

Enseignantes : A. NAJJAR- I. BEN OTHMEN - F. JENHANI TP : Z. ZOUAGHIA - I. BEN AYCHA	TP1 Machine Learning	Classe : 3ème GLSI
---	---------------------------------------	---------------------------

Partie 1- Rappel python

Exercice

- 1- Ecrire un programme qui permet de calculer le factoriel d'un nombre introduit au clavier. La valeur introduite par l'utilisateur ne doit être acceptée que si elle est positive.
- 2- Ecrire un programme qui permet de résoudre une équation de second degré sous la forme $ax^2+bx+c=0$. Il faut alors saisir trois nombres flottants correspondants aux coefficients et les stocker dans des variables a, b et c.
- 3- Écrire une fonction `nombreOccurrences(caractere, mot)` qui prend en argument un caractère x et une chaîne de caractère mot et qui renvoie le nombre de fois où le caractère x est présent dans mot. Par exemple, si mot est le mot «java», `nombreOccurrences('a', mot)` vaut 2.
- 4- Ecrire un programme qui saisit une suite d'entiers qui se termine par "=" et affiche leur somme et leur moyenne ?

Partie 2- Exploration des données

Plusieurs bases dédiées pour l'apprentissage automatique peuvent être trouvée sur le web. Consulter la page web suivante :

<https://archive.ics.uci.edu/ml/datasets.html>

Dans ce TP, on se basera sur la base "**pima-indians-diabetes.data.csv**".

Pima est un groupe d'indien de l'Amérique (**Native Americans**) vivant à Arizona. Ils sont sujets de plusieurs études car le changement dans leurs habitude alimentaire a favorisé qu'ils soient atteints de diabète de type 2.

La base inclus 767 femmes décrites par les 8 attributs suivants.

1. Nombre de grossesses (**NumTimesPrg**)
2. Concentration du plasma en glucose (**PIGlcConc**)
3. Tension artérielle (**BloodP**)
4. Épaisseur du pli cutané des triceps (**SkinThick**)
5. Taux d'insuline (**TwoHourSerIns**)
6. Indice de masse corporelle (**BMI**)

7. *Fonction pédigrée du diabète* : hérédité (**DiPedFunc**)
8. Age (**age**)

La dernière colonne "**HasDiabetes**" de la base (9^{ème} colonne) indique si la personne est diagnostiqué (1) de diabète ou pas (0).

On souhaite d'abord lire et inspecter les données dans la base. La base "**pima-indians-diabetes.data.csv**" est fournie avec le format **CSV (Comma Separated Values, valeurs séparées par des virgules)**. Il s'agit du format le plus commun dans l'import et l'export de feuilles de calculs et de bases de données.

Exercice

1. Lire le contenu du fichier "**pima-indians-diabetes.data.csv**".
2. Attribuer à chaque colonne le nom de l'attribut correspondant.
3. Afficher la structure de la base.
4. Afficher les 10 premières lignes de la base.
5. Afficher seulement les valeurs de la tension artérielle (attribut "BloodP") pour la totalité des individus.
6. Obtenir des statistiques sur les individus dans la base. Analyser certains résultats obtenus
7. Créer une matrice qui ne contient que les données des attributs (séparer les valeurs des attributs de leurs classes d'appartenances).