データサイエンス

問題１．スカラーの値を持つデータ点 x_i $(i=1,\cdots,n)$ が，確率密度関数 $p(x|C_1)$ と $p(x|C_2)$ のいずれか一方に従って生成される場合の確率推論について，以下の問いに答えよ．ただし，n は１以上の整数であり，C_1 と C_2 はそれぞれ異なるクラス１，クラス２を表すものとする．

$$p(x|C_1)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-\mu_1)^2}{2}\right),\quad p(x|C_2)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu_2)^2}{2\sigma^2}\right)$$

ここで，クラス１の事前確率は 3/7 であり，クラス２の事前確率は 4/7 であるとする．μ_1 と μ_2 は定数であり，σ は正の定数であるとする．$\exp(-0.5)\approx 0.61$ を用いてよい．

(1) ３つのデータ点 x_1，x_2，x_3 がクラス１に属しているものとする．このとき，μ_1 が従う尤度関数 $L_1(\mu_1)$ は $L_1(\mu_1)=p(x_1|C_1)p(x_2|C_1)p(x_3|C_1)$ となる．最尤推定法により，μ_1 の推定値を導出せよ．また，３つのデータ点が $x_1=0.1$，$x_2=1.1$，$x_3=-0.5$ である場合の μ_1 の推定値を求めよ．

(2) ３つのデータ点 x_4，x_5，x_6 がクラス２に属しているものとする．μ_2 と σ^2 が従う尤度関数 $L_2(\mu_2,\sigma^2)$ を求めよ．最尤推定法により，μ_2 と σ^2 の推定値を導出せよ．また，３つのデータ点が $x_4=1$，$x_5=2$，$x_6=3$ である場合の μ_2 と σ^2 の推定値を求めよ．

(3) $\mu_1=-0.1$，$\mu_2=2.9$，$\sigma=2$ とし，１つのデータ点 $x_7=0.9$ のみが与えられたとする．各クラスの事後確率を求めよ．また，事後確率最大化法により，データ点 x_7 が属するクラスを推定せよ．

問題２．２次元上の３点のデータ点 $\boldsymbol{r}_i=(x_i,y_i)^T$ $(i=1,2,3)$ からなるデータ集合が与えられた場合の主成分分析について，以下の問いに答えよ．ここで，$\boldsymbol{r}_1=(-1,-1)^T$，$\boldsymbol{r}_2=(0,0)^T$，$\boldsymbol{r}_3=(1,1)^T$ とし，T は転置を示す.

(1) データ集合における x_i の分散 σ_x^2，y_i の分散 σ_y^2，ならびに，x_i と y_i の共分散 σ_{xy} を求めることで，分散共分散行列 $\Sigma=\begin{pmatrix}\sigma_x^2 & \sigma_{xy}\\ \sigma_{xy} & \sigma_y^2\end{pmatrix}$ を求めよ．得られた分散共分散行列 Σ の固有値 λ_1，λ_2 と対応する固有ベクトル $\boldsymbol{u}_1$，$\boldsymbol{u}_2$ を求めよ．なお，$\lambda_1\geq\lambda_2$ とする．

(2) データ集合を主成分の空間に変換し，変換後のデータ点 $\tilde{\boldsymbol{r}}_1$，$\tilde{\boldsymbol{r}}_2$，$\tilde{\boldsymbol{r}}_3$ を求めよ．変換前の座標軸と変換後の座標軸（第１主成分，第２主成分の各軸）を用いて，変換後のデータ点 $\tilde{\boldsymbol{r}}_1$，$\tilde{\boldsymbol{r}}_2$，$\tilde{\boldsymbol{r}}_3$ を図示せよ．次元削減が可能であるかどうかを理由を付して述べよ.

(3) 変換後のデータ点 $\tilde{\boldsymbol{r}}_1$，$\tilde{\boldsymbol{r}}_2$，$\tilde{\boldsymbol{r}}_3$ の第１主成分のみが与えられる場合を考える．第１主成分のみを持つ変換後のデータ点を元の空間に逆変換せよ．この逆変換により，変換前のデータ集合 $\boldsymbol{r}_1,\boldsymbol{r}_2,\boldsymbol{r}_3$ を再現可能であるかどうかを理由を付して述べよ.