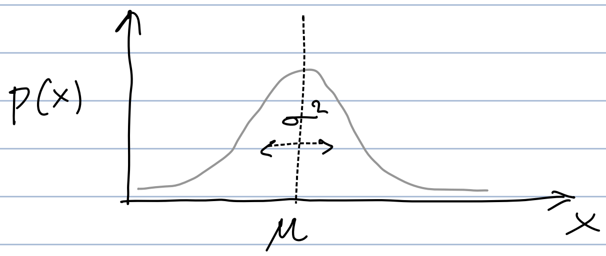
① Review of Gaussian Distributions

1) 1D gaussians



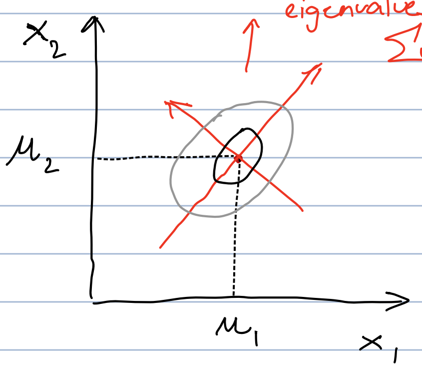
$$p(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

$\mu \to$ mean $\sigma \to$ std. dev.

$\Downarrow$ Generalizing this to d-D space results in a multi-variate gaussian $\Rightarrow$ variables are 'jointly' gaussian.

2) 2D gaussians

Extent & orientation of ellipse depends on eigenvalues of Σ



$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22}^2 \end{bmatrix}$$

$$p(x;\underline{\mu},\Sigma) \triangleq \mathcal{N}(x;\underline{\mu},\Sigma)$$

$$= (2\pi)^{-d/2}|\Sigma|^{-1/2}\exp\left(\frac{-1}{2}(\underline{x}-\underline{\mu})^T\Sigma^{-1}(\underline{x}-\underline{\mu})\right)$$

$\Sigma \to$ p.s.d matrix $\Rightarrow$ i.e all eigenvalues are ≥ 0.

More generally, for a d-Dimensional Gaussian.

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}_{d\times 1}, \quad \underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{bmatrix}_{d\times 1}, \quad \Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1d} \\ \vdots & \sigma_{22}^2 & \cdots & & \vdots \\ \vdots & & \ddots & & \vdots \\ \sigma_{d1} & \cdots & & \cdots & \sigma_{dd}^2 \end{bmatrix}_{d\times d}$$

$\Sigma = \Sigma^T \to$ p.s.d.

$$p(\underline{x};\underline{\mu},\Sigma) = (2\pi)^{-d/2}|\Sigma|^{-1/2}\exp\left(\frac{-1}{2}(\underline{x}-\underline{\mu})^T\Sigma^{-1}(\underline{x}-\underline{\mu})\right)$$

Covariance enters the above expression as an inverse:

Inverse covariance $\to$ "PRECISION" $(\Sigma^{-1} = \Lambda)$

3) Some key properties.

i) LINEARITY

If $x_1 \sim \mathcal{N}(\mu_1, \Sigma_1)$ and $x_2 \sim \mathcal{N}(\mu_2, \Sigma_2)$, then

$$z = x_1 + x_2 \sim \mathcal{N}(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2)$$

ii) AFFINE PROPERTY

constant offset!

If $x \sim \mathcal{N}(\mu, \Sigma)$, $y = Hx + d$, then

$$p(y) \sim \mathcal{N}(H\mu + d, H\Sigma H^T).$$

iii) BAYES RULE

If $x \sim \mathcal{N}(x; \cdot, \cdot)$ and $y|x \sim \mathcal{N}(y; \cdot, \cdot)$, then

$$p(x|y) \sim \mathcal{N}(x; \cdot, \cdot)$$

4) MARGINALIZING A GAUSSIAN:

$p(x_a, x_b)$ is jointly gaussian, then marginals $p(x_a)$ and $p(x_b)$ are also gaussian, where marginals are.

$$\int_{x_a} p(x_a, x_b)\, dx_a = p(x_b)$$

$$\int_{x_b} p(x_a, x_b)\, dx_b = p(x_a).$$

If $$p\left(\begin{bmatrix} x_a \\ x_b \end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix}, \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}\right).$$

then,

$$p(x_a) = \mathcal{N}(\mu_a, \Sigma_{aa}) \quad \text{and}$$

$$p(x_b) = \mathcal{N}(\mu_b, \Sigma_{bb}).$$

5) CONDITIONING A GAUSSIAN.

$p(x_a, x_b)$ is jointly gaussian, then $p(x_a | x_b)$ is also gaussian.

If $p\left(\begin{bmatrix} x_a \\ x_b \end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} x_a \\ x_b \end{bmatrix}, \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix}, \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}\right)$

then,

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

where,

$$\mu_{a|b} = \mu_a + \Sigma_{ab} \Sigma_{bb}^{-1} (x_b - \mu_b)$$

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

6) BAYES RULE FOR GAUSSIANS

Say $p(x) = \mathcal{N}(x; \mu, \Sigma)$ and

$p(y|x) = \mathcal{N}(y; Hx + d, \Gamma)$

then,

$p(x|y) = \mathcal{N}(x; \mu_{x|y}, \Sigma_{x|y})$, where

$$\mu_{x|y} = (\Sigma^{-1} + H^T \Gamma^{-1} H)^{-1} [H^T \Gamma^{-1} (y - d) + \Sigma^{-1} \mu]$$

$$\Sigma_{x|y} = (\Sigma^{-1} + H^T \Gamma^{-1} H)^{-1}$$