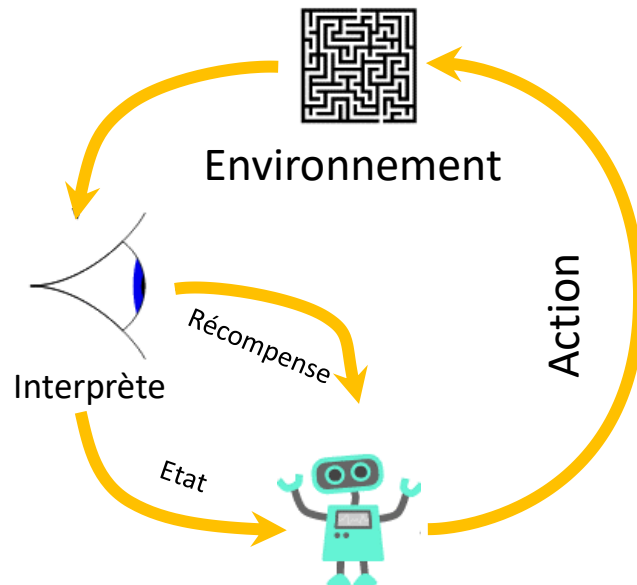


# TP : Introduction au Machine Learning

Télécharger/ installer WEKA si ce n'est pas déjà fait, en attendant

## Partie I : ML par Renforcement (30 min)



Se rendre sur [http://rednuht.org/genetic\\_cars\\_2/](http://rednuht.org/genetic_cars_2/)

Modifier les hyperparamètres et **laisser les simulations tourner pendant que vous faites les autres parties.**

1. Quels sont les paramètres :
  - a. du modèle ?
  - b. de l'algorithme ?
  - c. de l'environnement ?
2. Quelle est la récompense ?
3. Quels sont les effets attendus des paramètres de l'algorithme sur le résultat ?

## Partie II : Régression linéaire (30 min)

Données :

I		II		III		IV	
$x$	$y$	$x$	$y$	$x$	$y$	$x$	$y$
10,0	8,04	10,0	9,14	10,0	7,46	8,0	6,58
8,0	6,95	8,0	8,14	8,0	6,77	8,0	5,76
13,0	7,58	13,0	8,74	13,0	12,74	8,0	7,71
9,0	8,81	9,0	8,77	9,0	7,11	8,0	8,84
11,0	8,33	11,0	9,26	11,0	7,81	8,0	8,47
14,0	9,96	14,0	8,10	14,0	8,84	8,0	7,04
6,0	7,24	6,0	6,13	6,0	6,08	8,0	5,25
4,0	4,26	4,0	3,10	4,0	5,39	19,0	12,50
12,0	10,84	12,0	9,13	12,0	8,15	8,0	5,56
7,0	4,82	7,0	7,26	7,0	6,42	8,0	7,91
5,0	5,68	5,0	4,74	5,0	5,73	8,0	6,89

Dans un tableur, réaliser une régression linéaire sur chaque jeu de donnée.

1. Quelle est l'équation de la régression linéaire pour chaque série ?
2. Quel est le coefficient de corrélation pour chaque série ?
3. Que remarque-t-on ? Que peut-on proposer ?
4. Tracer la différence entre la prédiction et la valeur  $y$  de chaque série par rapport à  $x$
5. Que peut-on remarquer ?

## Partie III : WEKA (30 min)

Ouvrir [WEKA](#) aller dans l'Explorer :



Ouvrir un jeu de données (Open Files) à l'adresse C:\Program Files\Weka-3-8\data)

Données : **hypothyroid.arff**

### 1. Exploration des données :

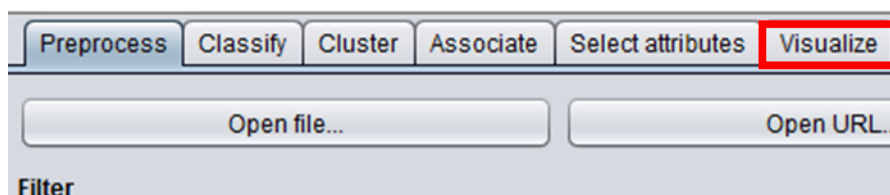
- Combien y a-t-il d'exemples ?
- Combien y a-t-il de classes ?

Cliquez sur les différents noms de la liste

- Combien y a-t-il de descripteurs et quel est le type des données ?
- Ces données sont-elles complètes ? Quels problèmes cela pourraient poser ?
- Si l'on regarde la répartition des classes pour chaque descripteur que peut-on remarquer ? Quels problèmes cela pourraient poser ?
- Si je veux réaliser une validation croisée avec 10-fold, combien d'exemple d'entraînement ai-je dans chaque fold ?

### 2. Visualisation :

Aller dans l'onglet « Visualize », et modifier la sélection des descripteurs, la taille des points.

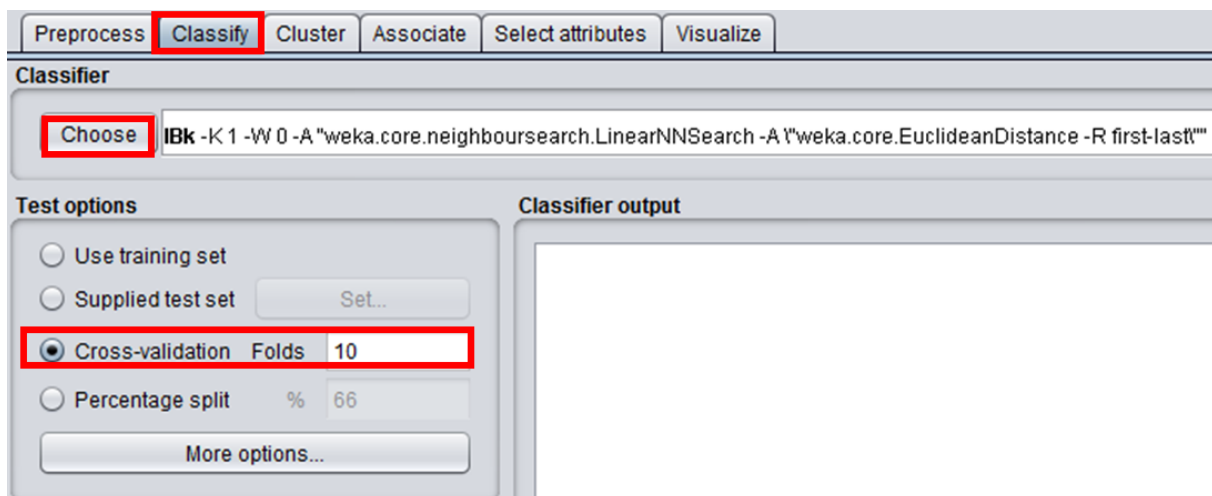




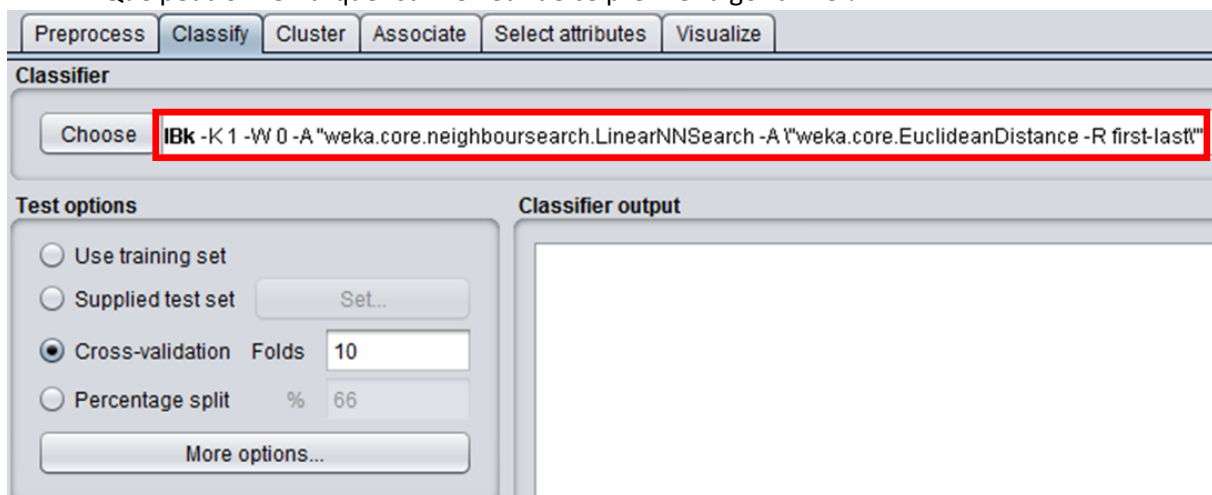
- Pouvez-vous trouver des séparations évidentes entre les classes ?
- Y a-t-il des descripteurs redondants ?

### 3. Classification :

Aller dans l'onglet « Classify », cliquer sur « Choose » et sélectionner un k plus proche voisin (lazy\IBk) puis choisir une cross validation à 10 folds et lancer l'entraînement (Start)



- Que peut-on remarquer sur l'erreur de ce premier algorithme ?



- Quels sont les hyperparamètres (double cliquer sur le champ à côté de « Choose ») de cet algorithme ?

Faire varier les hyperparamètres et réentraîner plusieurs fois :

- Quelle conséquence le choix de ces hyperparamètres peut avoir sur le résultat ?

Changer les hyperparamètres de l'algorithme pour essayer d'augmenter la sensibilité

Essayer différents algorithmes de classification (SMO, J48, REPTree) pour atteindre une sensibilité supérieure à 90% en suivant la méthodologie vue en cours.

- Est-ce que tous les descripteurs sont utilisés pour chaque algorithme ?

## Partie IV : Tracking de paupière avec dlib

Ouvrez un notebook colab :

<https://colab.research.google.com/drive/17FALFs-yF6H1IYr1PkhToC5T2m8FqYjt?usp=sharing>

Ajouter les éléments du dossier au notebook :

<https://drive.google.com/drive/folders/1Yk917KIUUs0lf49QAII3JHPVm0oY21w5?usp=sharing>

Suivre les instruction (et lancer les blocs)