TD1
Apprentissage Artificiel

Exercice 1

- D’après vous :

1. C’est quoi **l’intelligence** et comment peut-on la **matérialiser** ?
2. Différents types d’apprentissage existent, est-il possible de définir **plusieurs formes de l’intelligence** ?
3. Que faut-il intégrer à un **algorithme** pour lui **donner de l’intelligence** ?

Exercice 2

- Trouver, par **interpolation linéaire** ou polynômiale (Newton, Lagrange, …), la fonction qui passe le plus près possible par les points du plan 2D :
  - $p_1(15, 16)$
  - $p_2(3, 19)$
  - $p_3(30, 14)$
- Puis écrire un algorithme qui implémente cette fonction.

Exercice 3

Soit un vecteur descripteur $\boldsymbol{v} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \boldsymbol{x_4}, \boldsymbol{x_5})$ avec 5 variables décrivant les caractéristiques de la classe d'objets **voiture**.

1. Calculer la matrice de **covariance** et trouver **les liens entre les variables** de $\boldsymbol{v}$ pour les exemples suivants :

Objets \ Descripteur	x_1	x_2	x_3	x_4	x_5
Voiture 1 type Renault	1,4	0,7	3,5	1,62	1,02
Voiture 2 type Peugeot	1,6	0,7	3,1	1,60	1,06
Voiture 3 type Citroën	1,1	0,4	2,9	1,74	1,33
Voiture 4 type BMW	1,9	0,9	4,2	1,55	1,67
Voiture 5 type Audit	2,4	1,5	3,2	1,95	1,92

Dites quelle variable est redondante ou non.

Exercice 3

- La covariance est utilisée en statistique pour décrire la relation linéaire entre 2 variables, elle indique la *corrélation* entre ces 2 variables.
- Le calcul de la corrélation est compris entre –1 et +1, en dehors de ce rang le résultat est erroné.
- Plus la covariance :
  - est élevée (tant vers 1) entre 2 variables, plus leurs valeurs sont *parfaitement corrélées* et
  - *parfaitement inversement corrélées* lorsqu'elle tend vers –1,
  - par contre *aucune corrélation* n'existe entre ces 2 variables si elle vaut 0.

2. Calculer la distance de Mahalanobis entre des différents objets, puis interpréter ces valeurs de distances.

Exercice 3

Rappels de notions de statistiques

- Espérance d'une variable aléatoire statistique
  - Une loi de probabilité est caractérisée par les notions de *valeur centrale*, de *dispersion* et de *forme de distribution*.
  - L'espérance d'une v. a. *E(X)* est *la moyenne des valeurs possibles de X pondérées par les probabilités* associées à ces valeurs.
  - Cas discret : Si X est une v. a. discrète de loi de probabilité *(xi,pi)i* définit sur un nombre fini *n* d'évènements alors :

$$E(X)=\sum_{i=1}^{n} xi.pi$$

  - Cas continu : Si X est une v. a. continue de densité *f*, l'espérance de X est le réel E(X) défini par :

$$E(X)=\int_{-\infty}^{+\infty} x.f(x).dx$$ si cette intégrale est convergente.

  - Propriétés :

$$E(X+Y)=E(X)+E(Y) \quad \text{et} \quad E(aX)=aE(X)$$

Exercice 3

Rappels de notions de statistiques

- Variance d'une variable aléatoire

– La variance d'une v. a. *V(X)* est l'espérance mathématique du carré de l'écart à l'espérance mathématique.

– Si X est une v. a. ayant une espérance *E(X)* , alors la variance *V(X)* de *X* est le réel

$$V(X) = E(\ [X-E(X)]^2\ )$$

– Équivalente à : $V(X) = E(X2) - [E(X)]^2$

– <u>Cas discret</u> : Si X est une v. a. discrète de loi de probabilité (xi,pi)i définie sur un nombre fini *n* d'évènements alors la variance est :

$$V(X)=\sum_{i=1}^{n} (xi-E(X))^2.pi = \sum i=1n\ xi^2.pi - E(X)^2$$

– <u>Cas continu</u> : Si X est une v. a. continue donnée par sa densité de probabilité alors la variance de X est le nombre réel positif tel que :

$$V(X)=\int_{-\infty}^{+\infty} (x-E(X))^2 . f(x). dx = \int+\infty-\infty\ x^2 .f(x).dx - E(X)^2$$

Exercice 3

Rappels de notions de statistiques

- écart-type

✓ Si X est une v. a. ayant une variance V(X), on appelle écart-type de X, le réel :

$$\sigma(X)= \sqrt{V(X)}$$

✓ L’écart-type est un *paramètre de dispersion* qui s’exprime dans les mêmes unités que la v. a.

Exercice 3

<u>Rappels de notions de statistiques</u>

- Covariance de 2 variables aléatoires
  - La covariance de 2 v. a. *X* et *Y* est un indicateur de leur « liaison ».
  - Si X et Y sont 2 v. a. définies dans l'univers Ω, alors leur covariance est le réel :

$$Cov(X,Y) = E(X.Y) - E(X).E(Y)$$

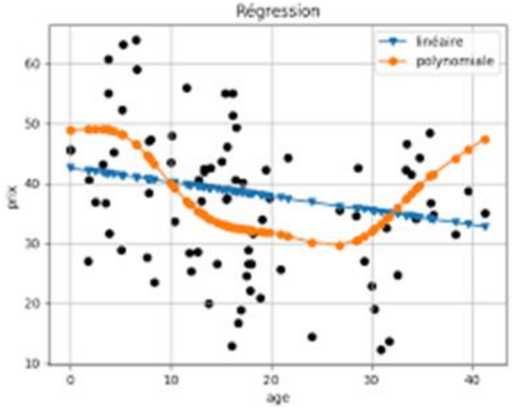
- Coefficient de corrélation entre 2 variables

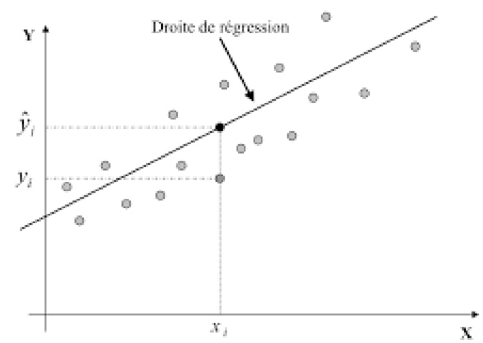
$$R(X,Y) = Cov(X,Y)/\sigma(X).\sigma(Y)$$

  - Si X et Y sont indépendantes, alors: $cov(X,Y) = 0$
  - $-1 \leq R(X,Y) \leq 1$
  - Si X et Y sont indépendantes, R=0, mais la réciproque est pas toujours vraie.

Exercice 4

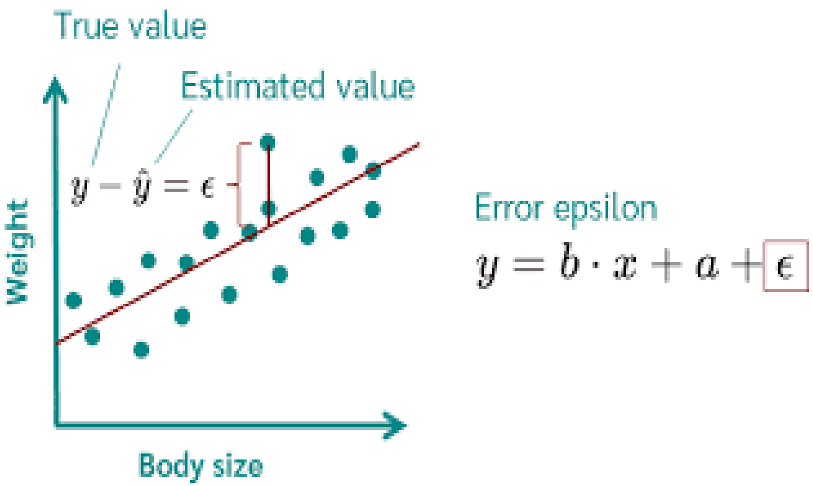
- La notion de distance est au cœur des statistiques et de l'analyse de données.
- Pour évaluer la proximité de variables (entières ou réelles) décrivant les objets, on peut représenter ces variables par des points sur un graphe et mesurer les distances entre les points (dist. géométrique), en utilisant, par exemple, la distance euclidienne.
- En statistique, la distance euclidienne n'est pas très praticable.
- Exemple, si on veut établir un modèle de régression linéaire, simple ou multiple, alors on essaye de minimiser des distances entre des observations ou prédictions (représentées par un nuage de points) et une droite (ou courbe) de régression qui résume au mieux les valeurs réelles à atteindre.

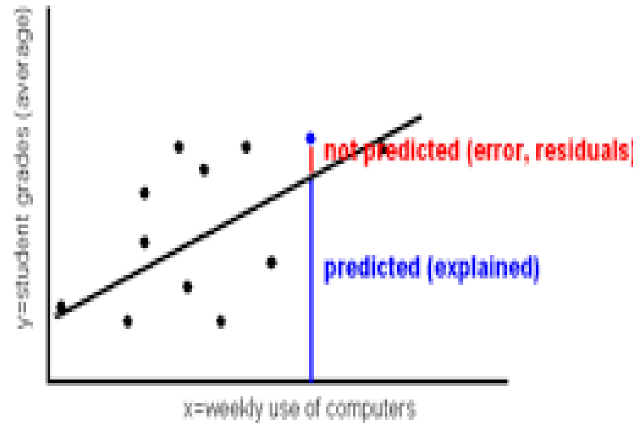


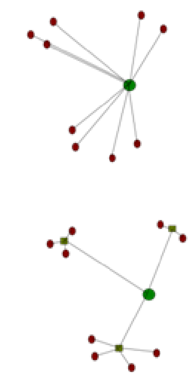


Exercice 4

- La mesure est alors la distance des moindres carrés.

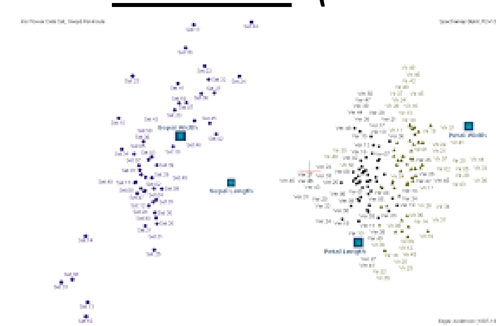






- En analyse de données, lorsque les points sont très nombreux et où plusieurs observations (nuage de points) se situent en un même endroit, nécessitant ainsi une pondération (*p*), on utilise la *notion très proche* *<u>d'inertie</u>* *(ou de moment d'inertie)*.

$$I = \sum_{i=1}^{n} p_i d_i^2$$



d_i est la distance d'un point x_i à son barycentre (du nuage) et p_i la pondération (poids ou la masse) de ce point.

Exercice 4

Evaluer la ressemblance entre des différents objets du tableau précédent à travers :

- ✓ la distance Euclidienne,
- ✓ la distance des moindres carrées,
- ✓ leurs valeurs d'inertie (en définissant une pondération adéquate),

Enfin, interpréter ces valeurs de distances.