

Projet-1 L2-Apprentissage

Université Paris Sud

Lynda ATTOUCHE

Yacine MOKHTARI

Groupe : 01

Introduction :

L'apprentissage automatique est au cœur de plusieurs domaines : intelligence artificielle, statistiques, théorie des probabilités, science des données,..etc. Lors de l'apprentissage sur les données, les tâches sont classées par catégories (classes) basées sur comment le feedback sur l'apprentissage est donné.

Deux méthodes sont adoptées pour ce type d'apprentissage : supervisé et non supervisé.

Dans ce projet nous traitons des données MNIST dont lesquelles nous disposons d'images de lettres manuscrites et dont nous voulons déterminer automatiquement les "vraies" lettres correspondantes. Pour ce faire, nous nous appuyons sur l'apprentissage supervisé avec un algorithme de perceptron Multi classes.

Question 01 :

En faisant varier la valeur de N_a entre 1000 à 10000 avec une marge de 200 nous avons obtenu le graphe ci-dessous, la partie de code qui nous a permis de l'obtenir correspond au bout de code dans la classe "Question_1"

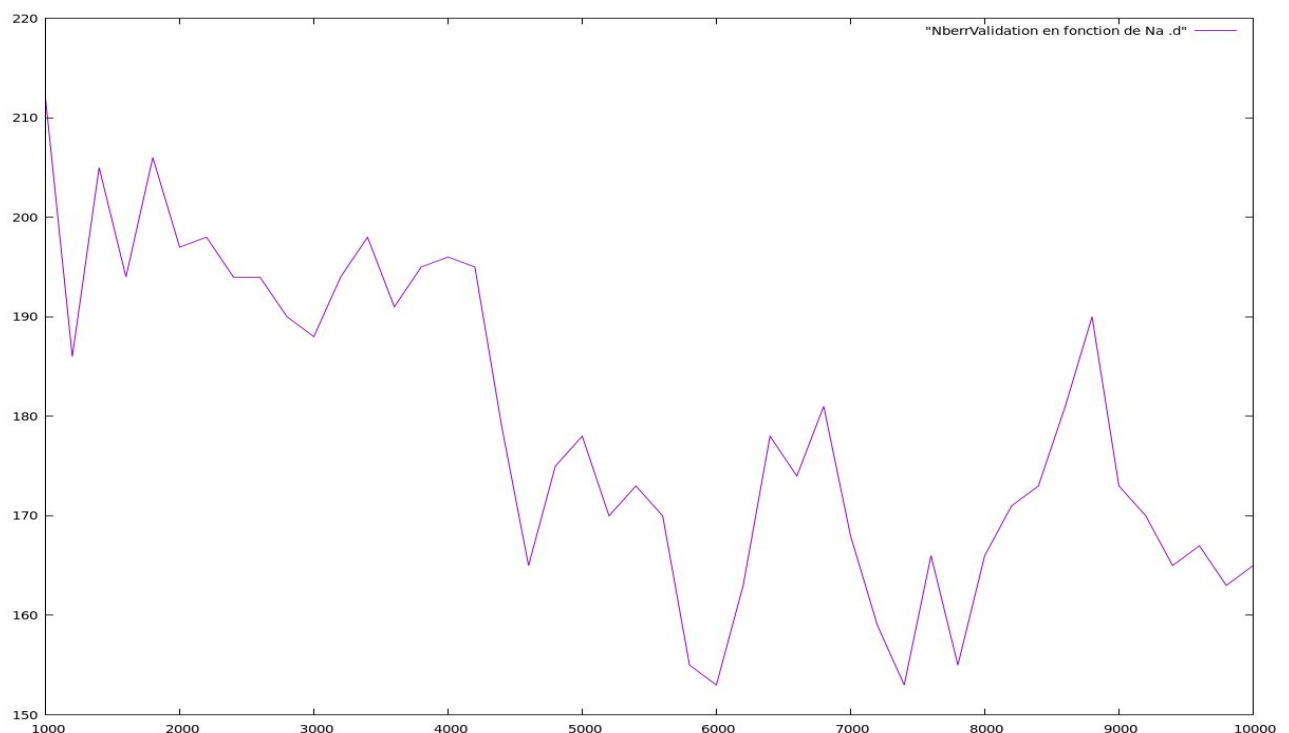


Figure 01 : Graphe illustrant le nombre d'erreurs sur l'ensemble de validation N_v en fonction de l'ensemble d'apprentissage N_a

Question 02 :

Pour choisir le meilleur taux d'apprentissage nous avons illustré le nombre d'erreurs sur l'ensemble de validation et choisi ainsi le η correspondant au minimum de tous les nombres d'erreurs. Le bout de code se trouve dans la classe "Question_2"

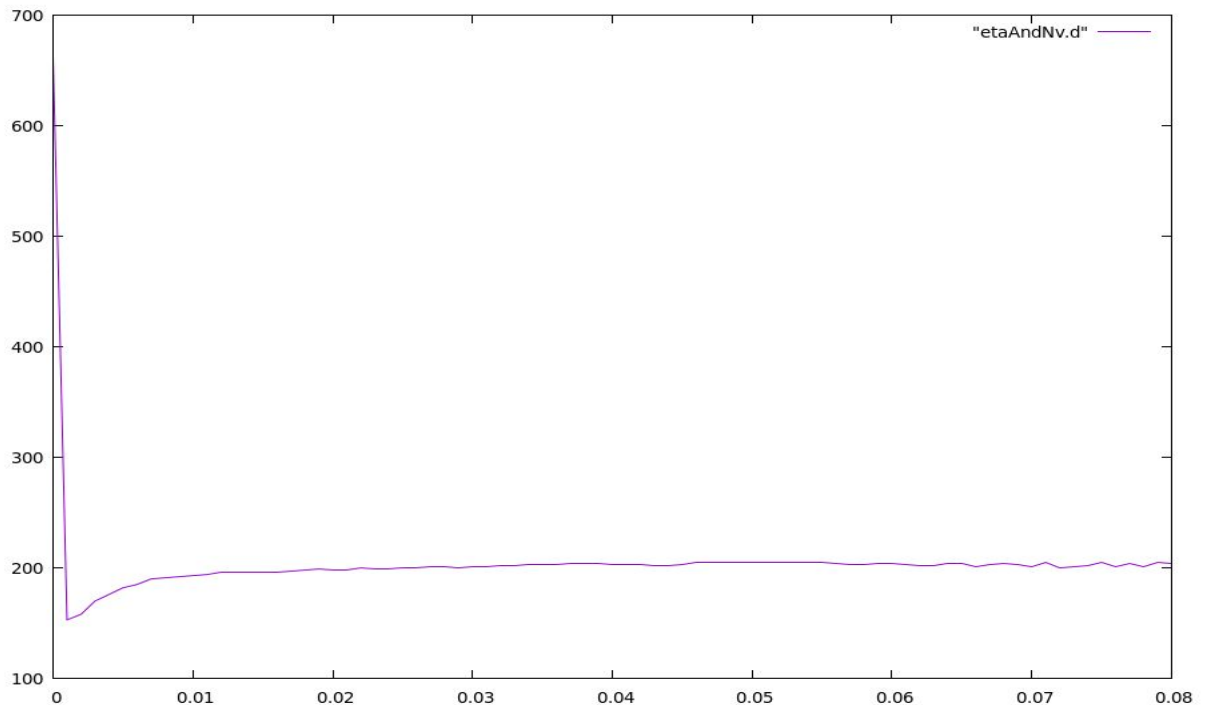


Figure 02 : Graphe illustrant le nombre d'erreurs sur l'ensemble de validation en fonction de η variant entre 0 et 0.08

Question 03 :

Dans cette question nous avons affiché les matrices de confusion de chacun des ensembles : apprentissage et validation. Nous avons obtenu les matrices ci-dessous :

	A	B	C	D	E	F	G	H	I	J	K	L
A	75	0	0	0	0	0	0	0	0	0	0	0
B	0	56	0	0	0	0	0	0	0	0	0	0
C	0	0	155	0	0	0	0	0	0	0	0	0
D	0	0	0	81	0	0	0	0	0	0	0	0
E	0	0	0	0	66	0	0	0	0	0	0	0
F	0	0	0	0	0	135	0	0	0	0	0	0
G	0	0	0	0	0	0	33	0	0	0	0	0
H	0	0	0	0	0	0	0	61	0	0	0	0
I	0	0	0	0	0	0	0	0	190	0	0	0
J	0	0	0	0	0	0	0	0	0	57	0	0
K	0	0	0	0	0	0	0	0	0	0	34	0
L	0	0	0	0	0	0	0	0	0	0	0	57

Figure 02 : Matrice de confusion pour l'ensemble Na

	A	B	C	D	E	F	G	H	I	J	K	L
A	77	0	0	0	2	5	0	7	3	0	2	0
B	4	37	0	1	3	1	1	1	2	3	2	1
C	0	0	105	2	4	6	5	1	0	3	0	2
D	0	2	1	54	0	0	2	1	0	4	0	0
E	2	2	7	0	55	4	1	0	2	1	5	1
F	7	1	0	0	5	121	0	1	5	2	0	0
G	1	1	3	1	7	4	15	0	0	1	0	2
H	7	2	0	3	0	3	1	28	1	0	0	0
I	0	0	0	1	1	5	0	1	172	3	0	2
J	0	0	3	1	0	0	1	1	5	42	0	0
K	1	1	1	0	1	4	0	1	1	0	25	0
L	0	0	3	1	0	0	0	0	4	0	0	73

Figure 03 : Matrice de confusion pour l'ensemble Nv

Commentaires :

En calculant le rappel (sensibilité) pour chaque classe, soit la capacité de couvrir correctement cette classe nous avons :

$r_A = 1$, $r_B = 0.9$, $r_C = 1$, $r_D = 0.94$, $r_E = 0.83$, $r_F = 0.9$, $r_G = 0.47$, $r_H = 0.67$, $r_I = 0.96$, $r_J = 0.79$, $r_K = 0.71$, $r_L = 0.9$

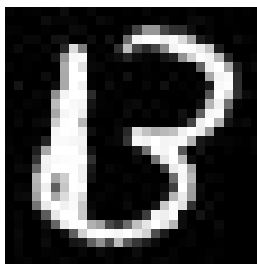
En calculant la précision pour chaque classe nous avons :

$p_A = 0.8$, $p_B = 0.71$, $p_C = 0.82$, $p_D = 0.88$, $p_E = 0.8$, $p_F = 0.93$, $p_G = 0.83$, $p_H = 0.96$, $p_I = 1$, $p_K = 1$, $p_L = 1$

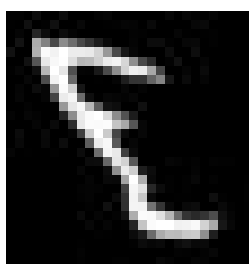
Comme la précision est assez élevée pour toutes les classes nous nous intéressons donc seulement à la sensibilité et ainsi nous remarquons que les classes les plus problématiques sont les classes avec des sensibilités minimales soit les classes : G, H.

Question 05 :

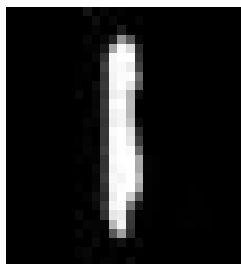
Lors de cette question, nous avons pu afficher les images bien classées par notre perceptron, nous avons pris comme exemple les classes correspondantes aux images suivantes :



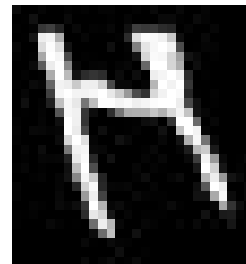
“Classe B”



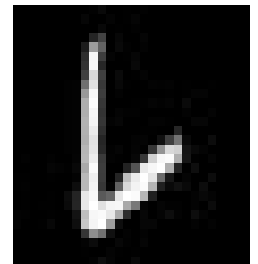
“Classe E”



“Classe I”



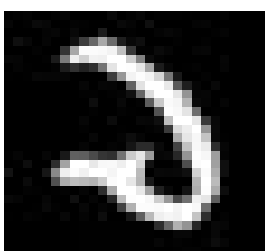
“Classe H”



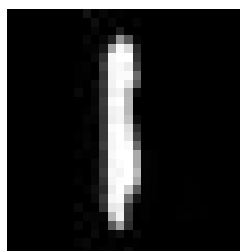
“Classe L”

Questions 06 :

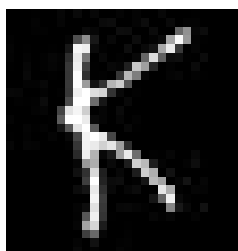
Contrairement à la question précédente, nous nous sommes intéressés ici aux images mal classées pour une classe donnée, nous avons sélectionné 5 classes :



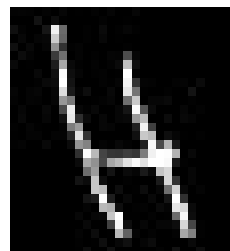
“Classe G”



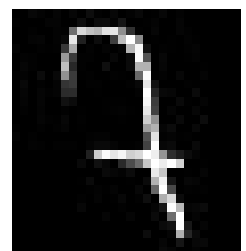
“Classe I”



“Classe K”



“Classe H”



“Classe F”

Question 07 :

Lors de cette question nous avons repris notre perceptron multi classe tout modifiant certains facteurs nous permettant de travailler ainsi sur un ensemble de classes plus large, de plus nous avons effectué une extraction d'un intervalle plus grand de labels (A-Z).

La partie de code correspondante :

Le perceptron : la classe "PerceptronMultiNew"

L'extraction et l'application du perceptron : la classe "Question_7"

En appliquant notre nouveau perceptron sur nos nouveaux ensembles nous obtenons la figure ci-dessous :

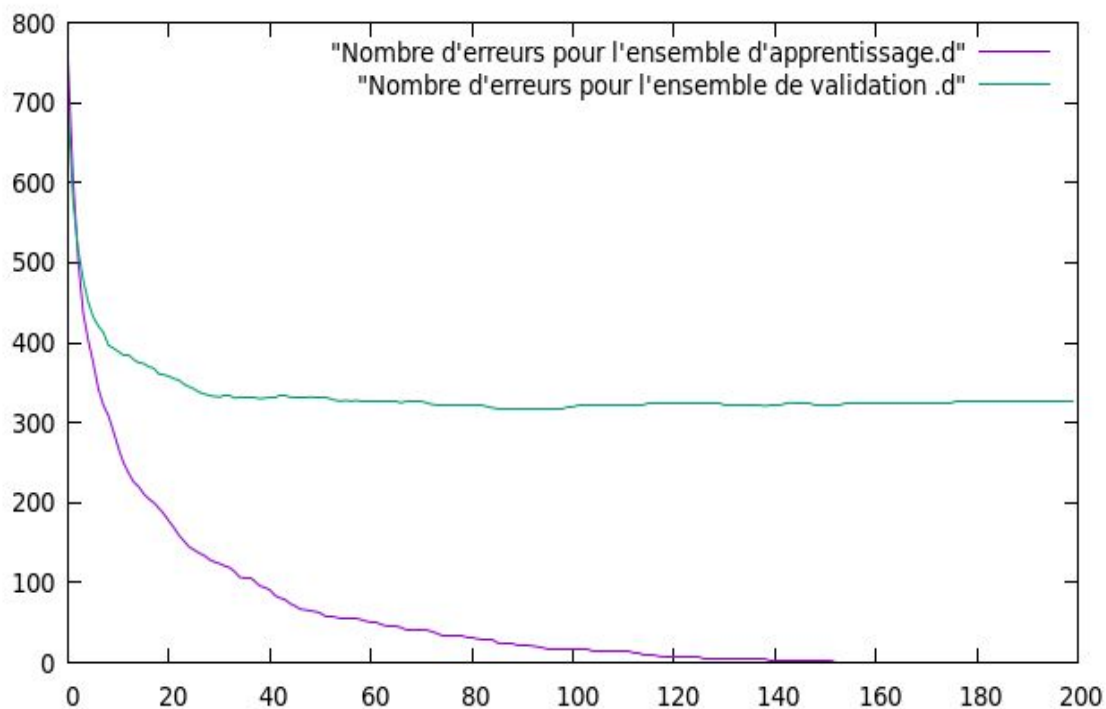


Figure 03 : Graphe illustrant le nombre d'erreurs pour l'ensemble d'apprentissage N_a et l'ensemble de validation N_v en fonction du nombre d'itération