

# ALGORITHME



DES K PLUS PROCHES VOISINS

# KNN



## Partie 1

Les 'k plus proches voisins' ou K-Nearest  
Neighbors en anglais (d'où l'appellation)



# KNN



C'est un algorithme d'apprentissage automatique (Machine Learning) utilisé pour la reconnaissance (visuelle ou vocale) et les problèmes de classification et de régression.

# Principe de l'algorithme



# Principe



Pour fixer les idées, nous allons prendre un exemple.

Nous voulons élaborer un logiciel capable de reconnaître la race d'un chien.

Du point de vue informatique, on appelle la race « CLASSE », et nous voulons faire une « CLASSIFICATION » des chiens. C'est à dire reconnaître la race d'un chien à partir de caractéristiques de ce chien.





# Principe



Avant tout, pour pouvoir décider de la classe du chien, il convient d'avoir une base de données :

Chien	Taille	Poids	Sexe	Longueur des poils	Forme des oreilles	Couleur	Longueur museau	Etc...
Berger blanc	58	44	M	48	triangle	blanc	10	
Bernois	71	58	M	65	arrondie	tricolore	9	
Berger blanc	60	45	M	50	triangle	blanc	11	

# Principe



Chien	Taille	Poids	Sexe	Longueur des poils	Forme des oreilles	Couleur	Longueur museau	Etc...
Berger blanc	58	44	M	48	triangle	blanc	10	
Bernois	71	58	M	65	arrondie	tricolore	9	
Berger blanc	60	45	M	50	triangle	blanc	11	

Plus la base de donnée est importante, meilleur sera le résultat.



# Principe



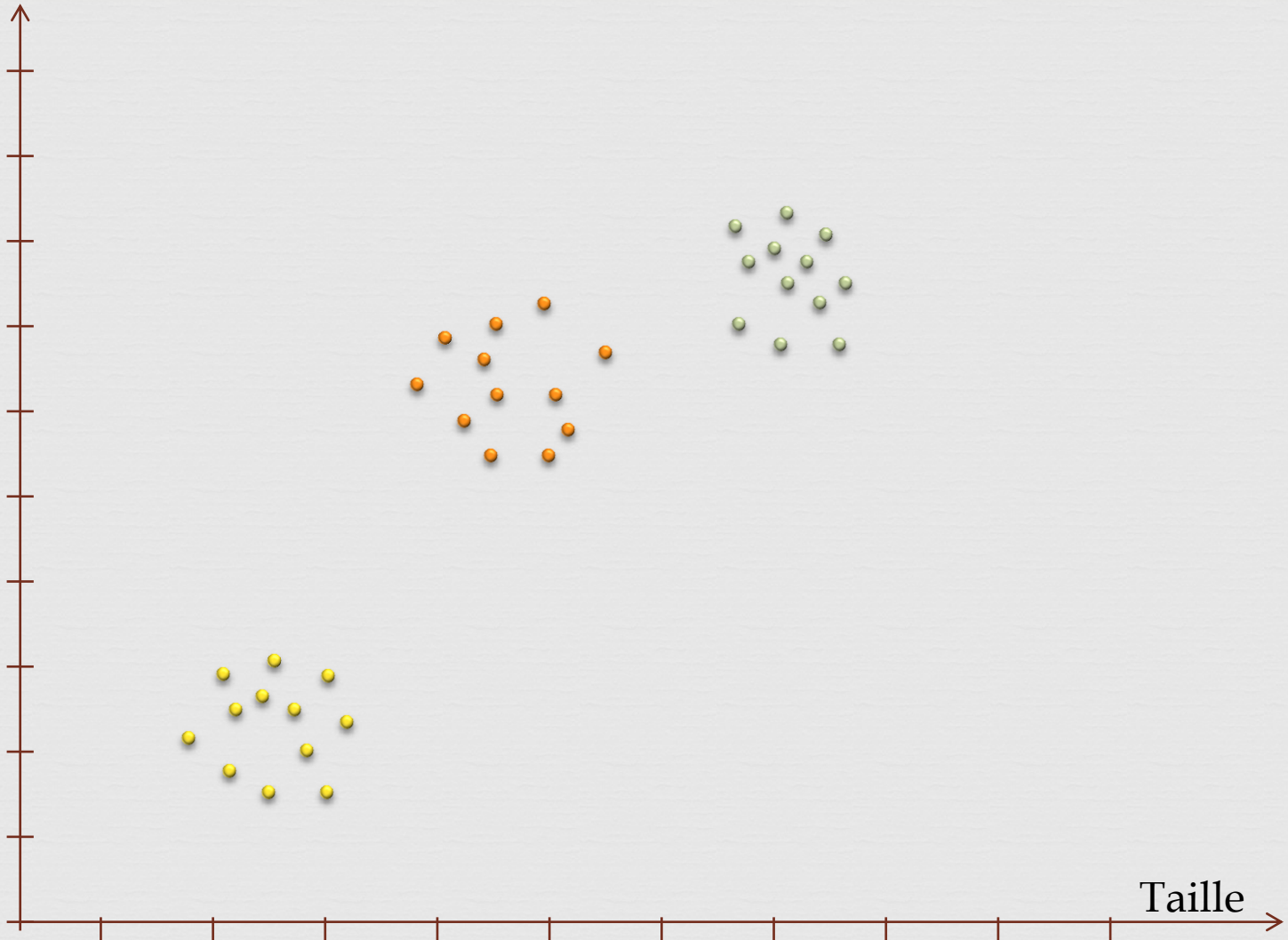
- ✧ Pour simplifier, nous allons nous occuper d'éléments ayant 2 caractéristiques :
- ✧ on peut par exemple trier des races de chien seulement par le poids et la taille

# Principe



☞ Regardons ce que la base de données donne sous forme d'un graphique

Poids



Taille

Teckel

Berger  
blanc

Bernois

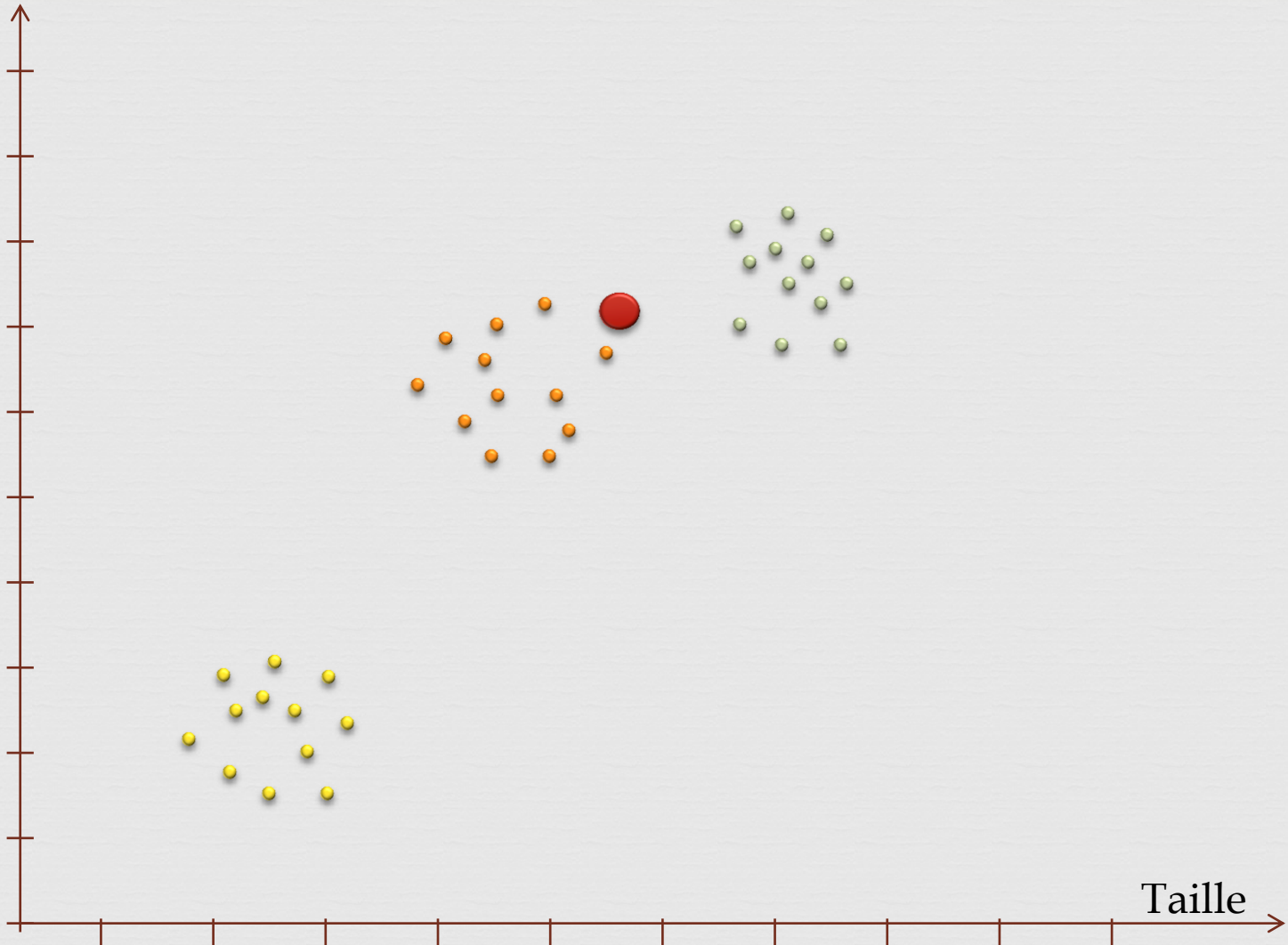


# Principe



☞ Une fois la base de données constituée, il est temps maintenant d'essayer de reconnaître la classe d'un nouveau chien.

Poids



Taille

Teckel

Berger blanc

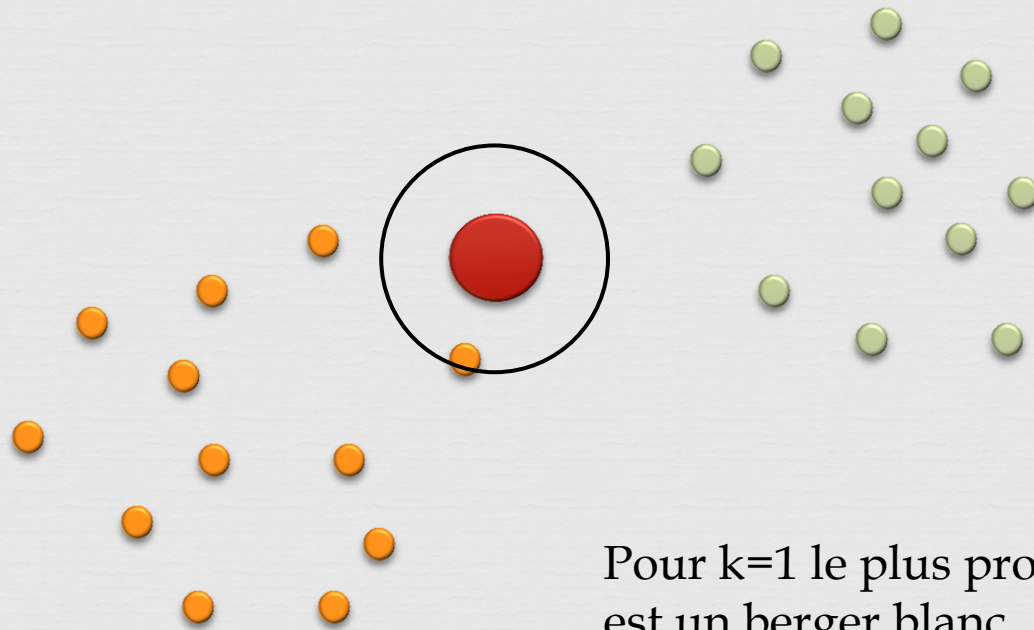
Bernois

# Principe

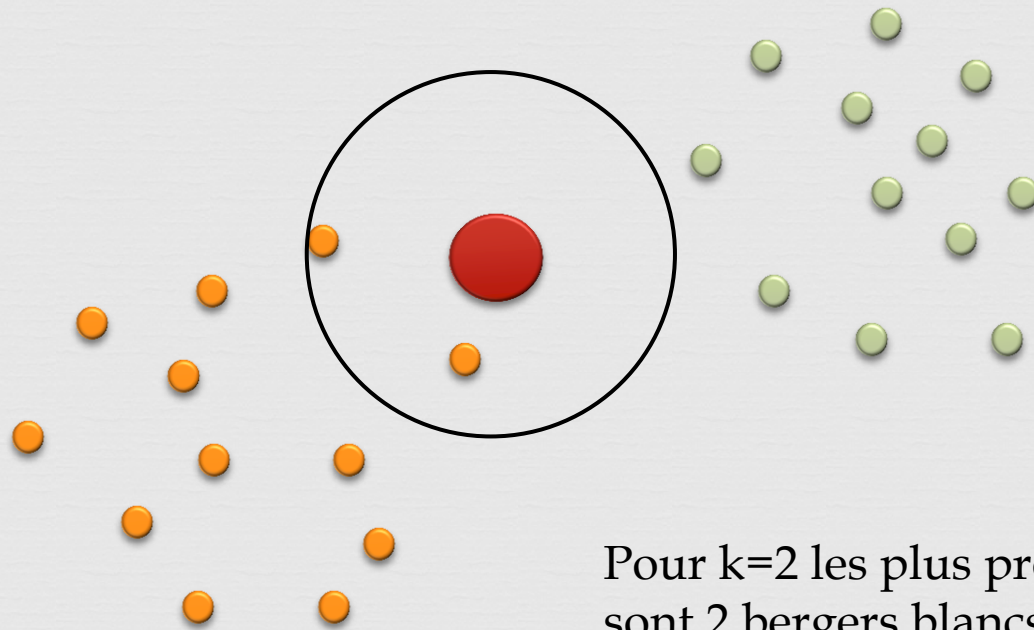


- Il convient maintenant de regarder les voisins graphiques de notre chien.
- Faisons un zoom et regardons les distances entre notre chien et ses voisins.

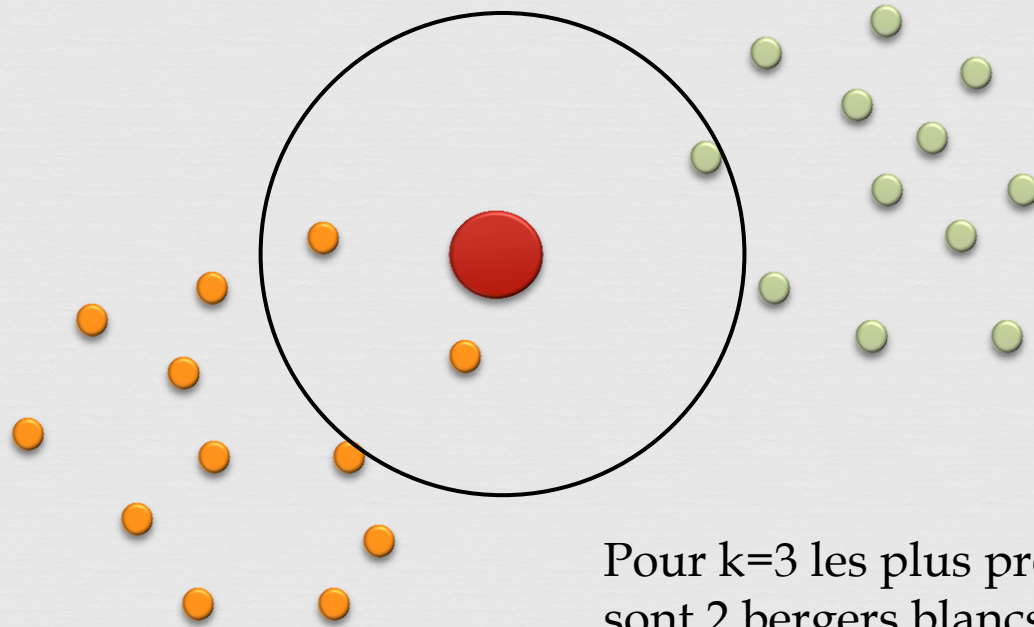




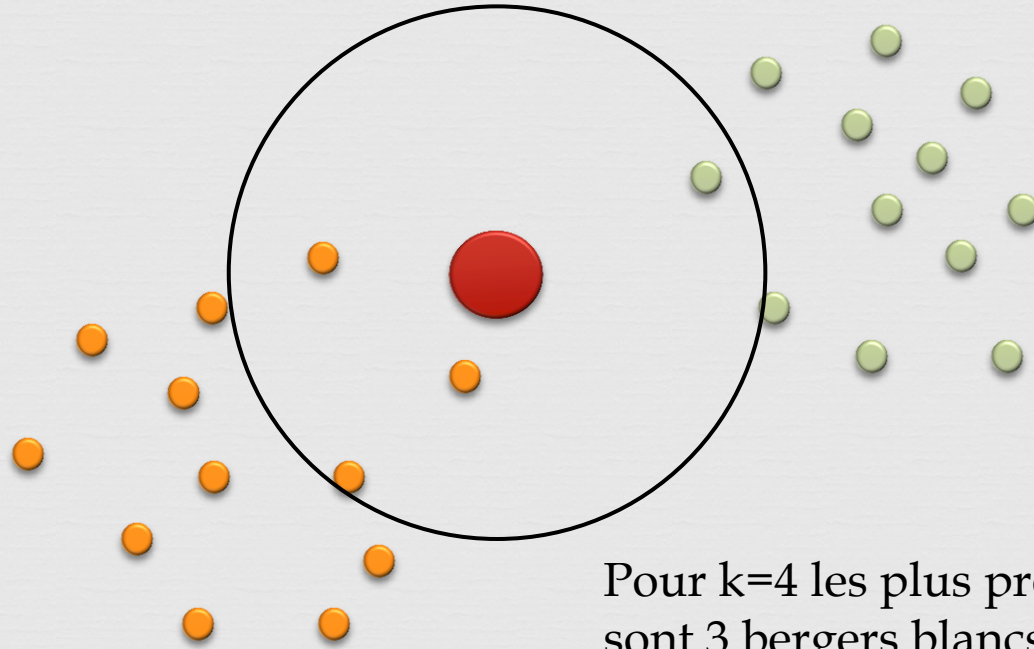
Pour  $k=1$  le plus proche voisin est un berger blanc



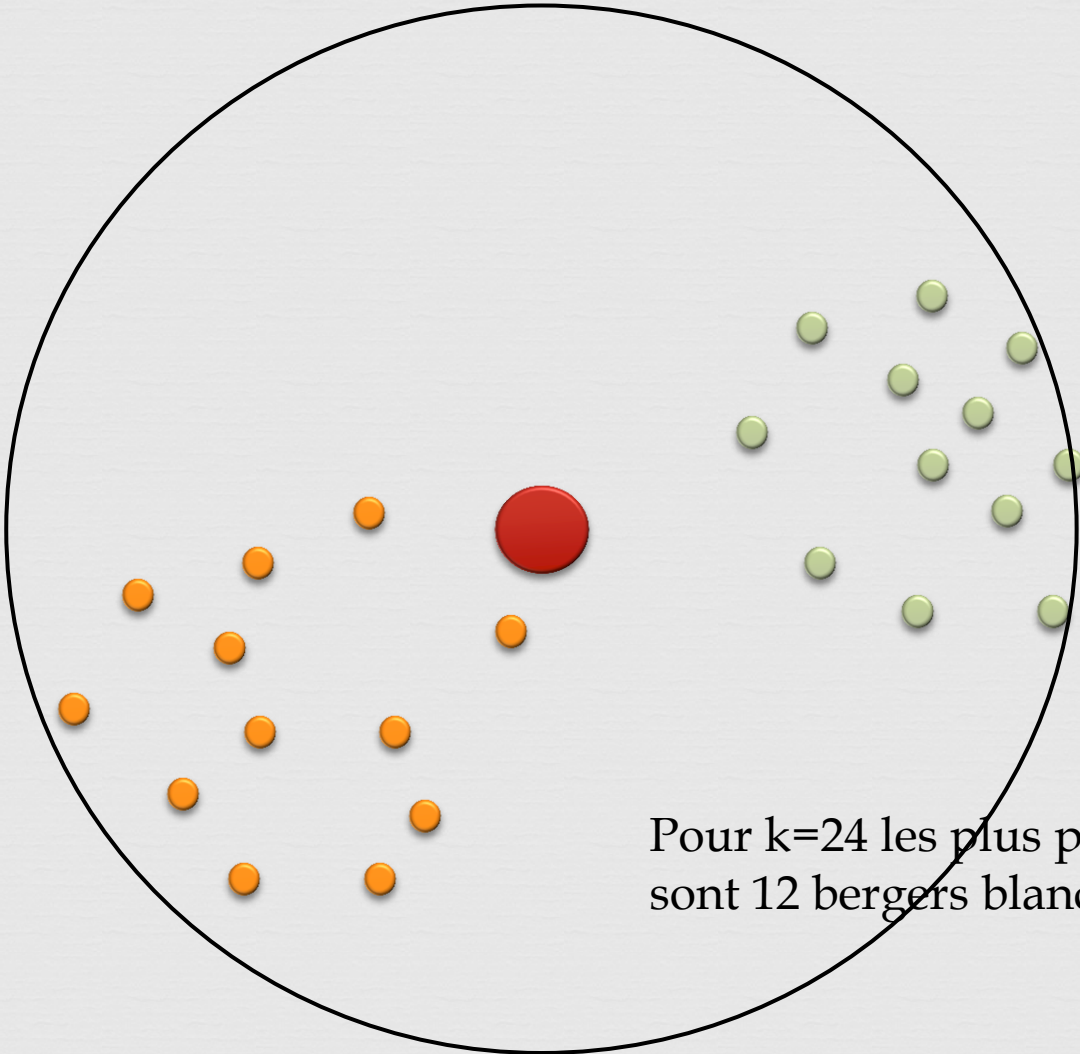
Pour  $k=2$  les plus proches voisins  
sont 2 bergers blancs







Pour  $k=4$  les plus proches voisins  
sont 3 bergers blancs, 1 Bernois



Pour  $k=24$  les plus proches voisins  
sont 12 bergers blancs, 12 Bernois

# Principe



⌘ Nous voyons qu'en fonction du nombre de voisins, le résultat change. Il va donc falloir trouver  $k$  pas trop petit et pas trop grand pour que la réponse soit la meilleur possible.



# Travail pour voir si cela est compris

---

## Question 1

Trouver une représentation des nuages de points et du chien à classifier qui donne :

- Pour  $k=1$  Bernois : 1    Berger : 0
- Pour  $k=3$  Bernois : 2    Berger : 1
- Pour  $k=5$  Bernois : 2    Berger : 3

# Travail pour voir si cela est compris

---

## Question 2

Calcul de la distance du chien aux autres.

Dans un problème plan la distance est simple à obtenir à l'aide du théorème de Pythagore.

Soit chien\_a\_classifier (taille=66,poids=52)

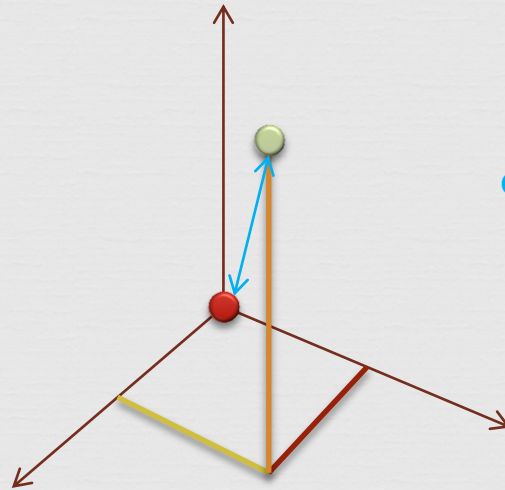
Et notre liste\_de\_chiens = [[bernois,70,58],  
[bernois,69,58],  
[bernois,70,55],  
[berger,65,45],  
[berger,64,41],  
[berger,63,42]]

Ecrire un programme python permettant de trouver la distance entre le chien à classifier et les chiens de la liste.

# Principe de la distance



- En fait la distance dans un plan (2 dimensions) est relativement simple, dans un espace de dimension 3 on peut toujours utiliser Pythagore



$$d^2 = \Delta x^2 + \Delta y^2 + \Delta z^2$$

# Principe de la distance



☞ Pythagore peut même se généraliser à des espaces de dimension plus grande :

$$d^2 = \Delta a^2 + \Delta b^2 + \Delta c^2 + \dots + \Delta y^2 + \Delta z^2$$

☞ Même si cela devient un peu abstrait



# Principe de la distance



- ❧ Là où cela devient plus complexe c'est par exemple lorsque l'on veut tenir compte de la forme des oreilles d'un chien. Là, Pythagore ne peut plus s'appliquer et il faudra trouver autre chose.
- ❧ Vous verrez cela plus tard si vous continuez dans l'informatique.

# KNN



Fin de la partie 1