

IFT3395/6390

Fondements de l'apprentissage machine

Notion de distributions de probabilité

Gaussienne multivariée

Professeur: Pascal Vincent

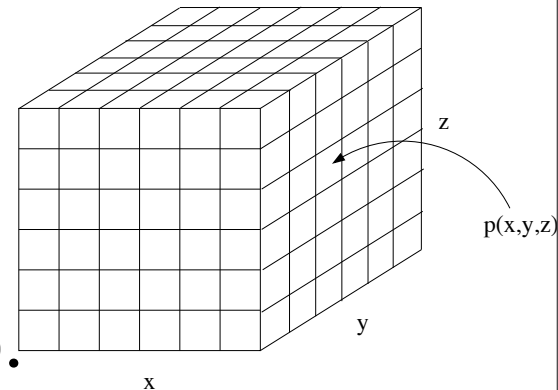
Distributions

Voir <http://www.techno-science.net/?onglet=glossaire&definition=6395>
et le rappel de proba de Balazs

- On associe naturellement une **loi de probabilité (ou distribution)** à une variable aléatoire pour décrire la *répartition des valeurs qu'elle peut prendre*.
- La distribution d'une variable aléatoire peut se caractériser par sa **fonction de distribution cumulative (c.d.f.)**:

$$F_X(x) = P(X \leq x) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$$

- La loi d'une **variable discrète** est déterminée par la **probabilité de chacune des valeurs** qu'elle peut prendre. => table de probabilités (qui doivent sommer à 1).



- La loi d'une **variable continue** peut être donnée sous la forme d'une fonction de **densité de probabilité (p.d.f.)** qui est la dérivée de la c.d.f. *La probabilité qu'un tirage de la variable tombe dans une certaine région de l'espace est l'intégrale de la densité sur cette région.*

Opérations avec les distributions

Avec une distribution, on peut vouloir:

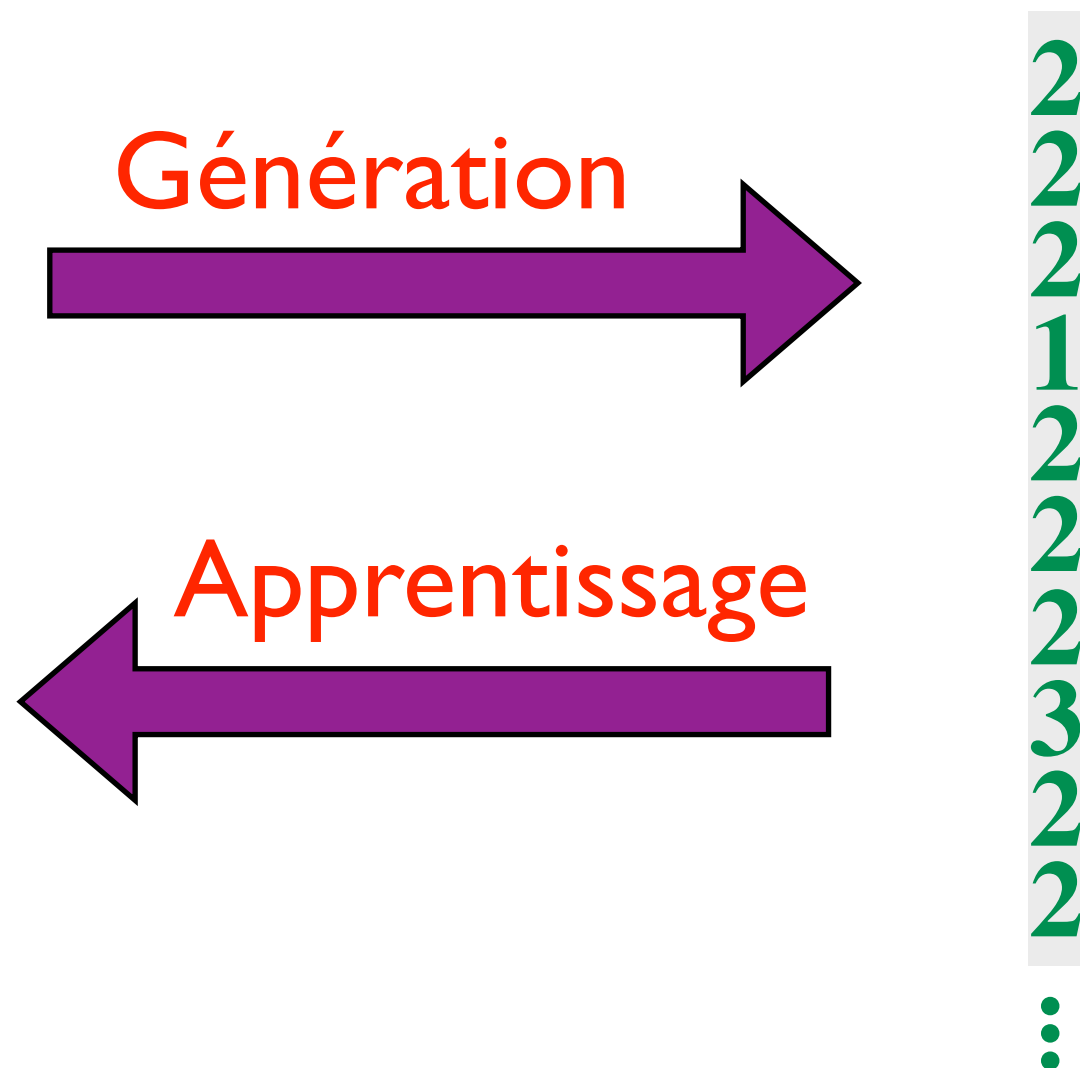
- **Générer des données**, c.a.d. tirer des échantillons selon la distribution.
- **Calculer la (log) probabilité d'une configuration** (sachant la valeur de certaines des variables et ayant marginalisé celles dont on ne connaît pas la valeur).
- **Inférence**: *inférer* la valeur la plus probable ou la valeur espérée d'un sous ensemble de variables sachant la valeur des autres.
- **Apprentissage** des paramètres de la distribution **à partir d'un ensemble de données** (de sorte à maximiser la probabilité que les données soient générées par cette distribution avec ces paramètres: principe du *maximum de vraisemblance*).

Ex. variable scalaire discrète X

Table de probabilité:

x	$P(X=x)$
1	0.10
2	0.80
3	0.10

Ensemble de données:



Ex. variable scalaire continue x

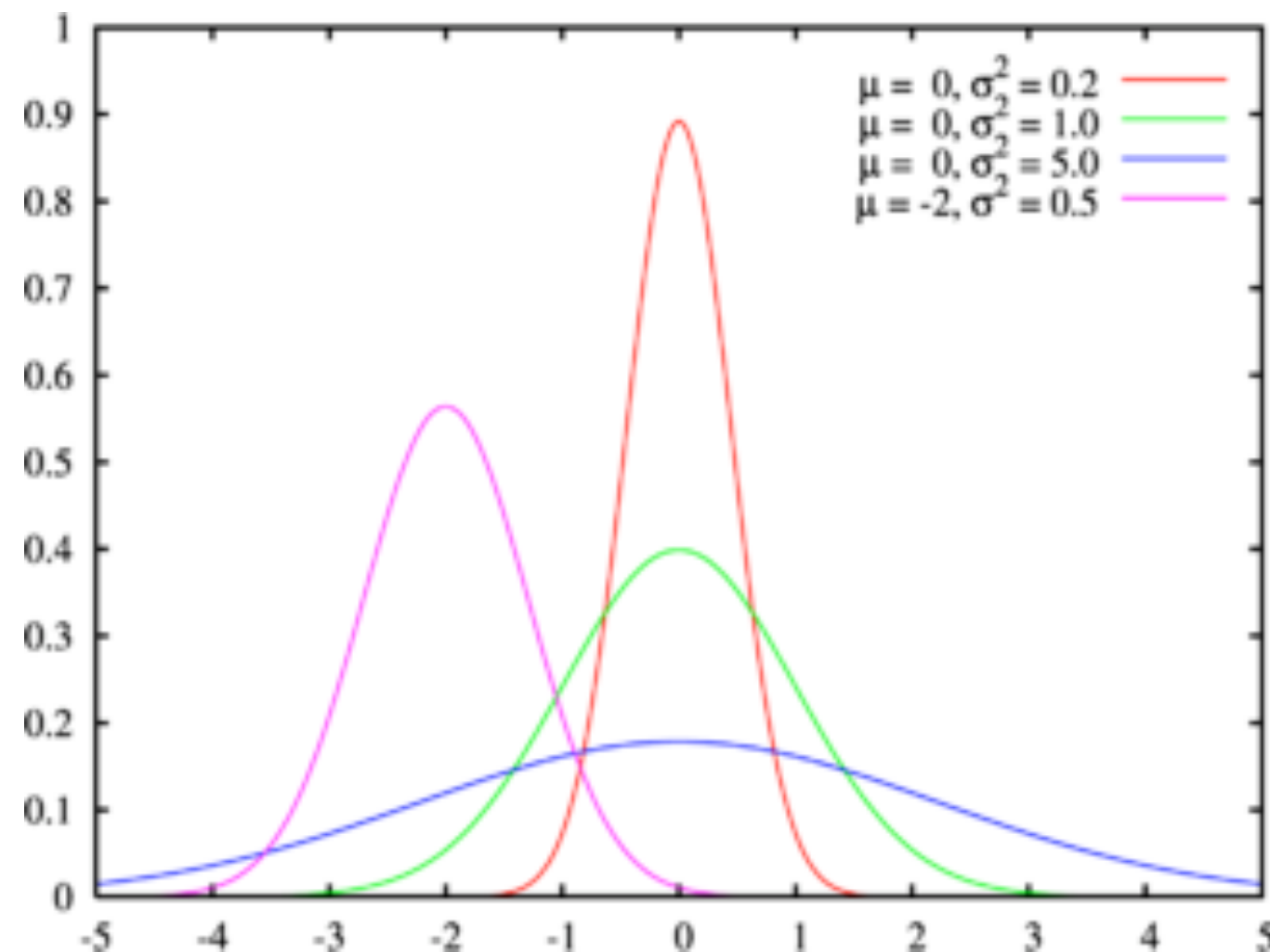
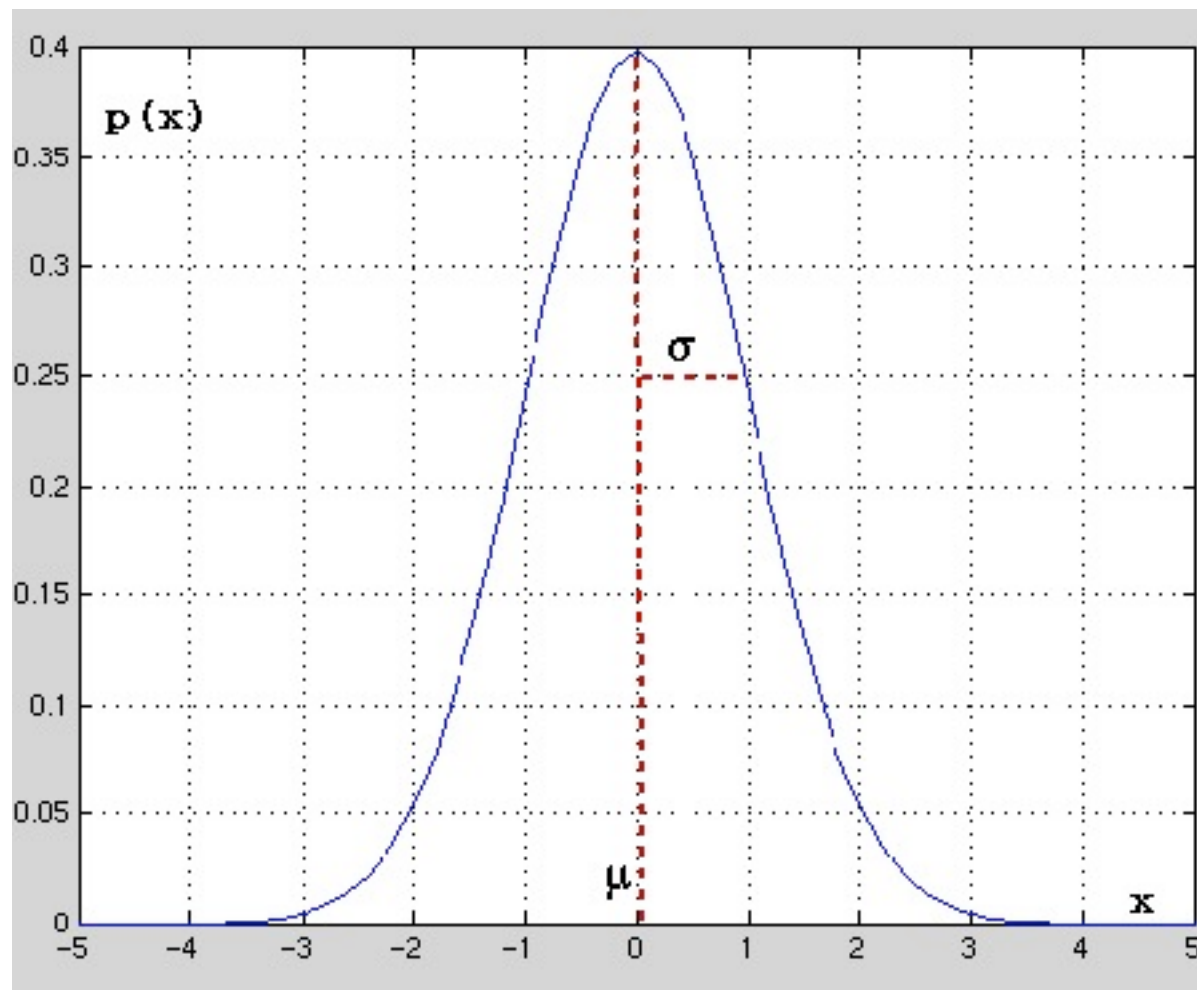
la densité Gaussienne univariée
(distribution dite Gaussienne ou Normale)

La densité Gaussienne univariée

(distribution dite Gaussienne ou Normale)

Densité Gaussienne **univariée** (c.a.d. en dimension 1) de moyenne μ et de variance σ^2 (écart type σ).

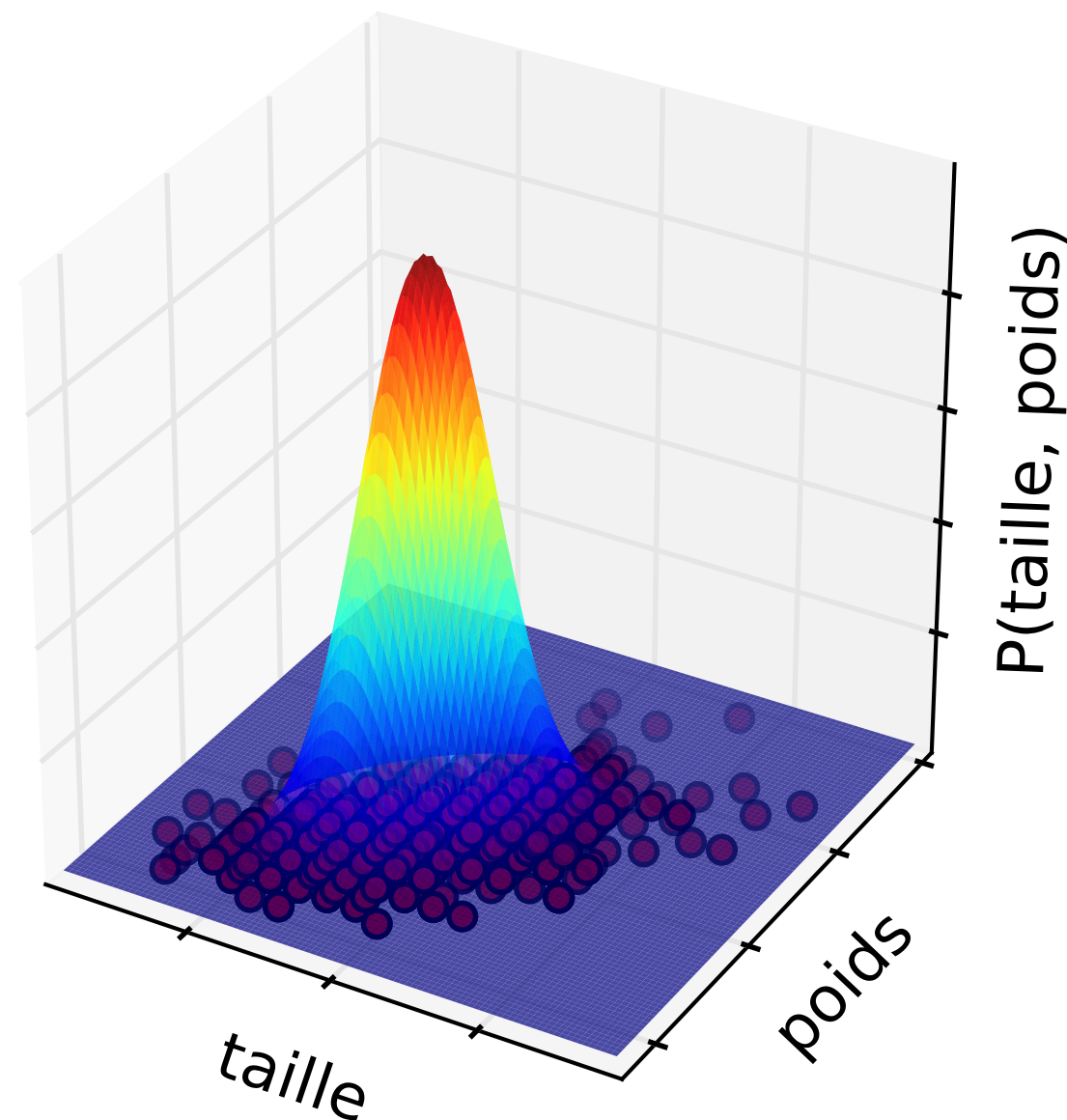
$$p(x) = \mathcal{N}_{\mu, \sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Ex. variable vectorielle continue x

la densité Gaussienne multivariée

(distribution dite Gaussienne ou Normale)

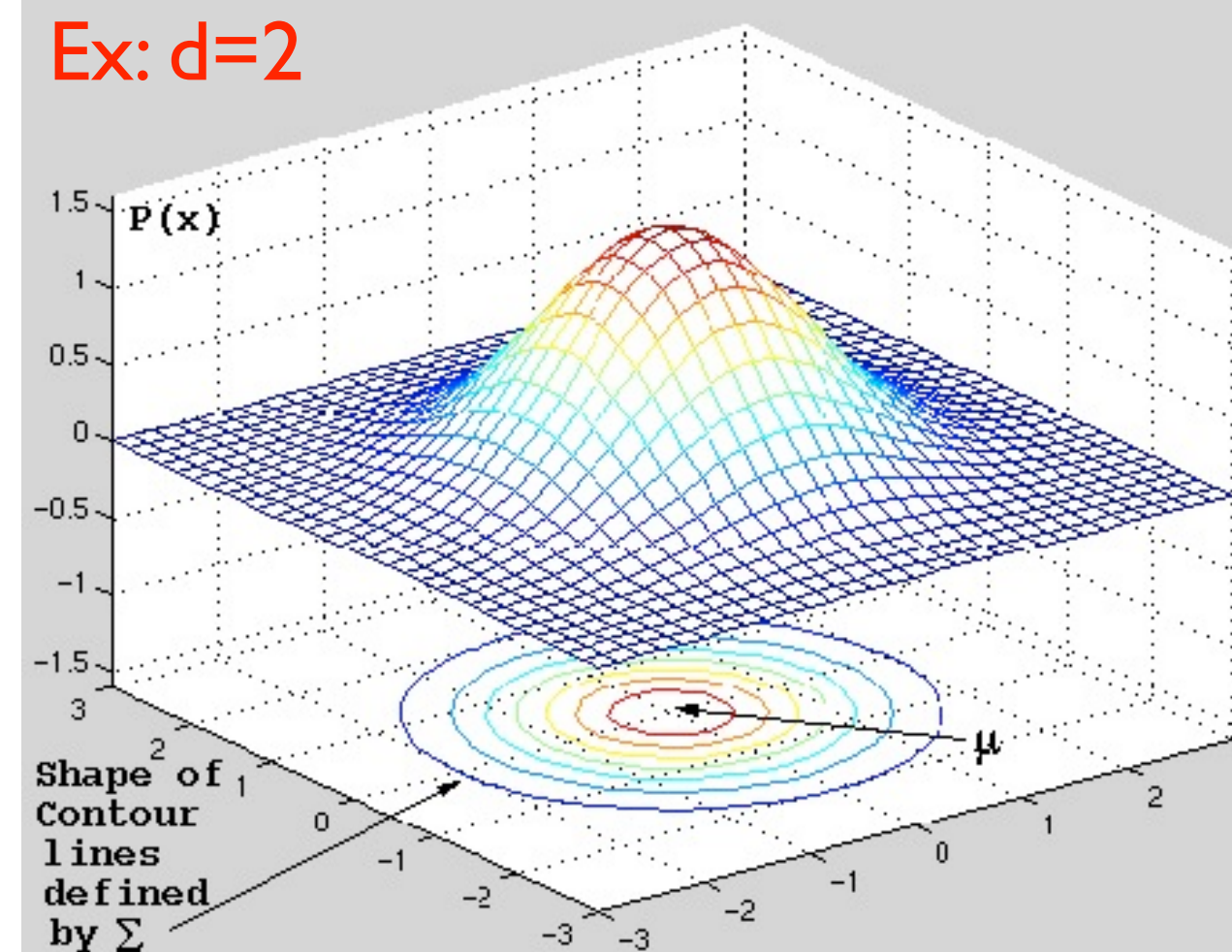


La densité Gaussienne multivariée

Gaussienne isotropique
("sphérique") en dimension d:

$$p(x) = \mathcal{N}_{\mu, \sigma^2}(x) = \frac{1}{(2\pi)^{\frac{d}{2}} \sigma^d} e^{-\frac{1}{2} \frac{\|x - \mu\|^2}{\sigma^2}}$$

Il s'agit d'une "bosse" Gaussienne
"**centrée**" en μ et de "**largeur**" σ , la même dans toutes les directions.



Gaussienne générale en dimension d, de moyenne μ et de matrice de covariance Σ .

$$p(x) = \mathcal{N}_{\mu, \Sigma}(x) = \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{|\Sigma|}} e^{-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu)}$$

$$\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_d \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1d} \\ \vdots & & \vdots \\ \sigma_{d1} & \dots & \sigma_{dd} \end{bmatrix}$$

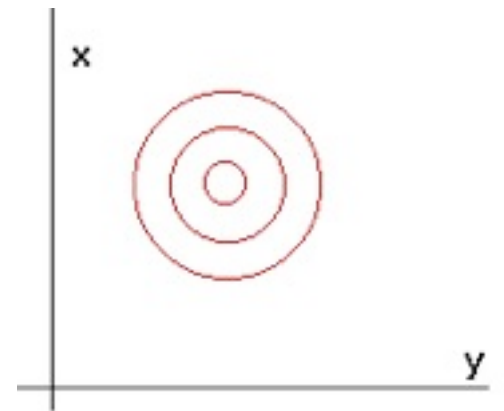
} déterminant de Σ

Note: le dénominateur est une simple normalisation qui assure que la densité intègre à 1, mais ne change rien à sa forme

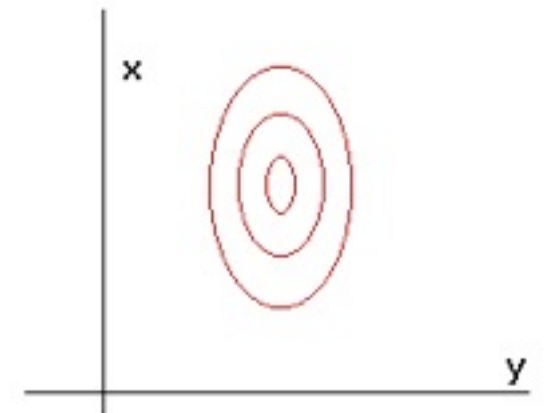
La densité Gaussienne multivariée

matrices de covariance particulières

- Gaussienne isotropique ou sphérique: $\Sigma = \sigma^2 I$
(avec I la matrice identité)



- Gaussienne diagonale: $\Sigma = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \ddots \\ & & & \sigma_d^2 \end{pmatrix}$

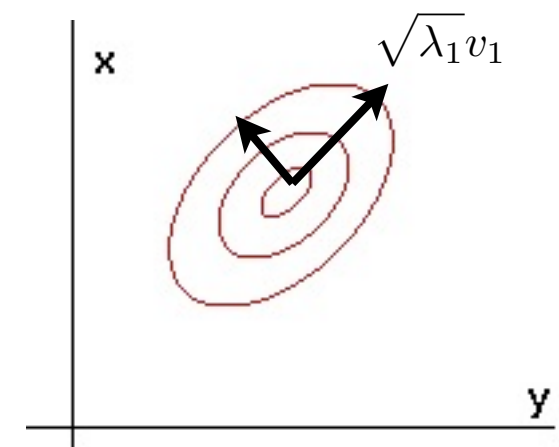


- Décomposition en valeurs propres/vecteurs propres:

$$\Sigma = \sum_{i=1}^d \lambda_i v_i v_i^T$$

Les vecteurs propres indiquent les directions des axes de l'ellipsoïde associée, les valeurs propres leur "longueur".

Le déterminant $|\Sigma| = \lambda_1 \lambda_2 \dots \lambda_d$ indique la "taille" de l'ellipsoïde.



La densité Gaussienne multivariée

apprentissage des paramètres

- On peut **apprendre les paramètres** d'une Gaussienne à partir d'un ensemble de données de manière simple:
- Pour μ on calcule la **moyenne empirique** des points (le “centroïde”).

$$\mu = \frac{1}{n} \sum_{t=1}^n \mathbf{x}_t$$

- Pour Σ on calcule la **matrice de covariance empirique** des données.

$$\Sigma_{ij} = \frac{1}{n} \sum_{t=1}^n (\mathbf{x}_{ti} - \mu_i)(\mathbf{x}_{tj} - \mu_j) \quad \text{ou bien} \quad \Sigma = \frac{1}{n} \sum_{t=1}^n (\mathbf{x}_t - \mu)(\mathbf{x}_t - \mu)'$$

- On verra comment dériver ces résultats dans un prochain cours (*principe du maximum de vraisemblance*).