Last time:

- Marginal
- conditional

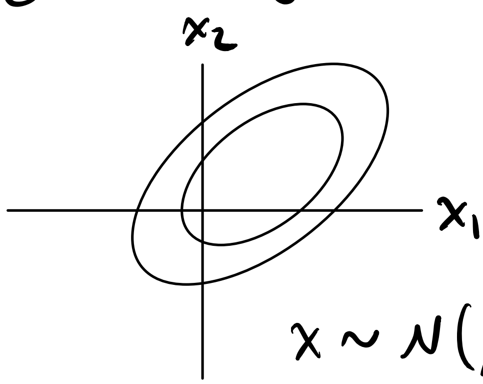
Multivariate Gaussian

Jointly Gaussian.

$x = (x_1 \dots x_n)$

is Jointly Gaussian if any linear comb is Gaussian

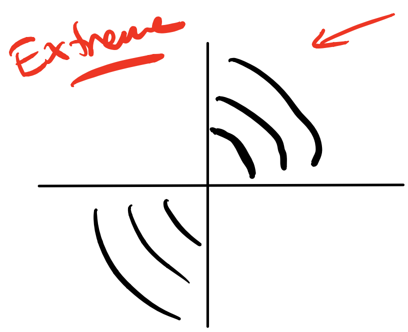
$$\mathbb{R} \ni \underbrace{y}_{\text{Gaussian}} = a^T \underbrace{x}_{\text{jointly Gaussian}} \qquad x \in \mathbb{R}^n$$

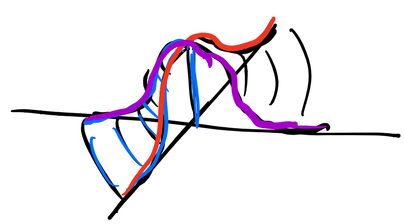


$x \sim N(\underbrace{\mu}_{\text{mean}}, \underbrace{\Sigma}_{\text{Covariance}})$

If you slice the distribution along any affine space (and renormalize) $\rightarrow$ Gaussian

Possible to have Gaussian marginal distributions but not be jointly Gaussian





there is some way to slice the distribution so you don't get a Gaussian (even tho the marginals were Gaussian)

the vector has Gaussian
Marginals under any
coordinate transformation
on x.

Joint Distributions & Conditional Distributions
marginal distributions

x_1, x_2 random variables ✓ *

$$p(x_1, x_2) = \underbrace{p(x_1|x_2)}_{\text{conditional of } x_1 \text{ given } x_2}\ \underbrace{p(x_2)}_{\text{marginal dist } x_2} = p(x_2|x_1)\,p(x_1)$$

Independence:

x_1 & x_2 if ✓

$$p(x_1|x_2) = p(x_1)$$

knowing x_2 doesn't give you any information about x_1

$$p(x_2|x_1) = p(x_2)$$ ✓

Another Defn of Ind:

$$p(x_1, x_2) = p(x_1)\,p(x_2)$$

Bayes Rule:

$$\boxed{p(x_1|x_2) = \frac{p(x_1)}{p(x_2)}\,p(x_2|x_1)}$$

$$p(x_1|x_2)\,p(x_2) = p(x_2|x_1)\,p(x_1)$$

$$p(x_1|x_2) = \frac{p(x_2|x_1)\,p(x_1)}{p(x_2)}$$

Describing Distributions:

Expected value: $E[f(x)] = \int_x \underbrace{f(x)p(x)}dx$

vector... / scalar

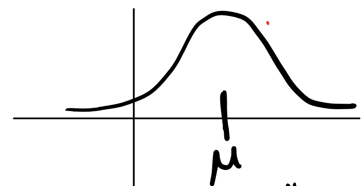
x: vector or scalar

on average, the value of f(x) you expect to see.

$x \in \mathbb{R}$

mean $f(x) = x$.

$$\mu = E[x] = \int_x x\,p(x)\,dx.$$

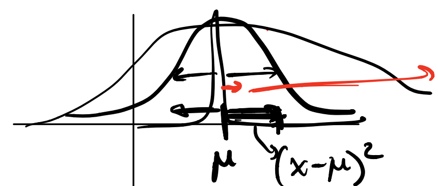


"center of mass"

covariance $f(x) = (x-\mu)^2$

$$\sigma^2 = E[(x-\mu)^2] = \int_x (x-\mu)^2 p(x)\,dx$$

σ^2: always positive



"how far away from the mean is most of the mass"

$x \in \mathbb{R}^n$

$\mu \in \mathbb{R}^n$ $\quad \mu = E[x] = \int_x x\,p(x)\,dx$

$$= \int_{x_1} \cdots \int_{x_n} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} p(x_1, \dots x_n)\, dx_1 \cdots dx_n$$

Covariance

$\Sigma \in \mathbb{R}^{n \times n}$

$$\Sigma = E[(x-\mu)(x-\mu)^T] = \int_x (x-\mu)(x-\mu)^T p(x)\,dx$$

$$= \int_{\mathcal{X}} \begin{bmatrix} x_1 - \mu_1 \\ \vdots \\ x_n - \mu_n \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 & \cdots & x_n - \mu_n \end{bmatrix} p(x)\, dx$$

$$= \int_{\mathcal{X}} \begin{bmatrix} (x_1-\mu_1)^2 & \cdots & (x_1-\mu_1)(x_n-\mu_n) \\ \vdots & & \vdots \\ (x_1-\mu_1)(x_n-\mu_n) & \cdots & (x_n-\mu_n)^2 \end{bmatrix} p(x)\, dx.$$

$$\Sigma_{ij} = \int_{\mathcal{X}} (x_i - \mu_i)(x_j - \mu_j)\, p(x)\, dx$$

$x \in \mathbb{R}^2$ $\quad x_1, x_2$ are indep. $\quad \Sigma \in \mathbb{R}^{2\times 2}$

$$\Sigma_{12} = \int_{\mathcal{X}} (x_1-\mu_1)(x_2-\mu_2)\, p(x)\, dx.$$

$$= \int_{x_1}\int_{x_2} (x_1-\mu_1)(x_2-\mu_2)\, p(x_1)p(x_2)\, dx_1\, dx_2$$

$$= \left(\int_{x_1} (x_1-\mu_1)\, p(x_1)\, dx_1\right) \int_{x_2} (x_2-\mu_2)\, p(x_2)\, dx_2$$

$$= \left(\overset{\mu_1}{\int_{x_1} x_1 p(x_1)\, dx_1} - \mu_1 \overset{1}{\int_{x_1} p(x_1)\, dx_1}\right)\left(\mu_2 - \mu_2\right)$$

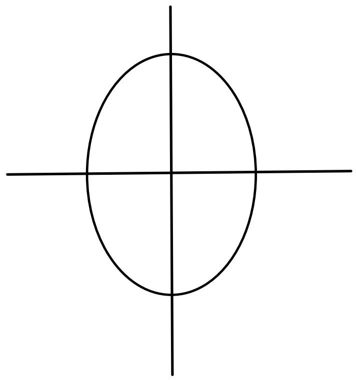
$\Sigma_{12} = 0 \Leftarrow x_1$ & x_2 indep.

In general if x_i & x_j are indep.

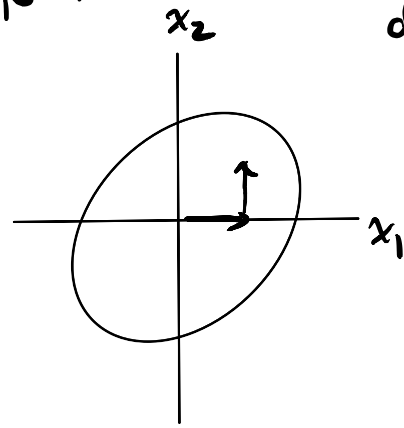
$\Sigma_{ij} = 0$

Multivariate Gaussians

x_i's ind. Σ diag

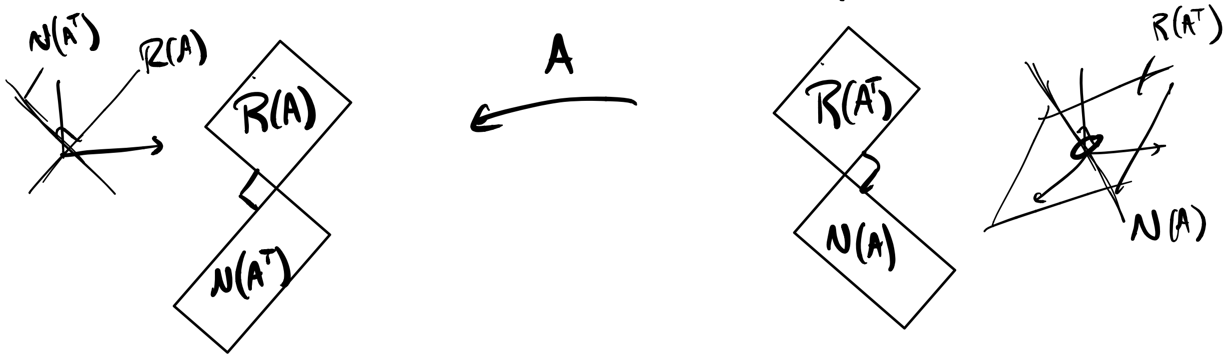


x_i's not ind Σ not diag



Fund. Thm Lin Alg. $A \in \mathbb{R}^{m \times n}$

$\mathbb{R}^m$ CODOMAIN $\quad y = Ax \quad$ $\mathbb{R}^n$ DOMAIN



$$\mathbb{R}^m = R(A) \overset{\perp}{\oplus} N(A^T) \qquad \mathbb{R}^n = R(A^T) \overset{\perp}{\oplus} N(A)$$

SVD:

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^T$$

$$= \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$

basis for $R(A)$ basis for $N(A^T)$ basis $R(A^T)$ basis $N(A^T)$

General system of Eqns:

$$y = Ax \qquad A \in \mathbb{R}^{m \times n}$$

computing an inverse step by step $\longrightarrow$ Gaussian Elimination

Row reduct. oper. E_i : elementary matrices

$E_i A \longrightarrow$ performs a row operation on A.

if A was invertible . .

$$\underbrace{E_k \cdots E_2 (E_1}_{A^{-1}} A) = I$$

for full row rank, A fat

$$E_k \cdots E_1 A = \left[\, I \mid * \,\right]$$

for full col rank, A tall

$$E_k \cdots E_1 A = \begin{bmatrix} I \\ 0 \end{bmatrix}$$

general case $\quad A \in \mathbb{R}^{m\times n}$ $\qquad A = [A_1 \; A_2]$

$$\to E_k \cdots E_1 A = \left[\begin{array}{cc} I & B \\ \hline 0 & 0 \end{array}\right]$$

lin dep rows

columns of A_2 are lin dep on cols of A_1

$$(E_k \cdots E_1)^{-1} = E_1^{-1} \cdots E_k^{-1}$$

$$A = (E_k \cdots E_1)^{-1} (E_k \cdots E_1) A$$

$$= (E_k \cdots E_1)^{-1} \begin{bmatrix} I & B \\ 0 & 0 \end{bmatrix}$$

$$= \left[\begin{array}{c|c} A_1 & M \end{array}\right] \left[\begin{array}{c|c} I & B \\ 0 & 0 \end{array}\right]$$

B is a matrix s.t. $\boxed{A_2 = A_1 B}$

M is a basis for space ⊥ to cols of A

ie. cols of B are coords of A_2 w.r.t. the cols of A_1

$$= [A_1 I + M0, \; A_1 B + M0]$$

$$= [A_1 \; A_2]$$

M depends on $E_k \cdots E_1 \to$ not unique

M spans $N(A^T)$

from this can easily get a basis for

$N(A)$ $\qquad N = \begin{bmatrix} -B \\ I \end{bmatrix}$

$$AN = \underbrace{\begin{bmatrix} A_1 & M \end{bmatrix}} \underbrace{\begin{bmatrix} I & B \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -B \\ I \end{bmatrix}}$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\begin{bmatrix} I \\ B^T \end{bmatrix}$ spans $R(A^T)$ AFTER CLASS

---

cols of A_1 basis for $R(A)$

" M basis for $N(A^T)$

" $\begin{bmatrix} I \\ B^T \end{bmatrix}$ basis for $R(A^T)$

" $\begin{bmatrix} -B \\ I \end{bmatrix}$ basis for $N(A)$

---

finding solution:

$$y = Ax$$

$$y = \begin{bmatrix} A_1 & M \end{bmatrix} \begin{bmatrix} \boxed{I} & B \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \boxed{x_1} \\ x_2 \end{bmatrix} = A_1 (x_1 + B x_2)$$

$y \in R(A_1)$: necessary

$$y = A_1 x_1 \Rightarrow A_1^T y = A_1^T A_1 x_1$$

$$x_1 = (A_1^T A_1)^{-1} A_1^T y$$

$$[A_1\ M] = (E_k \cdots E_1)^{-1}$$

$$E_k \cdots E_1 = \begin{bmatrix} (A_1^T A_1)^{-1} A_1^T \\ \star \end{bmatrix}$$

$$E_k \cdots E_1 [A \mid y] = \left[\begin{bmatrix} (A_1^T A_1)^{-1} A_1^T \\ \star \end{bmatrix} A \,\middle|\, \begin{bmatrix} (A_1^T A_1)^{-1} A_1^T \\ \star \end{bmatrix} y \right]$$

$$= \left[\begin{matrix} (A_1^T A_1)^{-1} A_1^T [A_1\ A_2] \\ \cancel{\star}\, 0\ [\cancel{A_1}\ A_2]\ 0 \end{matrix} \,\middle|\, \begin{matrix} (A_1^T A_1)^{-1} A_1^T y \\ \star\, y \end{matrix} \right]$$

$$= \left[\begin{matrix} I & \underbrace{(A_1^T A_1)^{-1} A_1^T A_2}_{B} \\ 0 & 0 \end{matrix} \,\middle|\, \begin{matrix} x_1 \\ \star y \end{matrix} \right]$$