

Data Science

Massih-Reza Amini[†], Renaud Blanch[†], Marianne Clause[‡], Jean-Baptiste Durand[‡],
Eric Gaussier[†], Jérôme Malick[‡], Christophe Picard[‡], Vivien Quéma[†], Georges Quénot[†]

Université Grenoble Alpes, Grenoble INP, CNRS

[†] Laboratoire d'Informatique de Grenoble

[‡] Laboratoire Jean Kunzmann

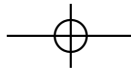
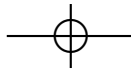
Abstract

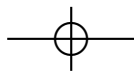
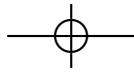
La data science, ou science des données, est la discipline qui traite de la collecte, de la préparation, de la gestion, de l'analyse, de l'interprétation et de la visualisation de grands ensembles de données complexes. Elle n'est pas seulement concernée par les outils et les méthodes pour obtenir, gérer et analyser les données ; elle consiste aussi à en extraire de la valeur et de la connaissance.

Cet ouvrage présente les fondements scientifiques et les composantes essentielles de la science des données, à un niveau accessible aux étudiants de master et aux élèves ingénieurs. Notre souci a été de proposer un exposé cohérent reliant la théorie aux algorithmes développés dans ces domaines. Il s'adresse aux chercheurs et ingénieurs qui abordent les problématiques liées à la science des données, aux data scientists de PME qui utilisent en profondeur les outils d'apprentissage, mais aussi aux étudiants de master, doctorants ou encore futurs ingénieurs qui souhaitent un ouvrage de référence en data science.

À qui s'adresse ce livre ?

- *Aux développeurs, statisticiens, étudiants et chefs de projets ayant à résoudre des problèmes de data science.*
- *Aux data scientists, mais aussi à toute personne curieuse d'avoir une vue d'ensemble de l'état de l'art du machine learning.*





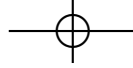
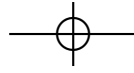


Table des matières

Table des figures	IX
Liste des algorithmes	XIII
CHAPITRE 1	
Introduction	1
CHAPITRE 2	
Prétraitement des données	5
2.1 Prétraitement des données	
textuelles	5
2.1.1 Segmentation	6
2.1.2 Normalisation et filtrage	7
2.1.3 Filtrage	10
2.2 Prétraitement des données image	
ou vidéo	11
2.2.1 Représentations globales	12
2.2.2 Représentations locales	16
2.2.3 Agrégation de représentations locales	17
2.2.4 Représentations apprises	19
2.3 Exercices	19
CHAPITRE 3	
Gestion de données large-échelle et systèmes distribués	21
3.1 Les limites des systèmes	
traditionnels de gestion de données	21

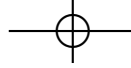


IV – DATA SCIENCE

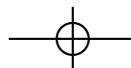
3.1.1	Les besoins liés au traitement de grandes masses de données	22
3.1.2	Limites des architectures incrémentales	23
3.2	L'architecture Lambda pour le traitement de grandes masses de données	24
3.3	La couche batch : traitement de données par lots	27
3.3.1	Caractéristiques du jeu de données principal	28
3.3.2	Stockage du jeu de données principal	29
3.3.3	Traitement de données par lots	30
3.4	La couche service : stockage et requêtes sur les vues batch	33
3.4.1	Remarque préliminaire sur le stockage des vues batch	33
3.4.2	Stockage des vues batch	35
3.5	La couche vitesse : traitement de flux de données, stockage et requêtes sur les vues temps réel	36
3.5.1	Traitement de flux de données	37
3.5.2	Stockage des vues temps réel	38
3.6	Conclusion	38

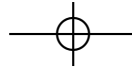
CHAPITRE 4

Calcul haute performance	41
4.1 Introduction	41
4.1.1 Motivations	42
4.1.2 Hiérarchies du parallélisme	44
4.1.3 Classification des plateformes	47
4.1.4 Coûts de communication	48
4.2 Principes de conception des algorithmes	50
4.2.1 Techniques de décomposition	51
4.2.2 Caractéristiques des tâches et des interactions	53
4.2.3 Équilibrage des ressources	54
4.2.4 Modèles d'algorithmes parallèles	56

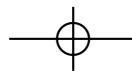


4.3 Modèles analytiques	57
4.3.1 Métriques de performances	58
4.3.2 Passage à l'échelle des systèmes parallèles	59
4.3.3 Effet de la granularité	61
4.3.4 Notion d'iso-efficacité	62
4.4 Conclusion	62
 CHAPITRE 5	
Optimisation pour l'analyse de données	65
5.1 Apprentissage et optimisation	66
5.2 Introduction à l'optimisation	72
5.2.1 Problèmes d'optimisation	72
5.2.2 Analyse convexe pour impatientes	74
5.2.3 Algorithmes d'optimisation	77
5.3 Algorithmes en science des données	82
5.3.1 Algorithmes incrémentaux	83
5.3.2 Algorithmes distribués	87
5.3.3 Au-delà de ce chapitre	91
5.4 Exercices	91
 CHAPITRE 6	
Décomposition matricielle/tensorielle	95
6.1 Motivations	95
6.2 La SVD	96
6.2.1 Quelques rappels d'algèbre linéaire	96
6.2.2 Approximation de rang faible	97
6.2.3 SVD et analyse en composantes principales	99
6.2.4 Algorithme pour déterminer la SVD d'une matrice	100
6.3 Décomposition en matrices non négatives	104
6.3.1 Algorithme de Seung et Lee	105
6.3.2 Algorithme des moindres carrés alternés	107
6.3.3 Comparaison de la SVD et de la NMF	107
6.4 Décomposition tensorielle	108
6.4.1 Décomposition canonique polyadique	109

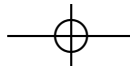




6.5	Conclusion	111
6.6	Exercices	111
CHAPITRE 7		
	Modèles génératifs	115
7.1	Motivations	115
7.1.1	Modèles graphiques	117
7.1.2	Modèles à variables latentes	117
7.2	Introduction à la statistique bayésienne	120
7.2.1	Généralités	120
7.2.2	Algorithmes génériques pour l'inférence bayésienne	123
7.3	Inférence dans les modèles à variables latentes	127
7.3.1	Modèles probabilistes graphiques	127
7.3.2	Mélanges	129
7.3.3	Analyse en composantes principales probabiliste	132
7.3.4	Chaînes de Markov cachées	134
7.3.5	Modèles hiérarchiques à variables latentes	136
7.4	Références	138
7.5	Exercices	140
CHAPITRE 8		
	Modèles discriminants	145
8.1	Approches supervisées	146
8.1.1	Modèles binaires	147
8.1.2	Modèles multi-classes	161
8.2	Approches semi-supervisées	164
8.2.1	Modèles graphiques	165
8.2.2	Modèles de mélange	171
8.2.3	Modèles discriminants	171
8.3	Exercices	172

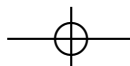
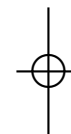


CHAPITRE 9	
Deep learning	177
9.1 Neurone formel	178
9.2 Réseaux simples	179
9.2.1 Perceptron	180
9.2.2 ADALINE	183
9.2.3 Perceptrons multicouches (PMC)	185
9.3 Réseaux à propagation avant	186
9.3.1 Composition de fonctions	186
9.3.2 Fonction-objectif et descente de gradient stochastique par mini-lots	187
9.3.3 Calcul des gradients par rétro-propagation de l'erreur	188
9.3.4 Architecture modulaire	190
9.3.5 Réseaux quelconques	195
9.3.6 Différentiation automatique	196
9.4 Réseaux convolutifs	197
9.4.1 Couche de convolution	197
9.4.2 Changements de résolution	199
9.4.3 Passage à des couches complètement connectées	200
9.4.4 Un exemple : AlexNet	200
9.5 Optimisations supplémentaires	202
9.5.1 Traitement par mini-lots	202
9.5.2 Moment	202
9.5.3 Fonctions d'activation	203
9.5.4 Dropout	204
9.5.5 Normalisation de lots	204
9.5.6 Augmentation de données	205
9.6 Réseaux pour la catégorisation d'images	206
9.7 Exercices	207
CHAPITRE 10	
Visualisation interactive d'information	211
10.1 Introduction	211
10.2 Des données au graphique	214
10.2.1 Les données	214



VIII – DATA SCIENCE

10.2.2 L'image	216
10.2.3 Encodage visuel	224
10.3 Encodages avancés	227
10.3.1 Utilisation multiple des variables graphiques	227
10.3.2 Encodage des liens entre individus	230
10.4 Pour aller plus loin	234
10.5 Exercices	235
Bibliographie	239
Index	253



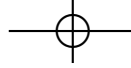
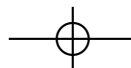
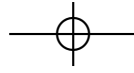


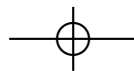
Table des figures

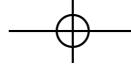
2.1	Filtres de Gabor circulaire (gauche) et elliptique (droite). La zone indiquée correspond à une extension d'un écart-type sur la gaussienne.	15
2.2	Banc de filtres de Gabor elliptiques avec 8 orientations et 4 échelles, partie réelle.	15
3.1	Architecture incrémentale.	23
3.2	L'architecture Lambda.	25
3.3	Exemple de stockage normalisé.	34
3.4	Exemple de stockage dénormalisé.	35
3.5	Architecture d'un système de traitement de flux de données.	37
4.1	Évolution des puissances de calcul des machines du Top 500 https://top500.org	43
4.2	Classification de Flynn	47
4.3	Méthodologie de conception d'une approche parallèle	50
4.4	Évolution de l'accélération pour différentes lois	60
5.1	Illustration de la convexité avec des fonctions de \mathbb{R} dans \mathbb{R} . À gauche, une fonction convexe; à droite, une fonction non convexe (dont on note qu'elle admet un minimum local en ω qui n'est pas un minimum).	75
5.2	Illustration de l'inégalité (5.12) pour une fonction convexe de \mathbb{R} dans \mathbb{R} . Une fonction convexe est toujours au-dessus de ses tangentes.	75



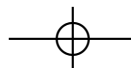


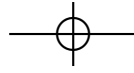
5.3	Dessin des lignes de niveau de deux fonctions de \mathbb{R}^2 dans \mathbb{R} avec les itérés de l'algorithme du gradient, illustrant la vitesse de convergence (5.15) et le rôle du conditionnement $\kappa = \mu/L$. Quand κ est grand (comme à droite), le problème est bien conditionné et l'algorithme converge vite ; plus le ratio est petit (comme à gauche), plus le problème est dur et plus l'algorithme patine.	78
5.4	Itérations des algorithmes du gradient (à gauche) et du gradient stochastique (à droite) sur les lignes de niveau d'une fonction de \mathbb{R}^2 dans \mathbb{R} . L'espérance des directions de descente correspond à la direction du gradient et on récupère des garanties de convergence si on peut aussi contrôler la variance des directions.	84
5.5	Représentation schématique du cadre d'optimisation distribuée : une machine-maître coordonne les calculs effectués en parallèle sur les machines-esclaves, stockant chacune une partie des données. Soulignons que les algorithmes ne communiquent que des résultats de calcul, pas des données brutes.	89
6.1	Tranches et fibres d'un tenseur du troisième ordre	108
7.1	Un graphe d'indépendance conditionnelle non orienté	128
7.2	Un DAG d'indépendance conditionnelle orienté acyclique.	129
7.3	Modèle graphique définissant les dépendances entre variables aléatoires d'une CMC.	135
7.4	Modèle graphique définissant les dépendances entre variables aléatoires du LDA.	137
8.1	Illustration du fonctionnement de l'algorithme d'Adaboost sur un problème jouet où la combinaison finale des classifieurs linéaires de base (ou faibles) conduit à un classifieur non linéaire. Le vecteur normal de chaque classifieur faible pointe vers le demi-espace des exemples positifs (cercles pleins et les exemples mal classés sont encerclés). À une itération donnée, le classifieur courant tente de bien classer les exemples mal classés par celui de l'itération précédente et les poids w de ces classifieurs, intervenant dans le vote majoritaire final, sont inversement proportionnels à leurs erreurs de classification.	150





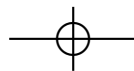
8.2	Hyperplans pour un problème de classification linéairement séparable en dimension 2. Les vecteurs de support appartenant aux hyperplans marginaux d'équations $\langle \bar{\omega}, \mathbf{x} \rangle + w_0 = \pm 1$ sont encerclés.	153
8.3	Hyperplans linéaires pour un problème de classification non linéairement séparable. Les vecteurs de support sont encerclés. Soit ces vecteurs reposent sur un des hyperplans marginaux, soit ce sont des points aberrants. La distance d'un point aberrant \mathbf{x} à l'hyperplan marginal associé à sa classe est $\frac{\xi}{\ \bar{\omega}\ }$	157
9.1	Illustration d'un neurone formel.	178
9.2	Architecture d'un perceptron inspiré du système perceptif et composé de quatre composantes principales : la rétine, les fonctions d'association, les poids synaptiques et l'unité à seuil.	181
9.3	Illustration de la règle de mise à jour de l'algorithme du perceptron (eq. 9.8) avec l'exemple $(\mathbf{x}_3, -1)$ choisi, qui est mal classé par l'hyperplan de vecteur normal $\omega^{(t)}$	183
9.4	Illustration des solutions trouvées par les algorithmes du perceptron (en pointillés) et de l'adaline (en trait plein) pour un problème de classification linéairement séparable.	184
9.5	Architecture d'un perceptron multicouche à une couche cachée (de profondeur 2). Sur cet exemple, les paramètres des biais sont introduits par des poids liés à deux unités supplémentaires associés à la couche d'entrée et à la première couche cachée ayant respectivement les valeurs fixées $x_0 = 1$ et $z_0 = 1$	186
9.6	Architecture modulaire d'un système d'entraînement d'un réseau multicouche (inspiré de LeCun et Ranzato (2013)). Propagation avant : du bas vers le haut ; rétro-propagation selon les données : du haut vers le bas ; rétro-propagation selon les paramètres : du haut vers la gauche.	191
9.7	Module « couche ». Propagation avant : du bas vers le haut ; rétro-propagation selon les données : du haut vers le bas ; rétro-propagation selon les paramètres : du haut vers la gauche.	192
10.1	Le <i>quartet</i> d'Anscombe (1973) visualisé.	212
10.2	La carte (à gauche) de l'épidémie de choléra à Londres en 1854 par le docteur J. Snow. À droite est visualisé un détail centré sur la fontaine infectée.	212

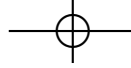




XII – DATA SCIENCE

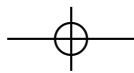
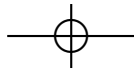
10.3	Les pertes humaines lors de la campagne de Russie par C. J. Minard (1869).	213
10.4	Photo tirée du n° 58 de février 1958 de <i>Life Magazine</i> par Gregory (1970) pour illustrer l' <i>émergence</i> des formes selon la <i>Gestalt</i> . Ce motif est aussi présent dans le tableau <i>le Torero Hallucinogène</i> de Salvador Dalí (1970).	217
10.5	Des groupes émergent du fait de leur proximité ou de leur ressemblance.	218
10.6	Trouver une marque foncée au milieu des claires (gauche) ou un carré au milieu de cercles (centre) demande un temps qui ne dépend pas du nombre de marques, alors que trouver le cercle foncé au milieu de cercles et de carrés clairs ou foncés (droite) prend un temps qui croît linéairement en fonction du nombre de marques.	219
10.7	Capture de l'interface de Gapminder.	225
10.8	Le jeu de données <i>Iris</i> , à quatre dimensions quantitatives, représenté par une matrice de nuages de points (gauche) et des coordonnées parallèles (droite).	228
10.9	Évolution des écarts à la moyenne des taux de chômage de divers états des États-Unis d'Amérique visualisée grâce à des <i>Horizon Graphs</i>	230
10.10	Deux représentations « nœuds-liens » d'un même arbre avec différents algorithmes de placement.	231
10.11	Deux représentations d'arbres utilisant l'emboîtement (de cercles et de rectangles) pour figurer la hiérarchie.	232
10.12	Un réseau visualisé sous forme de diagramme « nœuds-liens » (gauche) et d'une matrice d'adjacence (droite).	233
10.13	Visualisation d'un jeu de données de voitures.	236

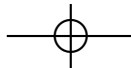




Liste des algorithmes

1	Réduction de Householder d'une matrice A sous forme bi-diagonale	102
2	Décomposition QR d'une matrice B	103
3	Version de l'algorithme de la NMF dans le cas simple où A est un vecteur y	105
4	Version générale de l'algorithme de Seung et Lee	106
5	Algorithme ALS	107
6	Algorithme des moindres carrés alternés (ALS) pour la décomposition polyadique	110
7	Algorithme d'acceptation-rejet	124
8	Algorithme de Métropolis Hastings	125
9	Algorithme de Gibbs	126
10	Algorithme de LDA	138
11	Algorithme d'Adaboost	149
12	SVM à marge dure	156
13	SVM à marge souple	160
14	Stratégie des codes correcteurs d'erreur pour la classification multi-classe	163
15	Propagation des étiquettes pour l'apprentissage semi-supervisé	167
16	Algorithme PEGASOS Shalev-Shwartz <i>et al.</i> (2011)	174
17	Algorithme de Perceptron	182





Bibliographie

W. AIGNER, S. MIKSCH, H. SCHUMANN et C. TOMINSKI : *Visualization of time-oriented data*. Springer, 2011.

E.L. ALLWIN, R.E. SCHAPIRE et Y. SINGER : Reducing multiclass to binary : a unifying approach for margin classifiers. *Journal of Machine Learning Research*, 1 :113 – 141, 2000.

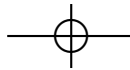
M.-R. AMINI : *Apprentissage machine : de la théorie à la pratique*. Eyrolles, 2015.

M.-R. AMINI et E. GAUSSIER : *Recherche d'information : applications, modèles et algorithmes*. 2^{de} édition, Eyrolles, 2017.

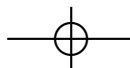
M.-R. AMINI, N. USUNIER et F. LAVIOLETTE : A transductive bound for the voted classifier with an application to semi-supervised learning. *In Advances in Neural Information Processing Systems 21*, pages 65–72, 2009.

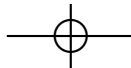
H. AMOUALIAN, M. CLAUSEL, E. GAUSSIER et M.-R. AMINI : Streaming-lda : A copula-based approach to modeling topic dependencies in document streams. *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 695–704, 2016.

D.P. ANDERSON : BOINC : a system for public-resource computing and storage. *In Fifth IEEE/ACM International Workshop on Grid Computing*, 2004.

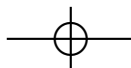
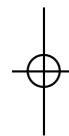


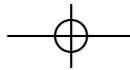
- J.A. ANDERSON : Logistic discrimination. *Handbook of Statistics*, 2 : 169–191, 1982.
- R. J. ANDREWS : An interactive timeline of the most iconic infographics, 2017.
<http://infowetrust.com/scroll/>.
- F.J. ANSCOMBE : Graphs in statistical analysis. *American Statistician*, 27(1) :17–21, 1973.
- L. BALDO, L. BRENNER, L.G. FERNANDES, P. FERNANDES et A. SALES : Performance models for master/slave parallel programs. *Electronic Notes in Theoretical Computer Science*, 2005.
- G. BALIKAS et M.R. AMINI : Twice at semeval-2016 task 4 : Twitter sentiment classification. *CoRR*, abs/1606.04351, 2016.
- S. BASU, A. BANERJEE et R.J. MOONEY : Semi-supervised clustering by seeding. *In Proceedings of the Nineteenth International Conference on Machine Learning*, 2002.
- A.G. BAYDIN, B.A. PEARLMUTTER et A. RADUL : Automatic differentiation in machine learning : a survey. *CoRR*, abs/1502.05767, 2015.
- M. BEHRISCH, B. BACH, R.N. HENRY, T. SCHRECK et J.-D. FEKETE : Matrix reordering methods for table and network visualization. *Computer Graphics Forum*, 35 :24, 2016.
- M.W. BERRY, D. MEZHER, B. PHILIPPE et A. SAMEH : *Handbook of parallel computing and statistics*, chapitre Parallel algorithms for the singular value decomposition. Statistics, textbooks and monographs. Chapman & Hall/CRC, 2005.
- J. BERTIN : *Sémiologie graphique*. Mouton/Gauthier-Villars, 1967.
- C. BISHOP : *Pattern recognition and machine learning*. Springer Verlag, 2006.
- C. BISHOP et J. LASSERE : Generative or discriminative? Getting the best of both worlds. *Bayesian Statistics*, 8 :3–24, 2007.



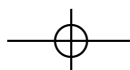


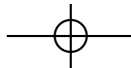
- R. BLANCH, R. DAUTRICHE et G. BISSON : Dendrogramix : a hybrid tree-matrix visualization technique to support interactive exploration of dendrograms. *In Proceedings of Pacific Vis 2015*, 2015.
- R. BLANCH et E. LECOLINET : Browsing zoomable treemaps : structure-aware multi-scale navigation techniques. *IEEE Transactions on Visualization and Computer Graphics*, 13(6) :1248–1253, 2007.
- L. BOTTOU, F.E. CURTIS et J. NOCEDAL : Optimization methods for large-scale machine learning. *arXiv preprint :1606.04838*, 2017.
- S. BOUCHERON, O. BOUSQUET et G. LUGOSI : Theory of classification : a survey of some recent advances. *ESAIM : Probability and Statistics*, 2005.
- S. BOYD et L. VANDENBERGHE : *Convex optimization*. Cambridge University Press, 2004.
- C. BREWER : ColorBrewer : color advice for cartography, 2018.
<http://www.ColorBrewer.org>.
- S. BUBECK : Convex optimization : algorithms and complexity. *Foundations and Trends in Machine Learning*, 8(3-4) :231–357, 2015.
- S.K. CARD et J. MACKINLAY : The structure of the information visualization design space. *In proc. InfoVis'97*, 1997.
- J.D. CARROLL et J.J. CHANG : Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition. *Psychometrika*, 35 :283–319, 1970.
- E.H. CHI et J.T. RIEDL : An operator interaction framework for visualization systems. *In proc. InfoVis'98*, 1998.
- D.M. CHITTY : Fast parallel genetic programming : multi-core CPU versus many-core GPU. *Soft Computing*, 2012.
- J. CHOI, D.W. WALKER et J.J. DONGARRA : Pumma : Parallel universal matrix multiplication algorithms on distributed memory concurrent computers. *Concurrency : Practice and Experience*, 1994.



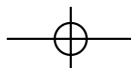
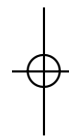


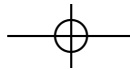
- W.S. CLEVELAND et R. MCGILL : Graphical perception : theory, experimentation and application to the development of graphical methods. *Journal of American Statistics Association*, 79(387) :531–554, 1984.
- I. COHEN, F.G. COZMAN, N. SEBE, M.C. CIRELO et T.S. HUANG : Semisupervised learning of classifiers : theory, algorithms and their application to human-computer interaction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(12) :1553–1567, 2004.
- M. CORDEIL, A. CUNNINGHAM, T. DWYER, B.H. THOMAS et K. MARRIOTT : Imaxes : immersive axes as embodied affordances for interactive multivariate data visualisation. *In Proc. ACM UIST*, 2017.
- F.G. COZMAN et I. COHEN : Unlabeled data can degrade classification performance of generative classifiers. *In Fifteenth International Florida Artificial Intelligence Society Conference*, 2002.
- G. CSURKA, C. R. DANCE, L. FAN, J. WILLAMOWSKI et C. BRAY : Visual categorization with bags of keypoints. *In Workshop on Statistical Learning in Computer Vision, ECCV*, 2004.
- M. DANELUTTO, T. DE MATTEIS, G. MENCAGLI et M. TORQUATI : A divide-and-conquer parallel pattern implementation for multicores. *In Proceedings of the 3rd International Workshop on Software Engineering for Parallel Systems*, 2016.
- S. DEERWESTER, S.T. DUMAIS, G.W. FURNAS, T.K. LANDAUER et R. HARSHMAN : Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6) :391–407, 1990.
- J. DIEBOLT et C.P. ROBERT : Estimation of finite mixture distributions through bayesian sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(2) :363–375, 1994.
- T.G. DIETTERICH et G. BAKIRI : Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 2 :263–286, 1995.
- S. DONNET et J.-M. MARIN : An empirical bayes procedure for the selection of gaussian graphical models. *Statistics and Computing*, 22 (5) :1113–1123, 2012.



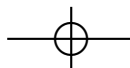
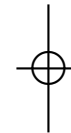


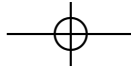
- R.O. DUDA, P.E. HART et D.G. STORK : *Pattern classification*. Wiley, 2001.
- N. ELMQVIST, P. DRAGICEVIC et J.-D. FEKETE : Rolling the dice : multidimensional visual exploration using scatterplot matrix navigation. *IEEE Transactions on Visualization and Computer Graphics*, 14 (6) :1141–1148, 2008.
- Y. EPHRAIM et N. MERHAV : Hidden Markov processes. *IEEE Transactions on Information Theory*, 48 :1518–1569, 2002.
- R.-E. FAN, K.-W. CHANG, C.-J. HSIEH, X.-R. WANG et C.-J. LIN : Liblinear : a library for large linear classification. *Journal of Machine Learning Research*, 9 :1871–1874, 2008.
- P. FEARNHEAD : Exact and efficient bayesian inference for multiple changepoint problems. *Statistics and Computing*, 16(2) :203–213, 2006.
- J.P. FISHBURN : *Analysis of speedup in distributed algorithms*. Ann Arbor, Mich. : UMI Research Press, 1984.
- M.J. FLYNN : Some computer organizations and their effectiveness. *IEEE Trans. Computers*, 1972.
- I.T. FOSTER : *Designing and building parallel programs : concepts and tools for parallel software engineering*. Addison-Wesley Longman Publishing Co., Inc., 1995.
- S.C. FRALICK : Learning to recognize patterns without a teacher. *IEEE Transactions on Information Theory*, 13(1) :57–64, 1967.
- Y. FREUND : Boosting a weak learning algorithm by majority. *Information and Computation*, 121 :256–285, 1995.
- K. FUKUNAGA : *Introduction to statistical pattern recognition*. Academic Press, New York, USA, 1972.
- E. GAUSSIER et C. GOUTTE : Relation between PLSA and NMF and implications. *In Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2005.



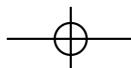
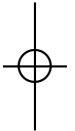


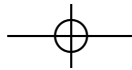
- M.R. GENESERETH et N.J. NILSSON : *Logical foundations of artificial intelligence*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987. ISBN 0-934613-31-1.
- M. GHONIEM, J.-D. FEKETE et P. CASTAGLIOLA : A comparison of the readability of graphs using node-link and matrix-based representations. *In Proceedings of InfoVis 2004*, 2004.
- N. GILLIS : Introduction to nonnegative matrix factorization. *arXiv preprint arXiv :1703.00663*, 2017.
- A.Y. GRAMA, A. GUPTA et V. KUMAR : Isoefficiency : measuring the scalability of parallel algorithms and architectures. *IEEE Parallel Distributed Technology : Systems Applications*, 1993.
- D.A. GREEN : A colour scheme for the display of astronomical intensity images. *Bull. Astr. Soc. India*, 39 :289–295, 2011.
- R.L. GREGORY : *The intelligent eye*. Weidenfeld and Nicolson, 1970.
- J. HAHNFELD, T. CRAMER, M. KLEMM, C. TERBOVEN et M.S. MÜLLER : A pattern for overlapping communication and computation with OpenMP target directives. *In Scaling OpenMP for Exascale Performance and Portability*, 2017.
- R.W. HAMMING : Error detecting and error correcting codes. *Bell System Technical Journal*, 29(2) :147–160, 1950.
- R.A. HARSHMAN : Foundations of the parafac procedure : models and conditions for an explanatory multi-modal factor analysis. Rapport technique, UCLA Working Papers in Phonetics, 16, 1970.
- K. HE, X. ZHANG, S. REN et J. SUN : Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- N. HENRY, J.-D. FEKETE et M.J. MCGUFFIN : NodeTrix : hybrid representation for analyzing social networks. *IEEE Transactions on Visualization and Computer Graphics*, 13(6) :1302–9, 2007.
- J.-B. HIRIART-URRUTY et C. LEMARÉCHAL : *Fundamentals of convex analysis*. Springer Verlag, Heidelberg, 2001.



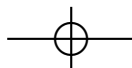
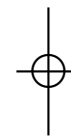


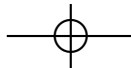
- F.L. HITCHCOCK : The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6 :164–189, 1927.
- T. HOFMANN : Probabilistic latent semantic indexing. *In Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1999.
- G. HUANG, Z. LIU, L. van der MAATEN et K.Q. WEINBERGER : Densely connected convolutional networks. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- A. INSELBERG : The plane with parallel coordinates. *The Visual Computer*, 1(2) :69–91, 1985.
- Sergey IOFFE et Christian SZEGEDY : Batch normalization : accelerating deep network training by reducing internal covariate shift. *In Proceedings of the 32nd International Conference on Machine Learning*, 2015.
- Michael J.M. et J.-M. ROBERT : Quantifying the space-efficiency of 2D graphical representations of trees. *Information Visualization*, 9 (2) :115–140, 2010.
- T. JOACHIMS : Making large-scale SVM learning practical. *In B. SCHÖLKOPF, C. BURGESS et A. SMOLA, éditeurs : Advances in kernel methods - support vector learning*, chapitre 11. MIT Press, Cambridge, MA, 1999.
- S. JOHNSON : *The Ghost Map : the Story of London's Most Terrifying Epidemic and How it Changed Science, Cities and the Modern World*. Riverhead, 2006.
- Y.-M. KIM, M.-R. AMINI, C. GOUTTE et P. GALLINARI : Multi-view clustering of multilingual documents. *In Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, pages 821–822, 2010.
- Y.M. KIM, J.F. PESSIOT, M.R. AMINI et P. GALLINARI : An extension of PLSA for document clustering. *In Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008*, pages 1345–1346, 2008.
- T.G. KOLDA et B.W. BADER : Tensor decompositions and applications. *SIAM Rev.*, 51(3) :455–500, 2009.



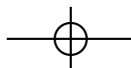


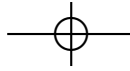
- Erricos John KONTOGHIORGES : *Handbook of parallel computing and statistics*. CRC Press, 2005.
- A. KRITHAM.RA, M.-R. AMINI, J.-M. RENDERS et C. GOUTTE : Semi-supervised document classification with a mislabeling error model. *In Advances in Information Retrieval , 30th European Conference on IR Research, ECIR 2008*, pages 370–381, 2008.
- A. KRIZHEVSKY, I. SUTSKEVER et G.E. HINTON : Imagenet classification with deep convolutional neural networks. *In Advances in Neural Information Processing Systems*, 2012.
- M. KUPPERMAN : Probabilities of hypotheses and information-statistics in sampling from exponential-class populations. *Annals of Mathematical Statistics*, 9(2) :571–575, 1958.
- I. LAPTEV : On space-time interest points. *International Journal on Computer Vision*, 64(2-3) :107–123, 2005.
- G. LATOUCHE et V. RAMASWAMI : *Introduction to matrix analytic methods in stochastic modeling*. ASA-SIAM Series on statistics and applied probability. Philadelphia, Pa. SIAM, Society for Industrial and Applied Mathematics Alexandria, Va. ASA, American Statistical Association, 1999. ISBN 0-89871-425-7.
- S.L. LAURITZEN : *Graphical models*. Clarendon Press, Oxford, United Kingdom, 1996.
- Y. LECUN et M. RANZATO : Deep learning tutorial, ICML, 2013.
- D.D. LEE et H.S. SEUNG : Algorithms for non-negative matrix factorization. *In Advances in neural information processing systems*, 2001.
- O. LEVY et Y. GOLDBERG : Neural word embedding as implicit matrix factorization. *In Advances in Neural Information Processing Systems 27*, 2014.
- J.B. LOVINS : Development of a stemming algorithm. *Mechanical Translation and Computational Linguistics*, 11 :22–31, 1968.



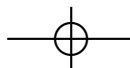


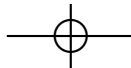
- D.G. LOWE : Distinctive image features from scale-invariant keypoints. *International Journal on Computer Vision*, 60(2) :91–110, novembre 2004. ISSN 0920-5691.
- H. P. LUHN : The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2) :159–165, 1958.
- J. MACKINLAY : Automating the design of graphical presentations of relational information. *ACM Trans. Graph.*, 5(2) :110–141, 1986.
- W.S. MCCULLOCH et Walter PITTS : A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4) :115–133, Dec 1943.
- E.W. MONTROLL : Random walks in multidimensional spaces, especially on periodic lattices. *Journal of the Society for Industrial and Applied Mathematics (SIAM)*, 4(4) :241 – 260, 1956.
- T. MUNZNER : *Visualization analysis and design*. A.K. Peters Visualization Series, 2014.
- V. NAIR et G.E. HINTON : Rectified linear units improve restricted Boltzmann machines. *In Proceedings of the 27th International Conference on Machine Learning*, 2010.
- S. NAKAJIMA, M. SUGIYAMA et D. BABACAN : On bayesian pca : automatic dimensionality selection and analytic solution. *In Proceedings of the 28th International Conference on Machine Learning*, 2011.
- Y. NESTEROV : *Introductory lectures on convex optimization : a basic course*, volume 87. Springer Science & Business Media, 2013.
- A.B. NOVIKOFF : On convergence proofs on perceptrons. *In Proceedings of the Symposium on the Mathematical Theory of Automata*, 1962.
- C.D. PAICE : Another stemmer. *SIGIR Forum*, 24(3) :56–61, 1990.
- Marjorie PATERNOSTRE, Pascal FRANCO, Julien LAMORAL, David WARTTEL et Marco SAERENS : Carry, un algorithme de désuffixation pour le français. Rapport technique, Information Technology, 2002.



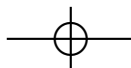
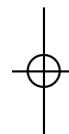


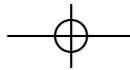
- E.A. PATRICK, J.P. COSTELLO et F.C. MONDS : Decision-directed estimation of a two-class decision boundary. *IEEE Transactions on Information Theory*, 9(3) :197–205, 1970.
- J. PEARL : Causal diagrams for empirical research. *Biometrika*, 82(4) : 669–710, 1995.
- W. PENG et T. LI : On the equivalence between nonnegative tensor factorization and tensorial probabilistic latent semantic analysis. *Applied Intelligence*, 35(2) :285–295, 2011.
- J.-F. PESSIOT, Y.-M. KIM, M.R. AMINI et P. GALLINARI : Improving document clustering in a learned concept space. *Information Processing & Management*, 46(2) :180–192, 2010.
- Boris POLYAK : Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4 :1–17, 12 1964.
- Martin F. PORTER : An algorithm for suffix stripping. *Program*, 14(3) : 130–137, 1980.
- M. RAJIH, P. COMON et R.A. HARSHMAN : Enhanced line search : a novel method to accelerate parafac. *SIAM Journal on Metric Analysis and Applications*, 30(3) :1128–1147, 2008.
- A.S. RAZAVIAN, H. AZIZPOUR, J. SULLIVAN et S. CARLSSON : CNN features off-the-shelf : an astounding baseline for recognition. *In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 512–519, 2014.
- H. REIJNER : The development of the horizon graph. *In proc. Vis08 Workshop From Theory to Practice : Design, Vision and Visualization*, 2008.
- C.P. ROBERT : *Le choix bayésien : principes et pratique*. Springer Science & Business Media, 2006.
- C.P. ROBERT, G. CELEUX et J. DIEBOLT : Bayesian estimation of hidden Markov chains : a stochastic implementation. *Statistics & Probability Letters*, 16 :77–83, 1993.



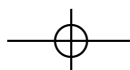


- F. ROSENBLATT : The perceptron : a probabilistic model for information storage and organization in the brain. *Psychological Review*, 1958.
- J.S. ROSENTHAL : Parallel computing and monte carlo algorithms. *In Far East Journal of Theoretical Statistics*, 1999.
- O. RUSSAKOVSKY, J. DENG, H. SU, J. KRAUSE, S. SATHEESH, S. MA, Z. HUANG, A. KARPATY, A. KHOSLA, M. BERNSTEIN, A.C. BERG et L. FEI-FEI : ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3) :211–252, 2015.
- T. SAITO, H.N. MIYAMURA, M. YAMAMOTO, H. SAITO, Y. HOSHIYA et T. KASEDA : Two-tone pseudo coloring : compact visualization for one-dimensional data. *In proc. Info Vis'05*, pages 173–180, 2005.
- J. SANCHEZ, F. PERRONNIN, T. MENSINK et J. VERBEEK : Image classification with the Fisher vector : theory and practice. *International Journal of Computer Vision*, 105(3) :222–245, 2013.
- R.E. SCHAPIRE : Theoretical views of boosting and applications. *In Proceedings of the 10th International Conference on Algorithmic Learning Theory*, pages 13–25, 1999.
- H. SCHMID : Probabilistic part-of-speech tagging using decision trees. *In Proceedings of the International Conference on New Methods in Language Processing*, 1994.
- B. SCHÖLKOPF et A.J. SMOLA : *Learning with kernels : support vector machines, regularization, optimization and beyond*. MIT Press, 2002.
- H.-J. SCHULZ : Treevis.net : a tree visualization reference. *IEEE Computer Graphics and Applications*, 31(6) :11–15, 2011.
- S. SHALEV-SHWARTZ, Y. SINGER, N. SREBRO et A. COTTER : Pegasos : primal estimated sub-gradient solver for SVM. *Mathematical Programming*, 127(1) :3–30, 2011.
- B. SHNEIDERMAN : Tree visualization with tree-maps : 2-d space-filling approach. *ACM Trans. Graph.*, 11(1) :92–99, 1992.
- B. SHNEIDERMAN : The eyes have it : a task by data type taxonomy for information visualizations. *In proc. Visual Languages*, 1996.

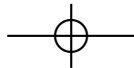
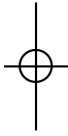
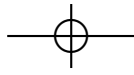




- K. SIMONYAN et A. ZISSERMAN : Very deep convolutional networks for large-scale image recognition. *CoRR*, 2014.
- J. SIVIC et A. ZISSERMAN : Video Google : a text retrieval approach to object matching in videos. *In Proceedings of the 9th IEEE International Conference on Computer Vision*, 2003.
- N. SRIVASTAVA, G. HINTON, A. KRIZHEVSKY, I. SUTSKEVER et R. SALAKHUTDINOV : Dropout : a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15 :1929–1958, 2014.
- R.J. STEELE et A.E. RAFTERY : Performance of bayesian model selection criteria for gaussian mixture models. *Frontiers of Statistical Decision Making and Bayesian Analysis*, 2 :113–130, 2010.
- I. SUTSKEVER, J. MARTENS, G. DAHL et G. HINTON : On the importance of initialization and momentum in deep learning. *In Proceedings of the 30th International Conference on Machine Learning*, 2013.
- C. SZEGEDY, W. LIU, Y. JIA, P. SERMANET, S. REED, D. ANGUELOV, D. ERHAN, V. VANHOUCHE et A. RABINOVICH : Going deeper with convolutions. *In Computer Vision and Pattern Recognition*, 2015.
- M. SZUMMER et T. JAAKKOLA : Partially labeled classification with Markov random walks. *In Advances in Neural Information Processing Systems*, 2002.
- A. TOSELLI et O.B. WIDLUND : *Domain decomposition methods : algorithms and theory*. Springer, 2005.
- A. TREISMAN et G. GELADE : A feature-integration theory of attention. *Cog. Psycho.*, 12 :97–136, 1980.
- T. TROUILLON, J. WELBL, S. RIEDEL, E. GAUSSIER et G. BOUCHARD : Complex embeddings for simple link prediction. *In Proceedings of the 33rd International Conference on Machine Learning*, 2016.
- J. TRUETT, J. CORNFIELD et W. KANNEL : A multivariate analysis of the risk of coronary heart disease in Framingham. *Journal of Chronic Diseases*, 20(7) :511–524, 1967.



- N. USUNIER, M.-R. AMINI et P. GALLINARI : A data-dependent generalisation error bound for the AUC. *In ICML'05 workshop on ROC Analysis in Machine Learning*, 2005.
- K.E.A. van de SANDE, T. GEVERS et C.G.M. SNOEK : Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9) :1582–1596, 2010.
- V.N. VAPNIK : *The nature of statistical learning theory (second edition)*. Springer-Verlag, 1999.
- P. VERGHESE : Visual search and attention : a signal detection theory approach. *Neuron*, 31 :523–535, 2001.
- J.-N. VITTAUT, M.-R. AMINI et P. GALLINARI : Learning classification with both labeled and unlabeled data. *In Proceedings of the 13th European Conference on Machine Learning (ECML'02) - Helsinki, Finland*, pages 468–476, 2002.
- B. WIDROW et M.E. HOFF : Adaptive switching circuits. *In 1960 IRE WESCON Convention Record, Part 4*, New York, 1960. IRE.
- T. ZHANG et F.J. OLES : A probability analysis on the value of unlabeled data for classification problems. *In 17th International Conference on Machine Learning*, 2000.
- S. ZHAO, M.J. MCGUFFIN et M.H. CHIGNELL : Elastic hierarchies : combining treemaps and node-link diagrams. *In Proc. IEEE InfoVis 2005*, October 2005.
- D. ZHOU, O. BOUSQUET, T.N. LAL, J. WESTON et B. SCHÖLKOPF : Learning with local and global consistency. *In Advances in Neural Information Processing Systems*, 2004.
- X. ZHU et Z. GHAHRAMANI : Learning from labeled and unlabeled data with label propagation. Rapport technique CMU-CALD-02-107, Carnegie Mellon University, 2002.
- X. ZHU, Z. GHAHRAMANI et J. LAFFERTY : Semi-supervised learning using gaussian fields and harmonic functions. *In 20th International Conference on Machine Learning*, 2003.



Index

- accélération, 58
- activation
 - Heaviside, 178
 - linéaire, 180
- adaptive boosting, *voir* algorithme Adaboost
- Adaptive Linear Neuron, *voir* algorithme ADALINE
- ADMM, 81, 90
- algorithme
 - Adaboost, 148, 149, 172
 - ADALINE, 179, 183
 - aléatoire, 84
 - CEM, 171
 - de Gibbs, 126
 - de gradient proximal, 80
 - des moindres carrés alternés, 107, 110
 - du gradient, 77
 - du gradient stochastique, 83
 - EM, 171
 - EM Monte Carlo, 127
 - LDA, 138
 - Metropolis Hastings, 125
 - NMF, 105
 - perceptron, 209
 - PESGASOS, 174
 - SAG, 86
 - SAGA, 86
- analyse en composantes principales
 - probabiliste, 118, 132
- apprentissage
 - non-supervisé, 66
 - profond, 177, 205
 - semi-supervisé, 66, 146, 164, 167, 171
 - modèles de mélange, 171
 - modèles discriminants, 171
 - modèles graphiques, 165
 - supervisé, 67, 145, 164
- attributs des données
 - agrégation, 215
 - attribut des données, 214
 - comparaison, 215
 - nominaux, 215, 224
 - ordonnés, 215, 224
 - quantitatifs, 216, 224
 - échelle d'intervalle, 216
 - échelle de ratio, 216
- base
 - d'entraînement, 145, 147, 151, 155–157, 159, 160, 162, 163, 183
- BIC, 132, 139
- boosting, 148
- chaînes de Markov cachées, 119, 134
- classe de fonctions, 145
- classification, 68
 - EM, *voir* algorithme CEM
 - multi-classe
 - mono-label, 161, 162
 - multi-label, 161, 162

- classifieur
 - binaire, 148
 - de base, 150
 - de vote, 148, 149
 - faible, 148, 172
 - linéaire, 150
 - non-linéaire, 155
- CMC, 119, 134
- condition
 - d’optimalité, 73
 - Karush-Kuhn-Tucker, 153, 158
- conditionnement, 78
- contrainte, 67
 - non saturée, 152
 - saturée, 152
- convergence, 78
 - exponentielle, 78
- convexité, 74
 - forte, 76
- coût
 - convexe, 207
- DAG, 127
- decision boundary, *voir* frontière de décision
- deep learning, *voir* apprentissage profond
- densité de probabilité, 171
- descente, *voir* gradient
- desired output, *voir* sortie désirée
- différentiation automatique, 196
- directed acyclic graphs, *voir* DAG
- direction de descente, *voir* gradient
- discrimination logistique, *voir* régression logistique
- distance de Hamming, 161, 163
- données redondantes, 84
- dual de Wolfe, 154
- décomposition
 - canonique de tenseur, 109
 - de données, 52
 - exploratoire, 52
 - réursive, 51
 - spéculative, 52
- déduction, 146
- développement de Taylor, 207, 208
- échantillonneur de Gibbs, 125
- efficacité, 59
- empirical risk minimization, *voir* minimisation du risque empirique
- équilibre des ressources, 54
- erreur
 - de classification, 175
 - de généralisation, 145
 - de Hinge, 175
 - empirique, 145
 - quadratique, 183
- espace
 - de Hilbert, 155
 - de plongement, 155
 - de redescription, 155, 157
- famille exponentielle, 122
- familles de lois conjuguées, 121
- feature space, *voir* espace de redescription
- feed forward, 185
- fonction
 - d’activation, 199
 - d’erreur, 145
 - de décision, 154
 - de projection, 155
 - de prédiction, 145, 164
 - objectif, 152, 154, 157, 159, 180
- forme logistique, 147
- fortement convexe, 175
- frontière de décision, 151
- gradient, 71
 - descente, *voir* algorithme du gradient
 - stochastique, 183, 187, 190, 202
- granularité, 51
- graphe moral, 128
- hard margin, *voir* marge dure
- hessienne, 152, 154, 166, 207, 208
 - symétrique positive définie, 154, 207
- hyperplan
 - marginal, 153–155, 159
 - séparateur, 151, 180, 183, 184
- hypersphère, 176

- hypothèse
 - de continuité, 166
 - de partition, 172
 - de variété, 166
- i.i.d., 131, 161
- identiquement et indépendamment
 - distribué, *voir* i.i.d.
- induction, 146
- iso-efficacité, 62
- isotrope, 119
- label spreading, *voir* propagation des étiquettes
- lagrangien, 152, 153, 158
- latent dirichlet allocation, *voir* LDA
- LDA, 136
- learning rate, *voir* pas de descente, *voir* pas de descente
- lemmatisation, 9
- ligne de niveau, 208
- linéairement séparable, 151, 153, 155, 157, 160, 183, 184
- localité, 55
- logistic regression, *voir* régression
- loi
 - a posteriori, 121
 - a priori, 121
 - d’Amdahl, 59
 - d’émission, 129
- maintenabilité, 42
- map-reduce, 88
- marche aléatoire, 169
- marge, 151, 155
 - d’un exemple, 180
 - dure, 151
 - souple, 156
- matrice
 - de Gram, 154
 - jacobienne, 152
 - laplacienne, 166
 - par blocs, 166
 - stochastique, 168
- max-pooling, 200
- minimisation du risque
 - empirique, 146, 162
- mode
 - batch, 181
- modèle
 - graphique, 117, 127
 - hiérarchique, 136
 - linéaire mixte gaussien, 143
 - maître-esclave, 57
 - probabiliste graphique, 117, 127
- mot de code, 163
- multi-layer perceptron, *voir* perceptron multicouche
- multistabilité, 218
- mélange de lois, 117, 129
- méthode
 - d’ensemble, 148
 - de Newton, 79
 - linéaire paramétrique, 147
 - à base de votes, 148
- namenode, 30
- normalisation textuelle, 7
- noyau
 - gaussien, 156
 - polynomial, 156
- objectif, 67
- optimisation
 - distribuée, 88
 - duale, 154, 159
 - sous contraintes, 152, 159
- opérateur proximal, 80–82, 87–90, 92, 93
 - soft-thresholding, 92
- outlier, *voir* point aberrant
- overfitting, *voir* surapprentissage
- parallélisme
 - d’instructions, 46
 - de bits, 45
 - de données, 56
 - de tâches, 46, 56
 - explicite, 46
 - implicite, 46
- parcimonie, 71
- pas d’apprentissage, *voir* pas de descente, 181, 184
- pas de descente, 77

- perception
 - Gestalt, 217
 - groupes, 218
 - invariance, 217
 - multistabilité, 218
 - perception, 217
 - pré-attentive, 218
 - réification, 218
 - stimulus, 217
 - émergence, 218
- perceptron, 179
 - multicouche, 179, 185
 - architecture, 186
- pipelining, 57
- PMC, *voir* perceptron multicouche
- poids synaptiques, 178
- point aberrant, 159
- prédiction, 68
- PReLU, 203
- problème d'optimisation, 67
- programmation quadratique, 153
- propagation
 - des étiquettes, 165
- prox-simple, *voir* opérateur proximal

- qualification de contraintes, 152

- racinisation, 9
- random walk, *voir* marche aléatoire
- rectified linear unit, *voir* ReLU
- ReLU, 203
- réduction de variance, 85
- régression, 68
 - logistique, 147
- régularisation, 70
- réseaux récurrents, 185

- scalabilité, 62
- segmentation, 6
- séparateurs à vaste marge, 150
 - algorithme (marge dure), 156
 - algorithme (marge souple), 160
- slack variables, *voir* variables d'écart
- soft margin, *voir* marge souple
- solution optimale, 67
- sortie désirée, 145
- statistique bayésienne, 120

- steepest descent, *voir* algorithme du gradient
- supervised learning, *voir* apprentissage supervisé
- support vector, *voir* vecteur de support
- support vector machine, *voir* séparateurs à vaste marge
- surapprentissage, 70, 146

- taux d'apprentissage, 77
- tenseur, 108
- théorème
 - de Schwarz, 207
- tokenisation, 6
- training set, *voir* base d'entraînement

- valeur propre, 208
- variable graphique
 - couleur, 221, 225
 - daltonisme, 222
 - dégradé, 222
 - encodage graphique, 214
 - forme, 221, 222
 - grain, 221, 225
 - marque, 219
 - orientation, 221
 - position, 220, 224
 - taille, 220, 223, 224
 - troisième dimension, 223
 - valeur, 220, 225
 - variable de l'image, 214
 - échelle de couleur, 222
- variables
 - d'écart, 157, 158
 - de Lagrange, 158, 159
 - primales, 153
- vecteur
 - de support, 151, 153–155, 157, 159
 - gradient, 181
 - indicateur
 - de classe, 161, 164
 - propre, 207–209
- visualisation interactive d'information, 211
- vraisemblance
 - classifiante, 147, 171