سوال ۱: (تجزیه Jordan-Chevalley)

الف) می دانیم همه ی ماتریس ها یک فرم جردن دارند که با انجام یک عملیات خطی ماتریس را به یک ماتریس بالا مثلثی تبدیل می کند و این تبدیل به صورت زیر است. $J = P^{-1}AP$

از آنجایی که این یک ماتریس بالا مثلثی است مقادیر ویژه ی آن به صورت زیر توصیف می شود

$(j_{11}-\lambda)(j_{22}-\lambda)\dots(j_{nn}-\lambda) = 0$

ماتریس J تبدیل ماتریس A است در نتیجه مقادیر ویژه یکسانی دارد

$tr(P^{-1}AP) = tr(PP^{-1}A) = tr(A)$

از طرفی $trace(J)$ را محاسبه می کنیم.

در نتیجه مجموع مقادیر ویژه برابر با $tr(J)$ برابر با $tr(A)$ است

ب) با توجه به فرمول ویژگی چندجمله ای برای ماتریس می دانیم:

$\det(A-\lambda I) = (-1)^n(\lambda-\lambda_1)(\lambda-\lambda_2)\dots(\lambda-\lambda_n)$

که در فرمول بالا $\lambda_1 \dots \lambda_n$ همان مقادیر ویژه ی ماتریس هستند.

حال اگر مقدار λ را برابر با صفر بگذاریم داریم:

$\det(A-0I) = (-1)^n(-1)^n(\lambda_1\dots\lambda_n)$

$\to \det(A) = \lambda_1 \dots \lambda_n$

ج) طبق تعریف معادله ی بردار ویژه داریم:

$ABv = \lambda v$

که v بردار ناصفر است

حال بردار جدید u را به صورت Bv تعریف می کنیم که باز هم ناصفر است.

$BAu = BA(Bv) = B\lambda v = \lambda Bv = \lambda u \to BAu = \lambda u$

در نتیجه این دو مقدار ویژگی یکسانی دارند.

د) طبق ویژگی های ماتریس ترانهاده می دانیم $\det(A^T) = \det(A)$. از طرفی مقدار ویژه با فرمول $\det(A-\lambda I)$ محاسبه می شود.

حال:

$\det(A-\lambda I) = \det(A-\lambda I)^T = \det(A^T-\lambda I^T)$

$= \det(A^T-\lambda I)$ ← ترانهاده ی ماتریس همانی دوباره ماتریس همانی است

$\longrightarrow \det(A-\lambda I) = \det(A^T-\lambda I)$

در نتیجه این دو مقدار ویژه های یکسانی دارند.

سوال ۲.

الف)

$$Cov(X,Y|Z) = E[(X - E(X|Z))(Y - E(Y|Z))|Z]$$

$$Var(X) = Var(\underbrace{E[X|Y]}_{A}) + E(\underbrace{Var[X|Y]}_{B})$$

طبق تعریف Var می‌دانیم:

$$B = Var[X|Y] = E(X^2|Y) - E(X|Y)^2$$

$$\rightarrow B = E(X^2|Y) - A^2$$

$$\rightarrow E(B) = E(E(X^2|Y)) - E(A^2)$$

$$var(A) = E(A^2) - (EA)^2 = E(A^2) - E(E(X|Y))^2$$

$$\rightarrow E(B) + Var(A) = E(E(X^2|Y)) - \cancel{E(A^2)} + \cancel{E(A^2)} - E(E(X|Y))^2$$

$$E(E(X|Y)) = \int_{-\infty}^{+\infty} E(X|Y=y) f_Y(y) dy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x \underbrace{f_{X|Y}(x|y) f_y}_{f_{x,y}} dx dy$$

$$\rightarrow \int_{-\infty}^{+\infty} x \underbrace{\int_{-\infty}^{+\infty} f_{(x,y)}(x,y) dy}_{f(x)} dx \rightarrow \int_{-\infty}^{+\infty} x f(x) dx = E(X)$$

$$\rightarrow E(E(X|Y)) = E(X)$$

$$\rightarrow E(B) + Var(A) = E(X^2) - E(X)^2 = Var(X)$$

ب)

$$Cov(X,Y|Z) = E[XY|Z] - E[X|Z]E[Y|Z]$$

فرمول گفته شده برای $Cov(X,Y|Z)$ را باز می‌کنیم. می‌دانیم $E(X_1 + \dots X_n) = EX_1 + \dots EX_n$

$$E(XY|Z] + E(E[X|Z)E[Y|Z]|Z) - E(XE(Y|Z)|Z) - E(YE(X|Z)|Z)$$

$$= E(XY|Z] + E(X|Z)\cancel{E(Y|Z)} - \cancel{E(X|Z)E(Y|Z)} - E(Y|Z)E(X|Z)$$

$$= E(XY|Z) - E(Y|Z)E(X|Z)$$

ج)

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

$$① \quad Cov(E(X|Z), E(Y|Z)) = E(E(X|Z)E(Y|Z)) - \underbrace{E(E(X|Z))}_{E(X)} \underbrace{E(E(Y|Z))}_{E(Y)}$$

$$② \quad E(Cov(X,Y|Z)) = \underbrace{E(E(XY|Z))}_{E(XY)} - E(E(X|Z)E(Y|Z))$$

$$\xrightarrow{①+②} E(XY) - E(X)E(Y)$$

سوال ۳.

الف)

$$\frac{\partial x^T A x}{\partial x} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j = f(x)$$

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k}\left[\underbrace{\sum_{i\neq k}\sum_{j\neq k} a_{ij}}_{0} + \sum_{i\neq k} a_{ik} x_i x_k + \sum_{j\neq k} a_{kj} x_k x_j + a_{kk} x_k^2\right]$$

$$\longrightarrow \sum_{i\neq k} a_{ik} x_i + \sum_{j\neq k} a_{kj} x_j + 2a_{kk}x_k^2 = \sum_{i=1}^{n} a_{ik}x_i + \sum_{j=1}^{n} a_{kj}x_j$$

$$= 2\sum_{i=1}^{n} a_{ik} x_i$$ اگر ماتریس A قرینه باشد

که این مقدار برابر ضرب ستون k ام از ماتریس A در X^T است در نتیجه اگر این مشتق را به ازای تمامی مقادیر x حساب کنیم $2X^TA$ بدست می‌آید.

ب) در این جا هر کدام از x های برابر است $[x_1\ x_2\ \dots\ x_n] \to$ ماتریس X

$$\frac{\partial(tr(X^TAX))}{\partial X} \Rightarrow \begin{bmatrix} x_1^T\\ x_2^T\\ \vdots\\ x_n^T\end{bmatrix}\begin{bmatrix} a_{11} & & \\ & & \\ & & \end{bmatrix}$$

برای ساخته شدن یک درایه قطری x_i^T در A ضرب شده و نتیجه‌ی آن در x_i ضرب می‌شود ⟵

$$tr(X^TAX) = \sum_{i=1}^{n} x_i A x_i$$

در صورتی که A متقارن نباشد $\partial \frac{xAx}{\partial x}$ از قسمت الف می‌دانیم برابر است با

$$\sum_{i=1}^{n} a_{ik}x_i + \sum_{j=1}^{n} a_{kj}x_j = x^T(A+A^T)$$

حالا به ازای تمامی $x_1 \dots x_n$ های ماتریس X این مقدار را حساب می‌کنیم که برابر می‌شود با

$$X^T(A+A^T)$$

سوال ۴. با مشتق گرفتن از CDF، pdf را حساب می‌کنیم

$X \sim U(0,1)$

الف)

$$P(\sqrt{X}<x) = P(X<x^2) = x^2 \to \frac{dx^2}{dx} = \boxed{2x}$$

ب)

$$P(X^2<x) = P(X<\sqrt{x}) = \sqrt{x} \to \frac{d\sqrt{x}}{dx} = \boxed{\frac{1}{2x}}$$

سوال ۵:

$P^{-1}MPv = \lambda v \xrightarrow[\text{ضرب می‌کنیم}]{\text{از چپ در } P} M(Pv) = P\lambda v = \lambda(Pv)$

با توجه به تعریف مقدار ویژه داریم
v مقدار ناصفر است
در نتیجه M مقادیر ویژه برابر با λ دارد

(ماتریس P^{-1} به دلیل full rank بودن P می‌توان ساخت)

دوباره با توجه به تعریف مقدار ویژه داریم

$Mv = \lambda v \to P^{-1}Mv = \lambda P^{-1}v$

$\to P^{-1}M(PP^{-1})v = \lambda P^{-1}(PP^{-1})v \to P^{-1}MP(P^{-1}v) = \lambda(P^{-1}v)$

در نتیجه با در نظر گرفتن تعریف مقدار ویژه برای هر کدام به دیگری می‌رسیم.

سوال ۶:

الف) در راستا نشان می‌دهیم این تجزیه وجود دارد و سپس نشان می‌دهیم که یکتاست.

با استقرا ثابت می‌کنیم:

پایه استقرا: برای ماتریس ۱×۱ داریم $a = \sqrt{a}^T$ و L

فرض استقرا: این در نظر می‌گیریم که برای ماتریس $A'_{(n-1)\times(n-1)}$ یک تجزیه به صورت $L'L'^T$ داریم:

با توجه به اینکه ماتریس باید possitive deffinit باشد ماتریس $A_{n\times n}$ را به صورت زیر می‌سازیم:

می‌دانیم $A' = L'L'^T$ است. L_1 را به صورت زیر تعریف می‌کنیم تا A را تجزیه کنیم

$$A = \begin{bmatrix} A' & a \\ a & \alpha \end{bmatrix} \qquad L_1 = \begin{bmatrix} L'^{-1} & 0 \\ 0 & 1 \end{bmatrix}$$

$$\to L_1^{-1}AL_1^{-T} = \begin{bmatrix} L'^{-1} & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} A' & a \\ a & \alpha \end{bmatrix}\begin{bmatrix} L'^{-T} & 0 \\ 0 & 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & L'^{-1}a \\ L'^{-T}a & \alpha \end{bmatrix}}_{B}$$

حالا سعی می‌کنیم $L'^{-1}a$ را حذف کنیم:

$$L_2^{-1}BL_2^{-T} = \begin{bmatrix} I & 0 \\ -L'^{-T}a & 1 \end{bmatrix}\begin{bmatrix} I & L'^{-1}a \\ L'^{-T}a & \alpha \end{bmatrix}\begin{bmatrix} I & -L'^{-1}a \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \alpha - L'^{-T}aL'^{-1}a \end{bmatrix}$$

$$= \begin{bmatrix} I & 0 \\ 0 & \underbrace{\alpha - aA'^{-1}}_{\beta} \end{bmatrix}$$

از آنجا که A possitive deffinit است جمله β را می‌توان به صورت یک x^2 نوشت $\to L_3 = \begin{bmatrix} I & 0 \\ 0 & x \end{bmatrix}$

$\to L_2^{-1}BL_2^{-T} = L_3L_3^T \to L_2^{-1}L_1^{-1}AL_1^{-T}L_2^{-T} = L_3L_3^T$

$\to A = LL^T$ که در آن $L = L_1L_2L_3$

در نتیجه حکم استقرا ثابت شد.

حال باید اثبات کنیم که تجزیه به دست آمده یکتاست برای این منظور از برهان خلف استفاده می‌کنیم و فرض می‌کنیم یکتا نباشد در این صورت یک $L'L'^T$ نیز وجود دارد که با LL^T برابر است حال از سمت راست طرفین را در L'^{-T} و از سمت چپ در L^{-1} ضرب می‌کنیم ←

$$L^T L'^{-T} = L^{-1} L'$$

می‌دانیم سمت چپ این معادله همواره بالا مثلثی و سمت راست همواره پایین مثلثی است در نتیجه این دو نمی‌توانند برابر باشند در نتیجه مقدار این تجزیه یکتاست.

ب) تجزیه‌ی SVD A را به شکل روبرو می‌نویسیم:

$$A = U\Sigma V^T = QR \to R = Q^{-1}U\Sigma V^T$$

حال باید ثابت کنیم که $Q^{-1}U$ نیز یک ماتریس unitary است تا تجزیه‌ی SVD صحیحی برای R با مقدار گفته شده در بالا داشته باشیم. (بقیه‌ی ماتریس‌های Σ و V^T شروط لازم برای SVD را دارند)

برای اثبات از ویژگی این بردارها استفاده می‌کنیم که $X^TX = XX^T = I$

$$\to (Q^{-1}U)^T(Q^{-1}U) = U^T \underbrace{Q^{-T} \underbrace{Q^{-1}}_{Q^T}}_{I} U$$

از طرفی می‌دانیم Q^{-1} و Q^T برابر هست

$$\to U^TU = I \to Q^{-1}U$$ هم یک بردار متعامد است

در نتیجه تفاوت تجزیه‌ی این دو در سمت U است که برای R در یک Q^{-1} نیز ضرب می‌شود

$$A^k = 0 \to (I - \overset{0}{A^k}) = (I-A)(I+A+A^2+\dots+A^{k-1}) = I$$

سه جمله‌ای فرض

سوال ۷:

در نتیجه ماتریس $(I-A)$ در یک ماتریس ضرب شده است در نتیجه معکوس دارد و برابر است با:

$$(I-A)^{-1} = I + A + A^2 + \dots + A^{k-1}$$

سوال ۸. طبق زیر می‌دانیم

الف)

$$f(x,y,z) = f(x,y|z)\, f(z) \to f(x|z)\, f(y|z)\, f(z)$$

$$\begin{cases} \frac{1}{2\pi} e^{\overset{\text{مستقل}}{\frac{-(x-\mu)^2}{2} - \frac{(y-\mu)^2}{2}}} \times 1 & [0,1] \\ 0 & \text{o.w} \end{cases} = f_{M,XY}(t,xy)$$

ب) برای به دست آوردن Map μ به ازای n x,y مختلف:

$$\operatorname*{argmax}_{\mu} f(\mu|x,y) \to \frac{f(\mu,x,y)}{f(x,y)} \to$$

$$\operatorname*{argmax}_{\mu} f(\mu,x,y) \xrightarrow{\text{با توجه به الف}} \left(\frac{1}{2\pi} e^{\frac{-(x_i-\mu)^2-(y_i-\mu)^2}{2}}\right)^n \xrightarrow[\text{به سادگی log را می‌گیریم}]{\text{برای سادگی}} n\log\frac{1}{2\pi} - \sum_{i=1}^{n}\frac{(x_i-\mu)^2+(y_i-\mu)^2}{2}$$

حال از عبارت بالا مشتق می‌گیریم تا μ را به دست آوریم

$$\to 0 = \sum \frac{+2(x_i-\mu)+2(y_i-\mu)}{\cancel{2}} \to 2n\mu = \sum x_i + y_i$$

$$\to \boxed{\mu = \sum_{i=1}^{n}\frac{x_i+y_i}{2n}}$$