

Nom :

Prénom :

Groupe :

Exercice 1 : Descente de gradient

Soit une fonction $f(x, y)$, $f'_x(x, y)$ sa dérivée par rapport à x et $f'_y(x, y)$ sa dérivée par rapport à y . Le vecteur gradient de f au point (x, y) est $(G_x = f'_x(x, y), G_y = f'_y(x, y))$. On cherche un minimum de la fonction f en utilisant l'algorithme de descente du gradient donné en partie ci-dessous :

$$x_{cur} = x_0$$

$$y_{cur} = y_0$$

$$G_x = f'_x(x_{cur}, y_{cur})$$

$$G_y = f'_y(x_{cur}, y_{cur})$$

Tant que $(\sqrt{G_x^2 + G_y^2} > \epsilon)$

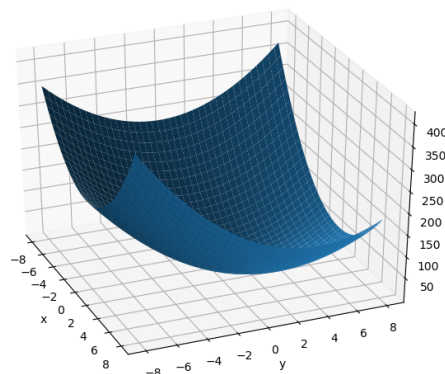
$$x_{cur} = \dots$$

$$y_{cur} = \dots$$

$$G_x = f'_x(x_{cur}, y_{cur})$$

$$G_y = f'_y(x_{cur}, y_{cur})$$

1. Quel est le rôle des paramètres x_0 et y_0 ?
2. Complétez l'algorithme avec la mise à jour de la position courante. Vous noterez ν le pas d'apprentissage. Expliquez le rôle du gradient dans cette mise à jour ainsi que le rôle du pas d'apprentissage. Que peut-il se passer si ν est trop petit ? S'il est trop grand ?
3. Expliquez la condition d'arrêt de la boucle en précisant le rôle du paramètre ϵ .
4. On étudie la fonction $f(x, y) = 3x^2 - xy + 2y^2 - 4x - 7y + 10$ dont la représentation graphique est la suivante :



- (a) Calculez son vecteur gradient.

- (b) En prenant comme point de départ $(x_0, y_0) = (2, 1)$, comme pas d'apprentissage $\nu = 0.2$, et en fixant $\epsilon = 0.2$, calculez les valeurs successives de (x_{cur}, y_{cur}) jusqu'à l'arrêt de l'algorithme.
- (c) Obtient-on le même résultat si on choisit $(x_0, y_0) = (10, -5)$? Justifiez votre réponse sans faire de nouveaux calculs.

Exercice 2 : k -NN

Supposons que l'on a un problème de classification qui consiste à déterminer la classe d'appartenance de nouvelles instances X_i . Le domaine de valeurs des classes possibles est $\{1, 2\}$.

Selon la base de connaissance suivante, déterminer à la main la classe de l'instance X_6 , dont les valeurs pour les attributs numériques A_1 à A_5 sont $\langle 3, 12, 4, 7, 8 \rangle$, à l'aide de l'algorithme des k -voisins les plus proches (k -NN) avec $k = 1$ puis $k = 3$. Montrez tous les calculs. Utiliser comme mesure la distance de Manhattan. Quel est l'impact du choix de k ?

Instances	A_1	A_2	A_3	A_4	A_5	Classe
X_1	3	5	4	6	1	1
X_2	4	6	10	3	2	2
X_3	8	3	4	2	6	1
X_4	1	1	4	3	6	2
X_5	2	5	1	4	8	2

Pour rappel, si $u = (u_1, u_2, \dots, u_k)$ et $v = (v_1, v_2, \dots, v_k)$ alors la distance de Manhattan entre u et v est définie comme suit :

$$d(u, v) = \sum_{i=1}^k |u_i - v_i|.$$

Exercice 3 : Régression linéaire

Dans un ensemble de 500 villes, on a collecté des informations relatives aux :

- pourcentage de la population se rendant au travail en vélo (**biking**),
- pourcentage de personnes fumant (**smoking**)
- pourcentage des personnes ayant eu une crise cardiaque (**heart.disease**) durant les 5 dernières années.

On souhaite étudier les liens entre les deux premières variables et la dernière. On fait donc une régression linéaire et on obtient le rapport suivant :

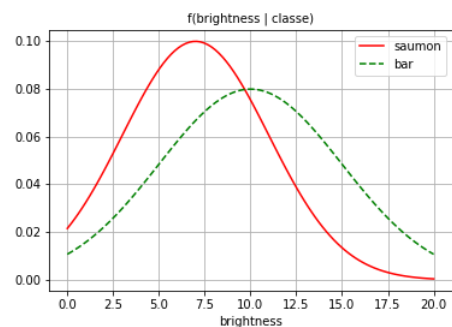
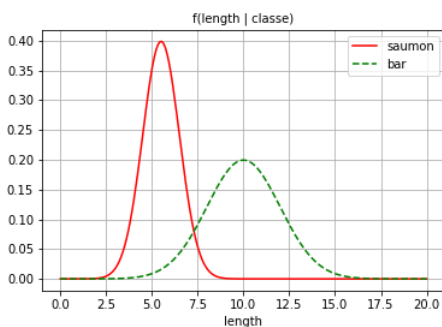
Dep. Variable:	heart.disease	R-squared:	0.980			
Model:	OLS	Adj. R-squared:	0.980			
Method:	Least Squares	F-statistic:	1.190e+04			
Date:	Tue, 09 Nov 2021	Prob (F-statistic):	0.00			
Time:	08:57:48	Log-Likelihood:	-493.68			
No. Observations:	498	AIC:	993.4			
Df Residuals:	495	BIC:	1006.			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t 	[0.025	0.975]
biking	-0.2001	0.001	-146.525	0.000	-0.203	-0.197
smoking	0.1783	0.004	50.387	0.000	0.171	0.185
intercept	14.9847	0.080	186.988	0.000	14.827	15.142
Omnibus:	2.794	Durbin-Watson:	1.917			
Prob(Omnibus):	0.247	Jarque-Bera (JB):	2.582			
Skew:	-0.141	Prob(JB):	0.275			
Kurtosis:	3.211	Cond. No.	125.			

1. Interpréter la première partie du rapport. Le modèle est-il pertinent ?
2. Les deux variables `biking` et `smoking` ont-elles un impact sur la variable `heart.disease` ? Justifiez votre réponse.
3. Quel est le modèle de régression obtenu ? Donnez les prévisions du pourcentage de personnes qui peuvent avoir une crise cardiaque dans une ville où 25% de la population se rend au travail à vélo et 30% fume.

Exercice 4 : Classifieur de Bayes

Dans cet exercice, on reconsidère le corpus vu en cours et en travaux dirigés : on cherche à distinguer entre les saumons et les bars en se basant sur deux descripteurs : la longueur (`length`) et la brillance (`brightness`).

On a donc estimé les paramètres des lois normales suivies par chacun des descripteurs et on a dessiné les courbes des densités de probabilités :



1. Si vous deviez choisir entre les deux descripteurs pour classifier les poissons, lequel est le plus pertinent, `length` ou `brightness` ? Justifiez votre réponse.
2. Si on ne considère que le descripteur `length`, comment serait classé un poisson de longueur 7.5 ? Quelle est la frontière de décision ?
3. Comment serait classé un poisson de longueur 7.5 si on sait que :
 - (a) $p = 30\%$ des poissons sont des bars.
 - (b) $p = 80\%$ des poissons sont des bars.
 - (c) Comment évolue la frontière de décision ? À partir de quelle valeur de p la classification d'un poisson de longueur 7.5 change-t-elle ?

FIN.