תרגיל 1

שאלה 1.

סעיף א'.

מקרה בדיד:

$$\begin{aligned}\mathbb{E}[XY] &= \sum_{x\in\Omega_1, y\in\Omega_2} xy\cdot P(X=x\wedge Y=y)\\ &\underbrace{=}_{X,Y \text{ are independant}} \sum_{x\in\Omega_1, y\in\Omega_2} xy\cdot P(X=x)\cdot P(Y=y)\\ &= \sum_{x\in\Omega_1}\sum_{y\in\Omega_2} xy\cdot P(X=x)\cdot P(Y=y)\\ &= \sum_{x\in\Omega_1} x\cdot P(X=x)\cdot \mathbb{E}[Y] = \mathbb{E}[X]\cdot\mathbb{E}[Y]\end{aligned}$$

מקרה רציף:

$$\begin{aligned}\mathbb{E}[XY] &= \int_{\mathbb{R}^2} xy\cdot f_{XY}(x,y)\ d(x,y)\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\cdot f_X(x)f_Y(Y)\ dx\ dy\\ &= \int_{-\infty}^{\infty} x\cdot f_X(x)\cdot\mathbb{E}[Y]\ d = \mathbb{E}[X]\cdot\mathbb{E}[Y]\end{aligned}$$

סעיף ב'.

נגדיר $X\sim N(0,1)$, X^2 . נשים לב ששניהם תלויים, אך נקבל:

$$Cov(X,X^2) = \underbrace{\mathbb{E}[X^3]}_{=0 \text{ due to symmetry}} - \underbrace{\mathbb{E}[X]}_{=\mu=0}\mathbb{E}[X^2] = 0$$

כלומר הם בלתי מתואמים.

סעיף ג'.

יהיו X,Y משתנים מקריים בלתי תלויים המייצגים הטלת מטבע ($P(X=0)=P(X=1)=P(Y=0)=P(Y=1)=\frac{1}{2}$). נגדיר משתנה שלישי: $Z=X\oplus Y$, כלומר XOR על התוצאות של X ו- Y . נראה ש- Z בלתי תלוי ב- X,Y :

$$P(Z=i,X=j)=P(X\oplus Y=i,X=j)$$
$$=\frac{1}{2}(P(X\oplus 1=i,X=j\mid Y=1)+P(X\oplus 0=i,X=j\mid Y=0))$$
$$\implies P(Z=0,X=0)=\frac{1}{2}(P(X\oplus 1=0,X=0\mid Y=1)+P(X\oplus 0=0,X=0\mid Y=0))$$
$$=\frac{1}{2}(P(X=1,X=0\mid Y=1)+P(X=0\mid Y=1))$$
$$=\frac{1}{2}\left(0+\frac{1}{2}\right)=\frac{1}{4}=P(Z=0)\cdot P(X=0)$$

בדומה עבור שאר המקרים ו- Y.
אבל, נשים לב שהם תלויים ביחד: נניח כי ידוע $X=i,Y=j$, בפרט ניתן לדעת את $Z=i\oplus j$.
בדומה, נניח כי ידוע $X=i,Z=j$, נקבל כי $i\oplus Y=j \implies Y=i\oplus j$ ובדומה ל- X.

סעיף ד'.

$$\mathbf{z}^\top C\mathbf{z}=\mathbf{z}^\top\left(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^\top\right)\mathbf{z}$$
$$=\sum_{i=1}^{n}\mathbf{z}^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{z}=\sum_{i=1}^{n}(\mathbf{z}^\top\mathbf{x}_i)(\mathbf{x}_i^\top\mathbf{z})$$
$$\underbrace{=}_{\mathbf{z}^\top\mathbf{x}_i=\mathbf{x}_i^\top\mathbf{z}}\sum_{i=1}^{n}(\mathbf{z}^\top\mathbf{x}_i)^2\geq 0$$

שאלה 2.

סעיף א'.

כאשר $d = 1$ מקבלים התפלגות נורמלית סטנדרטית (חד מימדית), כלומר $\sigma = 1$ ולכן נרצה לחשב:

$$P(-1 < X < 1) = \Phi(1) - \Phi(-1) = 0.8643 - 0.1587 = 0.7056$$

סעיף ב'.

לפי הרמז, נקבל שאם $\|\mathrm{x}\| \leq 1$ אז $\forall 1 \leq i \leq d : |x_i| \leq 1$, כלומר עברנו ממעגל היחידה לקוביית יחידה (חסמנו כל קואורדינטה בין -1 ל- 1 , וזה יוצר קובייה רב מימדית).

אם כן נקבל:

$$\begin{aligned}
P(\|\mathrm{x}\| \leq 1) &\leq P(|\mathrm{x}_1| \leq 1, |\mathrm{x}_2| \leq 1, \ldots, |\mathrm{x}_d| \leq 1) \\
&= \int_{-1}^{1}\int_{-1}^{1}\ldots\int_{-1}^{1} f_{\mathrm{X}}(x_1, x_2, \ldots, x_d)\, dx_d \ldots\, dx_2\, dx_1 \\
&= \int_{-1}^{1}\int_{-1}^{1}\ldots\int_{-1}^{1} \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\|\mathrm{x}\|^2\right) dx_d \ldots\, dx_2\, dx_1 \\
&= \frac{1}{(2\pi)^{d/2}} \int_{-1}^{1}\int_{-1}^{1}\ldots\int_{-1}^{1} \exp\left(-\frac{1}{2}(x_1^2 + x_2^2 + \ldots + x_d^2)\right) dx_d \ldots\, dx_2\, dx_1 \\
&= \frac{1}{(2\pi)^{d/2}} \int_{-1}^{1} \exp\left(-\frac{1}{2}x_1^2\right) \int_{-1}^{1} \exp\left(-\frac{1}{2}x_2^2\right) \ldots \int_{-1}^{1} \exp\left(-\frac{1}{2}x_d^2\right) dx_d \ldots\, dx_2\, dx_1 \\
&= \prod_{i=1}^{d} P(-1 \leq x_i \leq 1) = 0.7056^d \xrightarrow[d\to\infty]{} 0
\end{aligned}$$

שאלה 3.

נפתור משני הכיוונים:

$$Var(y) = Var(\mathrm{w}^\top X) = Var\left(\sum_{i=1}^{d} w_i X_i\right)$$

$$= \sum_{i=1}^{d} Var(w_i X_i) + 2\sum_{i<j} Cov(w_i X_i, w_j X_j)$$

$$= \sum_{i=1}^{d} w_i^2 Cov(X_i, X_i) + 2\sum_{i<j} w_i w_j Cov(X_i, X_j)$$

$$= \sum_{i=1}^{d}\sum_{j=1}^{d} w_i w_j Cov(X_i, X_j)$$

$$\mathrm{w}^\top C \mathrm{w} = \left(\sum_{i=1}^{d} w_i \cdot Cov(X_1, X_i), \ldots, \sum_{i=1}^{d} w_i \cdot Cov(X_d, X_i)\right) \cdot \mathrm{w}$$

$$= \sum_{i=1}^{d} w_i \sum_{j=1}^{d} w_j Cov(X_i, X_j) = Var(y)$$

שאלה 4.

סעיף א'.

מכיוון ש- $|A| = |A^\top|$ לכל מטריצה A , נקבל:

$$\begin{aligned}|A - \lambda I| &= |(A - \lambda I)^\top| = |A^\top - \lambda I^\top| \\ &= |A^\top - \lambda I|\end{aligned}$$

ולכן הפ"א של A ו- $A^\top$ שווים, ובפרט הע"ע שלהם שווים.

סעיף ב'.

מכיוון ש- A לכסינה נקבל שקיימת D אלכסונית ו- P הפיכה כך ש- $A = PDP^{-1}$.

$$\begin{aligned}|A| &= |PDP^{-1}| = |P| \cdot |D| \cdot |P^{-1}| \\ &= |P| \cdot |D| \cdot |P|^{-1} = |D|\end{aligned}$$

הערכים העצמיים של D נמצאים על האלכסון, ולכן נקבל:

$$|D| = \prod_{i=1}^{n} D_{i,i} = \prod_{i=1}^{n} \lambda_i$$

וכיוון של- A ול- D יש אותם ע"ע, נקבל את הנדרש.

שאלה 5.

סעיף א'.

נרשום מחדש את הפונקציה:

$$f(\mathrm{x}) = f(x_1, x_2, \ldots, x_d) = \begin{cases} 0 & \text{if } \sum_{i=1}^{d} w_i \cdot x_i \geq 1 \\ 1 - \sum_{i=1}^{d} w_i \cdot x_i & \text{otherwise} \end{cases}$$

לכל נקודה מחוץ לקבוצה $E = \{\mathrm{x} \quad \mathrm{w}^\top \mathrm{x} = 1\}$ נקבל כי קיימת סביבה פתוחה שלה כך שכל הנקודות בסביבה שייכות לאותו מקרה, כלומר 0 או $1 - \mathrm{w}^\top \mathrm{x}$. במקרים אלו קל לראות כי הפונקציה רציפה והנגזרות החלקיות (לכל רכיב) קיימות:

$$\frac{\partial(1 - \mathrm{w}^\top \mathrm{x})}{\partial x_i} = -w_i$$
$$\frac{\partial 0}{\partial x_i} = 0$$

ולכן בנקודות האלו הפונקציה דיפרנציאבילית. עבור הנקודות על E, נקבל שהיא לא דיפ' כיוון שהנגזרות החלקיות לא קיימות:
תהי נקודה $\mathrm{a} \in E$ ונניח בה"כ $w_i > 0, a_i > 0$. נקבל:

$$\lim_{x_i \to a_i^+} \frac{f(a_1, a_2, \ldots, x_i, \ldots, a_d) - f(\mathrm{a})}{x_i} = \frac{d}{dx_i} 0 = 0$$
$$\lim_{x_i \to a_i^-} \frac{f(a_1, a_2, \ldots, x_i, \ldots, a_d) - f(\mathrm{a})}{x_i} = \frac{d}{dx_i} 1 - \sum_{j \neq i} w_j \cdot a_j - w_i \cdot x_i = -w_i \neq 0$$

מכיוון שהגבולות הכיווניים שונים נקבל שהנגזרת החלקית לא קיימת, בפרט f לא דיפ' בנקודה a. המקרה עבור $w_i < 0$ או $a_i < 0$ דומה. אפשר להניח כי קיים $w_i \neq 0$ כי אחרת הפונקציה הייתה פשוט $f(\mathrm{x}) = 0$ ובפרט היא הייתה דיפ' בכל מקום.

נחשב את הגרדיאנטים לפי הנגזרות החלקיות שחישבנו מקודם:
במקרה ש- $\mathrm{w}^\top \mathrm{x} > 1$ נקבל $\nabla f(\mathrm{x}) = 0$. במקרה $\mathrm{w}^\top \mathrm{x} < 1$ נקבל $\nabla f(\mathrm{x}) = -\mathrm{w}$.

סעיף ב'.

הפונקציה היא הרכבה של פונקציות דיפ', רציפות בכל מקום ולכן היא גם דיפ' בכל מקום (כפל סקלרי זה פשוט סכום של מכפלות)

נחשב גרדיאנט:

$$
\begin{aligned}
\frac{\partial f}{\partial x_i} &= \left(1 + e^{-\sum w_i x_i}\right)' \cdot (-1) \cdot \left(1 + e^{-\sum w_i x_i}\right)^{-2} \\
&= -\left(-\sum w_i x_i\right)' \cdot \left(e^{-\sum w_i x_i}\right) \cdot \frac{1}{\left(1 + e^{-\sum w_i x_i}\right)^2} \\
&= w_i \cdot \left(\frac{1}{f(\mathrm{x})} - 1\right) \cdot f^2(\mathrm{x}) \\
&= w_i \cdot (1 - f(\mathrm{x}))f(\mathrm{x}) \\
\implies \nabla f(\mathrm{x}) &= (w_1(1 - f(\mathrm{x}))f(\mathrm{x}), \ldots, w_d(1 - f(\mathrm{x}))f(\mathrm{x})) \\
&= \mathrm{w} \cdot (1 - f(\mathrm{x}))f(\mathrm{x})
\end{aligned}
$$