Machine Learning - Theoretical exercise 1

Téo Bouvard

January 10, 2020

1 Problem 1

a) According to the sum rule, we have

$$\int_X \rho(x)\, dx = 1$$

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \rho(x_1, x_2)\, dx_1 dx_2 = 1$$

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2} c\, dx_1 dx_2 = 1 \qquad \text{p.d.f is zero outside of these bounds}$$

$$c(b_1 - a_1)(b_2 - a_2) = 1$$

$$\frac{1}{(b_1 - a_1)(b_2 - a_2)} = c$$

b) Using the formula to compute the expected value,

$$\begin{aligned}
\mathrm{E}(x) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rho(x_1, x_2)\, dx_1 dx_2 \\
&= c \int_{a_2}^{b_2}\int_{a_1}^{b_1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} dx_1 dx_2 \\
&= c \int_{a_2}^{b_2} \begin{bmatrix} \frac{b_1^2 - a_1^2}{2} \\ x_2(b_1 - a_1) \end{bmatrix} dx_2 \\
&= c \begin{bmatrix} \frac{b_1^2 - a_1^2}{2}(b_2 - a_2) \\ \frac{b_2^2 - a_2^2}{2}(b_1 - a_1) \end{bmatrix}
\end{aligned}$$

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.

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

We compute each term independently.

$$
\mathrm{E}(xx^T) = \mathrm{E}(\begin{bmatrix}x_1^2 & x_1x_2\\ x_1x_2 & x_2^2\end{bmatrix})
$$

Not so sure about that ?

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

d) identical diagonal?

e) diagonal ?

2 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

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

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 $v_2 = \sqrt{\lambda_2} u_2$.

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

b) DO SKETCH