Machine Learning Homework 3.1

专业：软件工程　　　　　　姓名：沈金龙　　　　　　学号：18214806

1. 作业题目

考虑三维正态分布 p($\mathbf{x}$|ω)~N($\boldsymbol{\mu}$,$\boldsymbol{\Sigma}$)，其中

$$\boldsymbol{\mu}=\begin{pmatrix}1\\2\\2\end{pmatrix}\quad 及\ \boldsymbol{\Sigma}=\begin{pmatrix}1&0&0\\0&5&2\\0&2&5\end{pmatrix}$$

(a) 求点 x_0 =(0.5,0,1)^t 处的概率密度；

(b) 构造白化变换矩阵 $A_\omega(A_\omega=\Phi\Lambda^{-1/2})$ ，计算分别表示本征向量和本征值的矩阵Φ和Λ；然后，将此分布转换为以原点为中心协方差矩阵为单位阵的分布，即 p($\mathbf{x}$|ω)~$\mathbf{N}(\mathbf{0},\mathbf{I})$；

(c) 将整个同样的转换过程应用于点 x_0 以产生一变换点 x_ω ；

(d) 通过详细计算，证明原分布中从 x_0 到均值$\boldsymbol{\mu}$的 Mahalanobis 距离与变换后的分布中从 x_ω 到 0 的 Mahalanobis 距离相等；

(e) 概率密度在某个一般线性变换下是否保持不变？换句话说，对于某线性变换 T，是否有 $p(x_0\mid N(\mu,\Sigma))=p(T^t x_0\mid N(T^t\mu,T^t\Sigma T))$ ？解释原因；

(f) 证明当把一个一般的白化变换 $A_\omega=\Phi\Lambda^{-1/2}$ 应用于一个高斯分布时可保证最终分布的协方差与单位阵 I 成比例，检查变换后的分布是否仍具有归一化特性。

2. 答案

(a) 多维高斯分布的联合概率密度函数为：

$$p(x)=\frac{1}{(\sqrt{2\pi})^d\,|\Sigma|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(x-\mu)^T\sum\nolimits^{-1}(x-\mu)\right]$$

代入 x_0 ，μ及Σ后可计算得　p =0.0513。

(b) 由定义，可根据下式计算协方差矩阵Σ的特征值，

$$(\Sigma-\lambda I)X=0\Rightarrow\lambda_1=1,\lambda_2=3,\lambda_3=7.$$

将 $\lambda_1=1,\lambda_2=3,\lambda_3=7$ 分别代入 $(\Sigma-\lambda I)X=0$ 可计算出对应的特征向量。由

特征向量为列向量形成的矩阵Φ以及特征值为对角元素形成的Λ为：

$$\Phi=\begin{bmatrix}1 & 0 & 0\\ 0 & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2}\\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2}\end{bmatrix},\Lambda=\begin{bmatrix}1 & 0 & 0\\ 0 & 3 & 0\\ 0 & 0 & 7\end{bmatrix}\Rightarrow A_{\omega}=\Phi\Lambda^{-1/2}=\begin{bmatrix}1 & 0 & 0\\ 0 & -\frac{\sqrt{6}}{6} & \frac{\sqrt{14}}{14}\\ 0 & \frac{\sqrt{6}}{6} & \frac{\sqrt{14}}{14}\end{bmatrix}.$$

构造变换和可将原分布转换为)。

(c) $x_{\omega}=A_{\omega}^{T}(x_0-\mu)=[-0.5\quad 0.4082\quad -0.8018]^{T}$ 。

(d) 原分布中从 x_0 到均值μ的 Mahalanobis 距离为 ：

$$D(x_0,\mu)=(x_0-\mu)^{T}\sum\nolimits^{-1}(x_0-\mu)=1.0595.$$

变换后的分布中从到 0 的 Mahalanobis 距离为：

$$D(x_{\omega},0)=x_{\omega}^{T}I^{-1}x_{\omega}=1.0595.$$

(e) 通常会变化。以高斯分布为例，分析 p(x)的表达式，尽管 e 的指数项上的值在变换前后不变，但|Σ|的值变化了，这将使得 p(x)的值变化。方差(或协方差矩阵)用于衡量总体与均值的偏离程度。在变换中总体的分布范围会改变，自然会造成方差(或协方差矩阵)的变化。

(f) 证明：令原始高斯分布的协方差矩阵为Σ，变换后高斯分布的协方差矩阵为，由协方差矩阵的定义，

$$\Sigma=E[(x-\mu)(x-\mu)^{T}],$$
$$\Sigma_{\omega}=E[(A_{\omega}{}^{T}x-A_{\omega}{}^{T}\mu)(A_{\omega}{}^{T}x-A_{\omega}{}^{T}\mu)^{T}]=E[(A_{\omega}{}^{T}x-A_{\omega}{}^{T}\mu)(x^{T}A_{\omega}-\mu^{T}A_{\omega})]$$
$$=E[A_{\omega}{}^{T}(x-\mu)(x-\mu)^{T}A_{\omega}]=A_{\omega}{}^{T}E[(x-\mu)(x-\mu)^{T}]A_{\omega}=A_{\omega}{}^{T}\Sigma A_{\omega}.$$

前面的解题中为了方便将Φ取为正交的，在一般情况下表示本证向量的矩阵不一定正交，不妨设为 $\Phi=k\Omega$ ，其中Ω为正交阵，k 为常数。由于Σ为对称阵，故有分解 $\Sigma=\Omega\Lambda\Omega^{T}$ ，将之代入 Σ_{ω} 中有：

$$\Sigma_{\omega}=A_{\omega}{}^{T}\Omega\Lambda\Omega^{T}A_{\omega}=(k\Omega\Lambda^{-1/2})^{T}\Omega\Lambda\Omega^{T}(k\Omega\Lambda^{-1/2})=k^{2}(\Lambda^{-1/2})^{T}\Omega^{T}\Omega\Lambda\Omega^{T}\Omega\Lambda^{-1/2}.$$

由于Λ为对角阵，因此：

$$\Sigma_{\omega} = k^2(\Lambda^{-1/2})^T\Omega^T\Omega\Lambda\Omega^T\Omega\Lambda^{-1/2} = k^2\Lambda^{-1/2}\Lambda\Lambda^{-1/2} = k^2 I \,.$$

下面验证变换后的分布是否具有归一化特性。为简单起见　仅考虑原多维正态分布中各维之间相互独立的情况。此时协方差矩阵为对角阵，可将多维情况分解为一维的组合分别进行处理。

假设对某维 i，原来的方差为 σ_i^2 ，Φ取为正交阵，变换后的方差为：

$$\sigma_i^{\,2}(A_{\omega}{}^T A_{\omega})_{ii} = \sigma_i^{\,2}[(\Phi\Lambda^{-1/2})^T(\Phi\Lambda^{-1/2})]_{ii} = \sigma_i^{\,2}\Lambda^{-1}{}_{ii} = \sigma_i^{\,2}\Sigma^{-1}{}_{ii} = \sigma_i^{\,2}\sigma_i^{\,-2} = 1 \,.$$

由于 $A_{\omega} = \Phi\Lambda^{-1/2} = \Lambda^{-1/2} = \Sigma^{-1/2}$，有 $[A_{\omega}]_i = \sigma_i^{-1}$ ，因此在该维可得 ：

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left\langle [A_{\omega}]_i^{\,T}(x_i-\mu_i)\right\rangle^T\left\langle [A_{\omega}]_i^{\,T}(x_i-\mu_i)\right\rangle\right]d[A_{\omega}]_i^{\,T}x_i$$

$$=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(x_i-\mu_i)^T[A_{\omega}]_i[A_{\omega}]_i^{\,T}(x_i-\mu_i)\right]d[A_{\omega}]_i^{\,T}x_i$$

$$=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(x_i-\mu_i)^T\sigma_i^{\,-2}(x_i-\mu_i)\right]d\sigma_i^{\,-1}x_i$$

$$=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\,\sigma_i}\exp\left[-\frac{(x_i-\mu_i)^T(x_i-\mu_i)}{2\sigma_i^{\,2}}\right]dx_i = 1$$