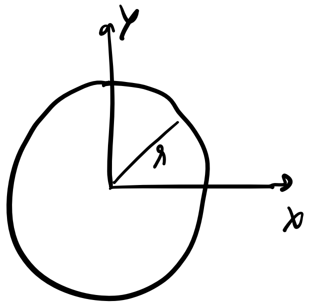
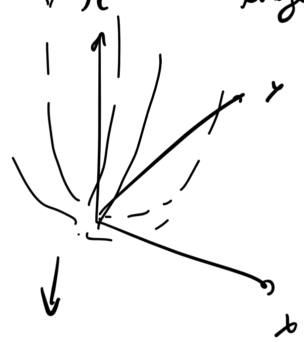
Gaussian Distribution

$x^2+y^2=r^2$



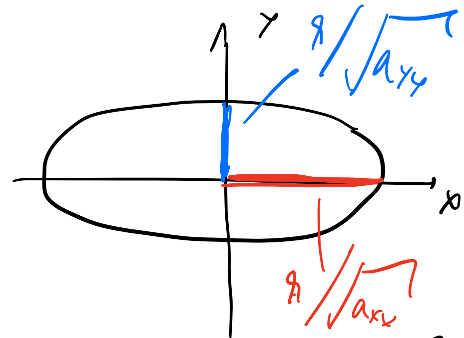
$\equiv$

$r \rightarrow$ height $\equiv$ radius



$\Longrightarrow$

scaling factors



$$a_{xx}x^2+a_{yy}y^2=r^2$$

paraboloid

slice

SLANTED CIRCLE $\equiv$ "turned" circle $\rightarrow$ $a_{xx}x^2+a_{xy}xy+a_{yy}y^2=r^2$

$$\parallel$$

$$[x\ y]\underbrace{\begin{bmatrix} a_{xx} & a_{xy}/2 \\ a_{xy}/2 & a_{yy} \end{bmatrix}}_{A}\begin{bmatrix} x \\ y \end{bmatrix}=r^2$$

has an EIGEN VALUE decomposition $\leftarrow$ simmetric matrix

cross-correlation coefficient

(the reason why ellipse is rotated)

$$A=R^T\begin{pmatrix} \lambda_1 & \emptyset \\ \emptyset & \lambda_2 \end{pmatrix}R$$

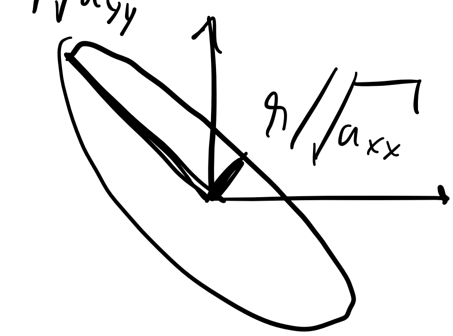
R^T $\downarrow$ eigen vectors

$\parallel$

orthogonal matrix $\rightarrow$ form a ROTATION MATRIX

λ $\downarrow$ eigen values (diagonal matrix)

$\Downarrow$



The ellipse may also be translated by $[x_c, y_c]^T$:

$$A = \left[R\begin{pmatrix}x-x_c\\ y-y_c\end{pmatrix}\right]^T \begin{pmatrix}\lambda_1 & \\ & \lambda_2\end{pmatrix}\left[R\begin{pmatrix}x-x_c\\ y-y_c\end{pmatrix}\right] = r^2$$

(generic equation of an ellipse)

Gaussian Distribution

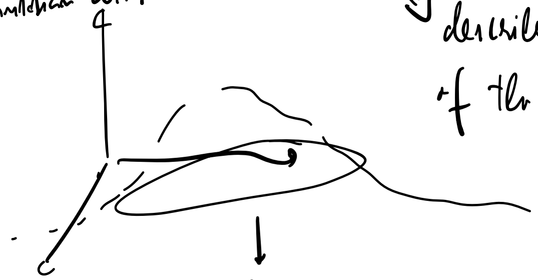
constant / normalizer $\Rightarrow$ ensures that integral of the second part sums up to 1 (defining a probability distribution)

$$p(x) = \mathcal{N}(x;\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \cdot \exp\left[-\frac{1}{2}\underbrace{(x-\mu)^T\Sigma^{-1}(x-\mu)}_{\text{paraboloid}}\right]$$

$\Leftarrow$ mean

covariance $\Downarrow$ describes the shape of the Gaussian distribution (generic symmetric positive definite matrix)

point of highest value of a Gaussian dist.



$x \in \mathbb{R}^2$ 2D ellipse

For example,

$$p(x) = 0.3 \Leftrightarrow$$

$$\Leftrightarrow 0.3 = \frac{1}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

$$\Leftrightarrow \ln(0.3) = \frac{1}{2\pi^{n/2}\sqrt{\det\Sigma}} \cdot \left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right] \longrightarrow \text{ellipsoid!}$$

Gaussian $\longrightarrow$ $\boxed{\text{sum; affine: } Ax+b\text{; conditioning}}$ $\longrightarrow$ Gaussian marginalization

and only require "2" parameters $\begin{pmatrix}\mu \\ \Sigma\end{pmatrix}$

Discrete case → we were using a histogram ....

→ describe how uncertainty we are along a certain direction

⟹ **MOMENT PARAMETRISATION** (w/ ↑↑ large set of samples)

$$\mu = \frac{1}{N}\sum x^{(i)}$$

$$\Sigma = \frac{1}{N}\sum \left(x^{(i)}-\mu\right)\left(x^{(i)}-\mu\right)^T$$

→ parameters obtained experimentally

⇓

assuming that we know that the process is governed by a gaussian distribution

IF $\Sigma = \emptyset$

⇓

each sample is exactly the same as the mean

⇓

EXPECTED VALUE

$$\mu = \int_{\Omega} x\, p(x)\, dx = \mathbb{E}(x)$$ —— 1^{st}-order moment

↳ event space

$$\Sigma = \int_{\Omega} (x-\mu)(x-\mu)^T p(x)\, dx = \mathbb{E}\left[(x-\mu)(x-\mu)^T\right]$$ —— 2^{nd}-order moment

of the distribution

↑ probability density

↳ in a discrete case, a sample is highly likely if it occurs more dense....

more likely (samples more dense)

⇕

CANONICAL PARAMETRIZATION → how certain we are

"stiffness of a string"

- information matrix $\Omega = \Sigma^{-1}$
- information vector $\nu = \Omega\mu$

$$p(x) = \mathcal{N}(x;\nu,\Omega) = \frac{\exp\left[\frac{1}{2}\cdot\nu^T\Omega^{-1}\nu\right]\sqrt{\det(\Omega)}}{(2\pi)^{n/2}} \cdot \exp\left[-\frac{1}{2}x^T\Omega x + x^T\nu\right]$$

!)

$$p(x) = \mathcal{N}(x;\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}\cdot\sqrt{\det\Sigma}}\cdot\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

$(x-\mu)^T \equiv (x^T-\mu^T)$

$$= \frac{1}{(2\pi)^{n/2}\cdot\sqrt{\det\Sigma}}\cdot\exp\left[-\frac{1}{2}(x^T-\mu^T)\cdot\left(\Sigma^{-1}x-\Sigma^{-1}\mu\right)\right] =$$

$$= \frac{1}{(2\pi)^{n/2}\cdot\sqrt{\det\Sigma}}\cdot\exp\left[-\frac{1}{2}x^T\Sigma^{-1}x+\frac{1}{2}x^T\Sigma^{-1}\mu+\frac{1}{2}\mu^T\Sigma^{-1}x-\frac{1}{2}\mu^T\Sigma^{-1}\mu\right]$$

$\frac{1}{2}x^T\nu$ $\quad$ $\frac{1}{2}\nu^T\Omega^{-1}\Lambda x = \frac{1}{2}\nu^T x$

$$= \frac{1}{(2\pi)^{n/2}\cdot 1/\sqrt{\det\Omega}}\cdot\exp\left[-\frac{1}{2}x^T\Omega x+\frac{1}{2}x^T\Omega\Omega^{-1}\nu+\frac{1}{2}(\Omega^{-1}\nu)^T\Omega\, x - \frac{1}{2}(\Omega^{-1}\nu)^T\Omega(\Omega^{-1}\nu)\right] =$$

$$\boxed{\begin{aligned}\Omega &= \Sigma^{-1}\\ \nu &= \Omega\mu\end{aligned}}$$

$\Sigma^T = \Sigma$

symmetric matrix $\to (\Omega^{-1})^T = \Omega^{-1}$

$$= \frac{1}{(2\pi)^{n/2}\cdot 1/\sqrt{\det(\Omega)}}\cdot\exp\left[-\frac{1}{2}x^T\Omega x+\frac{1}{2}x^T\nu+\frac{1}{2}\nu^T(\Omega^{-1})^T\Omega\, x-\frac{1}{2}\nu^T(\Omega^{-1})^T\Omega\,\Omega^{-1}\nu\right]$$

$$= \frac{1}{(2\pi)^{n/2}\cdot 1/\sqrt{\det(\Omega)}}\cdot\exp\left[-\frac{1}{2}x^T\Omega x+\frac{1}{2}x^T\nu+\frac{1}{2}\nu^T x-\frac{1}{2}\nu^T\Omega^{-1}\nu\right]$$

inner product : $x^T\nu = \nu^T x$ ✓

$$= \frac{\exp\left[-\frac{1}{2}\nu^T\Omega^{-1}\nu\right]\sqrt{\det(\Omega)}}{(2\pi)^{n/2}}\cdot\exp\left[-\frac{1}{2}x^T\Omega x+x^T\nu\right]$$

qed

$$x = \begin{pmatrix}x_a\\ x_b\end{pmatrix}\quad \mu = \begin{pmatrix}\mu_a\\ \mu_b\end{pmatrix}\quad \nu = \begin{pmatrix}\nu_a\\ \nu_b\end{pmatrix}$$

$$\Sigma = \begin{pmatrix}\Sigma_{aa} & \Sigma_{ab}\\ \Sigma_{ba} & \Sigma_{bb}\end{pmatrix}\quad \Omega = \begin{pmatrix}\Omega_{aa} & \Omega_{ab}\\ \Omega_{ba} & \Omega_{aa}\end{pmatrix}$$

the multivariate space may be partitioned...

but the density is over a joint distribution

3) Affine Transformation

$$x_a \sim \mathcal{N}(x_a;\mu_a,\Sigma_a)$$

LET $x_b = f(x_a) = A\,x_a + c \to$ Affine Transformation

x_b is Gaussian : $p(x_b) = \mathcal{N}(x_b;\mu_b,\Sigma_b)$

$$\begin{cases}\mu_b = A\mu_a + c\\ \Sigma_b = A\,\Sigma A^T\end{cases}$$

$$\Sigma_b = \mathbb{E}\left[(x_b-\mu_b)^T(x_b-\mu_b)\right] =$$

$$= \mathbb{E}\left[(Ax_a+\cancel{c}-A\mu_a-\cancel{c})^T(Ax_a+\cancel{c}-A\mu_a-\cancel{c})\right] =$$

$$= \mathbb{E}\left[(A\cdot(x_a-\mu_a))^T(A\cdot(x_a-\mu_a))\right] =$$

$$= \mathbb{E}\left[(x_a-\mu_a)^T A^T A\cdot(x_a-\mu_a)\right]$$

What if a Transformation is a generic function in the vector space?

Taylor Expansion

$$f(x) \approx f(x_0) + \underbrace{\frac{\partial f(x)}{\partial x}\bigg|_{x_0}}_{A}(x-x_0)$$

$$= A x + \underbrace{f(x_0) - A x_0}_{b} = \text{Affine transformation}$$

however, farther from linear $\to$ worse approximation

4) Marginalization $\longrightarrow$ "Kills a dimension" (collapses the axis)

LET $x = \begin{pmatrix}x_a\\ x_b\end{pmatrix} \sim \mathcal{N}(x;\mu,\Sigma)$

$$p(x_a) = \int_{x_b} P(x_a, x_b)\,dx_b \to \text{collapsing } x_b$$

$$p(x_a) = \mathcal{N}(x_a;\mu_a,\Sigma_{aa})$$

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$$

for symmetry, $\Sigma_{ba} = \Sigma_{ab}^T$

marginalization DOES NOT CHANGE THE MEAN

⇒ CONDITIONING → "slices the original distribution" → usually gives a smaller uncertainty distribution

$$p(x_a \mid x_b) = \frac{p(x_a, x_b)}{\int_{x_a} p(x_a, x_b)\, dx_a} \rightarrow \mu_{a|b} = \mu_a + \Sigma_{ab} \Sigma_{bb}^{-1} (x_b - \mu_b)$$

correction factor

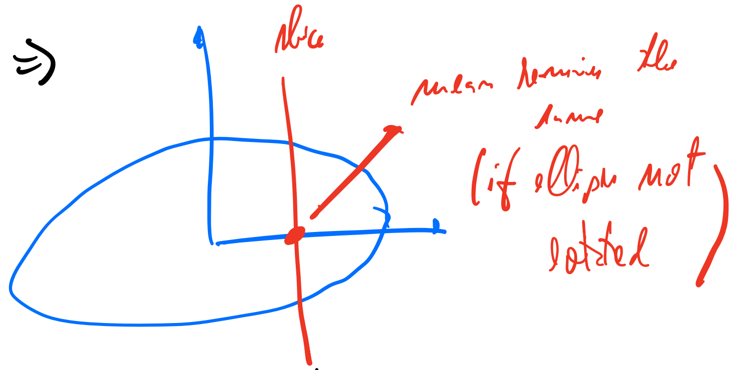
cross-correlation factor!

(if $\Sigma_{ab} = \emptyset$ (the ellipse not rotated)) → NO MATTER WHAT I DO, MEAN DOES NOT CHANGE

$$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba}$$

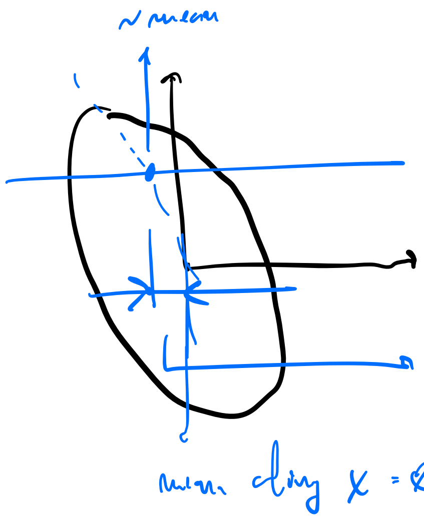
the more the 2 variables are correlated ⇒

⇒ ⇓ uncertainty (given more information)



if $\uparrow\uparrow \Sigma_{a|b}$ ...

2 variables are not correlated ⇐ if $\Sigma_{ab} = \emptyset$ ⇒ DO NOT CARE ABOUT EACH OTHER



this segment multiplied inverse slope, gives the new mean...

$$V_{a|b} = V_a - \Omega_{ab} X_b$$

$$\Omega_{a|b} = \Omega_{aa}$$

4) Chain Rule

$$p(x_a) = \mathcal{N}(x_a; \mu_a, \Sigma_a)$$

$$p(x_b | x_a) = \mathcal{N}(x_b; \underbrace{A x_a + c}_{\mu_{b|a}}, \Sigma_{b|a})$$

$$p(x_a, x_b) = \mathcal{N}(x_{a,b}; \mu_{a,b}, \Sigma_{a,b})$$

$$\mu_{a,b} = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} = \begin{pmatrix} \mu_a \\ A\mu_a + c \end{pmatrix}$$

$$\Sigma_{a,b} = \begin{pmatrix} \Sigma_a & \Sigma_a A^T \\ A\Sigma_a & \Sigma_{b|a} + A\Sigma_a A^T \end{pmatrix}$$

A is acting as a cross-correlation factor

Canonical parametrization

$$\mu_{a,b} = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} = \begin{pmatrix} \mu_a \\ A\mu_a + c \end{pmatrix}$$

$\longrightarrow$ why the mean and not the information vector?

$\Downarrow$

mean is smaller make useful ~~[illegible]~~ linearization

$$\Omega_{a,b} = \begin{pmatrix} A^T \Omega_{b|a} A + \Omega_a & -A^T \Omega_{b|a} \\ -\Omega_{b|a} A^T & \Omega_{b|a} \end{pmatrix}$$