Multivariate Probability Review

- Let $\vec{x} = (x_1, \ldots x_p)' \in \mathbb{R}^p$ be a random vector (column) with CDF

$$F(\vec{x}^\circ) = F(\vec{x} \le \vec{x}^\circ) = P(x_1 \le x_1^\circ, \ldots x_p \le x_p^\circ)$$

$$= \int_{-\infty}^{x_1^\circ} \cdots \int_{-\infty}^{x_p^\circ} f(x_1, \ldots x_p)\, dx_p \ldots dx_1 = \int_{-\infty}^{\vec{x}^\circ} f(\vec{x})\, d\vec{x} \quad \text{(continuous case)}$$

$$= \sum_{\vec{x} \le \vec{x}} P(\vec{x} = \vec{x}_0) \quad \text{(discrete case)}$$

our main focus

- Mean & Covariance

  - The mean of $\vec{x}$, denoted $E(\vec{x}), \mu, \mu_x$, is the p-dimensional vector

$$E(\vec{x}) = \int \vec{x} f(\vec{x})\, d\vec{x} = \begin{bmatrix} \int x_1 f(\vec{x})\, d\vec{x} \\ \vdots \\ \int x_p f(x)\, dx \end{bmatrix} = \begin{bmatrix} E(x_1) \\ \vdots \\ E(x_p) \end{bmatrix}$$

Given a $q \times p$ matrix A and $q \times 1$ vector $\vec{b}$

$$E[A\vec{x} + \vec{b}] = A\,E[\vec{x}] + \vec{b} = \begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{q1} & \cdots & a_{q1} \end{bmatrix} \begin{bmatrix} E(x_1) \\ \vdots \\ E(x_p) \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_q \end{bmatrix} \quad \text{(linearity)}$$

- The (auto-)covariance of $\vec{x}$, denoted Σ, Σ_x, is the $p \times p$ matrix

$$\Sigma = E\left[(\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})^T\right] = \begin{bmatrix} E((x_1-\mu_1)^2) & E[(x_1-\mu_1)(x_2-\mu_2)] & \cdots & E[(x_1-\mu_1)(x_p-\mu_p)] \\ \vdots & E[(x_2-\mu_2)^2] & & \vdots \\ \vdots & & \ddots & \\ E[(x_p-\mu_p)(x_1-\mu_1)] & \cdots & \cdots & E[(x_p-\mu_p)^2] \end{bmatrix}$$

$$\Sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)] = \mathrm{Cov}(x_i, x_j) \qquad * \ \mathrm{Cov}(x_i, x_i) = \mathrm{Var}(x_i)$$

$$= \mathrm{Cov}(x_j, x_i) = \Sigma_{ji} \quad \text{(symmetric)}$$

Q: Ask for factors about Σ

Q: What can we say about the eigenvalues of Σ?

A: All positive (non-negative)

NOTE:

- $$\Sigma_x = E\left[(\vec{x}-\vec{\mu})(\vec{x}-\vec{\mu})^T\right] = E\left[\vec{x}\vec{x}^T\right] - \vec{\mu}E\left[\vec{x}^T\right] - \underbrace{E\left[\vec{x}\right]}_{\mu}\vec{\mu}^T + \vec{\mu}\vec{\mu}^T$$
$$= E\left[\vec{x}\vec{x}^T\right] - \vec{\mu}\mu^T = E\left[\vec{x}\vec{x}^T\right] - E[\vec{x}]\,E[\vec{x}]^T$$
- $$E\left[(\vec{x}-\vec{\mu})^T(\vec{x}-\vec{\mu})\right] = E\left[(\vec{x}_1-\mu_1)^2 + \dots + (\vec{x}_p-\vec{\mu}_p)^2\right] = \sum_{i=1}^{p} Var(x_i) = tr\left(\Sigma_x\right) = \text{TOTAL VARIATION}$$
- GIVEN A $q \times p$ MATRIX A AND $q \times 1$ VECTOR $\vec{b}$

$$\Sigma_{A\vec{x}+\vec{b}} = E\left[\left(A\vec{x}+\vec{b}-(A\vec{\mu}+b)\right)\left(A\vec{x}+\vec{b}-(A\vec{\mu}+b)\right)^T\right] = E\left[(A\vec{x}-A\vec{\mu})(A\vec{x}-A\vec{\mu})^T\right]$$
$$= E\left[A(\vec{x}-\vec{\mu})(\vec{x}-\vec{\mu})^T A^T\right] = A\,\Sigma_{\vec{x}}\,A^T \in \mathbb{R}^{q\times q}$$

Ex: MULTIVARIATE NORMAL DIST.

GIVEN A MEAN VECTOR $\vec{\mu} \in \mathbb{R}^p$ AND COVARIANCE MATRIX $\Sigma \in \mathbb{R}^{p\times p}$ (W/ POSITIVE EIGENVALUES), THE MULTIVARIATE NORMAL DISTRIBUTION HAS DENSITY

$$f(\vec{x}) = \underbrace{\left|2\pi\Sigma\right|^{-1/2}}_{(2\pi)^{-p/2}|\Sigma|^{-1/2}} \exp\left(-\frac{1}{2}(x-\vec{\mu})^T\Sigma^{-1}(x-\vec{\mu})\right)$$

SHORTHAND: $\vec{x} \sim N(\vec{\mu}, \Sigma)$

SPECTRAL DECOMPOSITION THM

- ANY SYMMETRIC MATRIX $A \in \mathbb{R}^{p\times p}$ CAN BE WRITTEN AS

$$A = V\Lambda V^T$$

WHERE Λ IS A DIAGONAL MATRIX OF EIGENVALUES OF A $= \begin{bmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_p\end{bmatrix}$

V IS AN ORTHONORMAL MATRIX WHOSE COLUMNS ARE ~~[illegible]~~ NORMALIZED EIGENVECTORS $\left[\vec{v}_1, \dots \vec{v}_p\right]$

$$VV' = I$$

- IF EIGENVALUES OF Σ ARE ALL POSITIVE THEN

1) $\Sigma = V\Lambda V^T$ (THM)

2) $\Lambda^{-1/2} = \begin{bmatrix}\Lambda_{11}^{-1/2} & & 0\\ & \ddots & \\ 0 & & \Lambda_{pp}^{-1/2}\end{bmatrix}$ IS WELL DEFINED

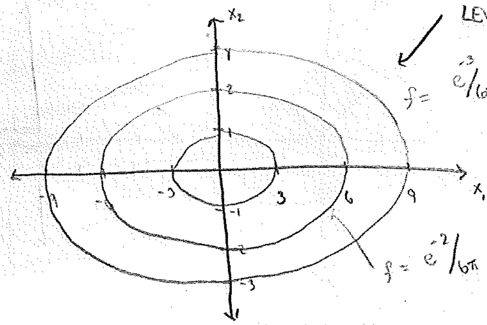
$\rightarrow$ MAY WRITE $\Sigma^{-1/2} = V\Lambda^{-1/2}V^T$

$\Sigma^{-1} = V\Lambda^{-1}V^T$ (DON'T NEED $\lambda_i > 0$ HERE)

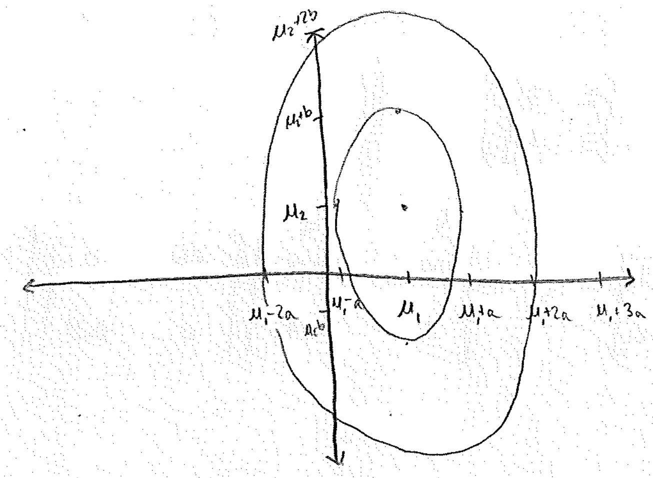
Geometric Interpretation of Σ in $N(\vec{\mu}, \Sigma)$

Ex: $\vec{x} \sim N\left(\vec{0}_2, \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix}\right)$

LEVEL CURVES OF $f(\vec{x}) = \frac{1}{6\pi} e^{-\frac{1}{2}\left(\frac{x_1^2}{9} + \frac{x_2^2}{1}\right)}$

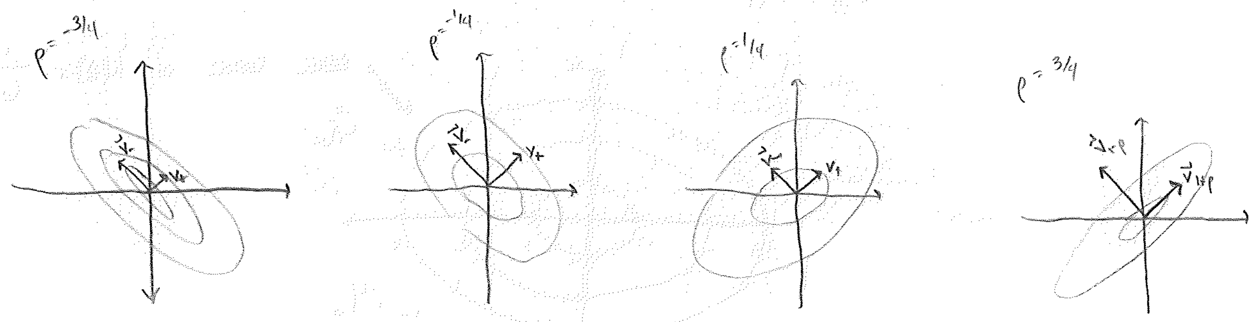


Ex: $\vec{x} \sim N\left(\vec{\mu}, \begin{bmatrix} a^2 & 0 \\ 0 & b^2 \end{bmatrix}\right)$



CONCENTRIC ELLIPSES W/ MAJOR MINOR AXES $\propto (a, b)$

Ex: $N\left(\vec{0}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$ $\quad -1 < \rho < 1$

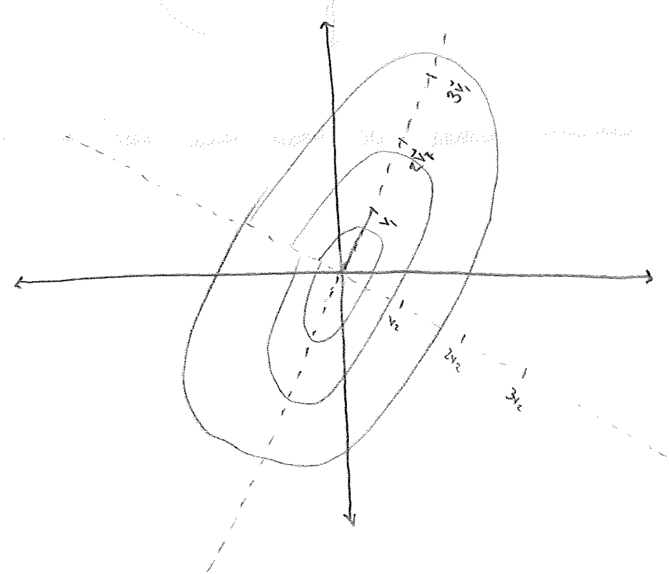


$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Ex:

$N\left(\vec{0}, \begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix}\right)$

$$\begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix} \approx \begin{array}{c} \begin{array}{cc} \vec{v}_1 & \vec{v}_2 \end{array} \\ \begin{bmatrix} .47 & -0.88 \\ .88 & 0.47 \end{bmatrix} \end{array} \begin{bmatrix} 7.61 & 0 \\ 0 & 0.39 \end{bmatrix} \begin{bmatrix} 0.47 & 0.88 \\ -0.88 & 0.47 \end{bmatrix}$$



* $\vec{v}_1, \vec{v}_2$ ... FORM ORTHOGONAL BASIS FOR $\mathbb{R}^2$ w $\vec{v}_1$ POINTING ALONG DIRECTION OF GREATEST VARIABILITY

Ex: SUPPOSE $\vec{x} \sim N(\vec{\mu}, \Sigma)$, $\vec{\mu} \in \mathbb{R}^p$, $\Sigma \in \mathbb{R}^{p \times p}$ HAS POSITIVE E. VALUES.

1) FIND DISTRIBUTION OF $\vec{y} = \Sigma^{-1/2}(\vec{x} - \vec{\mu})$

GIVEN ANY $y^0 \in \mathbb{R}^p$

$$F_y(y^0) = P(\vec{y} \le \vec{y}^0) = P(\Sigma^{-1/2}(\vec{x}-\vec{\mu}) \le \vec{y}^0)$$

$$= \int_{\{\vec{x}: \Sigma^{-1/2}(\vec{x}-\vec{\mu}) \le \vec{y}_0\}} |2\pi\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(x-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})\right)$$

$$[\Sigma^{-1/2}(\vec{x}-\mu)]^T[\underbrace{\Sigma^{-1/2}(\vec{x}-\vec{\mu})}_{\vec{y}}] \qquad \frac{d\vec{y}}{d\vec{x}} = \Sigma^{-1/2}$$

$$= \int_{\{\vec{y} \le y_0\}} |2\pi \Sigma^{-1/2}| \exp\left(-\frac{1}{2} y^T y\right) |\Sigma^{-1/2}| \, d\vec{y} = \int_{\{\vec{y} \le y_0\}} |2\pi|^{-p/2} \exp\left(-\frac{1}{2}\|y\|^2\right) dy$$

$$\vec{y} \sim N(\vec{0}_p, I_p) \qquad \vec{0}_p = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}\Big\}p, \quad I_p = \underbrace{\begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & 1 \end{bmatrix}}_{p}\Big\}p$$

2) FIND DISTRIBUTION OF $z = (\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu}) = y^T y$

$$P(z \le z_0) = P(y^T y \le z_0) = P(y_1^2 + y_2^2 + \dots + y_p^2 \le z_0) \rightarrow \boxed{z \sim \chi^2(p)}$$

$$= \int_{\{\vec{y}: \|y\|^2 \le z_0\}} |2\pi|^{-p/2} \exp\left(-\frac{1}{2}\|y\|^2\right) d\vec{y}$$

$$= C_p \cdot \int_0^{\sqrt{z_0}} r^{p-1} e^{-\frac{1}{2}\rho^2} \, d\rho$$

$$\rightarrow f_z(z) = C_p z^{\frac{p-1}{2}} e^{-z/2} \cdot \frac{1}{2\sqrt{z}} = C_p z^{\frac{p}{2}-1} e^{-z/2}$$

ESTIMATING $\vec{\mu}$ AND $\vec{\Sigma}$

GENERALLY, WE'LL BE WORKING W/ N IID OBSERVATIONS OF $\vec{x} \sim f$

$$\vec{x}_1, \ldots \vec{x}_N \sim f$$

SAMPLE MEAN & COVARIANCE

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} \vec{x}_i \quad , \quad S = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})(\bar{x}_i - \bar{x})^T = \frac{1}{N}\sum x_i x_i^T - \frac{1}{N}\bar{x}\sum \vec{x}_i^T + \bar{x}\bar{x}^T$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T - \bar{x}\bar{x}^T$$

$$S_{jk} = \frac{1}{N}\sum_{i=1}^{N} (x_{ij} - \bar{x}_j)(\vec{x}_{ik} - \bar{x}_k) = \frac{1}{N}\sum_{i=1}^{N} x_{ij}x_{ik} - \bar{x}_j\bar{x}_k$$

DATA MATRIX

$$X = \begin{bmatrix} \vec{x}_1^T \\ \vdots \\ \vec{x}_N^T \end{bmatrix} \Rightarrow \bar{x} = \frac{1}{N}X^T \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \frac{1}{N} X^T 1_N \qquad 1_N = \left.\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}\right\} N$$

$$S = \frac{1}{N}\sum x_i x_i^T - \bar{x}\bar{x}^T$$

$$= \frac{1}{N} X^T X - \frac{1}{N^2} X^T 1_N 1_N^T X$$

$$= \frac{1}{N} X^T \underbrace{\left[I_N - \frac{1}{N} 1_N 1_N^T \right]}_{\text{CENTERING MATRIX}} X$$

QUESTIONS:

WHAT HAPPENS TO S IF $J > N$?

$$\text{RANK}(S) \leq N \Rightarrow S \text{ IS NOT FULL RANK}$$

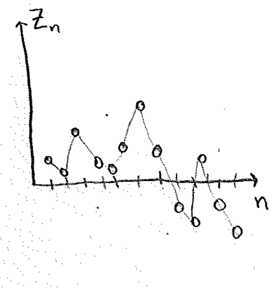
$$\Rightarrow S \text{ HAS AT LEAST ONE EIGENVALUE} = 0$$

$$\Rightarrow \det(S) = 0$$

Aside on Precision Matrices Σ^{-1}

Ex: Gaussian Random Walk

$$X_1, \ldots X_n \overset{iid}{\sim} N(0,1) \qquad X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \sim N(\vec{0}, I_n)$$



" $Z_0 = 0$ "

$$Z_1 = X_1$$
$$Z_2 = X_1 + X_2$$
$$Z_3 = X_1 + X_2 + X_3$$
$$\vdots$$
$$Z_n = X_1 + \ldots + X_n$$

$$\vec{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \vdots \\ 1 & 1 & 1 & \ddots \\ \vdots & & \ddots & 0 \\ 1 & \cdots & & 1 \end{bmatrix}}_{A} \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} = A\vec{x}$$

$$\Sigma_Z = AIA^T = AA^T = \begin{bmatrix} 1 & 1 & \cdots & & & 1 \\ 1 & 2 & 2 & \cdots & & 2 \\ 1 & 2 & 3 & 3 & 3 \cdots & 3 \\ \vdots & 2 & 3 & 4 & 4 \; 4 \cdots & 4 \\ & \vdots & & 4 & \ddots & \vdots \\ & 2 & & & \ddots & n-1 \\ 1 & & & & n-1 & n \end{bmatrix}$$ ← NOT SPARSE

$$\Sigma_Z^{-1} = \begin{bmatrix} -2 & -1 & & & \\ -1 & -2 & -1 & & 0 \\ & -1 & \ddots & \ddots & \\ & & \ddots & \ddots & -1 \\ & 0 & & -1 & -2 \end{bmatrix}$$

SPARSE

- In multivariate normal setting

$\Sigma^{-1}_{ij} = 0 \iff Z_i$ and Z_j are independent given all other components of $\vec{Z}$

- Does not generalize outside Normal setting!

Note:

1) $Z_3 = Z_2 + X_3$ ← independent
$Z_1 = Z_2 - X_2$ $\Rightarrow Z_3 \perp\!\!\!\perp Z_1 \mid Z_2$

2) Σ^{-1} and conditional independence are important in Linear Discriminant Analysis (LDA), a dimension reduction technique for classification.

(See Hastie Ch. 4)