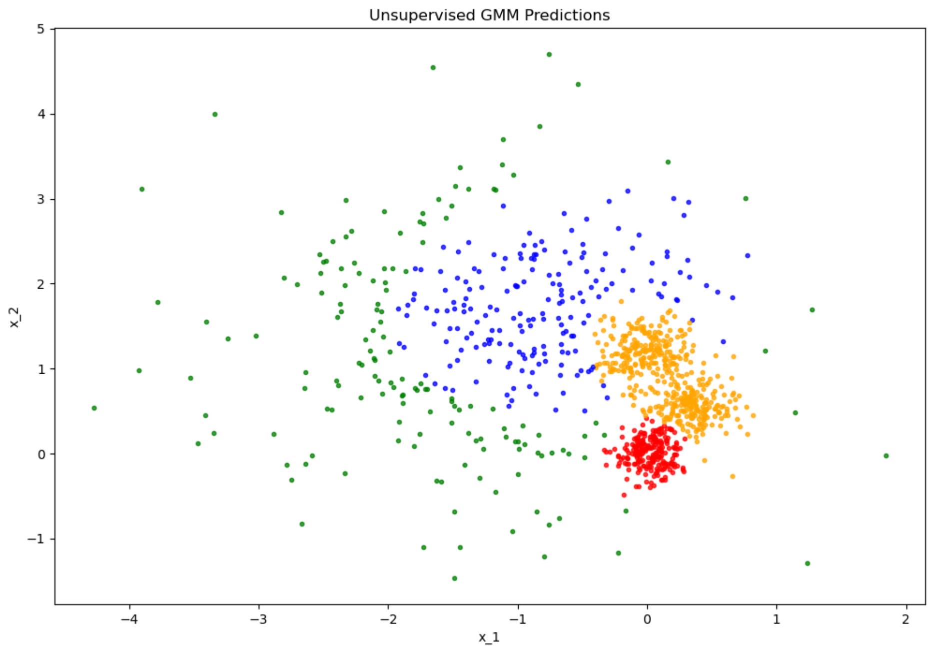
then,

$$
\phi_l^{(t+1)} = \frac{\sum_{i=1}^{m} w_l^{(i)} + \alpha\sum_{i=1}^{m} 1\{\tilde{z}^{(i)} = l\}}{\sum_j \big(\sum_{i=1}^{m} w_j^{(i)} + \alpha\sum_{i=1}^{m} 1\{\tilde{z}^{(i)} = j\}\big)}
$$

4 (d)

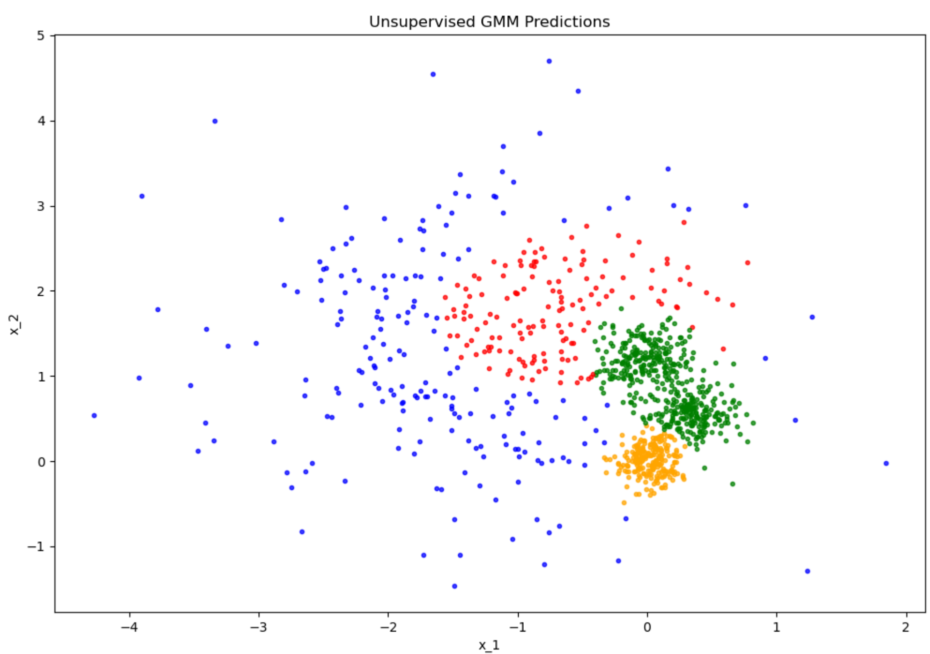
p04_pred_0.png

iteration number: 118



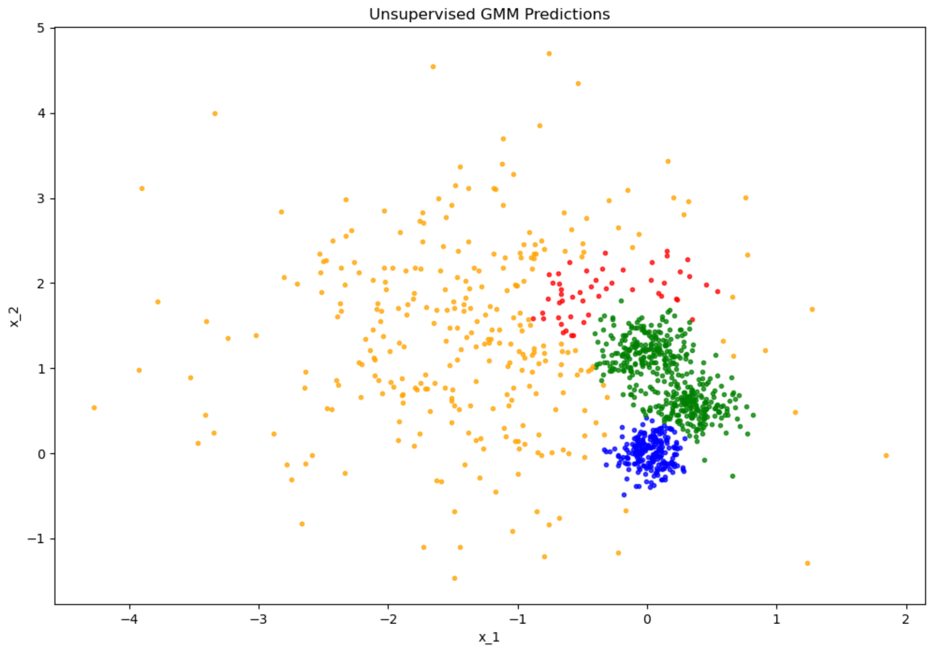
p04_pred_1.png

iteration number: 119



p04_pred_2.png

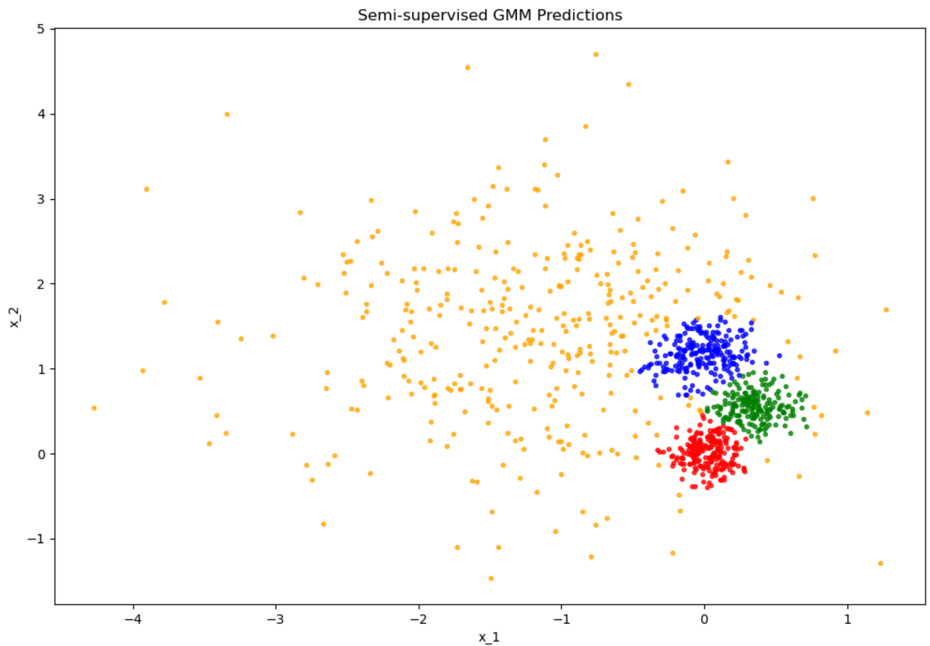
iteration number: 105



5 (e)

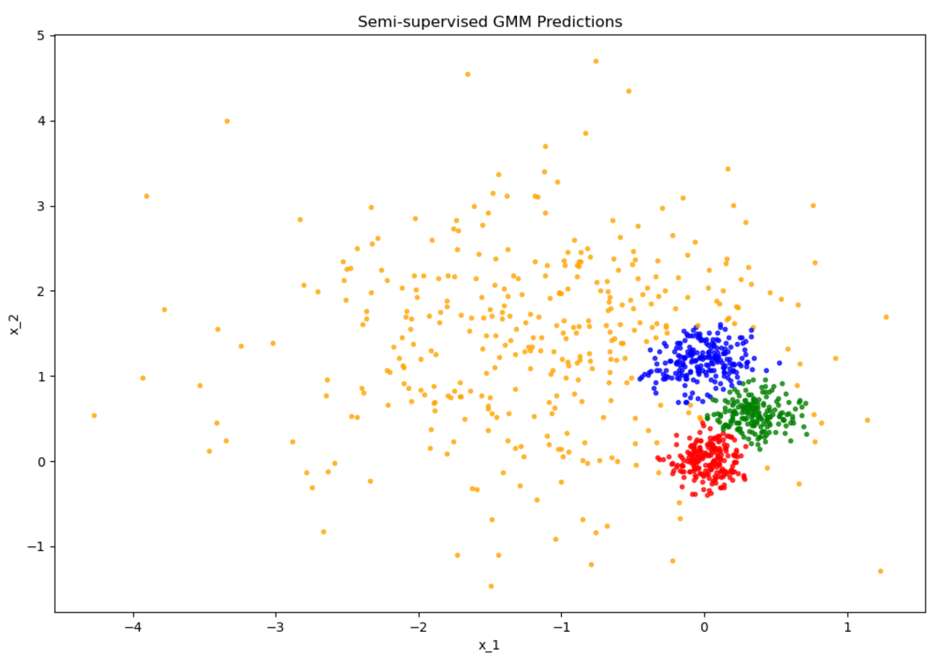
p04_pred_ss_0.png

iteration number: 22



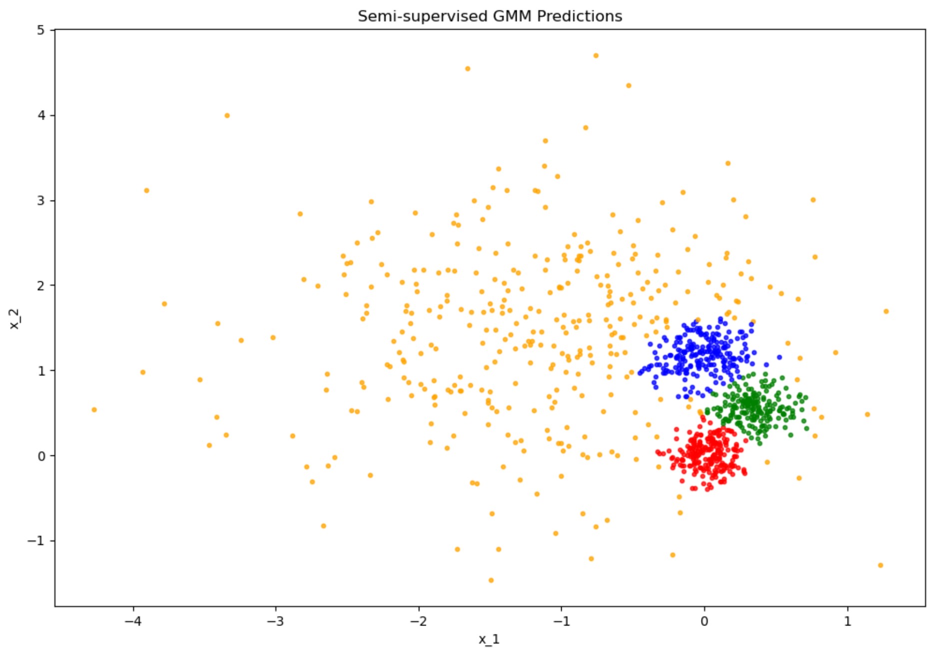
p04_pred_ss_1.png

iteration number: 36



p04_pred_ss_2.png

iteration number: 33



6 (f)

6.1 i

Unsupervised EM takes more iteration number to converge.

6.2 ii

Semi-supervised EM are more stable than unsupervised EM.

When started with different random initializations, unsupervised EM will change its assignments. But semi-supervised EM will keep its assignments same.

6.3 iii

The assignments quality of semi-supervised EM is better than unsupervised EM.

Semi-supervised EM can find three low-variance Gaussian distributions with one high-variance Gaussian distribution. Unsupervised EM can only find two low-variance Gaussian distributions with two high-variance Gaussian distributions.

```
[ ]:
```