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

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

(a) 求点$\mathbf{x}_0=(0.5,0,1)^t$处的概率密度;

(b) 构造白化变换矩阵$\mathbf{A}_\omega(\mathbf{A}_\omega=\boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2})$，计算分别表示本征向量和本征值的矩阵$\boldsymbol{\Phi}$和$\boldsymbol{\Lambda}$；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$p(\mathbf{x}|\omega)\sim \mathbf{N}(\mathbf{0},\mathbf{I})$;

(c) 将整个同样的转换过程应用于点$\mathbf{x}_0$以产生一变换点$\mathbf{x}_\omega$;

(d) 通过详细计算，证明原分布中从$\mathbf{x}_0$到均值$\boldsymbol{\mu}$的 Mahalanobis 距离与变换后的分布中从$\mathbf{x}_\omega$到 $\mathbf{0}$ 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 $\mathbf{T}$，是否有$p(\mathbf{x}_0|N(\boldsymbol{\mu},\boldsymbol{\Sigma}))=p(\mathbf{T}^t\mathbf{x}_0|N(\mathbf{T}^t\boldsymbol{\mu},\mathbf{T}^t\boldsymbol{\Sigma}\mathbf{T}))$？解释原因;

(f) 证明当把一个一般的白化变换$\mathbf{A}_\omega=\boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2}$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。

解：

a、　多维正态分布的密度函数公式为：

$$f(x;\mu,\Sigma)=\frac{1}{\sqrt{2\pi}^d\sqrt{det(\Sigma)}}exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

将 $x_0=(0.5,0,1)^T$带入上式可得密度函数为 $f_0 = 0.008157$

```
import numpy as np
import math


average = np.array([[1], [2], [2]])
variance = np.array([[1, 0, 0], [0, 5, 2], [0, 2, 5]])
def gauss(d, x, ave, var):
    f = 1/(((math.sqrt(2 * math.pi)) ** d) * math.sqrt(np.linalg.det(var))) * \
    math.exp(-0.5 * np.dot(np.dot(np.transpose(x - ave), np.linalg.inv(var)), (x - ave)))
    return f
x = np.array([[0.5], [0], [1]])
f = gauss(3, x, average, variance)
print(f)
```

```
0.008157327113891l9
```

b、 协方差矩阵的特征值为 7、3、1；

对应的特征向量分别为：

[0，0，1]、[0.707,0.707,0]、[0.707,-0.707,0]

特征向量按照列排列形成的特征矩阵 Φ 和以特征值为对角矩阵 Λ 分别为：

$$\Phi = \begin{bmatrix} 0 & 0 & 1 \\ 0.707 & 0.707 & 0 \\ 0.707 & -0.707 & 0 \end{bmatrix} \quad \Lambda = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

可得白化变换矩阵为：

$$A_w = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.408 & 0.267 \\ 0 & 0.408 & 0.267 \end{bmatrix}$$

构造变换 $T_\mu = A_w^T(x-\mu)$ 和 $T_\sum = A_w^T \sum A_w^T$

可将原分布转换为 $\mathrm{p}(\mathbf{x}|\omega) \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$

```
import numpy as np
import math


variance = np.array([[1, 0, 0], [0, 5, 2], [0, 2, 5]])
a, b = np.linalg.eig(variance)
print(a)
print(b)
```

```
[7. 3. 1.]
[[ 0.          0.          1.        ]
 [ 0.70710678  0.70710678  0.        ]
 [ 0.70710678 -0.70710678  0.        ]]
```

c、 $x_w = A_w^T(x_0 - \mu) = [-0.5, 0.408, -0.802]$

d、 原分布中从**x0**到均值**μ**的 Mahalanobis 距离为：

$$D(x_0, \mu) = (x_0 - \mu)^T \sum^{-1}(x_0 - \mu) = 1.060$$

变换后的分布中从**xω**到 0 的 Mahalanobis 距离为：

$$D(x_w, 0) = x_w^T I^{-1} x_w = 1.060$$

由此证得原分布中从**x0**到均值**μ**的Mahalanobis距离与变换后的分布中从**xω**到 0 的 Mahalanobis 距离相等

e、 通常会变化，以高斯分布为例，分析 $p(x|w)$ 的表达式，e 的指数项上的值再变换前后不变，但是 $|\sum|$ 的值变化了，这将使 $p(x|w)$ 的值同时发生变化。

f、 证明：

原始高斯分布的协方差为 $\sum$ ，变换后高斯分布的协方差矩阵为 $\sum_w$

由协方差矩阵的定义，有：

$$\sum = E[(x-\mu)(x-\mu)^T],$$

$$\begin{aligned}\sum\nolimits_w &= E[(A_w^T x - A_w^T \mu)(A_w^T x - A_w^T \mu)^T] \\ &= E[(A_w^T x - A_w^T \mu)(x^T A_w - \mu^T A_w)] \\ &= E[A_w^T (x-\mu)(x-\mu)^T A_w] \\ &= A_w^T E[(x-\mu)(x-\mu)^T] A_w \\ &= A_w^T \sum A_w\end{aligned}$$

假设 Φ = kF，其中 F 为正交矩阵，k 为常数。由于 $\sum$ 为对称矩阵，故有分解 $\sum = F\Lambda F^T$ ，代入 $\sum_w$ 后可得

$$\begin{aligned}\sum\nolimits_w &= A_w^T F\Lambda F^T A_w \\ &= (kF\Lambda^{-1/2})^T F\Lambda F^{-1/2} \\ &= k^2(\Lambda^{-1/2})^T F^T F\Lambda F^T F\Lambda^{-1/2}\end{aligned}$$

由于 Λ 为对称矩阵，因此有：

$$\begin{aligned}\sum\nolimits_w &= k^2(\Lambda^{-1/2})^T F^T F\Lambda F^T F\Lambda^{-1/2} \\ &= k^2\Lambda^{-1/2}\Lambda\Lambda^{-1/2} \\ &= k^2 I\end{aligned}$$

由此证明当把一个一般的白化变换$\mathbf{A\omega}=\Phi\mathbf{\Lambda-1/2}$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

验证变换后的分布是否具有归一特性。

考虑原多维正太分布中各维之间相互独立的情况。此时协方差矩阵为对角矩阵，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 j，原来的方差为 V_i^2，Λ 取为正交矩阵，变换后

$$V_i^2(A_w^T A_w)_{ii} = V_i^2[(\Phi\Lambda^{-1/2})^T(\Phi\Lambda^{-1/2})]_{ii} = V_i^2 \sum_i^{-1} i = V_i^2 V_i^{-2} = 1$$

由于 $A_w = \sum^{-1/2}$ 有 $[A_w]_i^T = V_i^{-1}$ ，因此可得出变换后的分布具有归一化的特性。