ENGG 5202 HOMEWORK #2 SOLUTION

March 6, 2014

Problem 1

1.

$$[\widetilde{\boldsymbol{S}}_W]_{i,j} = \boldsymbol{e_i}^t \boldsymbol{S}_W \boldsymbol{e_j} = \delta_{i,j}, \qquad [\widetilde{\boldsymbol{S}}_B]_{i,j} = \boldsymbol{e_i}^t \boldsymbol{S}_B \boldsymbol{e_j} = \lambda_i \delta_{i,j}$$
$$\widetilde{\boldsymbol{S}}_W = \boldsymbol{I}_{n\times n}, \qquad \widetilde{\boldsymbol{S}}_B = diag\{\lambda_1, ..., \lambda_n\}$$

2.

$$\boldsymbol{J} = \frac{|\widetilde{\boldsymbol{S}}_B|}{|\widetilde{\boldsymbol{S}}_W|} = \frac{|diag\{\lambda_1, ..., \lambda_n\}|}{|\boldsymbol{I}_{n\times n}|} = \prod_{i=1}^{n} \lambda_i$$

3. As $\boldsymbol{y'} = \boldsymbol{QDy} = \boldsymbol{QDW}^t\boldsymbol{x} = (\boldsymbol{WD}^t\boldsymbol{Q}^t)^t x$. After transformation, $\boldsymbol{J}$ changed to:

$$\begin{aligned}
\boldsymbol{J'} &= \frac{|\widetilde{\boldsymbol{S'}}_B|}{|\widetilde{\boldsymbol{S'}}_W|} = \frac{|(\boldsymbol{WD}^t\boldsymbol{Q}^t)^t\boldsymbol{S}_B(\boldsymbol{WD}^t\boldsymbol{Q})|}{|(\boldsymbol{WD}^t\boldsymbol{Q}^t)^t\boldsymbol{S}_W(\boldsymbol{WD}^t\boldsymbol{Q})|} = \frac{|\boldsymbol{QDW}^t\boldsymbol{S}_B\boldsymbol{WD}^t\boldsymbol{Q}^t|}{|\boldsymbol{QDW}^t\boldsymbol{S}_W\boldsymbol{WD}^t\boldsymbol{Q}^t|} \\
&= \frac{|\boldsymbol{Q}||\boldsymbol{D}||\boldsymbol{W}^t\boldsymbol{S}_B\boldsymbol{W}||\boldsymbol{D}^t||\boldsymbol{Q}^t|}{|\boldsymbol{Q}||\boldsymbol{D}||\boldsymbol{W}^t\boldsymbol{S}_W\boldsymbol{W}||\boldsymbol{D}^t||\boldsymbol{Q}^t|} = \frac{|\boldsymbol{W}^t\boldsymbol{S}_B\boldsymbol{W}|}{|\boldsymbol{W}^t\boldsymbol{S}_W\boldsymbol{W}|} = \prod_{i=1}^{n} \lambda_i = \boldsymbol{J}
\end{aligned}$$

Problem 2

1. From Page 29-30 of Lecture Note 3, we have:

$$\begin{aligned}
\mu_1 &= \mathbf{w}^T\boldsymbol{\mu}_1 \\
\mu_2 &= \mathbf{w}^T\boldsymbol{\mu}_2 \\
\sigma_1^2 &= \mathbf{w}^T\boldsymbol{\Sigma}_1\mathbf{w} \\
\sigma_2^2 &= \mathbf{w}^T\boldsymbol{\Sigma}_2\mathbf{w}
\end{aligned}$$

Then,

$$\boldsymbol{J}_1(\mathrm{w}) = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} = \frac{(\mathbf{w}^T\boldsymbol{\mu}_1 - \mathbf{w}^T\boldsymbol{\mu}_2)^2}{\mathbf{w}^T\boldsymbol{\Sigma}_1\mathbf{w} + \mathbf{w}^T\boldsymbol{\Sigma}_2\mathbf{w}} = \frac{\mathbf{w}^T(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\mathbf{w}}{\mathbf{w}^T(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\mathbf{w}}$$

This is generalized Rayleigh quotient. A vector $\mathbf{w}$ that maximizes $\boldsymbol{J}_1(\mathbf{w})$ must satisfy

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\mathbf{w} = \lambda(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\mathbf{w} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)a$$

It is always the direction $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. Therefore the $\mathbf{w}$ maximizing $\boldsymbol{J}_1(\mathbf{w})$ is

$$\mathbf{w} = (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

2.

$$\begin{aligned} \boldsymbol{J}_2(\mathbf{w}) &= \frac{(\mu_1 - \mu_2)^2}{P(\omega_1)\sigma_1^2 + P(\omega_2)\sigma_2^2} = \frac{(\mathbf{w}^T\boldsymbol{\mu}_1 - \mathbf{w}^T\boldsymbol{\mu}_2)^2}{\mathbf{w}^T P(\omega_1)\boldsymbol{\Sigma}_1\mathbf{w} + \mathbf{w}^T P(\omega_2)\boldsymbol{\Sigma}_2\mathbf{w}} \\ &= \frac{\mathbf{w}^T(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\mathbf{w}}{\mathbf{w}^T(P(\omega_1)\boldsymbol{\Sigma}_1 + P(\omega_2)\boldsymbol{\Sigma}_2)\mathbf{w}} \end{aligned}$$

This is generalized Rayleigh quotient. A vector $\mathbf{w}$ that maximizes $\boldsymbol{J}_2(\mathbf{w})$ must satisfy

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\mathbf{w} = \lambda[P(\omega_1)\boldsymbol{\Sigma}_1 + P(\omega_2)\boldsymbol{\Sigma}_2]\mathbf{w} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)a$$

It is always the direction $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. Therefore the $\mathbf{w}$ maximizing $\boldsymbol{J}_1(\mathbf{w})$ is

$$\mathbf{w} = [P(\omega_1)\boldsymbol{\Sigma}_1 + P(\omega_2)\boldsymbol{\Sigma}_2]^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

3. $\boldsymbol{J}_2(\mathbf{w})$ is more closely related to the Fisher criterion. S_i in Fisher criterion will consider sample number of each class which can be represented by $P(\omega_i)$.

Problem 3

1.

$$p\left(z_k, y_k | x_k, \boldsymbol{\theta}^{(t)}\right) = \frac{p\left(z_k, y_k, x_k | \boldsymbol{\theta}^{(t)}\right)}{p\left(x_k | \boldsymbol{\theta}^{(t)}\right)} = \frac{p_{z_k}^{(t)} q_{y_k}^{(t)} N(x_k; \mu_{z_k}^{(t)}, \sigma_{y_k}^{(t)})}{\sum_{j=1}^m \sum_{i=1}^l p_j^{(t)} q_i^{(t)} N(x_k; \mu_j^{(t)}, \sigma_i^{(t)})}$$

2.

$$\begin{aligned} & \mathcal{E}\left\{ \; l_c(x_1, ...x_n, z_1, ...z_n, y_1, ..., y_n; \boldsymbol{\theta}) | x_1, ...x_n, \boldsymbol{\theta}^{(t)} \right\} \\ =& \sum_{k=1}^n \mathcal{E}\left\{ \; \log\left(p_{z_k} q_{y_k} N(x_k; \mu_{z_k}, \sigma_{y_k}) | x_k, \boldsymbol{\theta}^{(t)} \right) \right\} \\ =& \sum_{k=1}^n \sum_{z_k=1}^m \sum_{y_k=1}^l p(z_k, y_k | x_k, \boldsymbol{\theta}^{(t)}) \log p(z_k, y_k, x_k | \boldsymbol{\theta}) \\ =& \sum_{k=1}^n \sum_{z_k=1}^m \sum_{y_k=1}^l p(z_k, y_k | x_k, \boldsymbol{\theta}^{(t)}) \log\left(p_{z_k} q_{y_k} N(x_k; \mu_{z_k}, \sigma_{y_k}) \right) \\ =& \sum_{k=1}^n \sum_{z_k=1}^m \sum_{y_k=1}^l p(z_k, y_k | x_k, \boldsymbol{\theta}^{(t)}) \left[\log \frac{p_{z_k} q_{y_k}}{\sqrt{2\pi}\sigma_{y_k}} - \frac{(x_k - \mu_{z_k})^2}{2\sigma_{y_k}^2} \right] \end{aligned}$$

3. When fixing σ_{yk}, the portion containing new mean parameters:

$$A = \sum_{k=1}^{n} \sum_{z_k=1}^{m} \sum_{y_k=1}^{l} p(z_k, y_k | x_k, \boldsymbol{\theta}^{(t)}) \left[-\frac{(x_k - \mu_{z_k})^2}{2\sigma_{y_k}^2} \right]$$

$$\begin{aligned} \frac{\partial A}{\partial \mu_j} &= \sum_{k=1}^{n} \sum_{y_k=1}^{l} p(z_k = j, y_k | x_k, \boldsymbol{\theta}^{(t)}) \left[\frac{(x_k - \mu_j)}{\sigma_{y_k}^2} \right] = 0 \\ \mu_j &= \frac{\sum_{k=1}^{n} \sum_{y_k=1}^{l} p(z_k = j, y_k | x_k, \boldsymbol{\theta}^{(t)}) \sigma_{y_k}^{-2} x_k}{\sum_{k=1}^{n} \sum_{y_k=1}^{l} p(z_k = j, y_k | x_k, \boldsymbol{\theta}^{(t)}) \sigma_{y_k}^{-2}} \end{aligned}$$

4. Figure 1 shows samples, true densities and estimated densities.

A model where components can have two different variance parameters cannot be sufficiently well approximated by a model consisting of a small number of components where all components have a single variance parameter. In other words, the number of components with a single common variance required to approximate the underlying two-variance distribution accurately is much larger than 4.

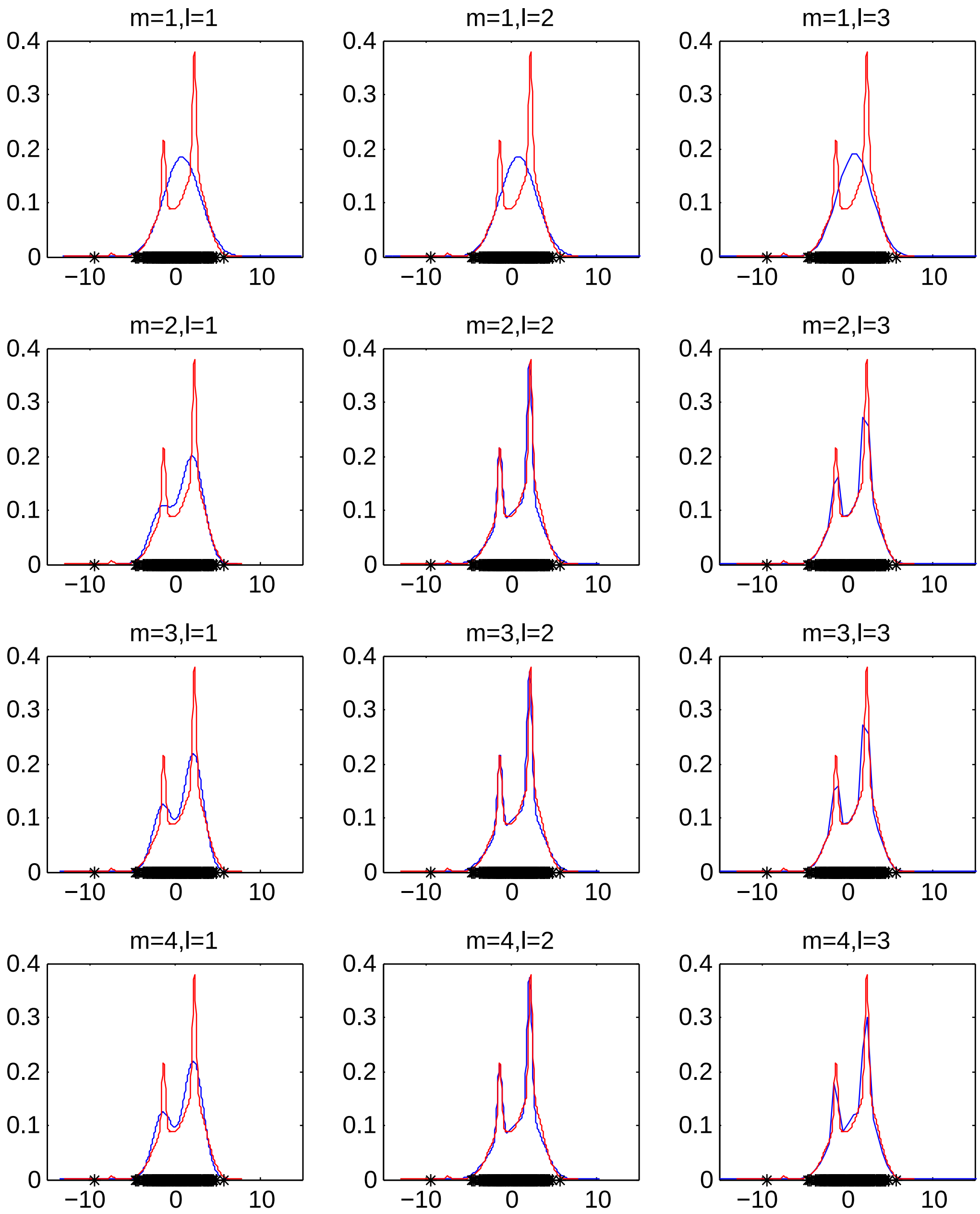

Figure 1: Samples, true densities (Red Curves) and estimated densities (Blue curves) for different m and l.

Contents

Setup parameters

```
n = 2;  % number of latent states
m = 3;  % number of obeserved states
T = 10;  % time in total

p = [0.6 0.4];
A = [0.7 0.3; 0.4 0.6];
B = [0.1 0.4 0.5; 0.6 0.3 0.1];
```

Sampling

```
randfrom = @(d) find(rand(1) <= cumsum(d(:)), 1);
x = zeros(T, 1);
z = zeros(T, 1);

for i = 1:T
    if i == 1
        z(i) = randfrom(p);
    else
        z(i) = randfrom(A(z(i-1), :));
    end
    x(i) = randfrom(B(z(i), :));
end

fprintf( '[Sampling] The latent sequence is '   );
for i = 1:T
    fprintf( '%d' , z(i));
end
fprintf( '\n[Sampling] The observed sequence is '   );
for i = 1:T
    fprintf( '%d' , x(i));
end
fprintf( '\n' );
```

```
[Sampling] The latent sequence is 2211222222
[Sampling] The observed sequence is 2123111221
```

Decoding

```
f = zeros(T, n);
g = zeros(T, n);
```

```
f(1, :) = p .* B(:, x(1))';
g(1, :) = (1:n);

for i = 2:T
    for j = 1:n
        [maxf, maxk] = max(f(i-1, :) .* A(:, j)');
        f(i, j) = maxf * B(j, x(i));
        g(i, j) = maxk;
    end
end

y = zeros(T, 1);
[~, j] = max(f(T, :));
y(T) = j;
for i = T-1:-1:1
    j = g(i+1, j);
    y(i) = j;
end

fprintf( '[Viterbi] The probability is %e\n'   , f(T, y(T)));
fprintf( '[Viterbi] The latent sequence is '   );
for i = 1:T
    fprintf( '%d' , y(i));
end
fprintf( '\n' );
```

```
[Viterbi] The probability is 6.582590e-007
[Viterbi] The latent sequence is 1211222222
```