L03: Gaussians II

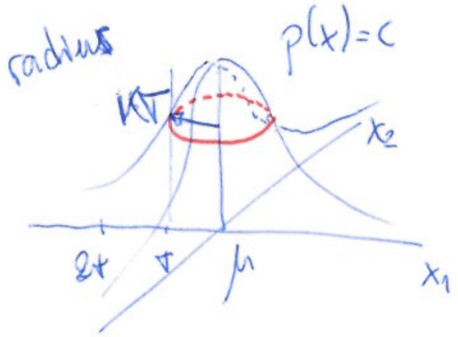
* Visualizing Gaussians (2D)

$$p(x) = \alpha \cdot e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

find contours of constant $p(x)$

$$\kappa^2 = (x-\mu)^T \Sigma^{-1} (x-\mu) \quad , \quad \kappa = 1, 2, \ldots$$

$\kappa = 1 \to 1\text{-}\sigma$



Isocontours are eqs of a quadratic curve.

Ellipse.

* Mahalanobis distance.

$$\|x-y\|_\Sigma = \sqrt{(x-y)^T \Sigma^{-1} (x-y)}$$

* Non-spherical Σ: we use covariance projection and L.A.

(recall:) if $y = Ax + b$ \ $x \sim N(0, \mathbb{I})$

then $\Sigma_y = A \Sigma_x A^T$

Problem: find A such that projects into Σ_y.

given $\Sigma_x = \mathbb{I}$ (standard $N(0, \mathbb{I})$).

$$\Sigma_y = A \cdot \Sigma_x \cdot A^T = A \cdot \mathbb{I} \cdot A^T = AA^T$$

1) SVD decomposition:

$$\underset{\text{symetric.}}{\Sigma_y = U \cdot D V^T} = U \cdot D \cdot U^T = \underbrace{U \cdot D^{1/2}}_{A} \cdot \underbrace{D^{T/2} U^T}$$

2) Cholesky decomposition. (More efficient.)

$$\Sigma_y = L \cdot L^T$$

Ex:

$$\Sigma = \begin{bmatrix} 4 & -2 \\ -2 & 10 \end{bmatrix}$$ isocontour for $K=1$? (1-sigma)

$$= \underbrace{\begin{bmatrix} a & 0 \\ b & c \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} a & b \\ 0 & c \end{bmatrix}}_{L^T} = \begin{bmatrix} 4 & -2 \\ -2 & 10 \end{bmatrix}$$

$a^2 = 4 \Rightarrow a = \pm 2$

$ba = ab = -2 \Rightarrow b = -1$

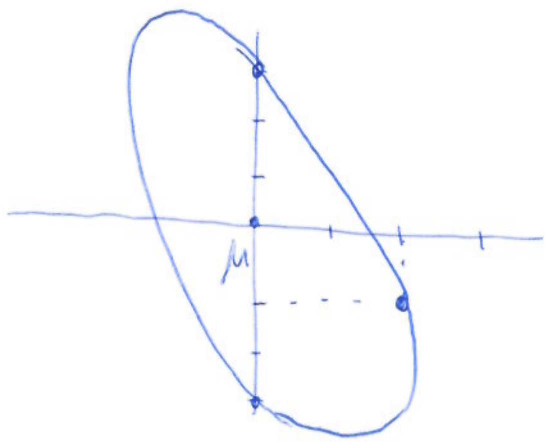
$b^2 + c^2 = 10 \Rightarrow c = \sqrt{10 - b^2} = \pm 3$

We choose positive values for diagonal $\Rightarrow \exists! L$
(unique solution)

Project points from the circumference $r=1$

$$\begin{bmatrix} 2 & 0 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

$$\begin{bmatrix} & L & \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$



note: Solution is centered at μ, so we need to translate.

* Sampling from Gaussians

Most functions sample $x \sim N(0,1)$

Problem: how to sample $y \sim N(\mu_+, \Sigma_y)$?

Solution: use covariance projection. Affine $y = Ax + b$

1) Select $x \sim N(0, I_D)$. $x = \begin{bmatrix} N(0,1) \\ N(0,1) \\ \vdots \\ N(0,1) \end{bmatrix}$ } iid

2) $y \sim N(A\mu_x + b, A\Sigma_x A^T)$

$\Rightarrow \boxed{b = \mu_y}$

3) Find A such as $\boxed{A \cdot A^T = \Sigma_y}$ (Cholesky)

4) $x \sim N(0, I)$. then $y = A \cdot x + b$

equiv. $y \sim N(\mu_y, \Sigma_y)$.

* Conditioning a joint Gaussian distribution.

$$p(x_a, x_b) = \alpha \exp \underbrace{\left\{ -\frac{1}{2} \begin{bmatrix} x_a - \mu_a \\ x_b - \mu_b \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} x_a - \mu_a \\ x_b - \mu_b \end{bmatrix} \right\}}_{\Delta}$$

Problem: $p(x_a | x_b)$ $\quad x_a \in \mathbb{R}^n, \quad x_b \in \mathbb{R}^m$

$$\Sigma = \begin{bmatrix} \Sigma_a & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_b \end{bmatrix} \qquad \Sigma_{ab} = \Sigma_{ba}^T \quad (\Sigma \text{ symetric})$$

Information matrix $\Lambda = \Sigma^{-1}$, $\quad \Lambda = \begin{bmatrix} \Lambda_a & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_b \end{bmatrix}$

Expand the exponent Δ:

$$\Delta = -\frac{1}{2}(x_a - \mu_a)^T \Lambda_a (x_a - \mu_a) - \frac{1}{2}(x_a - \mu_a)^T \Lambda_{ab} (x_b - \mu_b)$$
$$- \frac{1}{2}(x_b - \mu_b)^T \Lambda_{ba} (x_a - \mu_a) - \frac{1}{2}(x_b - \mu_b)^T \Lambda_b (x_b - \mu_b).$$

Solution: Completing the square

intuition → we want an exponent only dependent on x_a since x_b is conditioning ("given").

$$\Delta = -\frac{1}{2} x_a^T \Sigma_{a|b}^{-1} x_a + x_a^T \Sigma_{a|b}^{-1} m - \frac{1}{2} m^T \Sigma_{a|b}^{-1} m + \text{ctant.}$$

Q: what happens to x_b and constant terms?

* $\boxed{\Sigma_{a|b}^{-1} = \Lambda_a}$ I (2nd order term) $x_a^T \Sigma_{a|b}^{-1} x_a$

* 1st order terms:

$$x_a^T \cdot \Big(\underbrace{\Lambda_a \mu_a - \Lambda_{ab}(x_b - \mu_b)}_{\Sigma_{a|b}^{-1} \mu_{a|b}} \Big)$$

$$\Rightarrow \mu_{a|b} = \Sigma_{a|b} \left(\Lambda_a \mu_a - \Lambda_{ab}(x_b - \mu_b) \right)$$

I

$$= \boxed{\mu_a - \Sigma_{a|b} \Lambda_{ab}(x_b - \mu_b)} \quad \text{II}$$

We'll use the foll. matrix equality:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1}CMBD^{-1} \end{bmatrix}$$

$$M = (A - BD^{-1}C)^{-1} \quad (M^{-1} \text{ Schur complement})$$

$$\begin{bmatrix} \Sigma_a & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_b \end{bmatrix}^{-1} = \begin{bmatrix} \Lambda_a & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_b \end{bmatrix}$$

$$\rightarrow \Lambda_a = (\overbrace{\Sigma_a - \Sigma_{ab}\Sigma_b^{-1}\Sigma_{ba}}^{\Sigma_{a|b}})^{-1} \quad \text{(Schur complement)}$$

$$\rightarrow \Lambda_{ab} = -\Lambda_a \cdot \Sigma_{ab}\Sigma_b^{-1}$$

$$\Rightarrow \mu_{a|b} = \mu_a - \underbrace{\cancel{\Sigma_{a|b}}(-\cancel{\Lambda_a}}_{I}\Sigma_{ab}\Sigma_b^{-1})(x_b - \mu_b)$$

$$= \mu_a + \Sigma_{ab}\Sigma_b^{-1}(x_b - \mu_b)$$

$$\boxed{p(x_a|x_b) = N(x_a; \mu_a + \Sigma_{ab}\Sigma_b^{-1}(x_b - \mu_b), \Sigma_a - \Sigma_{ab}\Sigma_b^{-1}\Sigma_{ba})}$$

* Marginalizing a joint Gaussian distribution.

problem: $p(x_a) = \int p(x_a, x_b)\, dx_b$

(idea) solution: same as before, we will expand the exponent and complete the square, now twice

$$\int e^{\Delta} dx_b = e^{\Delta(x_a)} \underbrace{\int e^{\Delta(x_b)} dx_b}_{\eta} \cdot \underbrace{e^{\Delta(\text{ctant})}}_{\eta'}$$

3.7

$$\Delta = -\frac{1}{2}(x_a-\mu_a)^T\Lambda_a(x_a-\mu_a) \;-\; \frac{1}{2}(x_a-\mu_a)^T\Lambda_{ab}(x_b-\mu_b)$$

$$-\frac{1}{2}(x_b-\mu_b)^T\Lambda_{ba}(x_a-\mu_b) \;-\; \frac{1}{2}(x_b-\mu_b)^T\Lambda_b(x_b-\mu_b)$$

$$= \Delta(x_b)\left\{\begin{aligned} & -\frac{1}{2}x_b^T\Lambda_b x_b \qquad (1) \\ & + x_b^T\left(\Lambda_b\mu_b - \Lambda_{ba}(x_a-\mu_a)\right) \qquad (2) \\ & -\frac{1}{2}m^T\Lambda_b\cdot m \end{aligned}\right.$$

$$\hookrightarrow \Lambda_b\cdot m \Rightarrow m = \mu_b - \Lambda_b^{-1}\Lambda_{ba}(x_a-\mu_a)$$

$$\Delta(x_a)\left\{\begin{aligned} & +\frac{1}{2}m^T\Lambda_b\cdot m \\ & -\frac{1}{2}x_a^T\Lambda_a x_a \qquad (3) \\ & + x_a^T\left(\Lambda_a\mu_a + \Lambda_{ab}\mu_b\right) \qquad (4) \\ & + \text{ctant} \end{aligned}\right.$$

$$(\Lambda_b^{-1})^T = \Lambda_b^{-1}$$

- $$\frac{1}{2}m^T\Lambda_b m = \frac{1}{2}\left(\mu_b - \Lambda_b^{-1}\Lambda_{ba}(x_a-\mu_a)\right)^T\cdot\Lambda_b\cdot(\cdot) \stackrel{?}{=}$$

$$= \frac{1}{2}\, x_a^T\Lambda_{ab}\cancel{\Lambda_b^{-1}}\cdot\cancel{\Lambda_b}\cdot\Lambda_b^{-1}\Lambda_{ba}x_a$$

$$- x_a^T\Lambda_{ab}\cancel{\Lambda_b^{-1}}\cdot\cancel{\Lambda_b}\left(\mu_b + \Lambda_b^{-1}\Lambda_{ba}\mu_a\right)$$

$$+ \text{ctant.}$$

$$\Delta(x_a) = -\frac{1}{2} x_a^T \Lambda_a x_a + \frac{1}{2} x_a^T \Lambda_{ab} \Lambda_b^{-1} \Lambda_{ba} x_a$$

$$+ x_a^T \left(\Lambda_a \mu_a + \cancel{\Lambda_{ab} \mu_b} \right) - x_a^T \cancel{\Lambda_{ab} (\mu_b} + \Lambda_b^{-1} \Lambda_{ba} \mu_a)$$

$$+ \text{const}$$

$$= -\frac{1}{2} x_a^T \Big(\underbrace{\Lambda_a - \Lambda_{ab} \Lambda_b^{-1} \Lambda_{ba}}_{\Sigma_a} \Big) x_a \qquad \text{(Schur complement!)}$$

$$+ x_a^T \left(\Lambda_a - \Lambda_{ab} \Lambda_b^{-1} \Lambda_{ba} \right) \mu_a$$

$$+ \text{const.}$$

$$\boxed{p(x_a) = \int p(x_a, x_b)\, dx_b = \mathcal{N}(x_a; \mu_a, \Sigma_a)}$$

Gaussian again!

Marginalizing a Gaussian is as <u>simple</u> as selecting the submatrix inside Σ and the corresponding mean!

Gaussians are their self conjugate priors.

Prob Rob 2.4, 3.1, 3.2