

UNIVERSITÉ NOUVEAUX HORIZONS



Machine Learning: Étude de la minimisation d'erreur dans l'apprentissage supervisé, avec une application dans la technologie ANPR

Auteur :

TSHELEKA KAJILA Hassan

Directeur :

Prof. MASAKUNA Jordan

*Mémoire présenté à la Faculté des Sciences Informatiques en vue de
l'obtention du grade de Licencié en informatique.*

en

Calcul Scientifique

6 avril 2022

RÉSUMÉ

Au cours de la dernière décennie, la taille des données a augmenté plus rapidement que la vitesse des processeurs. Dans ce contexte, faire un traitement de reconnaissance des formes dans des images et vidéos, les ensembles de données d'entraînement pour les problèmes de détection d'objets sont généralement très volumineux et les capacités des méthodes d'apprentissage automatique statistique sont limitées par le temps de calcul plutôt que par la taille de l'échantillon.

Le cas des problèmes d'apprentissage à grande échelle implique la complexité de calcul de l'algorithme d'optimisation sous-jacent de manière non triviale. Des algorithmes d'optimisation improbables tels que la **descente de gradient stochastique** (en anglais : **Stochastic Gradient Descent** ou SGD) montre des performances étonnantes pour les problèmes à grande échelle, lorsque l'ensemble d'apprentissage est volumineux.

En particulier, les variants du SGD n'utilisent qu'un seul nouvel échantillon d'apprentissage à chaque itération, sont asymptotiquement efficaces après un seul passage sur l'ensemble d'apprentissage.

Ce travail vise à proposer une méthode intelligente, basée sur l'intelligence artificielle, qui permet aux ordinateurs et aux systèmes informatiques de dériver des informations significatives à partir d'images numériques, de vidéos et d'autres entrées visuelles, avec un coût plus bas que possible. Dans notre contexte la reconnaissance des plaques d'immatriculation des véhicules à l'aide d'un classificateur de la famille de descente de gradient stochastique. Pour minimiser la **fonction coût** du classificateur, la SGD adopte un modèle d'optimisation convexe. De plus, pour augmenter la vitesse de convergence du classificateur, la descente de gradient stochastique, à chaque étape, elle tire un échantillon aléatoire de l'ensemble des fonctions (f_i), de la fonction objectif, constituant la somme.

Mots clés : Apprentissage supervisé, vision par ordinateur, Descente de gradient stochastique, Adaline, ANPR, ALPR.

ABSTRACT

Over the past decade, data size has grown faster than processor speeds. In this context, doing pattern recognition processing in real-time videos, training datasets for object detection problems are usually very large, and the capabilities of statistical machine learning methods are limited by computation time rather than sample size.

The case of large scale learning problems involves the computational complexity of the underlying optimization algorithm in a nontrivial way.

Improbable optimization algorithms such as **Stochastic Gradient Descent** (SGD) show amazing performance for large scale problems, when the training set is bulky.

In particular, SGD variants use only one new training sample at each iteration, are asymptotically efficient after a single pass over the training set.

This work aims to provide an intelligent method, based on artificial intelligence, that allows computers and computer systems to derive meaningful information from digital images, videos and other visual inputs, with a lower cost. as possible. In our context the recognition of vehicle license plates using a classifier of the family of stochastic gradient descent. To minimize the **cost function** of the classifier, the SGD adopts a convex optimization model. Moreover, to increase the speed of convergence of the classifier, the stochastic gradient descent, at each step, it draws a random sample from the set of functions (f_i), of the objective function, constituting the sum.

Key words : Supervised learning, computer vision, Stochastic gradient descent, Adaline, ANPR, ALPR.

TABLE DES MATIÈRES

Résumé	ii
o Introduction	2
0.1 Présentation (généralités)	2
0.2 Contexte et problématique de notre recherche	3
0.3 Objectifs de notre étude	4
I État des connaissances (Background material)	
1 Les bases mathématiques pour le Machine Learning	9
1.1 Eléments de calcul différentiel	9
1.1.1 Convexité	9
1.1.2 Développement limité	9
1.1.3 Fonctions dérivables	10
1.2 Statistique & probabilité	14
1.3 Série statistique	14
1.3.1 Echantillonnage	14
1.3.2 Analyse bayésienne	14
1.4 Méthode d'optimisation et de minimisation d'erreur	15
1.4.1 Erreur et fonction coût	15
1.4.2 Moindres carrés linéaires	17
1.4.3 Descent de gradient	17
1.4.4 Descente de gradient stochastique	18
2 Modélisation et Apprentissage automatique	19
2.1 Régression Linéaire	19
2.1.1 Le problème de la régression linéaire	19
2.2 Régression Logistique	20
2.2.1 Le problème de classification	20
2.2.2 Le cas séparable	20
2.2.3 Le cas non séparable	20
2.2.4 Le modèle de la régression logistique	20
2.3 Réseaux de neurones	21
2.3.1 Perceptron	21
2.3.2 Neurones	21
2.3.3 Neurone linéaire adaptatif (ADALINE)	21
2.3.4 Réseau neuronal convolutif (CNN)	21
Bibliographie	22

LISTE DES ACRONYMES

ML Machine Learning
CV Computer Vision
OCR Optical character recognition
ANPR Automatic number-plate recognition
ALPR Automatic license plate recognition
API Application Programming Interface
UML Unified Modeling Language



INTRODUCTION

0.1 PRÉSENTATION (GÉNÉRALITÉS)

L'intelligence désigne communément le potentiel des capacités mentales et cognitives d'un individu, animal ou humain, lui permettant de résoudre un problème ou de s'adapter à son environnement. L'intelligence nous fait ressentir ce besoin d'apprendre pour arriver à nos fins, extrinsèquement l'intelligence c'est l'apprentissage. Pour que nous puissions dire qu'une machine est intelligente, premièrement elle doit passer par une phase d'apprentissage. Apprendre à résoudre des problèmes ou à réaliser des tâches par lui-même d'une façon autonome. Dans le IA nous parlons de l'apprentissage automatique (en anglais : machine Learning, ML), nous utilisons plusieurs paradigmes d'apprentissage automatique : apprentissage supervisé, apprentissage non supervisé, apprentissage par renforcement, apprentissage en profondeur.

L'apprentissage supervisé représente une grande partie de l'activité de recherche en apprentissage automatique (ML) et de nombreuses techniques d'apprentissage supervisé ont trouvé une application dans le traitement de contenu multimédia. La caractéristique qui définit l'apprentissage supervisé est la disponibilité de données d'apprentissage annotées[6]. Le nom évoque l'idée d'un **superviseur** qui instruit le système d'apprentissage sur les étiquettes à associer à des modèles ¹ d'entraînement.

L'application de cette étude est orientée vers la reconnaissance automatique d'objet dans les vidéos et images, une des applications intéressantes, parmi tant d'autres, dans l'intelligence artificielle.

La reconnaissance automatique d'objet est un problème important dans la vision par ordinateur (Computer Vision ²) et en traitement d'images. Cette tâche est très utile vue l'accroissement du nombre de vidéos générées par des smartphones, des systèmes de sécurité, des caméras de circulation et autres dispositifs dotés d'instruments visuels. La reconnaissance automatique des objets en vidéo peut ainsi renforcer la sécurité, faciliter la gestion des vidéos ainsi que permettre de nouvelles applications en interaction homme/machine.

¹ Un modèle de machine learning est le résultat généré lorsque vous entraînez votre algorithme d'apprentissage automatique avec des données.

² La vision par ordinateur est un domaine de l'intelligence artificielle (IA) qui permet aux ordinateurs et aux systèmes de dériver des informations significatives à partir d'images numériques, de vidéos et d'autres entrées visuelles, et de prendre des mesures ou de faire des recommandations sur la base de ces informations.

Par ailleurs, les images numériques et la vidéo sont devenues indispensables pour divers domaines d'application, tels que la détection d'intrusions pour la sécurité, la surveillance du trafic routier, la médecine pour l'imagerie médicale, ou encore lors des événements sportifs (ex., renforcement de l'arbitrage, création automatique de résumés). Des contraintes d'exploitation découlent des observations citées ci-dessus, parmi lesquelles nous citerons celles qui sont liées à la reconnaissance des objets en mouvement dans les vidéos. Par exemple, de nos jours, un très grand nombre de caméras est déployé exclusivement pour la surveillance vidéo [1]. Souvent, le contenu de ces vidéos est interprété par des opérateurs humains qui engendrent des coûts exorbitants pour le suivi et l'analyse du contenu, sans mentionner les erreurs qui peuvent être induites par la fatigue et l'inattention humaine. Une des interrogations importantes abordées lors l'apprentissage supervisé appliqué dans la surveillance vidéo est la reconnaissance des types d'objets en mouvement et leurs actions. Afin de détecter, par exemple, des menaces potentielles (ex., vols, attentats, accidents), ou tout simplement pour des fins de statistiques (ex., compter le nombre d'individus, de voitures dans une entrée de parc). Les applications du monde réel démontrent l'importance de la vision par ordinateur pour les entreprises, les secteurs du divertissement, des transports, des soins de santé et dans la vie quotidienne. L'un des principaux moteurs de la croissance de ces applications est le flot d'informations visuelles provenant des médias numériques (ex., internet, la télévision, les vidéos personnelles, la surveillance vidéo).

0.2 CONTEXTE ET PROBLÉMATIQUE DE NOTRE RECHERCHE

Ce travail présente les résultats d'une étude approfondie sur les algorithmes de minimisation d'erreur, la fonction coût³ (en anglais : loss function).

Dans ce contexte, faire une application dans le traitement de reconnaissance des formes dans des vidéos, les ensembles de données d'entraînement pour les problèmes de détection d'objets sont généralement très volumineux et les capacités des méthodes d'apprentissage automatique statistique sont limitées par le temps de calcul plutôt que par la taille de l'échantillon[2].

Par exemple, pour entraîner une machine à reconnaître des plaques d'immatriculation de voiture, elle doit recevoir de grandes quantités d'images de plaques d'immatriculation et d'éléments liés aux plaques pour apprendre les différences et reconnaître une plaque, en particulier la voiture qui porte une plaque sans défaut. Plus nous avons des données, plus nous gagnons en précision et plus la complexité en temps augmente.

Une analyse plus précise révèle des compromis qualitativement différents pour le cas des problèmes d'apprentissage à petite et à grande échelle [2]. La complexité de calcul de l'algorithme d'apprentissage devient le facteur limitant critique

³ Dans l'optimisation mathématique et en statistique, une fonction de perte ou une fonction de coût est généralement utilisée pour l'estimation des paramètres, et l'événement en question est une fonction de la différence entre les valeurs estimées et vraies pour une instance de données.

lorsque l'on envisage de très grands ensembles de données. C'est à ce point critique qu'entre en jeu cette étude, la minimisation des erreurs sans alourdir la complexité en temps et espace de l'algorithme d'apprentissage. Minimiser les erreurs dans les modèles d'apprentissage a toujours été une tâche très importante pour renforcer la fiabilité de notre Machine Learning Model[10]. Établir un algorithme d'apprentissage qui s'adapte au mieux à notre modèle, selon la nature du problème métier traité, il existe différentes approches qui varient selon le type et le volume des données. Dans cette section, nous discutons des algorithmes de descente de gradient stochastique parce qu'ils montrent des performances d'optimisation incroyables pour les problèmes à grande échelle [2].

Le travail de léon bottou et al (e. g., [2] [13] [3]), présente *la descente de gradient stochastique comme un algorithme d'apprentissage fondamental*. L'un des piliers de l'apprentissage automatique est l'optimisation mathématique [Jorge Nocedal dans 4, page : 3], qui, dans ce contexte, implique le calcul numérique de minimisation des paramètres d'un système conçu pour prendre des décisions basées sur des données actuellement disponibles, ces paramètres sont choisis pour être optimaux par rapport à un problème d'apprentissage donné.

Dans l'ensemble, ce document tente d'apporter des réponses aux questions suivantes.

1. Comment les problèmes de minimisation surviennent-ils dans les applications d'apprentissage automatique et qu'est-ce qui les rend difficiles ?
2. Quelles ont été les méthodes de minimisation les plus efficaces pour l'apprentissage supervisé à grande échelle et pourquoi ?
3. Comment des algorithmes d'apprentissage supervisé arrivent-ils à résoudre le problème de la reconnaissance automatique d'objet ?
4. Quelles avancées récentes ont été réalisées dans la conception d'algorithmes d'apprentissage et quelles sont les questions ouvertes dans ce domaine de recherche ?

0.3 OBJECTIFS DE NOTRE ÉTUDE

Le but de cette étude est de fournir une revue et un commentaire sur le passé, le présent et le futur de l'utilisation des algorithmes d'optimisation numérique, précisément de minimisation, dans le contexte des applications d'apprentissage automatique qui permet aux ordinateurs et aux systèmes informatiques de dériver des informations significatives à partir d'images numériques, de vidéos et d'autres entrées visuelles, avec un coût plus bas que possible.

Expressément, nous faisons la reconnaissance des plaques d'immatriculation des véhicules à l'aide d'un classificateur de la famille de descente de gradient stochastique (SGD) [?]. Pour minimiser la fonction de coût du classificateur, le SGD adopte un modèle d'optimisation convexe. De plus, pour augmenter la vitesse de convergence du classificateur, la descente de gradient stochastique, à

chaque étape, elle tire un échantillon aléatoire de l'ensemble des fonctions (f_i), de la fonction objectif, constituant la somme.

Pour chaque algorithme, nous examinons l'efficacité et comparons le score pour différents cas. Les objectifs de notre travail regroupent les points suivants :

1. État des connaissances, c'est sont les éléments sur lequel je me base pour constituer ce travail, nous parlons des base mathématique essentiel pour le Machine Learning : les éléments différentiel, statistique, l'optimisation de modèle linéaire convexe, etc.
2. La méthodologie utilisée parmi tant d'autres, pour entraîner les modèles d'apprentissage automatique (Machine learning Model) de façon optimale. Pour la minimisation de la fonction coût nous utilisons des algorithmes comme SGD, ADAM, ADAGRAD, ADADELTA, ASGD, NAG. Puis faire une étude comparative de leurs performances. Et aussi des méthodes intelligentes de classification des images pour application reconnaissance automatique d'objet.
3. Nous construirons des modèles à partir d'une base de données annotée pour l'apprentissage et pour les tests de reconnaissance d'objets. Les résultats concluants de cette étude pourront conduire à un déploiement de notre système dans les domaines comme celui de la surveillance vidéo de voitures dans une entrée de parking. Des métriques connues pour mesurer les erreurs et en déduire le score du classificateur seront utilisées pour évaluer la qualité de la reconnaissance automatique des plaques d'immatriculation (en anglais : Automatic Number Plate Recognition ou ANPR) par notre approche.

Première partie

ÉTAT DES CONNAISSANCES (BACKGROUND MATERIAL)

État des connaissances, c'est sont les éléments sur lequel je me base pour constituer ce travail, nous parlons des base mathématique essentiel pour le Machine Learning : les éléments différentiel, statistique, l'optimisation numérique de modèle linéaire convexe, etc.

LES BASES MATHÉMATIQUES POUR LE MACHINE LEARNING

1.1 ÉLÉMENTS DE CALCUL DIFFÉRENTIEL

Cette section est inspirée des notes écrites par le Professeur TSHIMANGA [voir 9, page :45-82] et de consignes données par [voir ?, page :...-...].

1.1.1 Convexité

DÉFINITION : (ENSEMBLE CONVEXE) Une partie $\mathcal{C} \subset \mathbb{R}^n$ est dite convexe si et seulement si pour tout $(x, y) \in \mathcal{C}^2$, et pour tout $\alpha \in [0, 1]$,

$$\alpha x + (1 - \alpha)y \in \mathcal{C}$$

combinaison convexe.

DÉFINITION : (FONCTION CONVEXE) Une fonction f d'un intervalle réel $I \in \mathcal{C}$ est dite fonction convexe lorsque, $\forall (x, y)$ de I tel que $(x, y) \in \mathcal{C}^2$ et tout $\alpha \in [0, 1]$ on a :

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) \quad (1)$$

et si

$$f(\alpha x + (1 - \alpha)y) < \alpha f(x) + (1 - \alpha)f(y) \quad (2)$$

on dit que la fonction est strictement convexe dans \mathcal{C}

Ex :

- La fonction $f(x) = x^2$ est convexe.
- La fonction $f(x) = x^T x$ est convexe.
- La fonction $f(x) = x^T A x$ est convexe, ssi A est symétrique semi-définie positive.

1.1.2 Développement limité

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

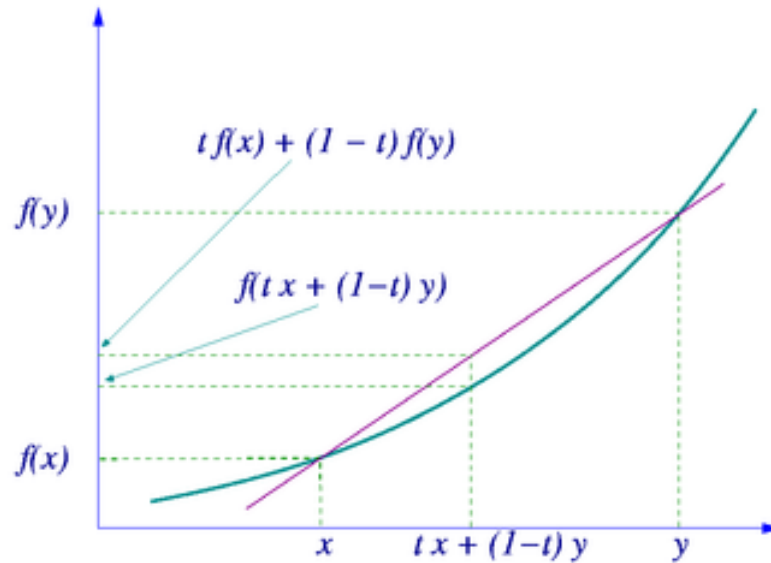


FIGURE 1 : Fonction convexe (image de Wikipédia)

A Différentiabilité au sens de Frechet

??? parler de son implication dans le gradient Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

1.1.3 Fonctions dérivables

A Gradient

DÉFINITION : Le gradient d'une fonction de plusieurs variables en un certain point est un vecteur qui caractérise la variabilité de cette fonction au voisinage de ce point. Défini en tout point où la fonction est différentiable, il définit un champ de vecteurs, également dénommé gradient. Le gradient est la généralisation à plusieurs variables de la dérivée d'une fonction d'une seule variable.

DÉFINITION MATHÉMATIQUE : Dans un système de coordonnées cartésiennes, le gradient d'une fonction $f(x_1, x_2, \dots, x_n)$ est le vecteur de composantes $\partial f / \partial x_i$ ($i = 1, 2, \dots, n$), c'est-à-dire les dérivées partielles de f par rapport aux coordonnées.

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} \in \mathbb{R}^n$$

GRADIENT SOUS FORME DE DÉVELOPPEMENT LIMITÉ : Si une application admet un gradient en un point, alors on peut écrire ce développement limité du premier ordre (voir le point 1.1.2)

$$f(x+h) = f(x) + \langle \nabla f(x) | h \rangle + o(h)$$

ou

$$f(x-h) = f(x) - \langle \nabla f(x) | h \rangle + o(h)$$

Numériquement, il est très intéressant de faire ensuite la demi-différence des deux développements pour obtenir la valeur du gradient et on note que celui-ci ne dépend pas en fait de la valeur de la fonction au point $x : f(x)$. Cette formule a l'avantage de tenir compte des gradients du 2e ordre et est donc beaucoup plus précise et numériquement robuste. L'hypothèse est, en pratique, de connaître les valeurs "passé" et "futur" de la fonction autour d'un petit voisinage du point x .

DÉFINITION NUMÉRIQUE : Une fonction multivariée (a variable vectorielle) $f(x) : \mathbb{R}^n \rightarrow \mathbb{R} : x \rightarrow f(x)$ définie sur un ouvert $O \in \mathbb{R}^n$ est dite dérivable (au sens de Fréchet ??) en x ssi il existe un vecteur noté $\nabla f(x) \in \mathbb{R}^n$ tel que

$$f(x+h) = f(x) + \nabla f(x)^T h + o(\|h\|) \quad (3)$$

$\nabla f(x) \in \mathbb{R}^n$ et où l'on a posé que le reste $o(\|h\|) = \|h\| \epsilon(h) \in \mathbb{R}^n$, avec $h \in \mathbb{R}^n$

$$\epsilon(h) : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \lim_{\|h\| \rightarrow 0} \epsilon(h) = 0.$$

Le vecteur $\nabla f(x)$ est unique et s'appelle **gradient** de $f(x)$ en x . Le gradient s'adresse aux fonctions scalaires à variables vectorielles.

A PROPOS DE LA NOTATION $o(\|h\|)$: La notation de Landau $o(\|h\|)$ traduit le comportement d'une fonction de h qui [est ??] tend vers 0 d'un ordre de grandeur plus vite que $\|h\|$.

Elle est infiniment plus petit que h dans le voisinage de 0

B Hessienne

DÉFINITION MATHÉMATIQUE : Etant donnée une fonction f à valeurs réelles

$$f : \mathbb{R}^n \rightarrow \mathbb{R}; (x_1, \dots, x_n) \mapsto f(x_1, \dots, x_n)$$

dont toutes les dérivées partielles secondes existent, le coefficient d'indice i, j de la **matrice hessienne**¹ $H(f)$ vaut $H_{ij}(f) = \frac{\partial^2 f}{\partial x_i \partial x_j}$.

Autrement dit,

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

DÉFINITION NUMÉRIQUE : Supposons que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ définie sur un ouvert $\mathcal{O} \in \mathbb{R}^n$. La fonction $f(x)$ est dite 2 fois continûment dérivable (au sens de Fréchet??) si en tout $x \in \mathcal{O}$ on a

$$f(x+h) = f(x) + \nabla f(x)^T h + \frac{1}{2} h^T \nabla^2 f(x) h + o(\|h\|^2) \quad (4)$$

avec $\nabla f(x) \in \mathbb{R}^{n \times n}$ et où on a posé que le reste $o(\|h\|^2) = \|h\| \epsilon(h) \in \mathbb{R}$ avec $\lim_{\|h\| \rightarrow 0} \epsilon(h) = 0$. La matrice carrée symétrique $\nabla^2 f(x)$ appelée **Hessien** de $f(x)$ en x .

Remarque :

$$\lim_{\|h\| \rightarrow 0} \frac{o(\|h\|^2)}{\|h\|} = 0 \in \mathbb{R}$$

La Hessienne s'adresse aux fonctions scalaires à variables vectorielles.

C Jacobienne

DÉFINITION MATHÉMATIQUE : Soit F une fonction d'un ouvert de \mathbb{R}^n à valeurs dans \mathbb{R}^m ($F : \mathbb{R}^n \rightarrow \mathbb{R}^m$). Une telle fonction est définie par ses m fonctions composantes à valeurs réelles :

$$F : \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{pmatrix}.$$

¹ En mathématiques, la matrice hessienne (ou simplement la hessienne) d'une fonction numérique f est la matrice carrée, notée $H(f)$, de ses dérivées partielles secondes.

Les dérivées partielles de ces fonctions en un point M , si elles existent, peuvent être rangées dans une matrice à m lignes et n colonnes, appelée **matrice jacobienne**² de F :

$$J_F(M) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

La case sur la ligne i et la colonne j contient $\frac{\partial f_i}{\partial x_j}$ qui est la dérivée partielle de f_i selon la variable x_j . Cette matrice est notée :

$$J_F(M), \quad \frac{\partial (f_1, \dots, f_m)}{\partial (x_1, \dots, x_n)} \quad \text{ou} \quad \frac{D(f_1, \dots, f_m)}{D(x_1, \dots, x_n)}$$

Pour $i = 1, \dots, m$, la i -ème ligne de cette matrice est la transposée du vecteur **gradient** (voir le point A) au point M de la fonction f_i , lorsque celui-ci existe. La matrice jacobienne est également la matrice de la différentielle de la fonction, lorsque celle-ci existe.

DÉFINITION NUMÉRIQUE : Soit $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ définie sur un ouvert $\mathcal{O} \subset \mathbb{R}$. On dit que $f(x)$ est dérivable (au sens de Fréchet) en x , si chacune des composantes $f_i(x)$ est dérivable en x . On a alors

$$f(x+h) = f(x) + D_f(x)h + o(\|h\|) \quad (5)$$

avec $D_f(x) \in \mathbb{R}^{n \times m}$ et/ou $o(\|h\|) = \|h\|\epsilon(h) \in \mathbb{R}^m$ avec $\lim_{\|h\| \rightarrow 0} \epsilon(h) = 0$. Remarque :

$$\lim_{\|h\| \rightarrow 0} \frac{o(\|h\|^2)}{\|h\|} = 0 \in \mathbb{R}$$

$$\begin{aligned} \text{Soient } x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n \text{ et } f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{bmatrix} \in \mathbb{R}^m \\ D_f(x) = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla f_1(x)^T \\ \nabla f_2(x)^T \\ \vdots \\ \nabla f_m(x)^T \end{bmatrix} \in \mathbb{R}^{n \times m}, \end{aligned}$$

² En analyse vectorielle, la matrice jacobienne est la matrice des dérivées partielles du premier ordre d'une fonction vectorielle en un point donné.

La matrice $D_f(x) \in \mathbb{R}^{n \times m}$ est appelée **Jacobienne** de $f(x)$ en x . La Jacobienne s'adresse aux fonctions vectorielles à variables vectorielles.

NOTE : Lorsque $m = 1$ la Jacobienne est la même que le gradient car il s'agit d'une généralisation du gradient.

1.2 STATISTIQUE & PROBABILITÉ

1.3 SÉRIE STATISTIQUE

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

1.3.1 Echantillonnage

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

1.3.2 Analyse bayésienne

La statistique bayésienne est une théorie dans le domaine des statistiques basée sur l'interprétation bayésienne de la probabilité où la probabilité exprime un degré de croyance en un événement. Le degré de croyance peut être basé sur des connaissances antérieures sur l'événement, telles que les résultats d'expériences

précédentes, ou sur des croyances personnelles sur l'événement. Cela diffère d'un certain nombre d'autres interprétations de la probabilité, telles que l'interprétation fréquentiste qui considère la probabilité comme la limite de la fréquence relative d'un événement après de nombreux essais.

Les statistiques bayésiennes portent le nom de Thomas Bayes³, qui a formulé un cas spécifique du théorème de Bayes dans un article publié en 1763.

Theorem 1 (Théorème de Bayes) *Le théorème de Bayes est utilisé dans les méthodes bayésiennes pour mettre à jour les probabilités, qui sont des degrés de croyance, après avoir obtenu de nouvelles données. Compte tenu de deux événements A et B, la probabilité conditionnelle de A étant donné que B est vrai s'exprime comme suit :*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \quad (6)$$

où $\mathbb{P}(B) \neq 0$ Bien que le théorème de Bayes soit un résultat fondamental de la théorie des probabilités, il a une interprétation spécifique dans les statistiques bayésiennes.

1.4 MÉTHODE D'OPTIMISATION ET DE MINIMISATION D'ERREUR

1.4.1 Erreur et fonction coût

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

A Erreur d'apprentissage

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada

³ Thomas Bayes était un Anglais statisticien, philosophe et ministre presbytérien qui est connu pour la formulation d'un cas spécifique du théorème qui porte son nom : théorème de Bayes.

fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

B Fonction coût ℓ cas de la régression linéaire

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

C Fonction coût ℓ cas de la classification

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec

varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

1.4.2 Moindres carrés linéaires

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

1.4.3 Descent de gradient

Il a souvent été proposé (e. g., [?]) de minimiser le risque empirique [E] en utilisant la descente de gradient (GD). Chaque itération met à jour les poids w en fonction du gradient de [E] [3].

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

$$A = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & \cdots & x_{mn} \end{pmatrix}$$

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc

elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

1.4.4 *Descente de gradient stochastique*

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

MODÉLISATION ET APPRENTISSAGE AUTOMATIQUE

2.1 RÉGRESSION LINÉAIRE

2.1.1 *Le probleme de la régression linéaire*

Lorem $O(h^3)$ Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

A *Le cas de la régression linéaire*

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

B *La fonction d'erreur*

2.2 RÉGRESSION LOGISTIQUE

2.2.1 *Le problème de classification*

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

2.2.2 *Le cas séparable*

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

2.2.3 *Le cas non séparable*

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

2.2.4 *Le modèle de la régression logistique*

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

2.3 RÉSEAUX DE NEURONES

2.3.1 *Perceptron*

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam,

2.3.2 *Neurones*

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

2.3.3 *Neurone linéaire adaptatif (ADALINE)*

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

2.3.4 *Réseau neuronal convolutif (CNN)*

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada

fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

BIBLIOGRAPHIE

- [1] Yaovi AHADJITSE. "Reconnaissance d'objets en mouvement dans la vidéo par description géométrique et apprentissage supervisé". Thèse de doct. Université du Québec en Outaouais, 2013.
- [2] Léon BOTTOU. "Large-scale machine learning with stochastic gradient descent". In : *Proceedings of COMPSTAT'2010*. Springer, 2010, p. 177-186.
- [3] Léon BOTTOU. "Stochastic gradient descent tricks". In : *Neural networks: Tricks of the trade*. Springer, 2012, p. 421-436.
- [4] Léon BOTTOU, Frank E CURTIS et Jorge NOCEDAL. "Optimization methods for large-scale machine learning". In : *Siam Review* 60.2 (2018), p. 223-311.
- [5] Rich CARUANA et Alexandru NICULESCU-MIZIL. "An empirical comparison of supervised learning algorithms". In : *Proceedings of the 23rd international conference on Machine learning*. 2006, p. 161-168.
- [6] Pádraig CUNNINGHAM, Matthieu CORD et Sarah Jane DELANY. "Supervised learning". In : *Machine learning techniques for multimedia*. Springer, 2008, p. 21-49.
- [7] Natarajan DEEPA, B PRABADEVI, Praveen Kumar MADDIKUNTA, Thippa Reddy GADEKALLU, Thar BAKER, M Ajmal KHAN et Usman TARIQ. "An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier". In : *The Journal of Supercomputing* 77 (2021), p. 1998-2017.
- [8] Kary FRÄMLING. "Scaled Gradient Descent Learning Rate". In : *Reinforcement Learning With Light-Seeking Robot, Proceedings of ICINCO* (2004), p. 1-8.
- [9] Jean Tshimanga ILUNGA. "Optimisation Numérique". In : *Jorge Nocedal and Steve Wright, (2000), Numerical Optimization, Springer Verlag*. T. 72. UNH. 2021, p. 21-23.
- [10] Daniel Kirsch JUDITH HURWITZ. *Machine Learning For Dummies*. IBM Limited Edition. 111 River St. Hoboken, NJ 07030-5774 : John Wiley et Sons, Inc., 2018.
- [11] Agnes LYDIA et Sagayaraj FRANCIS. "Adagrad an optimizer for stochastic gradient descent". In : *Int. J. Inf. Comput. Sci* 6.5 (2019), p. 566-568.
- [12] Praneeth NETRAPALLI. "Stochastic gradient descent and its variants in machine learning". In : *Journal of the Indian Institute of Science* 99.2 (2019), p. 201-213.
- [13] Rob GJ WIJNHOFEN et PHN de WITH. "Fast training of object detection using stochastic gradient descent". In : *2010 20th International Conference on Pattern Recognition*. IEEE. 2010, p. 424-427.