Assignment 2

Name: Vuong Minh Phu

Student ID: 202050798

Class: Computer Engineering

Instructor: Prof. Seung-Hoon Na

Problem 4.5

Prove that the normalization constant for a d-dimensional Gaussian is given by

$$(2\pi)^{d/2}|\Sigma|^{\frac{1}{2}} = \int exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))dx$$

We have: $\Sigma = U\Lambda U^T$

$$\int exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))dx = \int exp(-\frac{1}{2}(x-\mu)^T U\Lambda^{-1}U^T(x-\mu))dx \tag{1}$$

where U is an orthogonal matrix $(U^{-1} = U^T)$, we can derive Eq. 1. Then we perform coordinate changing $y = U^T(x-\mu)$

$$\begin{gathered}\int exp(-\frac{1}{2}(x-\mu)^T U\Lambda^{-1}U^T(x-\mu))dx = \int exp(-\frac{1}{2}y^T\Lambda^{-1}y)dx = \\ \int exp(-\frac{1}{2}\sum_i \frac{y_i^2}{\lambda_i})dx = \int exp(-\frac{1}{2}\sum_i \frac{y_i^2}{\lambda_i})\frac{\partial(x1,x2,\ldots,xn)}{\partial(y1,y2,\ldots,yn)}dy\end{gathered} \tag{2}$$

To change the vector of intergration from dx to dy, we need to compute the Jacobian.

$$\begin{gathered}y = U^T(x-\mu) \Rightarrow x = Uy + \mu \\ J_{ij} = \frac{\partial x_i}{\partial y_i} = u_{ij}\end{gathered} \tag{3}$$

Thus, $J = U$ and $\frac{\partial(x1,x2,\ldots,xn)}{\partial(y1,y2,\ldots,yn)} = \det(U)$. Because U is orthogonal, we have $\det(U) = \pm 1$. We also know that $\exp(f(y))$ is always positive $\Rightarrow \det(U) = 1$.

Therefore, the expression becomes:

$$\int \exp(-\frac{1}{2}\sum_i \frac{y_i^2}{\lambda_i})dy = \prod_i \int \exp(-\frac{1}{2}\frac{y_i^2}{\lambda_i})dy_i \tag{4}$$

From the Gaussian properties $\int exp(-\frac{x^2}{2\sigma^2})dx = \sqrt{(2\pi\sigma^2)}$. Then the Eq.4 becomes:

$$\prod_i \int exp(-\frac{1}{2}\frac{y_i^2}{\lambda_i})dy_i = \prod_{i=1}^{d}\sqrt{(2\pi\lambda_i)} = \sqrt{(2\pi)^d}\prod_{i=1}^{d}\sqrt{\lambda_i} = (2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}} \tag{5}$$

where $|\Sigma| = \prod \lambda_i$

Problem 4.7

Consider a bivariate Gaussian distribution $p(x_1, x_2) = N(x|\mu, \Sigma)$ where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} = \sigma_1\sigma_2 \begin{pmatrix} \frac{\sigma_1}{\sigma_2} & \rho \\ \rho & \frac{\sigma_1}{\sigma_2} \end{pmatrix} \tag{4.269}$$

where the correlation coefficient is given by

$$\rho \triangleq \frac{\sigma_{12}}{\sigma_1\sigma_2} \tag{4.270}$$

**

***a)** Since x_1, x_2 have a multivariate Gaussian distribution, the conditional $p(x_2|x_1)$ is also Gaussian. We have:*

$$\begin{gathered}p(x_2|x_1) = N(x_2|\mu_{2|1}, |\Sigma_{2|1}) \\ \mu_{2|1} = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1) \\ \Sigma_{2|1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\end{gathered} \tag{1}$$

For the bivariate Gaussian, the vectors μ_2, μ_1 and the matrix partitions Σ_{ij} become scalars. Therefore:

$$\begin{gathered}\mu_{2|1} = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1) \\ \Sigma_{2|1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \sigma_2^2 - \frac{\sigma_{21}\sigma_{12}}{\sigma_1^2} = \sigma_2^2(1-\rho^2)\end{gathered} \tag{2}$$

Using (2), we get:

$$p(x_2|x_1) = N(x_2|\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1), \sigma_2^2(1-\rho^2)) \tag{3}$$

***b)** Assuming $\sigma_1 = \sigma_2 = 1$ we can write Eq.3 as follow:*

$$p(x_2|x_1) = N(x_2|\mu_2 + \rho(x_1-\mu_1), (1-\rho^2))$$

Problem 4.11

Derive $p(\mu, \Sigma|D) = NIW(\mu, \Sigma|m_N, \kappa_N, \nu_N, S_N)$ by showing that

$$\begin{gathered}N(\bar{x}-\mu)(\bar{x}-\mu)^T + \kappa_0(\mu-m_0)(\mu-m_0)^T \\ = \kappa_N(\mu-m_N)(\mu-m_N)^T + \frac{\kappa_0 N}{K_N}(\bar{x}-m_0)(\bar{x}-m_0)^T\end{gathered}$$

***Solution:** The likelihood is:*

$$p(D|\mu,\Sigma)\alpha|\Sigma|^{-\frac{N}{2}}\exp(-\frac{N}{2}(\mu-\bar{x})^T\Sigma^{-1}(\mu-\bar{x}))\exp(-\frac{1}{2}tr(\Sigma^{-1}S_x)) \tag{1}$$

We also know that the conjugate prior is:

$$\begin{gathered}p(\mu,\Sigma) = NIW(\mu,\Sigma|m_0,\kappa_0,\nu_0,S_0)\ \alpha \\ |\Sigma|^{-\frac{\nu_0+D+2}{2}}\exp(-\frac{\kappa_0}{2}(\mu-m_0)^T\Sigma^{-1}(\mu-m_0) - \frac{1}{2}tr(\Sigma^{-1}S_0))\end{gathered} \tag{2}$$

Therefore, the posterior is proportional to:

$$\begin{gathered}p(\mu,\Sigma|D)\ \alpha\ p(D|\mu,\Sigma)p(\mu,\Sigma)\ \alpha \\ |\Sigma|^{-\frac{N}{2}}\exp(-\frac{N}{2}(\mu-\bar{x})^T\Sigma^{-1}(\mu-\bar{x}))\exp(-\frac{1}{2}tr(\Sigma^{-1}S_{\bar{x}})) \\ |\Sigma|^{-\frac{\nu_0+D+2}{2}}\exp(-\frac{\kappa_0}{2}(\mu-m_0)^T\Sigma^{-1}(\mu-m_0) - \frac{1}{2}tr(\Sigma^{-1}S_0)) = \\ |\Sigma|^{-\frac{\nu_0+D+2+N}{2}}\exp(-\frac{N}{2}(\mu-\bar{x})^T\Sigma^{-1}(\mu-\bar{x}) - \frac{\kappa_0}{2}(\mu-m_0)^T\Sigma^{-1}(\mu-m_0)) \\ \exp(-\frac{1}{2}tr(\Sigma^{-1}S_x) - \frac{1}{2}tr(\Sigma^{-1}S_0))\end{gathered} \tag{3}$$

In the psterior, we have $|\Sigma|^{-\frac{\nu_0+D+2+N}{2}}$, because D is fixed, then the only parameter left to update is ν_0. Thus $\nu_0 = \nu_0 + N$.

After that is the quadratic form expressions, we only need to derive $\exp(-\frac{N}{2}(\mu-\bar{x})^T\Sigma^{-1}(\mu-\bar{x}) - \frac{\kappa_0}{2}(\mu-m_0)^T\Sigma^{-1}(\mu-m_0))$.

From 4.6.3.1, we know that:

$$\sum_{i=0}^{N}(x_i-\mu)\Sigma^{-1}(x_i-\mu) = tr(\Sigma^{-1}S_{\bar{x}}) + N(\bar{x}-\mu)^T\Sigma^{-1}(\bar{x}-\mu) \tag{4}$$

Thus

$$\begin{gathered}N(\mu-\bar{x})^T\Sigma^{-1}(\mu-\bar{x}) + \kappa_0(\mu-m_0)^T\Sigma^{-1}(\mu-m_0) = \\ tr(\Sigma^{-1}S_{m_N}) + (N+\kappa_0)(m_N-\mu)^T\Sigma^{-1}(m_N-\mu) \\ m_N = \frac{sumAllTerms}{numberOfTerms} = \frac{\kappa_0 m_0 + N\bar{x}}{\kappa_0 + N} \\ S_{m_N} = N(\bar{x}-m_N)(\bar{x}-m_N)^T = \\ \frac{N_{\kappa_0}}{N+\kappa_0}(\bar{x}-m_0)(\bar{x}-m_0)^T\end{gathered} \tag{5}$$

We also have $tr(\Sigma^{-1}S_0) + tr(\Sigma^{-1}S_{\bar{x}}) + tr(\Sigma^{-1}S_{m_N}) = tr(\Sigma^{-1}(S_0 + S_{\bar{x}} + S_{m_N}))$

Finally, we need to combine all the new terms and update the remaining parameters:

$$\begin{gathered}p(\mu,\Sigma)\ \alpha\ |\Sigma|^{-\frac{\nu_N+D+2}{2}}\exp(-\frac{N+\kappa_0}{2}(m_N-\mu)^T\Sigma^{-1}(m_N-\mu)) \\ exp(-\frac{1}{2}tr\Sigma^{-1}(S_0 + S_{m_N} + S_{\bar{x}})) = \\ |\Sigma|^{-\frac{\nu_N+D+2}{2}}\exp(-\frac{\kappa_N}{2}(\mu-m_N)^T|\Sigma|^{-1}(\mu-m_N) - \frac{1}{2}tr(\Sigma^{-1}S_N))\end{gathered} \tag{6}$$

The posterior is given by the following distributions:

$$\begin{gathered}p(\mu,\Sigma) = NIW(\mu,\Sigma|m_N,\kappa_N,\nu_N,S_N) \\ m_N = \frac{\kappa_0 m_0 + N\bar{x}}{\kappa_0 + N} \\ \kappa_N = \kappa_0 + N \\ \nu_N = \nu_0 + N \\ S_N = S_0 + S_{\bar{x}} + S_{m_N}\end{gathered}$$

Problem 4.22

Given the prior probabilities of thress category classification:

$$P(Y=1) = P(Y=2) = P(Y=3)$$

The class-condiional densities are NVM with parameters:

$$\mu_1 = [0,0]^T, \mu_2 = [1,1]^T, \mu_3 = [-1,1]^T$$

and

$$\Sigma_1 = \begin{bmatrix} 0.7 & 0 \\ 0 & 0.7 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 0.8 & 0.2 \\ 0.8 & 0.8 \end{bmatrix}$$

Classify

a. $x = [-0.5, 0.5]$

b. $x = [0.5, 0.5]$

Solution:

In order to classify the points we need to calculate the NVM distributions for each of them. The classification of each point is based on which class produce the bigger posterior distribution. Since the prior of three classes are the same, the Gaussian class conditional is responsible to determine the bigger score.

We have the pdf for NVM is:

$$f(x) = \frac{1}{\sqrt{(2\pi)^k\det(\Sigma)}}\exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))$$

For each point we need to calculate the pdf for it with each μ and Σ values Hence, we can use numpy and scipy libraries to calculate it

In [1]:
```
import numpy as np
from scipy.stats import multivariate_normal
mu_1 = np.array([0, 0])
mu_2 = np.array([1, 1])
mu_3 = np.array([-1, 1])

Sigma_1 = np.array([[0.7, 0], [0, 0.7]])
Sigma_2 = np.array([[0.8, 0.2], [0.2, 0.8]])
Sigma_3 = np.array([[0.8, 0.2], [0.2, 0.8]])

x_1 = np.array([-0.5, 0.5])
x_2 = np.array([0.5, 0.5])
```

In [2]:
```
fx1=  multivariate_normal.pdf(x_1, mu_1, Sigma_1)
fx2=  multivariate_normal.pdf(x_1, mu_2, Sigma_2)
fx3=  multivariate_normal.pdf(x_1, mu_3, Sigma_3)

print("the pdf for x1 are: ",fx1,fx2,fx3)
```

```
the pdf for x1 are:  0.15908048981271555 0.04983035606132821 0.13545295138715246
```

*Therefore, for question a we will classify x as **Class 1***

*We will do the same for **question b** and get the result*

In [3]:
```
fx1=  multivariate_normal.pdf(x_2, mu_1, Sigma_1)
fx2=  multivariate_normal.pdf(x_2, mu_2, Sigma_2)
fx3=  multivariate_normal.pdf(x_2, mu_3, Sigma_3)

print("the pdf for x1 are: ",fx1,fx2,fx3)
```

```
the pdf for x1 are:  0.15908048981271555 0.1600187545745967 0.03022363873559281
```

*We classifiy x as **Class 2***