

Les Mains à la pâte (suite)

→ Rappel situationnel

→ Algorithme contraint

↳ aller plus vers "l'apprentissage" réflexion 4

→ réécrire le cœur de l'algo

→ évaluer la performance de la reconnaissance

Matrice de confusion et accuracy

→ vers le ML → réflexion 2

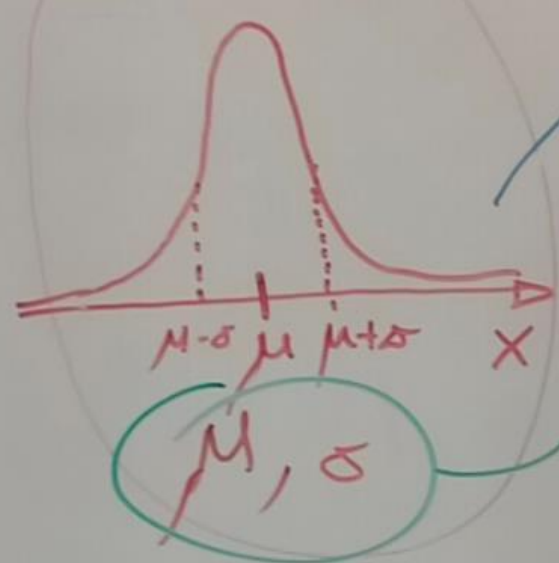
→ reconnaissance zéro fruits [orange, pêche, raisins]
(vous) + correction

→ ~~jeu~~ Biber science

X : VA d'étude $\mu \in \mathbb{R}$

$$\text{Bin F} = \mu - \varepsilon \sigma$$

$$\text{Bsup} = \mu + \varepsilon \sigma$$



Fisher externe consulté



Population $\mu - \varepsilon \sigma$ μ $\mu + \varepsilon \sigma$

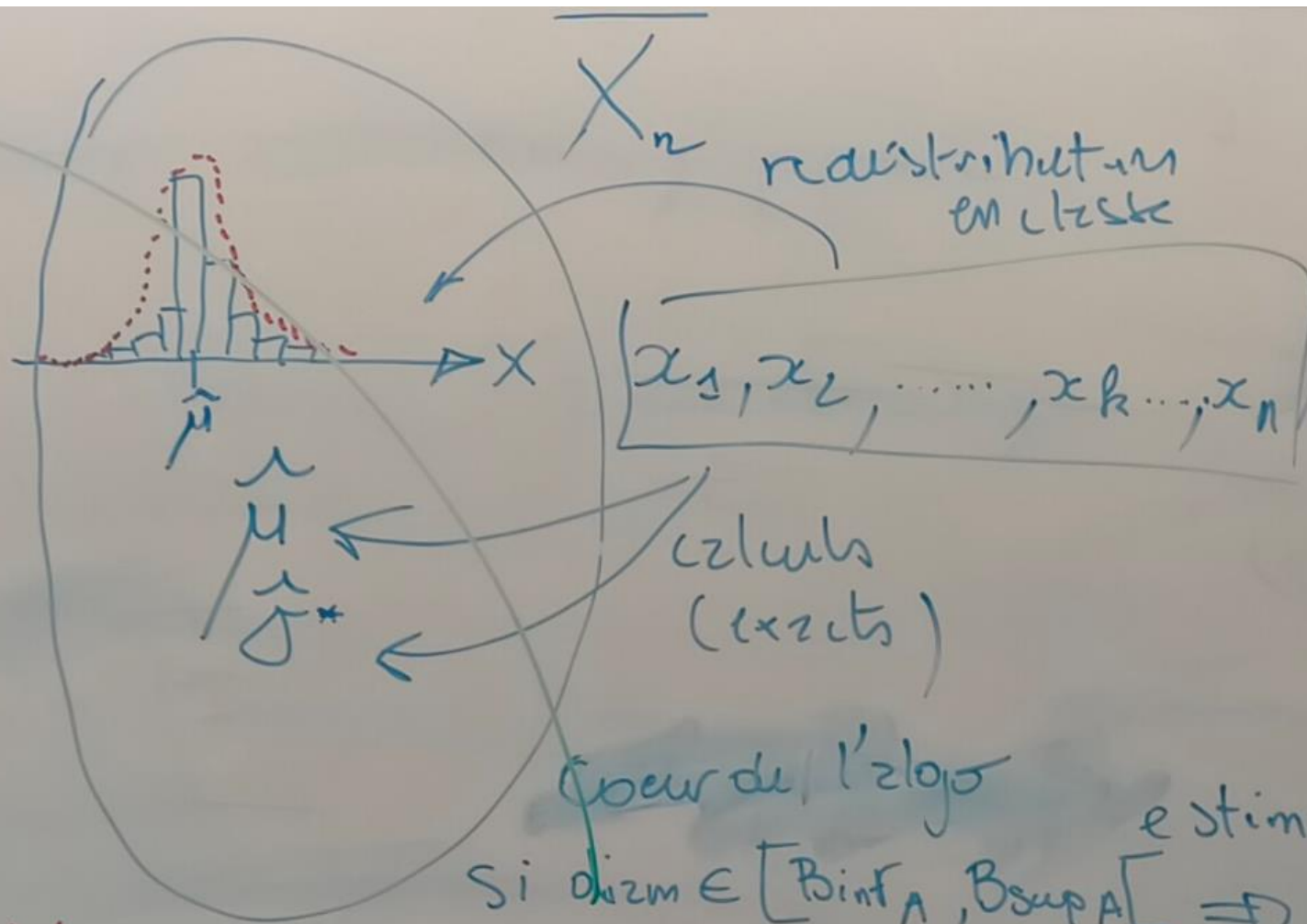
(+ caractéristique)

permet position ex μ

paramètre de dispersion : σ ou σ^2

forme de la distribution (normale?)

N inconnue



X_n redistribution en classe

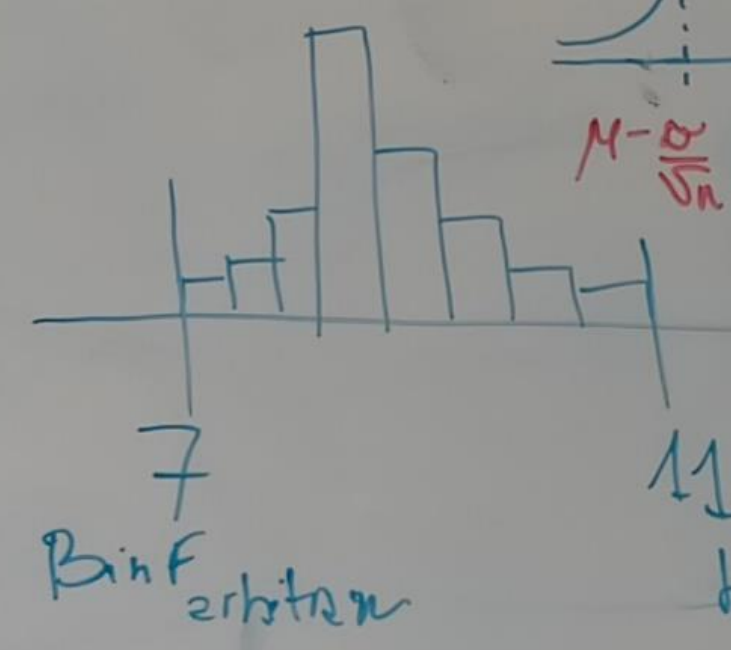
$x_1, x_2, \dots, x_k, \dots, x_n$

calculs (exacts)

pour de l'algo

Si $\mu \in [\text{Bin F}, \text{Bsup}] \Rightarrow$ orange

n échantillon de taille n



Bin F arbitraire

Bsup arbitraire

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left[\frac{x - \mu}{\sigma} \right]^2}$$

μ : estimation non biaisée (continue) de μ

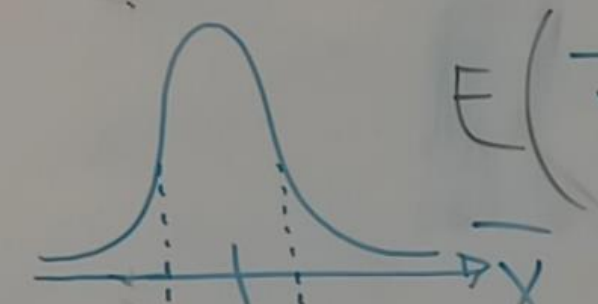
$n \geq 30$

$$\hat{\sigma}^2 = \frac{n}{n-1} s^2$$

variance échantillonnée de l'échantillon

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

estimateur



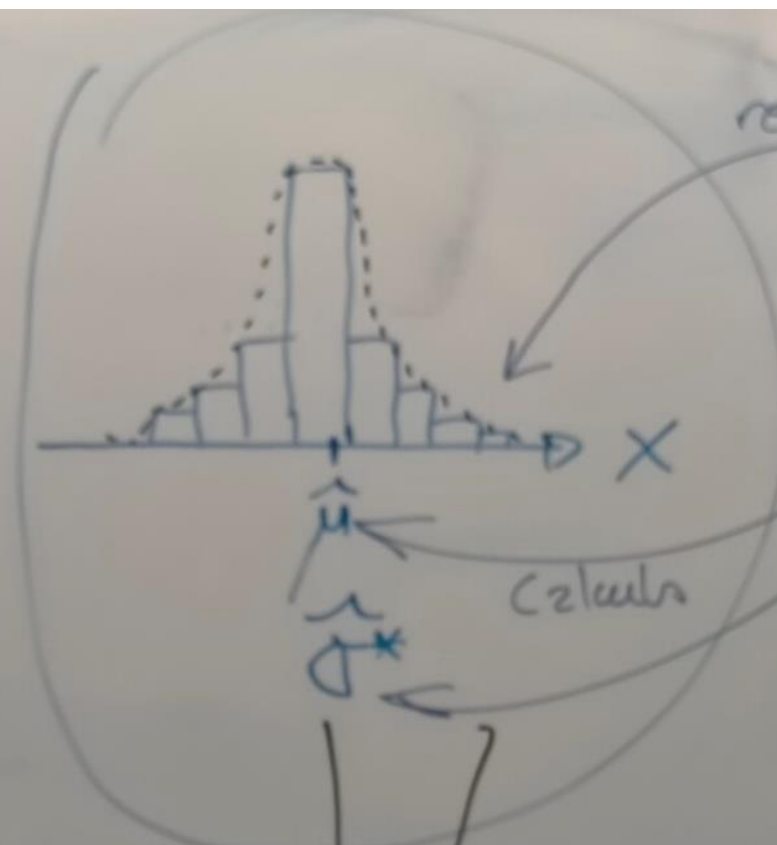
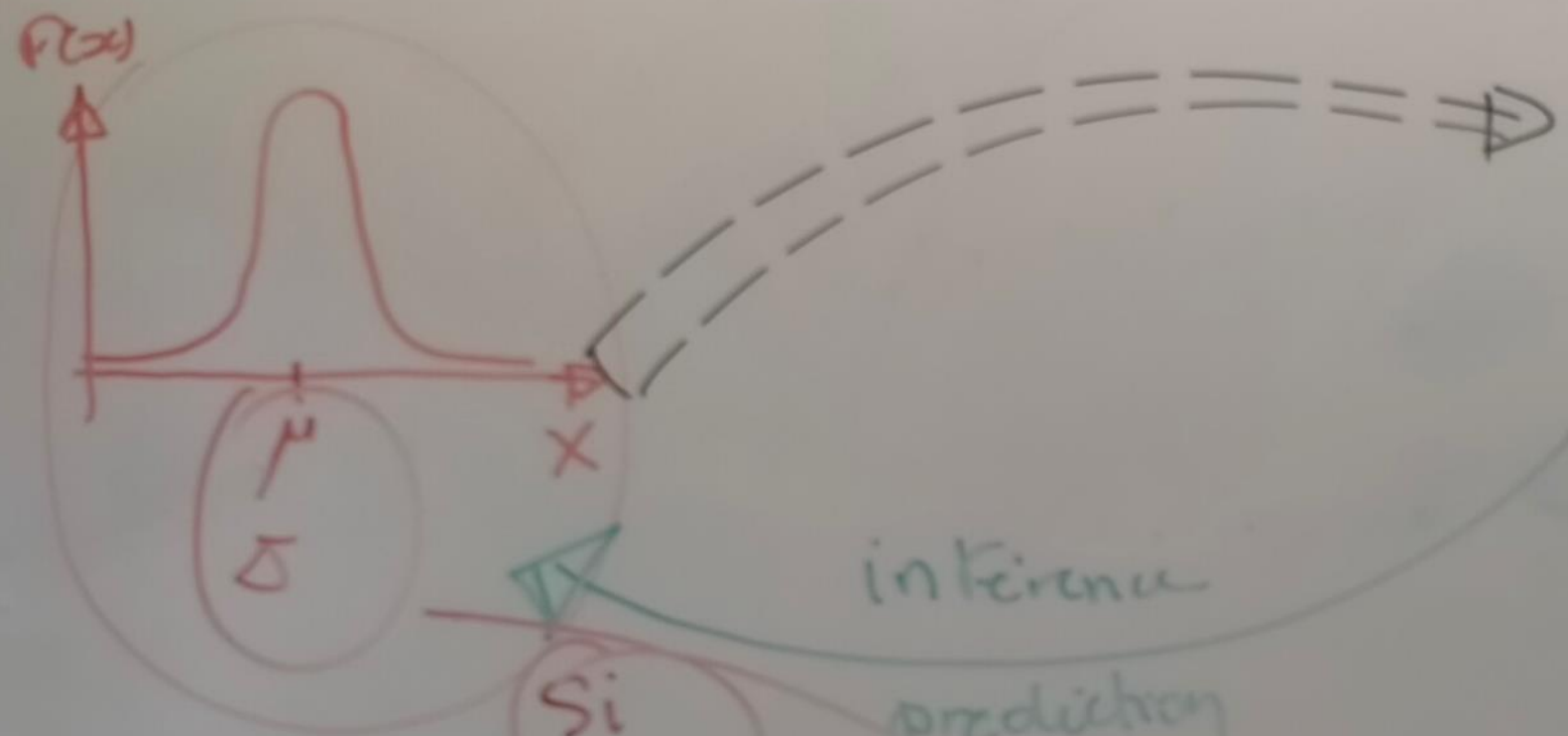
$$E(\bar{X}) - \mu = 0$$

estimateur non biaisé de μ

On peut construire un intervalle de confiance en utilisant $\bar{X}$

X ex : ϕ orange en cm

X : VA d'étude



redistribution en classes $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}$

$[x_1, x_2, \dots, x_n]$

$\hat{\mu}$ estimation ponctuelle de μ ($\hat{\mu} \approx \mu$)

$\hat{\sigma}$ estimation ponctuelle de σ

correction de bias

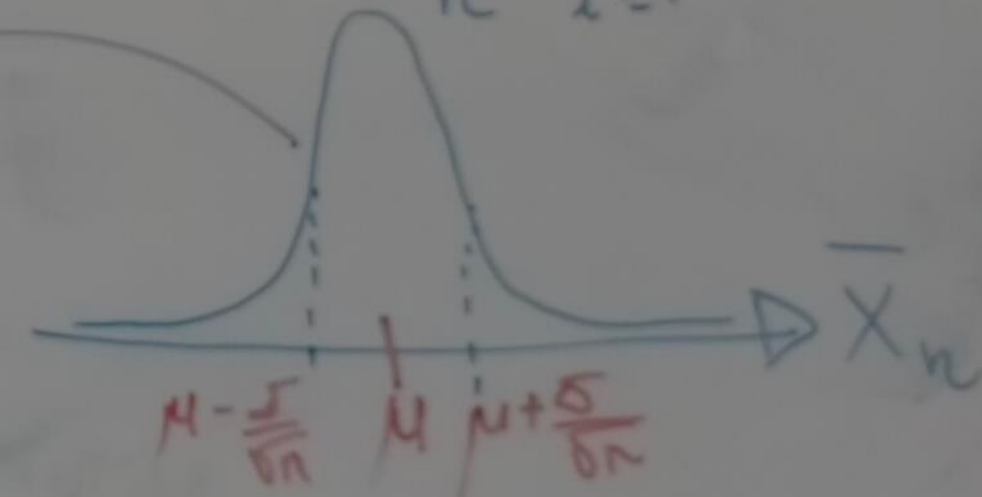
$\hat{\sigma}^2 = \frac{n}{n-1} \hat{\sigma}^2$ $E(\bar{X}_n) - \mu = 0$

echantillon de taille n
 $n \geq 30$

(Objet de travail en ML)

$\bar{X}_n$ estimateur

IC $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n x_i$



inférence
prédiction
modélisation
Généralisation

Si connue

! Biais (revolte, Stress, Rythme, etc...)

Population (Caractère)

N inconnue (très grande)

$X \sim N(\mu, \sigma)$

μ, σ inconnue

Position: mode, min, max, moy

Dispersion: σ, σ^2

Forme: Kurtosis, Skewness

Coeur algo ML

Enrichissement de μ

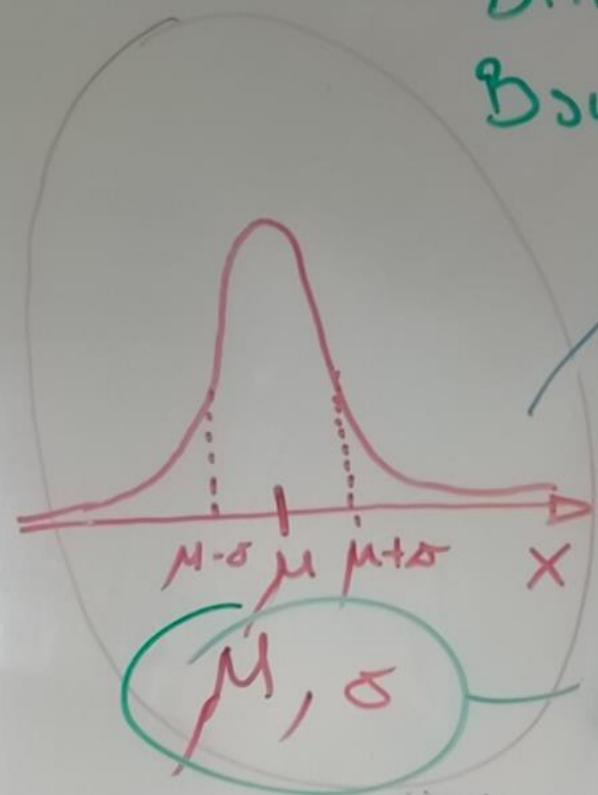
$7 < \text{dizim} < 11$ contours
 $\min < \text{dizim} < \max$
 $F_1(\mu, \sigma) < \text{dizim} < F_2(\mu, \sigma)$
 $F_1(?) < \text{dizim} < F_2(?)$

IC α
 $\mu \in [\text{BinF}, \text{Bsup}]$

X : VA d'étude indépendant

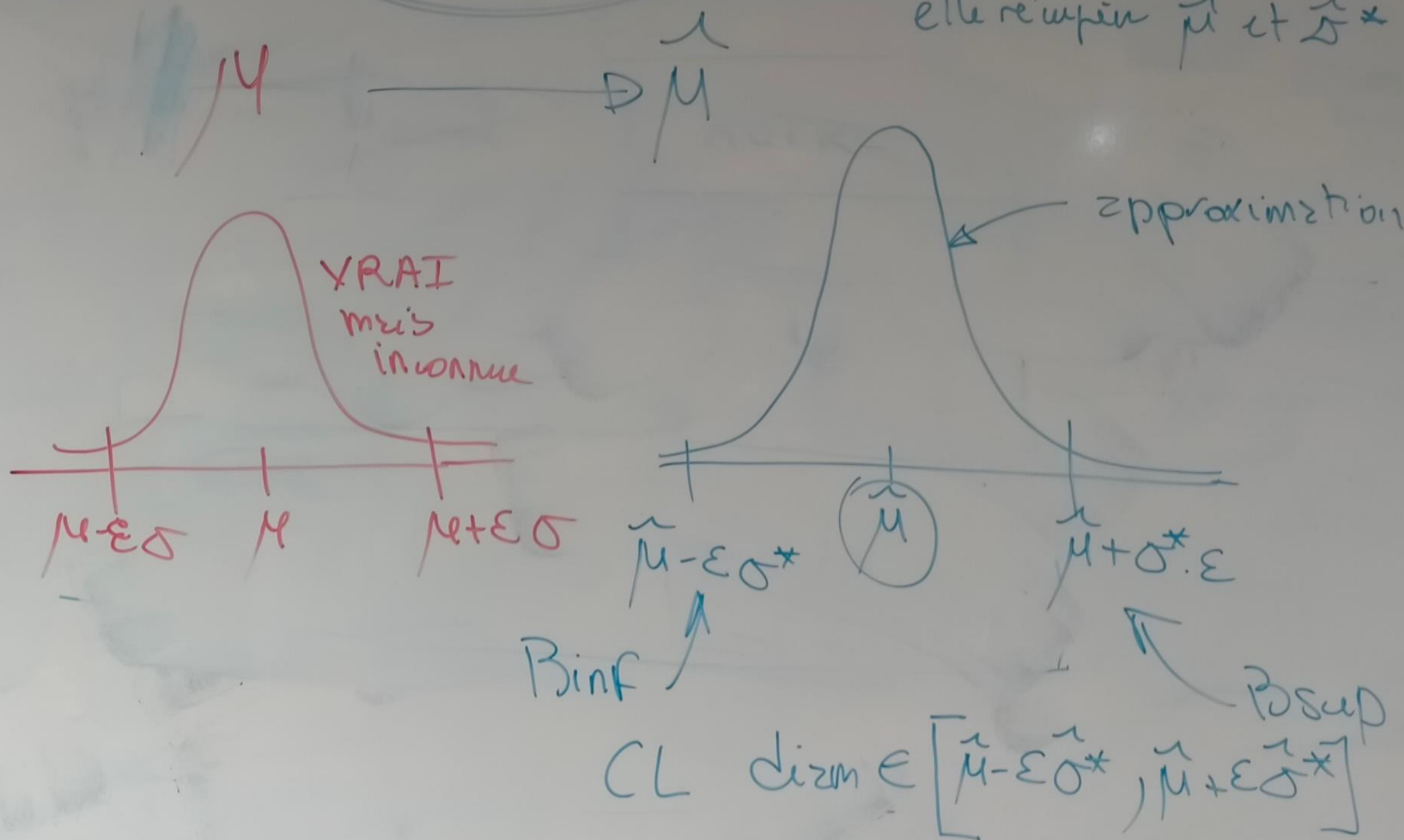
$$\text{BinF} = \mu - \epsilon \sigma$$

$$\text{BinS} = \mu + \epsilon \sigma$$

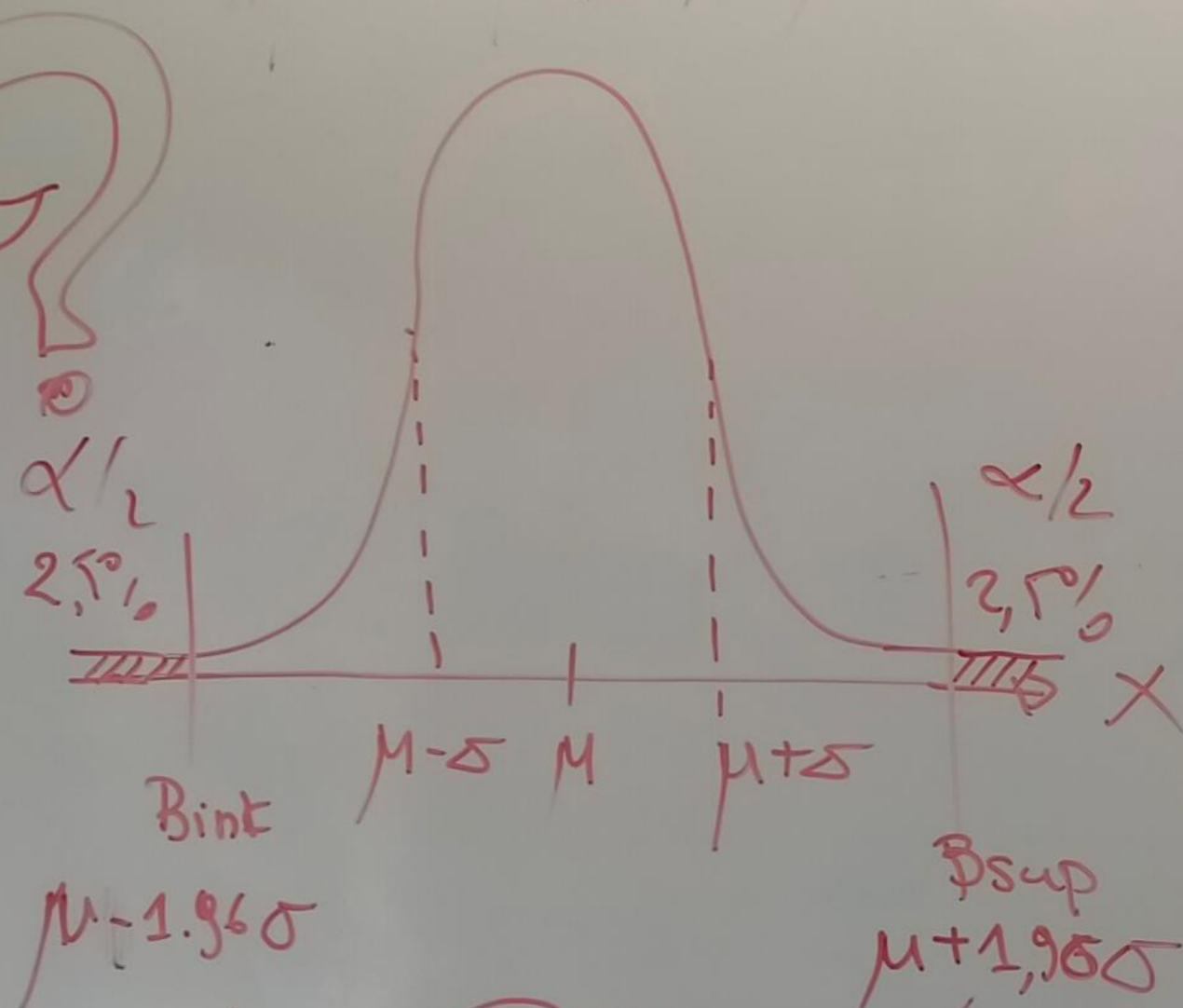


Machine Learning
 Bleu = Bizi
 travail sur échantillon

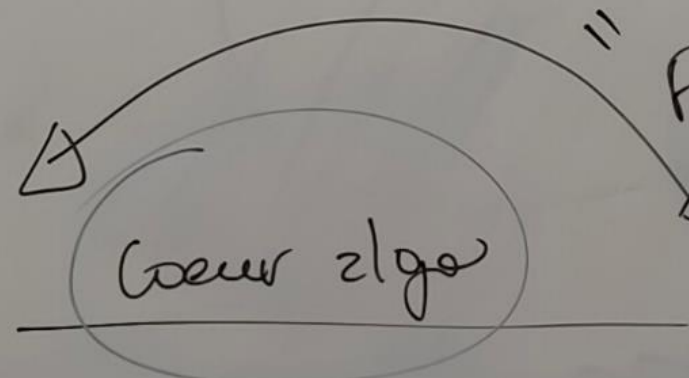
La machine lit la Fiche
 elle résume $\hat{\mu}$ et $\hat{\sigma}^*$



dirn 2D $N(\mu, \sigma)$



95% des orgs
présentent un dirnité
des cas limites



$\hat{\mu} - \epsilon \hat{\sigma}^* < \text{dirn} < \hat{\mu} + \epsilon \hat{\sigma}^*$
(tout provient de l'échantillon)
⚠ si échantillon trop petit ou non représentatif de la pop / 2D

Grosse échelle
pour garantir succès
(+) homogénéité
des données

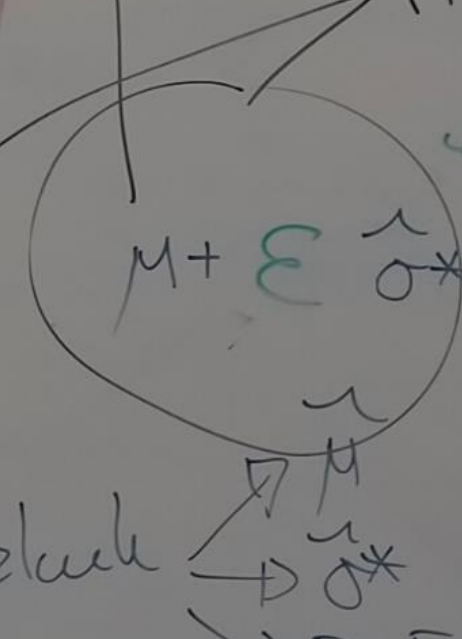
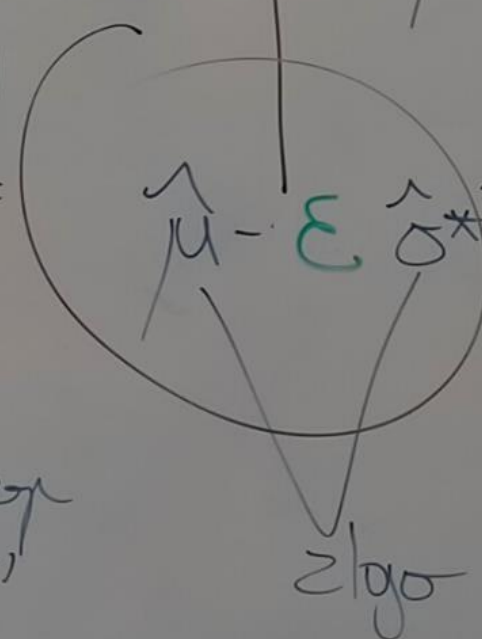
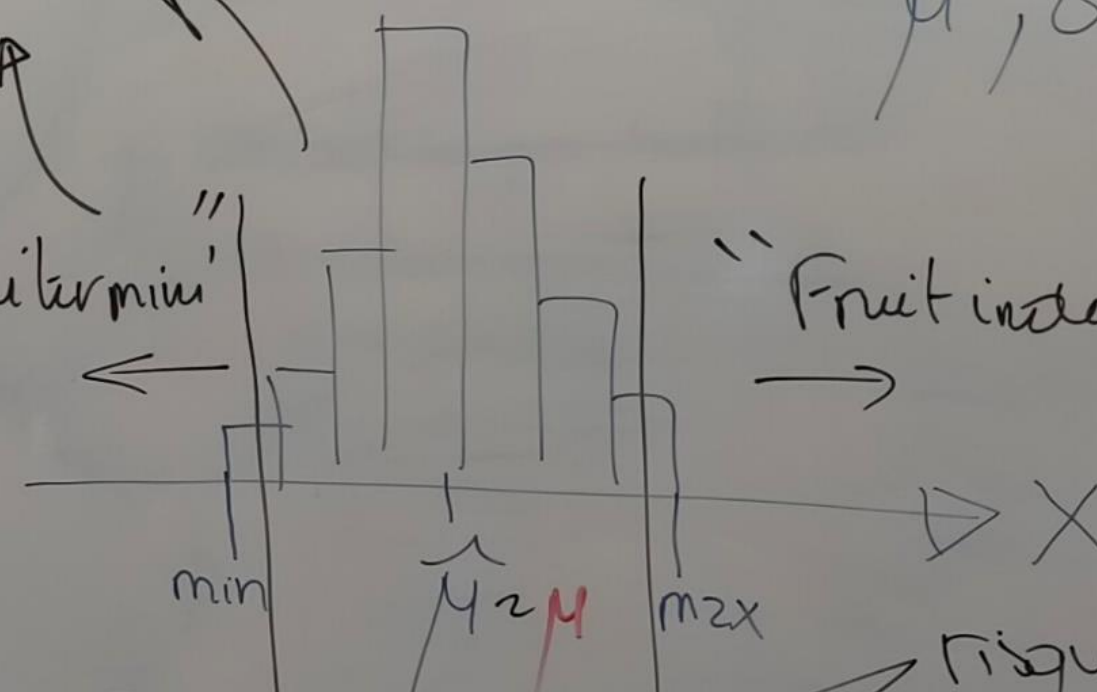
→ on commence à
résoudre l'équation de
partir des seuls données
de l'échantillon
→ et on évalue in fine
la pertinence de cette approximation

Cela est de probz de passer
d'un fruit
à l'autre (cas)

$\hat{\mu}, \hat{\sigma}^* \approx \sigma$

"Fruit
indéterminé"

"Fruit indéterminé"



algo calcul
→ μ
→ σ*
→ Binf
→ Bsup

variable
 $\epsilon = 1.96$
 $\alpha = 5\%$
 $\epsilon = 2.54$
 $\alpha = 1\%$

Algorithme de détection fruits (sélection de l'information retenue)

```
df3[df3['fruit'] == "orange"]
```

```
epsilon=2.54
```

```
df3[df3['fruit'] == "orange"].diametre.mean()
```

```
-epsilon*df3[df3['fruit'] == "orange"].diametre.std()
```

Algorithme de détection fruits (vers un apprentissage automatique)

epsilon=2.54

for i in range(len(df3)):

if (df3.couleur[i]=="orange" and

df3.diametre[i]>=df3[df3['fruit'] == "orange"].diametre.mean()

-epsilon*df3[df3['fruit'] == "orange"].diametre.std()

and df3.diametre[i]<df3[df3['fruit'] == "orange"].diametre.mean()

+epsilon*df3[df3['fruit'] == "orange"].diametre.std())) :

df3.predicted[i]="orange"

Algorithme de détection fruits (vers un apprentissage automatique)

```
epsilon=2.54
for i in range(len(df3)):
    if (df3.couleur[i]=="orange" and
        df3.diametre[i]>=df3[df3['fruit'] == "orange"].diametre.mean()
        -epsilon*df3[df3['fruit'] == "orange"].diametre.std()
        and df3.diametre[i]<df3[df3['fruit'] == "orange"].diametre.mean()
        +epsilon*df3[df3['fruit'] == "orange"].diametre.std()) :
        df3.predicted[i]="orange"
```

Quels sont vos commentaires et autres remarques ?

[Objectifs](#)[Data Sciences](#)[Documents Classe](#)[Bases de Données](#)[Analyse Statistique](#)[Un peu d'R](#)[Machine Learning](#)[Le coin des cracks](#)[Fichiers de Données](#)[Pour les plus curieux](#)[Le Coin des Jeux](#)[Les Défis du L3](#)[Projets](#)[Je vous recommande](#)[Divers](#)

Machine Learning

Description

Cette page s'alimentera des documents et de vidéos concernant spécifiquement le Machine Learning, l'algorithmique, le langage Python et la programmation en Python..

Téléchargement de Logiciels et installation de configurations

- Fichier d'installation d'environnement et d'éditeurs de code python user friendly, dont JupyterLab > Winpython
- Fichier d'installation d'environnement et d'éditeurs de code python user friendly, dont JupyterLab > Anaconda

Prise en main de l'environnement et du langage python (*analyse du fichier Arbres.csv*)

Etape de prise en main de Python : un peu de stats descriptives sur les données du fichier Arbres.

- Deux mots concernant la formation en Machine Learning et Python
- Dias "mains à la pâte" - Premiers pas et prise en main de Python
- Le fichier Arbres.csv
- Notre premier programme python (terminaison .ipynb , avec le résultat de l'exécution des cellules)
 - Notre premier programme python
 - Deuxième version du programme : reconnaissance des oranges; matrice de confusion et accuracy
- Notre deuxième programme python (terminaison .ipynb) **avec des commentaires détaillés**

Machine Learning / Apprentissage Automatique > *Reconnaissance de fruits*

Nous allons comprendre ce qu'est le Machine Learning en travaillant par étapes, les mains à la pâte. Soyez patiente. Paris ne s'est pas fait en un jour!

Etape 4 - L'apprentissage automatique commence à apparaître > l'IA que vous réalisez va se servir des données du dataframe pour construire l'intervalle de fluctuation du diamètre des oranges (cette démarche peut ensuite aisément s'étendre aux autres fruits). Dans un deuxième temps, nous allons apprendre à scinder en deux le dataset : le jeu d'apprentissage sur lequel sera échafaudé notre modèle de détection des fruits et le jeu test permettant de confirmer les performances de l'IA.

- Version construisant l'intervalle de fluctuation du diamètre des oranges à partir des données du dataframe
- Scission du dataset en jeu d'apprentissage et jeu test, le modele étant construit sur le jeu d'apprentissage