N°d'ordre NNT : ?

THÈSE de DOCTORAT DE L'UNIVERSITÉ DE LILLE

École Doctorale MADIS N° 631

Mathématiques-Sciences du numérique et de leurs interactions

Spécialité / discipline de doctorat : Mathématique

Soutenue par :

Maxime Haddouche

A PAC-Bayes Approach of Generalisation In Machine Learning: From Heavy-Tailed Martingales To Optimisation

Acknowledgements

TODO

Quote 1

Author

Quote 2

Author

Contents

List of Figures

Preamble

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of Algorithms

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 7

List of Theorems

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Appendix

List of Notations

General

a	A scalar (integer or real)
$\mathbf{a}$	A vector
$\mathbf{A}$	A matrix
$\mathbb{A}, \mathbbm{a}$	A set
$\mathbb{R}$	The set of real numbers
$\mathbb{R}_*$	The set of real numbers excluding 0
$\mathbb{R}_*^+$	The set of positive real numbers excluding 0
$\mathbb{N}$	The set of natural numbers
$\mathbb{N}_*$	The set of natural numbers excluding 0
$\mathrm{card}(\cdot)$	The cardinal of a set
a_i	i-th element of the vector $\mathbf{a}$

Statistical Learning Theory

$\mathbb{X}$	Set of d-dimensional inputs ($\subseteq \mathbb{R}^d$)
$\mathbb{Y}$	Set of labels
$\mathbf{x}$	A real-valued input $\mathbf{x} \in \mathbb{X}$
y	A label $y \in \mathbb{Y}$ associated to the input $\mathbf{x}$
$\mathcal{D}$	Unknown data distribution on $\mathbb{X} \times \mathbb{Y}$
$\mathcal{D}^m$	Unknown data distribution on the m-samples, *i.e.*, on $(\mathbb{X}\times\mathbb{Y})^m$
$\mathbb{S}$	Learning sample $\mathbb{S} = \{(\mathbf{x}_i, y_i)\}_{i=1}^m$ drawn from $\mathcal{D}^m$
$\mathcal{S}$	Uniform distribution on $\mathbb{S}$
$\mathbb{T}$	Test set drawn from $\mathcal{D}^m$

$\mathcal{T}$	Uniform distribution on $\mathbb{T}$
$\mathbb{H}$	The set of hypotheses
h	A hypothesis $h \in \mathbb{H}$
$\ell(\cdot,\cdot)$	Loss function
$\mathrm{R}^{\ell}_{\mathcal{D}'}(h)$	Risk of the hypothesis $h \in \mathbb{H}$ *w.r.t.* the loss function $\ell()$ on $\mathcal{D}'$
$\mathrm{R}_{\mathcal{D}'}(h)$	Risk of the hypothesis $h \in \mathbb{H}$ *w.r.t.* the 0-1 loss on $\mathcal{D}'$

Probability Theory

$\mathbb{E}_{X\sim\mathcal{X}}\left[\cdot\right]$	The expectation *w.r.t.* the random variable $X \sim \mathcal{X}$
$\mathbb{P}_{X\sim\mathcal{X}}\left[\cdot\right]$	The probability *w.r.t.* the random variable $X \sim \mathcal{X}$
$\mathrm{I}\left[a\right]$	Indicator function; returns 1 if a is true and 0 otherwise
$\mathbb{M}(\mathbb{H})$	Set of Probability densities *w.r.t.* the reference measure on $\mathbb{H}$
ρ	Posterior distribution $\rho \in \mathbb{M}(\mathbb{H})$ on $\mathbb{H}$
π	Prior distribution $\pi \in \mathbb{M}^*(\mathbb{H})$ on $\mathbb{H}$
$\mathrm{KL}(\rho\\|\pi)$	Kullback-Leibler (KL) divergence between ρ and π
$\mathrm{D}_\alpha(\rho\\|\pi)$	Rényi Divergence between ρ and π
$\mathrm{Uni}(\mathbb{A})$	Uniform distribution on $\mathbb{A}$
$\mathrm{Dir}(\boldsymbol{\alpha})$	Dirichlet distribution of parameters $\boldsymbol{\alpha} \in \mathbb{R}_*^+$

Majority Vote

$\mathrm{MV}_\rho(\cdot)$	The majority vote classifier
$m_\rho(\mathbf{x},y)$	Majority vote's margin for the example $(\mathbf{x},y) \in \mathbb{X}\times\mathbb{Y}$
$\widehat{m_\rho}(\mathbf{x},y)$	$\frac{1}{2}$-Margin for the example $(\mathbf{x},y) \in \mathbb{X}\times\mathbb{Y}$
$\mathrm{sign}(a)$	Sign function; returns $+1$ if $a \geq 0$ and -1 otherwise
$r_{\mathcal{D}'}(\rho)$	Gibbs risk on the distribution $\mathcal{D}'$ associated to the majority vote $\mathrm{MV}_\rho()$
$e_{\mathcal{D}'}(\rho)$	Joint Error on the distribution $\mathcal{D}'$ associated to the majority vote $\mathrm{MV}_\rho()$
$d_{\mathcal{D}'}(\rho)$	Disagreement on the distribution $\mathcal{D}'$ associated to the majority vote $\mathrm{MV}_\rho()$

Preamble: Generalisation and Optimisation in Machine Learning

Detail broadly what generalisation is, to what kind of structures it is applied (neural nets or linear classfier eg). Details on the other hand what optimisation is doing (ERM eg) and explain that interestingly in various methods, reaching minimisers of empirical objectives is enough to ensure a good generalisation ability. From this, discuss about the current limitations of generalisaiton: either not going so often beyound light-tailed assumptions or noticing that the interplays between generalisation (statistical arguments) and optimisation (geometric ones) remains uncharted for a vast range of cases.
CHALLENGE HERE: being very rigorous on the lit review.

List of Publications

Conference article

Paul Viallard, Maxime Haddouche, Umut Simsekli, and Benjamin Guedj. Learning via Wasserstein-Based High Probability Generalisation Bounds. *Advances in Neural Information Processing Systems (NeurIPS)*. (2023).

Maxime Haddouche and Benjamin Guedj. Online PAC-Bayes Learning. *Advances in Neural Information Processing Systems (NeurIPS)*. (2022).

Journal article

Maxime Haddouche and Benjamin Guedj. PAC-Bayes Generalisation Bounds for Heavy-Tailed Losses through Supermartingales. *Transactions on Machine Learning Research*. (2023).

Research Report

Maxime Haddouche and Benjamin Guedj. Wasserstein PAC-Bayes Learning: A Bridge Between Generalisation and Optimisation. *arXiv*. abs/2304.07048. (2023).

Maxime Haddouche, Benjamin Guedj, and Olivier Wintenberger. Optimistic Dynamic Regret Bounds. (2023). arXiv: `2301.07530 [cs.LG]`.

Pierre Jobic, Maxime Haddouche, and Benjamin Guedj. Federated Learning with Nonvacuous Generalisation Bounds. (2023). arXiv: `2310.11203 [cs.LG]`.

Maxime Haddouche, Benjamin Guedj, and John Shawe-Taylor. Upper and Lower Bounds on the Performance of Kernel PCA. (2020). arXiv: `2012.10369 [cs.LG]`.

Part I

Background

1 An introduction to Statistical Learning Theory and PAC-Bayes

Contents

Abstract

This chapter provides a brief introduction of statistical learning theory and various modern kind of generalisation bounds (go from uniform convergence bounds to algorithmic stability, search for other kinds of non PAC-Bayes bounds). Need to recall the historical and modern shapes of generalisation bounds in ML.

1.1 A brief survey of generalisation bounds

1.2 PAC-Bayes learning

2 A Brief Reminder on Optimisation for Batch and Online Learning

Contents

Abstract

Detail optimisation for GD, SGD and variants, list a lot of convergence guarantees under various assumptions (covnexity smoothness etc). On the measure spaces part, detail the required OT background to introduce Wasserstein distances. Detail what is online learning.

2.1 Convergence guarantees for optimisation on predictor spaces

2.2 Wasserstein distances and optimisation on measure spaces

2.3 Optimisation in Online Learning

Additionnal Tools to Link Generalisation and Optimisation 3

Contents

Abstract

Put here additionnal background on differential privacy and log-Sobolev inequalities.

3.1 Differential privacy

3.2 Log-Sobolev inequalities

Part II

Generalisation bounds for Martingales and Online Learning allowing Heavy-Tailed Losses

4 PAC-Bayesian Bounds for Martingales and Heavy-Tailed losses

This chapter is based on the following paper

Maxime Haddouche and Benjamin Guedj. PAC-Bayes Generalisation Bounds for Heavy-Tailed Losses through Supermartingales. *Transactions on Machine Learning Research*. (2023)

Abstract

TODO: put general bounds for martingales and batch learning as corollary

Online PAC-Bayes Learning for Bounded Losses and Beyond 5

This chapter is based on the following papers

Maxime Haddouche and Benjamin Guedj. Online PAC-Bayes Learning. *Advances in Neural Information Processing Systems (NeurIPS)*. (2022)
Maxime Haddouche and Benjamin Guedj. PAC-Bayes Generalisation Bounds for Heavy-Tailed Losses through Supermartingales. *Transactions on Machine Learning Research*. (2023)
Paul Viallard, Maxime Haddouche, Umut Simsekli, and Benjamin Guedj. Learning via Wasserstein-Based High Probability Generalisation Bounds. *Advances in Neural Information Processing Systems (NeurIPS)*. (2023)

Abstract

Put OPB here. Precise in the intro that the martingale bounds allow to go beyond batch learning but that this has never been made for OL. Put the supermartingale OPB bound in a supplementary section and the Online WPB bound after the main results of OPB to reach heavy-tailed losses.

Part III

Towards A Better Understanding of Generalisation through Optimisation

6 Undestanding Generalisation through Flat Minimas

This chapter is based on the following paper

TODO

Contents

Abstract

This is the PLS paper, precise that the supermartingales bounds are richer than simply recovering classical batch guarantees: we can incorporate gradient norms, which explains generalisation when a flat minima is reached.

6.1 Introduction

7 Wasserstein PAC-Bayes Learning: Exploiting Optimisation Guarantees to Explain Generalisation

This chapter is based on the following papers

Maxime Haddouche and Benjamin Guedj. Wasserstein PAC-Bayes Learning: A Bridge Between Generalisation and Optimisation. *arXiv*. abs/2304.07048. (2023)

Paul Viallard, Maxime Haddouche, Umut Simsekli, and Benjamin Guedj. Learning via Wasserstein-Based High Probability Generalisation Bounds. *Advances in Neural Information Processing Systems (NeurIPS)*. (2023)

Contents

Abstract

Put WPB here, precise that beyond the somewhat vague understanding of generalisation through flat minima, it is possible, for a certain optimisation algorithm to directly incorporate a sound geometric optimisation guarantee into a generalisation bound, trading the hope to reach a flat minima with a sound convergence guarantees. However, this comes at the cost of the explicit impact of the dimension. Also put the paper with Paul(batch bounds) as a supplementary content

7.1 Introduction

Part IV

Conclusion and Perspectives

Part V

Appendix

Some Mathematical Tools A

A.1 Jensen's Inequality

Theorem A.1.1 (Jensen's Inequality)**.** Let $X \in \mathbb{X}$ a random variable following a probability distribution $\mathcal{X}$ with $f : \mathbb{X} \to \mathbb{R}$ a measurable convex function, we have

$$f\left(\mathop{\mathbb{E}}_{X\sim\mathcal{X}}[X]\right) \leq \mathop{\mathbb{E}}_{X\sim\mathcal{X}}\Big[f(X)\Big].$$

Proof. Since $f()$ is a convex function, the following inequality holds, *i.e.*, we have

$$\forall X' \in \mathbb{X}, \quad a\left(X' - \mathop{\mathbb{E}}_{X\sim\mathcal{X}}[X]\right) \leq f(X') - f\left(\mathop{\mathbb{E}}_{X\sim\mathcal{X}}[X]\right),$$

where a is the tangent's slope. By taking the expectation to both sides of the inequality, we have

$$\underbrace{a\left(\mathop{\mathbb{E}}_{X\sim\mathcal{X}}[X] - \mathop{\mathbb{E}}_{X\sim\mathcal{X}}[X]\right)}_{=0} \leq \mathop{\mathbb{E}}_{X\sim\mathcal{X}}[f(X)] - f\left(\mathop{\mathbb{E}}_{X\sim\mathcal{X}}[X]\right).$$

Hence, by rearranging the terms, we prove the claimed result. ■

A.2 Markov's Inequality

Theorem A.2.1 (Markov's Inequality)**.** Let $X \in \mathbb{X}$ a non-negative random variable following a probability distribution $\mathcal{X}$ and $\tau > 0$, we have

$$\mathop{\mathbb{P}}_{X\sim\mathcal{X}}[X \geq \tau] \leq \frac{\mathbb{E}_{X\sim\mathcal{X}}[X]}{\tau}.$$

Proof. First of all, remark that we have the following inequality for any $X \in \mathbb{X}$

$$\tau\, \mathrm{I}[X \geq \tau] \;\leq\; X\, \mathrm{I}[X \geq \tau] \;\leq\; X. \tag{A.1}$$

Indeed, on the one hand, if $X < \tau$, $\mathrm{I}[X \geq \tau] = 0$, the inequality holds trivially. On the other hand, if $X \geq \tau$, $\mathrm{I}[X \geq \tau] = 1$ and the inequality becomes $\tau \leq X$, which is true. By taking the expectation of Equation (A.1), we have

$$\mathop{\mathbb{E}}_{X\sim\mathcal{X}}\Big[\tau\, \mathrm{I}[X \geq \tau]\Big] \leq \mathop{\mathbb{E}}_{X\sim\mathcal{X}}\Big[X\Big].$$

From the fact that the expectation of a constant is the constant and by definition of the probability, we have

$$\tau \mathop{\mathbb{P}}_{X\sim\mathcal{X}}[X \geq \tau] \leq \mathop{\mathbb{E}}_{X\sim\mathcal{X}}\Big[X\Big] \quad\iff\quad \mathop{\mathbb{P}}_{X\sim\mathcal{X}}[X \geq \tau] \leq \frac{\mathbb{E}_{X\sim\mathcal{X}}[X]}{\tau},$$

which is the desired result. ■

A.3 2nd Order Markov's Inequality

Theorem A.3.1 (2nd Order MARKOV's Inequality)**.** Let X a non-negative random variable following a probability distribution $\mathcal{X}$ and $\tau > 0$, we have

$$\mathop{\mathbb{P}}_{X\sim\mathcal{X}}[X \geq \tau] \leq \frac{\mathbb{E}_{X\sim\mathcal{X}}[X^2]}{\tau^2}.$$

Proof. We apply MARKOV's inequality (Theorem A.2.1) to have

$$\mathop{\mathbb{P}}_{X\sim\mathcal{X}}\left[X^2 \geq \tau^2\right] \leq \frac{\mathbb{E}_{X\sim\mathcal{X}}[X^2]}{\tau^2}.$$

Moreover, since $\mathrm{I}\left[X \geq \tau\right] = \mathrm{I}\left[X^2 \geq \tau^2\right]$, we have

$$\mathop{\mathbb{P}}_{X\sim\mathcal{X}}[X \geq \tau] = \mathop{\mathbb{P}}_{X\sim\mathcal{X}}\left[X^2 \geq \tau^2\right],$$

which proves the desired result. ■

A.4 Chebyshev-Cantelli Inequality

Theorem A.4.1 (CHEBYSHEV-CANTELLI Inequality)**.** Let X a random variable following a probability distribution $\mathcal{X}$ and $\tau > 0$, we have

$$\mathop{\mathbb{P}}_{X\sim\mathcal{X}}\left[X - \mathop{\mathbb{E}}_{X'\sim\mathcal{X}} X' \geq \tau\right] \leq \frac{\mathbb{V}_{X'\sim\mathcal{X}}\, X'}{\mathbb{V}_{X'\sim\mathcal{X}}\, X' + \tau^2}.$$

Proof. First of all, remark that we have

$$\begin{aligned}\mathop{\mathbb{P}}_{X\sim\mathcal{X}}\left[X - \mathop{\mathbb{E}}_{X'\sim\mathcal{X}} X' \geq \tau\right] &= \mathop{\mathbb{P}}_{X\sim\mathcal{X}}\left[X - \mathop{\mathbb{E}}_{X'\sim\mathcal{X}} X' + \frac{\mathbb{V}_{X'\sim\mathcal{X}}\, X'}{\tau} \geq \tau + \frac{\mathbb{V}_{X'\sim\mathcal{X}}\, X'}{\tau}\right]\\ &\leq \mathop{\mathbb{P}}_{X\sim\mathcal{X}}\left[\left[X - \mathop{\mathbb{E}}_{X'\sim\mathcal{X}} X' + \frac{\mathbb{V}_{X'\sim\mathcal{X}}\, X'}{\tau}\right]^2 \geq \left[\tau + \frac{\mathbb{V}_{X'\sim\mathcal{X}}\, X'}{\tau}\right]^2\right],\end{aligned} \tag{A.2}$$

where $\mathbb{V}_{X\sim\mathcal{X}}\, X$ is the variance of the random variable $X \sim \mathcal{X}$. From Equation (A.2) and MARKOV's Inequality (Theorem A.2.1), we can deduce that

$$\begin{aligned}\mathop{\mathbb{P}}_{X\sim\mathcal{X}}\left[X - \mathop{\mathbb{E}}_{X'\sim\mathcal{X}} X' \geq \tau\right] &\leq \frac{\mathbb{E}_{X\sim\mathcal{X}}\left[X - \mathbb{E}_{X'\sim\mathcal{X}}\, X' + \frac{\mathbb{V}_{X'\sim\mathcal{X}}\, X'}{\tau}\right]^2}{\left[\tau + \frac{\mathbb{V}_{X'\sim\mathcal{X}}\, X'}{\tau}\right]^2}\\ &= \frac{\mathbb{V}_{X'\sim\mathcal{X}}\, X'}{\mathbb{V}_{X'\sim\mathcal{X}}\, X' + \tau^2}.\end{aligned}$$

■

A.5 Hölder's Inequality

In order to prove HÖLDER's inequality, we first prove the following lemma.

Lemma A.5.1 (YOUNG's Inequality)**.** For any $\alpha > 1$ and $\beta > 1$ such that $\frac{1}{\alpha}+\frac{1}{\beta} = 1$, for any $a \geq 0$ and $b \geq 0$, we have

$$ab \leq \frac{a^\alpha}{\alpha} + \frac{b^\beta}{\beta}.$$

Proof. We first develop $\ln[ab]$ and we apply JENSEN's inequality (Theorem A.1.1) since the logarithm is concave and $\frac{1}{\alpha}+\frac{1}{\beta}=1$. Indeed, we have

$$\ln[ab] = \ln a + \ln b = \frac{\alpha}{\alpha}\ln a + \frac{\beta}{\beta}\ln b = \frac{\ln a^{\alpha}}{\alpha} + \frac{\ln b^{\beta}}{\beta} \le \ln\left[\frac{a^{\alpha}}{\alpha}+\frac{b^{\beta}}{\beta}\right].$$

Then, we take the exponential to both sides of the inequality and we are done. ■

We are now ready to prove HÖLDER's inequality.

Theorem A.5.1 (HÖLDER's Inequality). For any measurable function $f()$ and $g()$, for any $\alpha > 1$ and $\beta > 1$ such that $\frac{1}{\alpha}+\frac{1}{\beta}=1$, we have

$$\mathop{\mathbb{E}}_{X\sim\mathcal{X}}|f(X)g(X)| \le \left[\mathop{\mathbb{E}}_{X\sim\mathcal{X}}|f(X)|^{\alpha}\right]^{\frac{1}{\alpha}}\left[\mathop{\mathbb{E}}_{X\sim\mathcal{X}}|g(X)|^{\beta}\right]^{\frac{1}{\beta}}.$$

Proof. For convenience of notation, let $\|f\|_{\alpha} = [\mathbb{E}_{X\sim\mathcal{X}}|f(X)|^{\alpha}]^{\frac{1}{\alpha}}$ and $\|g\|_{\beta} = \left[\mathbb{E}_{X\sim\mathcal{X}}|g(X)|^{\beta}\right]^{\frac{1}{\beta}}$. If $\|f\|_{\alpha}=0$ or $\|g\|_{\beta}=0$, then $\mathbb{E}_{X\sim\mathcal{X}}|f(X)g(X)|=0$, hence, the inequality holds in this case. Then for $\|f\|_{\alpha}>0$ and $\|g\|_{\beta}>0$, we upper-bound the term $\frac{|f(X)g(X)|}{\|f\|_{\alpha}\|g\|_{\beta}}$ with YOUNG's inequality (Lemma A.5.1), *i.e.*, we have

$$\frac{|f(X)g(X)|}{\|f\|_{\alpha}\|g\|_{\beta}} \le \frac{|f(X)|^{\alpha}}{\alpha\|f\|_{\alpha}^{\alpha}} + \frac{|f(X)|^{\beta}}{\beta\|f\|_{\beta}^{\beta}}.$$

By taking the expectation *w.r.t.* $X\sim\mathcal{X}$, we have

$$\begin{aligned}\frac{\mathbb{E}_{X\sim\mathcal{X}}|f(X)g(X)|}{\|f\|_{\alpha}\|g\|_{\beta}} &\le \frac{\mathbb{E}_{X\sim\mathcal{X}}|f(X)|^{\alpha}}{\alpha\|f\|_{\alpha}^{\alpha}} + \frac{\mathbb{E}_{X\sim\mathcal{X}}|f(X)|^{\beta}}{\beta\|f\|_{\beta}^{\beta}}\\ &= \frac{\|f\|_{\alpha}^{\alpha}}{\alpha\|f\|_{\alpha}^{\alpha}} + \frac{\|f\|_{\beta}^{\beta}}{\beta\|f\|_{\beta}^{\beta}}\\ &= \frac{1}{\alpha}+\frac{1}{\beta}\\ &= 1.\end{aligned}$$

This concludes the proof since

$$\frac{\mathbb{E}_{X\sim\mathcal{X}}\,|f(X)g(X)|}{\|f\|_\alpha\|g\|_\beta} \leq 1 \iff \mathop{\mathbb{E}}_{X\sim\mathcal{X}}|f(X)g(X)| \leq \|f\|_\alpha\|g\|_\beta.$$

■

Abstract. In machine learning, a model is learned from data to solve a task automatically. In the supervised classification setting, the model aims to predict the label associated with an input. The model is learned using a limited number of examples, each consisting of an input and its associated label. However, the model's performance on the examples, computed by the empirical risk, does not necessarily reflect the performance on the task, which is represented by the true risk. Moreover, since it is not computable, the true risk is upper-bounded by a generalization bound that mainly depends on two quantities: the empirical risk and a complexity measure. One way to learn a model is to minimize a bound by a type of algorithm called self-bounding. PAC-Bayesian bounds are well suited to the derivation of this type of algorithm. In this context, the first contribution consists in developing self-bounding algorithms that minimize PAC-Bayesian bounds to learn majority votes. If these bounds are well adapted to majority votes, their use for other models becomes less natural. To overcome this difficulty, a second contribution focuses on the disintegrated PAC-Bayesian bounds that are natural for more general models. In this framework, we provide the first empirical study of these bounds. In a third contribution, we derive bounds that allow us to incorporate complexity measures defined by the user.

Keywords. Machine Learning, Generalization, PAC-Bayesian Bound, Disintegrated PAC-Bayesian Bound, Self-Bounding Algorithm, Majority Vote, Neural Network, Complexity Measure.

Résumé. En apprentissage automatique, un modèle est appris à partir de données pour résoudre une tâche de manière automatique. Dans le cadre de la classification supervisée, le modèle vise à prédire la classe associée à une entrée. Le modèle est appris à l'aide d'un nombre limité d'exemples, chacun étant constitué d'une entrée et de sa classe associée. Cependant, la performance du modèle sur les exemples, calculée par le risque empirique, ne reflète pas nécessairement la performance sur la tâche qui est représentée par le risque réel. De plus, n'étant pas calculable, le risque réel est majoré pour obtenir une borne en généralisation qui dépend principalement de deux quantités : le risque empirique et une mesure de complexité. Une façon d'apprendre un modèle est de minimiser une borne par un type d'algorithme appelé auto-certifié (ou auto-limitatif). Les bornes PAC-Bayésiennes sont bien adaptées à la dérivation de ce type d'algorithmes. Dans ce contexte, la première contribution consiste à développer des algorithmes auto-certifiés qui minimisent des bornes PAC-Bayésiennes pour apprendre des votes de majorité. Si ces bornes sont bien adaptées aux votes de majorité, leur utilisation pour d'autres modèles devient moins naturelle. Pour pallier cette difficulté, une seconde contribution se concentre sur les bornes PAC-Bayésiennes désintégrées qui sont naturelles pour des modèles plus généraux. Dans ce cadre, nous apportons la première étude empirique de ces bornes. Dans une troisième contribution, nous dérivons des bornes permettant d'incorporer des mesures de complexité pouvant être définies par l'utilisateur.

Mot-clés. Apprentissage Automatique, Généralisation, Borne PAC-Bayésienne, Borne PAC-Bayésienne Désintégrée, Algorithme Auto-certifié, Algorithme Auto-limitatif, Vote de Majorité, Réseau de Neurones, Mesure de Complexité.