模式识别作业二

数据科学与计算机学院　17大数据与人工智能

17341015　陈鸿峥

问题 1 (§2 Q23)**.** 考虑三维正态分布$p(\mathbf{x}\mid\omega)\sim N(\boldsymbol{\mu},\Sigma)$，其中

$$\boldsymbol{\mu}=\begin{bmatrix}1\\2\\2\end{bmatrix},\ \Sigma=\begin{bmatrix}1&0&0\\0&5&2\\0&2&5\end{bmatrix}$$

(a) 求点$\mathbf{x}_0=\begin{bmatrix}0.5&0&1\end{bmatrix}^{\mathrm{T}}$处的概率密度。

(b) 构造白化变换$A_\omega=\phi\Lambda^{-1/2}$，计算分别表示本征向量和本征值的矩阵$\phi$和$\Lambda$；接下来，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$p(\mathbf{x}\mid\omega)\sim N(0,I)$。

(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) 概率密度在一个一般的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$p(\mathbf{x}_0\mid N(\boldsymbol{\mu},\Sigma))=p(T^{\mathrm{T}}\mathbf{x}_0\mid N(T^{\mathrm{T}}\boldsymbol{\mu},T^{\mathrm{T}}\Sigma T))$？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然具有归一化特性。

解答.　(a) 我们有$d=3$维，

$$|\Sigma|=\begin{vmatrix}1&0&0\\0&5&2\\0&2&5\end{vmatrix}=21$$

$$\Sigma^{-1}=\begin{bmatrix}1&0&0\\0&5&2\\0&2&5\end{bmatrix}^{-1}=\begin{bmatrix}1&0&0\\0&5/21&-2/21\\0&-2/21&5/21\end{bmatrix}$$

又有平方Mahalanobis距离

$$
\begin{aligned}
&(\mathbf{x}_0-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}_0-\boldsymbol{\mu}) \\
=&\begin{bmatrix}0.5-1\\0-2\\1-2\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}1&0&0\\0&5/21&-2/21\\0&-2/21&5/21\end{bmatrix}\begin{bmatrix}0.5-1\\0-2\\1-2\end{bmatrix} \\
=&1.06
\end{aligned}
$$

将上述数值代入得到概率密度

$$
p(\mathbf{x}_0\mid\omega)=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{x}_0-\boldsymbol{\mu})^{\mathrm{T}}\right]\Sigma^{-1}(\mathbf{x}_0-\boldsymbol{\mu})=0.008155
$$

(b) 先求出Σ的特征值，考虑特征方程$|\Sigma-\lambda I|=0$

$$
\begin{vmatrix}1-\lambda&0&0\\0&5-\lambda&2\\0&2&5-\lambda\end{vmatrix}=(1-\lambda)(3-\lambda)(7-\lambda)=0
$$

对于不同特征值，回代求其特征向量

$$
\lambda_1=1:\quad\begin{bmatrix}1&0&0\\0&5&2\\0&2&5\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\implies\phi_1=\begin{bmatrix}1\\0\\0\end{bmatrix}
$$

$$
\lambda_2=3:\quad\begin{bmatrix}1&0&0\\0&5&2\\0&2&5\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}3x_1\\3x_2\\3x_3\end{bmatrix}\implies\phi_2=\begin{bmatrix}0\\1/\sqrt{2}\\-1/\sqrt{2}\end{bmatrix}
$$

$$
\lambda_3=7:\quad\begin{bmatrix}1&0&0\\0&5&2\\0&2&5\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}7x_1\\7x_2\\7x_3\end{bmatrix}\implies\phi_3=\begin{bmatrix}0\\1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix}
$$

最终得到本征值矩阵

$$
\phi=\begin{bmatrix}1&0&0\\0&1/\sqrt{2}&1/\sqrt{2}\\0&-1/\sqrt{2}&1/\sqrt{2}\end{bmatrix}
$$

以及

$$A_\omega = \phi\Lambda^{-1/2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{7} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{6} & 1/\sqrt{14} \\ 0 & -1/\sqrt{6} & 1/\sqrt{14} \end{bmatrix}$$

进而

$$Y = A_\omega^{\mathrm{T}}(\mathbf{x} - \boldsymbol{\mu}) \sim N(0, I)$$

(c) 将$\mathbf{x}_0$代入有

$$\mathbf{x}_\omega = A_\omega^{\mathrm{T}}(\mathbf{x}_0 - \boldsymbol{\mu}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{6} & -1/\sqrt{6} \\ 0 & 1/\sqrt{14} & 1/\sqrt{14} \end{bmatrix} \begin{bmatrix} -0.5 \\ -2 \\ -1 \end{bmatrix} = \begin{bmatrix} -0.5 \\ -1/\sqrt{6} \\ -3/\sqrt{14} \end{bmatrix}$$

(d) 由(a)知原距离$r^2 = 1.06$，而新的距离

$$r_\omega^2 = \mathbf{x}_\omega^{\mathrm{T}}\mathbf{x}_\omega = 1.06$$

因此两者相等

(e) 设线性变换后的向量为$\mathbf{x}' = T^{\mathrm{T}}\mathbf{x}$，进而有均值

$$\boldsymbol{\mu}' = \sum_{k=1}^{n} \mathbf{x}'_k \Big/ n = \sum_{k=1}^{n} T^{\mathrm{T}}\mathbf{x}_k \Big/ n = T^{\mathrm{T}} \sum_{k=1}^{n} \mathbf{x}_k \Big/ n = T^{\mathrm{T}}\boldsymbol{\mu}$$

和协方差

$$\Sigma' = \sum_{k=1}^{n} (\mathbf{x}'_k - \boldsymbol{\mu}')(\mathbf{x}'_k - \mu')^{\mathrm{T}} = T^{\mathrm{T}} \left[\sum_{k=1}^{n} (\mathbf{x}_k - \boldsymbol{\mu})(\mathbf{x}_k - \mu)^{\mathrm{T}} \right] T = T^{\mathrm{T}}\Sigma T$$

又$|\Sigma'| = |T^{\mathrm{T}}\Sigma T| = |\Sigma|$，故

$$p(\mathbf{x}_0 \mid N(\boldsymbol{\mu}, \Sigma)) = p(T^{\mathrm{T}}\mathbf{x}_0 \mid N(T^{\mathrm{T}}\boldsymbol{\mu}, T^{\mathrm{T}}\Sigma T))$$

(f) 由图2.8上面的公式，我们有

$$\mathbf{y} = A_\omega^{\mathrm{T}}\mathbf{x} \sim N(A_\omega^{\mathrm{T}}\boldsymbol{\mu}, A_\omega^{\mathrm{T}}\Sigma A_\omega)$$

进而可得协方差矩阵

$$\begin{aligned} A_\omega^{\mathrm{T}}\Sigma A_\omega &= (\phi\Lambda^{-1/2})^{\mathrm{T}}\phi\Lambda\phi^{\mathrm{T}}(\phi\Lambda^{-1/2}) \\ &= \Lambda^{-1/2}(\phi^{\mathrm{T}}(\phi\Lambda\phi^{\mathrm{T}})\phi)\Lambda^{-1/2} \\ &= I \end{aligned}$$

故变换后的分布依然具有归一化特性