Zhiqian Zhou
Hw4 Question 1

Hw4. Zhiqian Zhou.

3.19. for MAP, we maximize $l(\theta)p(\theta)$.

(1) $l(\mu) = \sum_{k=1}^{n} \ln p(x_k|\mu)$.

$$= -\frac{n}{2}\ln[(2\pi)^d|\Sigma|] - \sum^{n}(x_k-\mu)^t\Sigma^{-1}(x_k-\mu).$$

$$p(\mu) = \frac{1}{(2\pi)^{d/2}|\Sigma_0|^{\frac{1}{2}}} e^{[-\frac{1}{2}(\mu-m_0)^t\Sigma_0^{-1}(\mu-m_0)]}$$

$$\hat{\mu} = \arg\max_{\mu} l(\mu)p(\mu).$$

(2) $\mu' = \mathcal{E}[x'] = \mathcal{E}[Ax] = A\,\mathcal{E}[x] = A\mu$.

$$\Sigma' = \mathcal{E}[(x'-\mu')(x'-\mu')^t]$$
$$= \mathcal{E}[(Ax-A\mu)(Ax-A\mu)^t]$$
$$= A\,\mathcal{E}[(x-\mu)(x-\mu)^t]A^t$$
$$= A\Sigma A^t$$

Therefore, for $l(\mu')p(\mu')$ we have.

$$l(\mu') = \sum_{k=1}^{n} \ln p(x_k'|\mu')$$

$$= \sum_{k=1}^{n} \ln p(Ax_k|A\mu)$$

$$= -\frac{n}{2}\ln[(2\pi)^d|A\Sigma A^t|] - \sum^{n}(Ax_k-A\mu)^t(A\Sigma A^t)^{-1}(Ax_k-A\mu)$$

$$= -\frac{n}{2}\ln[(2\pi)^d|A\Sigma A^t|] - \sum^{n}(\cancel{A}x_k-\mu)^t\Sigma^{-1}(x_k-\mu).$$

$$p(\mu') = \frac{1}{(2\pi)^{d/2}|A\Sigma_0A^t|^{\frac{1}{2}}} e^{[-\frac{1}{2}(A\mu-Am_0)^t(A\Sigma_0A^t)^{-1}(A\mu-Am_0)]}$$

$$= \frac{1}{(2\pi)^{d/2}|A\Sigma_0A^t|^{\frac{1}{2}}} e^{[-\frac{1}{2}(\mu-m_0)^t\Sigma_0^{-1}(\mu-m_0)]}$$

$$\hat{\mu}' = \arg\max_{\mu'} l(\mu')p(\mu').$$

3.30. $O(g(x)) = f(x)$ where $0 \le f(x) \le c\,g(x)$ for all $x \ge x_0$

(1) Yes. $a^{n+1} = a \cdot a^n = O(a^n)$

$C = a, \; g(x) = a^n$

(2) No. $g(x) = a^n$ $h(x) = a^{(b-1)n}$ which is not a constant. Thus, $f(x) > c \cdot g(x)$.

(3) Yes $g(x) = a^n$ $c = a^b$ $a^n \cdot a^b = O(a^n)$

(4) $f(n) = O(f(n))$ let $c = 1$

$f(n) = c f(n)$ for all x

Therefore $0 \le f(n) \le c f(n)$, $f(n) = O(f(n))$.

Question 2

a) Maximum-likelihood values mu and sigma

```
% a) maximum-likelihood
function [mu,sigma] = ML_estimate(x)
    len = length(x);
    mu = sum(x) / len;
    sigma = sum((x - mu).^2) / len;
end
```

Three features xi of category w1

```
% a) three features xi of category w1
[mu1,sigma1] = ML_estimate(pxw1(:,1));
[mu2,sigma2] = ML_estimate(pxw1(:,2));
[mu3,sigma3] = ML_estimate(pxw1(:,3));
disp('a) estimation for mean and variance for w1')
disp(['mean for x1 = ', num2str(mu1)]);
disp(['variance for x1 = ',num2str(sigma1)]);
disp(['mean for x2 = ', num2str(mu2)]);
disp(['variance for x2 = ',num2str(sigma2)]);
disp(['mean for x3 = ', num2str(mu3)]);
disp(['variance for x3 = ',num2str(sigma3)]);
```

Result:

```
a) estimation for mean and variance for w1
mean for x1 = -0.0709
variance for x1 = 0.90618
mean for x2 = -0.6047
variance for x2 = 4.2007
mean for x3 = -0.911
variance for x3 = 4.5419
```

b/c) Multi-dimension estimation

```
% b/c) multi-dimensional Gaussian
function [mu,sigma] = ML_estimate2(x)
    len = size(x,1);
    mu = sum(x) / len;
    tmp = x - repmat(mu,len,1);
    sigma = (tmp'*tmp) / len;
end
```

b) Two-dimension

```
% b) two-dimension
[mu1,sigma1] = ML_estimate2(pxw1(:,1:2));
[mu2,sigma2] = ML_estimate2(pxw1(:,2:3));
[mu3,sigma3] = ML_estimate2(pxw1(:,[1,3]));
disp('b) 2-dimension estimation for mean and variance for w1')
disp('mean for x1,x2 = '); disp(mu1);
disp('variance for x1,x2 = '); disp(sigma1);
disp('mean for x2,x3 = '); disp(mu2);
disp('variance for x2,x3 = '); disp(sigma2);
disp('mean for x1,x3 = '); disp(mu3);
disp('variance for x1,x3 = '); disp(sigma3);
```

Result:

```
b) 2-dimension estimation for mean and variance for w1
mean for x1,x2 =
   -0.0709   -0.6047

variance for x1,x2 =
    0.9062    0.5678
    0.5678    4.2007

mean for x2,x3 =
   -0.6047   -0.9110

variance for x2,x3 =
    4.2007    0.7337
    0.7337    4.5419

mean for x1,x3 =
   -0.0709   -0.9110

variance for x1,x3 =
    0.9062    0.3941
    0.3941    4.5419
```

c) Three-dimension

```
% c) three-dimension
[mu,sigma] = ML_estimate2(pxw1);
disp('c) 3-dimension estimation for mean and variance for w1')
disp('mean for x1,x2,x3 of w1 = '); disp(mu);
disp('variance for x1,x2,x3 of w1 = '); disp(sigma);
```

Result:

```
c) 3-dimension estimation for mean and variance for w1
mean for x1,x2,x3 of w1 =
   -0.0709   -0.6047   -0.9110

variance for x1,x2,x3 of w1 =
    0.9062    0.5678    0.3941
    0.5678    4.2007    0.7337
    0.3941    0.7337    4.5419
```

d) Diagonal of the covariance

```
% d) diagonal component of coveriance
function [mu,sigma] = ML_diagonal(x)
    len = size(x,1);
    mu = sum(x) / len;
    tmp = x - repmat(mu,len,1);
    sigma = (tmp'*tmp) / len;
    sigma = diag(sigma);
end
```

Result:

```
d) mean and diagnoal variance for w2
mean for x1,x2,x3 of w2 =
   -0.1126    0.4299    0.0037

variance for x1,x2,x3 of w2 =
    0.0539
    0.0460
    0.0073
```

e/f) compare the mean and variance calculated in above ways.

```
% e/f) mean and variance of each feature
[mu1,sigma1] = ML_estimate2([pxw1(:,1),pxw2(:,1),pxw3(:,1)]);
[mu2,sigma2] = ML_estimate2([pxw1(:,2),pxw2(:,2),pxw3(:,2)]);
[mu3,sigma3] = ML_estimate2([pxw1(:,3),pxw2(:,3),pxw3(:,3)]);
disp('e/f) mean and variance of each feature x1,x2,x3 using function ML_estimate2')
disp('mean for x1 = '); disp(mu1);
disp('variance for x1 = '); disp(sigma1);
disp('mean for x2 = '); disp(mu2);
disp('variance for x2 = '); disp(sigma2);
disp('mean for x3 = '); disp(mu3);
disp('variance for x3 = '); disp(sigma3);
[mu1,sigma1] = ML_diagonal([pxw1(:,1),pxw2(:,1),pxw3(:,1)]);
[mu2,sigma2] = ML_diagonal([pxw1(:,2),pxw2(:,2),pxw3(:,2)]);
[mu3,sigma3] = ML_diagonal([pxw1(:,3),pxw2(:,3),pxw3(:,3)]);
disp('e/f) mean and variance of each feature x1,x2,x3 using function ML_diagonal')
disp('mean for x1 = '); disp(mu1);
disp('variance for x1 = '); disp(sigma1);
disp('mean for x2 = '); disp(mu2);
disp('variance for x2 = '); disp(sigma2);
disp('mean for x3 = '); disp(mu3);
disp('variance for x3 = '); disp(sigma3);
```

Result:

```
e/f) mean and variance of each feature x1,x2,x3 using function ML_estimate2
mean for x1 =
   -0.0709   -0.1126    0.2747

variance for x1 =
    0.9062    0.0753   -0.2760
    0.0753    0.0539   -0.0718
   -0.2760   -0.0718    0.3019

mean for x2 =
   -0.6047    0.4299    0.3001

variance for x2 =
    4.2007    0.1320   -0.5137
    0.1320    0.0460    0.0047
   -0.5137    0.0047    0.6450

mean for x3 =
   -0.9110    0.0037    0.6786

variance for x3 =
    4.5419   -0.0650    0.5060
   -0.0650    0.0073    0.0344
    0.5060    0.0344    1.2621
```

```
e/f) mean and variance of each feature x1,x2,x3 using function ML_diagonal
mean for x1 =
   -0.0709   -0.1126    0.2747

variance for x1 =
    0.9062
    0.0539
    0.3019

mean for x2 =
   -0.6047    0.4299    0.3001

variance for x2 =
    4.2007
    0.0460
    0.6450

mean for x3 =
   -0.9110    0.0037    0.6786

variance for x3 =
    4.5419
    0.0073
    1.2621
```

Conclusion:

The value of estimation of mean and covariance are the same.

For the mean, using the maximum-likelihood estimation, the estimated mean is the sample mean. Therefore, it doesn't matter the variance is diagonal matrix or not.

As for the variance, the diagonal one is actually the diagonal value of the covariance matrix in the former function. They are calculated by the same way and thus they are the same.