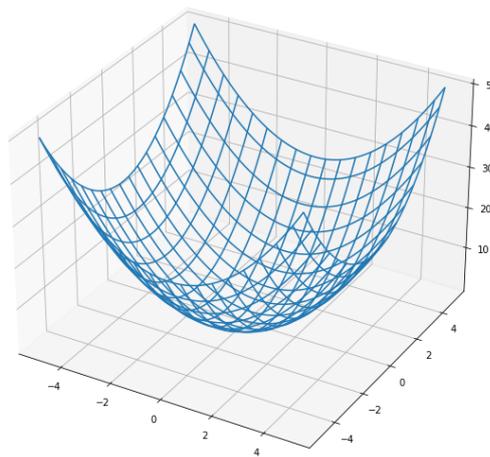


Nom :
 Prénom :
 Groupe :

Exercice 1 : Descente de gradient

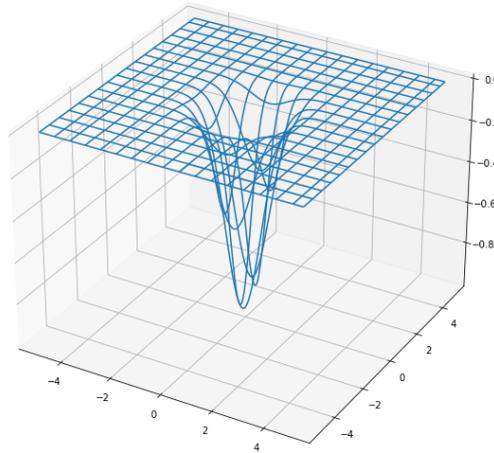
1. Quel est l'objectif de l'algorithme de la descente de gradient ? Est-ce que cet objectif est toujours atteint ? Rappeler les différentes étapes de la descente de gradient en donnant l'algorithme en pseudo-code. Détailler l'intérêt du pas d'apprentissage et justifier la condition d'arrêt.
2. On se donne la fonction $f(x, y) = x^2 + y^2$, calculer les dérivées partielles de cette fonction. Donner une implémentation de l'algorithme de la descente de gradient pour cette fonction. Faites tourner cet algorithme en commençant au point $p(0.1, 0.2)$, avec $\epsilon = 0.04$ et $\eta = 0.15$. On utilisera la norme euclidienne du gradient atténué pour calculer le critère d'arrêt. Portez un soin à la modularité, en écrivant la fonction `gradientF(x,y)`.

FIGURE 1 – $f(x, y) = x^2 + y^2$



3. On voudrait maintenant appliquer cette technique à la fonction $g(x, y) = -\exp(-(x^2 + y^2))$. Donner les dérivées partielles de cette fonction. On rappelle que la dérivée de $\exp(u)$ est $u' \exp(u)$. Adapter le code de la fonction précédente.
4. Bonus : A votre avis, est-il nécessaire de reprogrammer la descente de gradient pour la fonction $g(x, y)$ ou bien existe-t-il une solution basée sur la fonction $f(x, y)$? Justifiez votre réponse.

FIGURE 2 – $g(x, y) = -\exp(-(x^2 + y^2))$



Exercice 2 : k -NN

On dispose d'un jeu de données composé de 5 observations $X_i, i \in \{1, 5\}$. Pour chacune de ces observations, 5 attributs numériques $A_i, i \in \{1, 5\}$ sont mesurés. Ceux-ci sont rapportés dans le tableau ci-dessous avec la classe d'appartenance de chaque observation :

Instances	X_1	X_2	X_3	X_4	X_5	X_6
A_1	2	1	3	1	0	3
A_2	7	5	8	3	3	4
A_3	3	5	5	7	2	1
A_4	7	5	6	10	9	0
Class	1	2	1	2	2	1

On souhaite déterminer la classe d'appartenance d'une nouvelle observation X_7 dont les attributs numériques A_1 à A_4 sont $\langle 0, 7, 4, 7 \rangle$ grâce à l'algorithme des k plus proches voisins (k -NN). On utilisera la distance de Minkowski à l'ordre $p = 1$.

Pour rappel, si $u = (u_1, u_2, \dots, u_n)$ et $v = (v_1, v_2, \dots, v_n)$ alors la distance de Minkowski à l'ordre p entre u et v est définie comme suit :

$$d(u, v) = \left(\sum_{i=1}^n |u_i - v_i|^p \right)^{\frac{1}{p}}$$

1. Appliquez l'algorithme du k -NN avec $k = 1$ et $k = 5$, en donnant le détail des calculs.
2. Discutez de l'impact de k sur la prédiction.

Exercice 3 : Régression linéaire

Afin d'évaluer les liens entre un certain nombre de caractéristiques et le score (`score_jeux`) obtenu à un jeu, on a collecté les informations suivantes :

- Age du joueur (`age`),
- Nombre d'heures (par jour) passées au travail (`heures_travaillees`)

- Nombre d'heures (par jour) passées à jouer (`nombre_heures_jeux`)
 - Puissance de l'équipement utilisé pour jouer, il s'agit d'une évaluation calculée sur la base des caractéristiques de l'équipement (RAM, CPU, GPU , etc) utilisé pour jouer (`puissance_equipement`).
- On fait donc une régression linéaire et on obtient le rapport suivant :

OLS Regression Results

Dep. Variable:	score_jeu	R-squared:	0.352
Model:	OLS	Adj. R-squared:	0.325
Method:	Least Squares	F-statistic:	12.92
Date:	Tue, 25 Oct 2022	Prob (F-statistic):	1.94e-08
Time:	13:40:35	Log-Likelihood:	-635.37
No. Observations:	100	AIC:	1281.
Df Residuals:	95	BIC:	1294.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
age	36.5633	16.511	2.215	0.029	3.785	69.341
nombre_heures_jeux	94.7956	14.360	6.601	0.000	66.287	123.304
heures_travaillées	11.1547	13.565	0.822	0.413	-15.775	38.084
puissance_equipement	41.2936	16.052	2.572	0.012	9.425	73.162
intercept	-22.9032	14.448	-1.585	0.116	-51.587	5.781

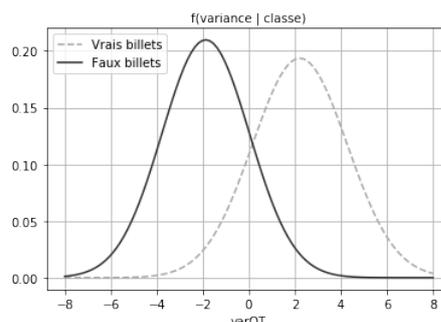
1. Interpréter la première partie du rapport. Le modèle est-il pertinent ?
2. Quelles sont les variables qui ont un impact sur la variable `score_jeux` ? Justifiez votre réponse.
3. Quel est le modèle de régression obtenu ? Donnez les prévisions du score pour un joueur âgé de 18 ans, ne travaillant pas, jouant en moyenne 2h par jour et ayant un équipement dont la puissance est estimée à 10.

Exercice 4 : Classifieur de Bayes

Dans cet exercice, on considère un dataset¹ dont les données ont été calculées à partir d'images de vrais et de faux billets de banque.

Ce dataset contient 4 descripteurs par billet ainsi que la classe de chaque billet (vrai billet / faux billet). Nous allons considérer un seul de ces descripteurs, la variance de la transformée en ondelette, appelé `varOT` dans la suite. On cherche donc à distinguer les vrais des faux billets en se basant sur cet unique descripteur.

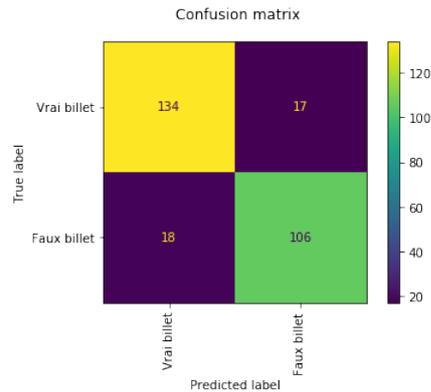
1. On veut tester le classifieur ML (maximum de vraisemblance) sur le dataset présenté plus haut en utilisant 20% des exemples pour le test et 80% pour l'entraînement.
On estime donc sur l'ensemble d'entraînement les paramètres de la loi normale suivie par `varOT` ce qui donne les courbes de densités de probabilités ci-dessous :



1. <https://archive.ics.uci.edu/ml/datasets/banknote+authentication>

Comment sera classé un billet avec $varOT = -4$? Un billet avec $varOT = 4$? Quelle est la frontière de décision ?

2. En appliquant ce classifieur à l'ensemble de test, on obtient les résultats exprimés par la matrice de confusion ci-dessous :



Calculez la valeur de l'accuracy à partir de ces résultats.

3. Quel est environ le nombre d'exemples du dataset ? Quelle est approximativement la proportion de faux billets dans ce dataset ? En réalité le nombre de faux billets en circulation est (très) inférieur à 0.1%. Le dataset est-il représentatif de la répartition réelle des vrais et faux billets ?
4. On utilise maintenant le fait que 0.1% des billets sont faux et on fait les tests sur une base respectant cette proportion. Comment évolue la frontière de décision par rapport au premier classifieur ? Le classifieur obtenu sera-t-il plus performant, au sens de l'accuracy, que le premier ? Concluez sur la pertinence de l'accuracy quand on a deux classes déséquilibrées.

FIN.