Machine Learning - Theoretical exercise 1

Téo Bouvard

January 10, 2020

Problem 1

a) According to the sum rule,

$$\begin{aligned}
\iint_X \rho(x)\,dx &= 1 \\
\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \rho(x_1,x_2)\,dx_1dx_2 &= 1 \\
\int_{a_2}^{b_2}\int_{a_1}^{b_1} c\,dx_1dx_2 &= 1 && \text{p.d.f is zero outside of these bounds} \\
c(b_1-a_1)(b_2-a_2) &= 1 \\
\frac{1}{(b_1-a_1)(b_2-a_2)} &= c
\end{aligned}$$

b) Using the formula to compute the expected value,

$$\begin{aligned}
\mathrm{E}(x) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rho(x_1,x_2)\,dx_1dx_2 \\
&= c\int_{a_2}^{b_2}\int_{a_1}^{b_1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} dx_1dx_2 \\
&= c\int_{a_2}^{b_2} \begin{bmatrix} \frac{b_1^2-a_1^2}{2} \\ x_2(b_1-a_1) \end{bmatrix} dx_2 \\
&= c\begin{bmatrix} \frac{b_1^2-a_1^2}{2}(b_2-a_2) \\ \frac{b_2^2-a_2^2}{2}(b_1-a_1) \end{bmatrix}
\end{aligned}$$

Using $a^2-b^2=(a-b)(a+b)$, we can factor factor this expression as

$$\mathrm{E}(x) = c\begin{bmatrix} \frac{(b_1-a_1)(b_1+a_1)(b_2-a_2)}{2} \\ \frac{(b_1-a_1)(b_2+a_2)(b_2-a_2)}{2} \end{bmatrix}$$

Substituting c with the value computed in the previous question, we can simplify it further as

$$\begin{aligned}
\mathrm{E}(x) &= \frac{1}{(b_1-a_1)(b_2-a_2)} \begin{bmatrix} \frac{(b_1-a_1)(b_1+a_1)(b_2-a_2)}{2} \\ \frac{(b_1-a_1)(b_2+a_2)(b_2-a_2)}{2} \end{bmatrix} \\
&= \frac{1}{2}\begin{bmatrix} b_1+a_1 \\ b_2+a_2 \end{bmatrix}
\end{aligned}$$

DO SKETCH

c) We compute the covariance matrix.

$$\begin{aligned}\mathrm{Cov}(x) &= \mathrm{E}((x-\mu)(x-\mu)^T) \\ &= \mathrm{E}(xx^T) - \mu\mu^T\end{aligned}$$

Let's compute each term independently.

$$\begin{aligned}\mathrm{E}(xx^T) &= \mathrm{E}(\begin{bmatrix} x_1^2 & x_1x_2 \\ x_1x_2 & x_2^2\end{bmatrix}) \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \begin{bmatrix} x_1^2 & x_1x_2 \\ x_1x_2 & x_2^2\end{bmatrix} \rho(x_1,x_2)\, dx_1dx_2 \\ &= c\int_{a_2}^{b_2}\int_{a_1}^{b_1} \begin{bmatrix} x_1^2 & x_1x_2 \\ x_1x_2 & x_2^2\end{bmatrix} dx_1dx_2 \\ &= c\int_{a_2}^{b_2} \begin{bmatrix} \frac{b_1^3-a_1^3}{3} & x_2\frac{b_1^2-a_1^2}{2} \\ x_2\frac{b_1^2-a_1^2}{2} & x_2^2(b_1-a_1)\end{bmatrix} dx_2 \\ &= c\begin{bmatrix} \frac{b_1^3-a_1^3}{3}(b_2-a_2) & \frac{b_2^2-a_2^2}{2}\frac{b_1^2-a_1^2}{2} \\ \frac{b_2^2-a_2^2}{2}\frac{b_1^2-a_1^2}{2} & \frac{b_2^3-a_2^3}{3}(b_1-a_1)\end{bmatrix}\end{aligned}$$

Substituting c by its value and using $a^3-b^3=(a-b)(a^2+ab+b^2)$,

$$\begin{aligned}\mathrm{E}(xx^T) &= \frac{1}{(b_1-a_1)(b_2-a_2)}\begin{bmatrix} \frac{(b_1-a_1)(a_1^2+a_1b_1+b_1^2)(b_2-a_2)}{3} & \frac{(b_2-a_2)(b_2+a_2)(b_1-a_1)(b_1+a_1)}{4} \\ \frac{(b_2-a_2)(b_2+a_2)(b_1-a_1)(b_1+a_1)}{4} & \frac{(b_2-a_2)(a_2^2+a_2b_2+b_2^2)(b_1-a_1)}{3}\end{bmatrix} \\ &= \begin{bmatrix} \frac{a_1^2+a_1b_1+b_1^2}{3} & \frac{(b_2+a_2)(b_1+a_1)}{4} \\ \frac{(b_2+a_2)(b_1+a_1)}{4} & \frac{a_2^2+a_2b_2+b_2^2}{3}\end{bmatrix}\end{aligned}$$

$$\begin{aligned}\mu\mu^T &= \frac{1}{2}\begin{bmatrix} a_1+b_1 \\ a_2+b_2\end{bmatrix} \times \frac{1}{2}\begin{bmatrix} a_1+b_1 & a_2+b_2\end{bmatrix} \\ &= \frac{1}{4}\begin{bmatrix} (a_1+b_1)^2 & (a_1+b_1)(a_2+b_2) \\ (a_1+b_1)(a_2+b_2) & (a_2+b_2)^2\end{bmatrix}\end{aligned}$$

Using the previous results to find the covariance,

$$\begin{aligned}\mathrm{Cov}(x) &= \mathrm{E}(xx^T) - \mu\mu^T \\ &= \begin{bmatrix} \frac{a_1^2+a_1b_1+b_1^2}{3} & \frac{(b_2+a_2)(b_1+a_1)}{4} \\ \frac{(b_2+a_2)(b_1+a_1)}{4} & \frac{a_2^2+a_2b_2+b_2^2}{3}\end{bmatrix} - \frac{1}{4}\begin{bmatrix} (a_1+b_1)^2 & (a_1+b_1)(a_2+b_2) \\ (a_1+b_1)(a_2+b_2) & (a_2+b_2)^2\end{bmatrix} \\ &= \frac{1}{12}\begin{bmatrix} 4(a_1^2+a_1b_1+b_1^2)-3(a_1+b_1)^2 & 3(b_2+a_2)(b_1+a_1)-3(b_2+a_2)(b_1+a_1) \\ 3(b_2+a_2)(b_1+a_1)-3(b_2+a_2)(b_1+a_1) & 4(a_2^2+a_2b_2+b_2^2)-3(a_2+b_1)^2\end{bmatrix} \\ &= \frac{1}{12}\begin{bmatrix} a_1^2-2a_1b_1+b_1^2 & 0 \\ 0 & a_2^2-2a_2b_2+b_2^2\end{bmatrix} \\ &= \frac{1}{12}\begin{bmatrix} (a_1-b_1)^2 & 0 \\ 0 & (a_2-b_2)^2\end{bmatrix}\end{aligned}$$

d) $\mathrm{Cov}(x) = \begin{bmatrix} \sigma_{x_1}^2 & \sigma(x_1,x_2) \\ \sigma(x_2,x_1) & \sigma_{x_2}^2\end{bmatrix}$, so identical elements on the diagonal $\implies \sigma_{x_1} = \sigma_{x_2}$. In other terms, x_1 and x_2 spread in a similar manner around their respective expected value.
In this problem, having identical elements on the diagonal would mean that $a_1-b_1 = a_2-b_2$, or that the probability density function is a square.

e) $\mathrm{Cov}(x)$ diagonal $\implies \sigma(x_1,x_2)=0$ and $\sigma(x_2,x_1)=0$, which means that x_1 and x_2 are uncorrelated.

Problem 2

a) We first find the eigenvalues by computing the characteristic equation.

$$\begin{aligned} \det(\Sigma - \lambda I) &= 0 \\ \begin{vmatrix} 5-\lambda & 3 \\ 3 & 5-\lambda \end{vmatrix} &= 0 \\ (5-\lambda)^2 &= 3^2 \end{aligned}$$

Which implies

$$\begin{cases} 5-\lambda_1 &= -3 \\ 5-\lambda_2 &= 3 \end{cases}$$
$$\begin{cases} \lambda_1 &= 8 \\ \lambda_2 &= 2 \end{cases}$$

Let $u_1 = \begin{bmatrix} e_{11} \\ e_{21} \end{bmatrix}$ and $u_2 = \begin{bmatrix} e_{12} \\ e_{22} \end{bmatrix}$ be the two eigenvectors of Σ.

$$\begin{aligned} \Sigma u_1 &= \lambda_1 u_1 \\ (\Sigma - \lambda_1 I) u_1 &= 0 \\ \begin{bmatrix} -3 & 3 \\ 3 & -3 \end{bmatrix} \begin{bmatrix} e_{11} \\ e_{21} \end{bmatrix} &= 0 \end{aligned}$$

Which leads to the following system of equations

$$\begin{cases} -3e_{11} + 3e_{21} = 0 \\ 3e_{11} - 3e_{21} = 0 \end{cases} \implies e_{11} = e_{21} \implies u_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

We use the same process to find u_2.

$$\begin{aligned} \Sigma u_2 &= \lambda_2 u_2 \\ (\Sigma - \lambda_2 I) u_2 &= 0 \\ \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} e_{12} \\ e_{22} \end{bmatrix} &= 0 \end{aligned}$$

Which leads to the following system of equations

$$\begin{cases} 3e_{12} + 3e_{22} = 0 \\ 3e_{12} + 3e_{22} = 0 \end{cases} \implies e_{12} = -e_{22} \implies u_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Therefore, we have

$$\begin{aligned} \Phi &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ \Lambda &= \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix} \end{aligned}$$

We can check our results by verifying that $\Phi\Lambda\Phi^T = \Sigma$.

$$\begin{aligned}
\Phi\Lambda\Phi^T &= \frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}8 & 0\\ 0 & 2\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\\
&= \frac{1}{2}\begin{bmatrix}8 & 2\\ 8 & -2\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\\
&= \frac{1}{2}\begin{bmatrix}10 & 6\\ 6 & 10\end{bmatrix}\\
&= \begin{bmatrix}5 & 3\\ 3 & 5\end{bmatrix}\\
&= \Sigma
\end{aligned}$$

Let v_1 and v_2 be the principal axes of the probability density function. According to the previous decomposition, we have $v_1 = \sqrt{\lambda_1}u_1$ and $v_2 = \sqrt{\lambda_2}u_2$.

$$\begin{aligned}
v_1 &= 2\sqrt{2}\begin{bmatrix}1\\ 1\end{bmatrix}\\
v_2 &= \sqrt{2}\begin{bmatrix}1\\ -1\end{bmatrix}
\end{aligned}$$

b) DO SKETCH