

# Index

- Équations normales, 242
- Échantillonnage d'importance biaisé, 588
- Échantillonnage préférentiel, 686
- Échelonnage des poids, 270
- Écart-type, 82, 143
- Écart-type de la moyenne, 144
- Échantillonnage ancestral, 575
- Échantillonnage d'importance, 616
- Échantillonnage d'importance par recuit, 704
- Élément linéaire adaptatif, 41, 48, 49
- Équation d'Euler-Lagrange, 637
- Équations normales, 126, 129
- Équivariance, 342
- Étape E, 626
- Étape M, 626
- Échantillonnage de Gibbs, 576
- Deep learning*, 28
- Knowledge-based approach*, 28
  
- Mélange (Chaîne de Markov), 596
  
- Accélération, 302
- Accélération de Nesterov, 305
- ACI topographique, 491
- ACP, Analyse en composantes principales 70
- ACP probabiliste, 489, 624
- AdaGrad, 312
- ADALINE, Élément linéaire adaptatif 41
- Adam, 315, 423
- Adaptation de domaine, 534
- Adressage fondé sur l'emplacement, 416
- Adressage fondé sur le contenu, 416
- AEV, Auto-encodeur variationnel 684
- Affine, 126
- Agglomération stochastique, 273
- Agrégation, Pooling 344
  
- AIS, Recuit par échantillonnage d'importance 618
- Ajustement fin discriminant, 652
- Ajustement fin supervisé, 652
- Algorithme glouton, 327
- Analyse de caractéristiques lentes, 492
- Analyse de perturbation, 677
- Analyse en composantes indépendantes, 489
- Analyse en composantes principales, 70, 162, 488, 623
- Analyse en sous-espaces indépendants, 491
- Analyse Factorielle, 488
- Ancestral sampling*, 590
- APC probabiliste, 488
- Appariement des rapports, 611
- Appariement des scores, 610
- Apprentissage automatique, 29
- Apprentissage de représentation, 29
- Apprentissage de structure, 577
- Apprentissage de variété, 175
- Apprentissage dirigé, 383, 384
- Apprentissage en un coup, 536
- Apprentissage multi-tâches, 536
- Apprentissage multimodal, 537
- Apprentissage multitâche, 252
- Apprentissage non supervisé, 122, 161
- Apprentissage par renforcement, 51, 123, 477, 678
- Apprentissage par transfert, 534
- Apprentissage profond, 28, 31
- Apprentissage sans exemple, 536
- Apprentissage supervisé, 122
- Approche fondée sur la connaissance, 28
- Arbres de décision, 544
- Arrêt précoce, 254, 255, 257, 423
- ASR, Reconnaissance de la parole 455
- Astuce de centrage (MBP), 663

Astuce de reparamétrisation, 677  
 Astuce du noyau, 156  
 Asymptotiquement sans biais, 141  
 Audio, 119, 362, 455  
 Augmentation de données, 455  
 Augmentation du jeu de données, 279  
 Auto-encodeur, 30, 358, 501  
 Auto-encodeur contractif, 518  
 Auto-encodeur débruiteur, 509  
 Auto-encodeur pondéré par AIS, 686  
 Auto-encodeur variationnel, 684  
 Auto-encodeur débruiteur, 678  
 Auto-encodeur variationnel, 678  
*Automatic Speech Recognition*, Reconnaissance de la parole automatique 455  
  
*Bagging*, 263  
 Bandits contextuels, 477  
 Base de connaissances, 480  
 Base de données relationnelles, 480  
 Bayésien naïf, 29  
 BFGS, 320  
 Bi-gramme, 459  
 Biais, 141, 238  
 Blanchiment, Whitening 454  
 Borne inférieure variationnelle, 624  
 Boucle, 572  
 BPTT, Rétropropagation temporelle 385  
  
 CAE, Auto-encodeur contractif 518  
 Calcul bayésien approximatif, 703  
 Calcul des variations, 192  
 Capacité de représentation, 129  
 Capacité effective, 129  
 Caractéristique/variable, 116  
 Cellule complexe, 367  
 Cellule simple, 367  
 Chaîne de Markov, 590  
*Chain rule (calculus)*, Règle de la chaîne, 216  
 Champ moyen, 629, 630  
 Champ récepteur, 341  
 Champs aléatoires de Markov, Modèle non dirigé 562  
 Classification, 117  
 Classification temporelle connexionniste, 458  
 Classifieur tangentiel entre variétés, 279  
 Clique, 97  
 CNN, Réseau de neurones convolutif 261  
 Codage parcimonieux, 325, 358, 494, 623, 681  
 Collisionneur, Explication par un autre facteur 570  
 Combinaison linéaire, 61  
*Conditional computation*, Structure dynamique 446  
 Conditions de Karush-Kuhn-Tucker, 112, 245  
 Connexionisme, 43, 441  
 Consistance, 146, 510  
 Constante de Lipschitz, 110  
 Contrainte active, 112  
 Contrainte d'égalité, 111  
 Contrainte d'inégalité, 111  
 Contraste, 452  
 Convergence presque sûre, 146  
 Convolution, 335, 672  
 Convolution pavée, 335  
 Convolution séparable, 364  
 Corde, 572  
 Corrélation, 83  
 Corrélation croisée, 337  
 Correspondance de moment, 691  
 Correspondance de score, 510  
 Cortex visuel primaire, 366  
 Couche (Réseau de neurones), 181  
 Couche cachée, 32, 181  
 Couche de détection, 344  
 Couche de sortie, 181  
 Couche visible, 32  
 Covariance, xxiii, 82  
*Coverage*, Recouvrement 422  
*Critical temperature*, 598  
 CTC, Classification temporelle connexionniste 458  
 Cyc, 28  
  
 D-séparation, 568  
 Décomposition en valeurs propres, 65  
 Décomposition en valeurs singulières, 67  
 Définie négative, 107  
 Dégradation des pondérations, 240  
 Dépendance linéaire, 61  
 Dérivée, xxiii, 102  
 Dérivée directionnelle, 104  
 Dérivée partielle, 104  
 Dérivée seconde, 105  
 Désambiguïsation lexicale, 481  
 Déterminant, xxii  
 Débruitage de l'appariement des scores, 612  
 Décodeur, 30

Décomposition en valeurs singulières, 163, 476  
 Décomposition parcimonieuse prévisionnelle, 521  
 Dérivée fonctionnelle, 636  
 Dérivée variationnelle, Dérivée fonctionnelle 636  
 Dérive de concept, 536  
 Détection d'objet, 450  
 Détection de spams, 29  
 DAE, Auto-encodeur débruiteur 509  
 DC, Divergence contrastive 603  
 DCGAN, 548, 549  
 DCP, Maximum de vraisemblance stochastique 605  
 DCPR, 607  
 Deep Blue, 28  
 Densité de probabilité, 79  
 Descente de gradient, 102, 104  
 Descente de gradient stochastique, 41, 166, 286, 300, 662  
 Descente par coordonnée, 325  
 DGS, descente de gradient stochastique 286  
 Diagonale principale, 57  
 Différences finies, 437  
 Différenciation numérique, Différences finies 437  
 Diffusion, 58  
 Dimension de Vapnik-Chervonenkis, 131  
 Disparition du gradient, 296  
 Distance tangentielle, 277  
 Distribution de Boltzmann, 566  
 Distribution de Gibbs, 564  
 Distribution de Laplace, 494, 495  
 Distribution de probabilité non normalisée, 563  
 Distribution de probabilité *a priori*, 151  
 Distribution de probabilités locales conditionnelles, 561  
 Distribution génératrice de données, 127, 147  
 Divergence contrastive, 297, 603, 662  
 Divergence contrastive persistante, Maximum de vraisemblance stochastique 605  
 Divergence de Kullback-Leibler, xxiii, 94  
 Divergence KL, Divergence de Kullback-Leibler 94  
 Données étiquetées, 45  
 Données manquantes, 117  
 Données volumétriques, 362  
 Double rétropropagation, 279  
 DPP, Décomposition parcimonieuse prévisionnelle 521  
 DropConnect, 273  
 Dropout, 266, 423, 428, 662, 678  
 EBM, Modèle fondé sur l'énergie 566  
 Echo state network, 403  
 ELBO, Borne inférieure variationnelle 624  
 EM, Espérance-Maximisation 626  
 Encodeur, 30  
 Energie libre, 567, 669  
 Energie libre de Helmholtz, Borne inférieure variationnelle 624  
 Energie libre variationnelle, Borne inférieure variationnelle 624  
 Enjambement, Pas 351  
 Ensemble, xxii  
 Entraînement par antagonisme, 276, 279, 529  
 Entropie croisée, 95, 148  
 Entropie de Shannon, xxiii, 94  
 Entropie différentielle, 94, 637  
 Epoch, 254  
 EQM, Erreur quadratique moyenne 125  
 Erreur d'entraînement, 127  
 Erreur de Bayes, 133  
 Erreur quadratique moyenne, 125  
 Erreur standard de la moyenne, 285  
 ESN, Echo state network 377  
 Espérance, 82  
 Espérance-Maximisation, 626  
 Espace d'hypothèse, 128, 135  
 Estimateur ponctuel, 139  
 Estimation contrastive bruitée, 613  
 Estimation de densité, 120  
 Estimation de fonction de masse, 120  
 Estimation MAP, 154  
 Exemple, 116  
 Exemple antagoniste, 276  
 Exemples antagonistes virtuels, 276  
 Explication par un autre facteur, 570, 623, 635  
 Exploding gradient, 296  
 Exploitation, 478  
 Exploration, 478  
 Explosion du gradient, 296  
 F-score, 421  
 Facteur (modèle graphique), 563  
 Facteurs de variation, 31  
 Feature, 29  
 Filtrage collaboratif, 475  
 Fine-tuning, Réglage fin 327  
 Fonction d'énergie, 566

Fonction d'activation, 184  
 Fonction d'activation Leaky ReLU, 205  
 Fonction d'erreur, Fonction objectif 101  
 Fonction de base radiale, 208  
 Fonction de coût, Fonction objectif 101  
 Fonction de Dirac, 86  
 Fonction de Gabor, 370  
 Fonction de Lagrange généralisée, Lagrangien généralisé 111  
 Fonction de masse, 78  
 Fonction de partition, 564, 599, 659  
 Fonction de perte, Fonction objectif 101  
 Fonction de perte de substitution, 283  
 Fonction logistique sigmoïde, 33, 89  
 Fonction objectif, 101  
 Forget gate, 311, Porte d'oubli 409  
 Fovéa, 358  
**Freebase**, 480  
  
 Généralisation, 127  
 GAN, Réseaux génératifs par antagonisme 541  
 GCN, Normalisation de contraste global 452  
**GeneOntology**, 480  
 GPU, 442  
 Gradient, 104  
*Gradient clipping*, Seuillage du gradient 295  
 Graphe, xxii  
 Graphe cordal, 572  
 Graphe de calcul, 215  
 Graphe de facteurs, 573  
 Graphe moralisé, 572  
 Graphe triangulaire, Graphe cordal 572  
*Greedy supervised pretraining*, 327  
  
 Hétéroscédastique, 200  
 Hachage sémantique, 523  
*Hard tanh*, 208  
 Harmonium, Machine de Boltzmann restreinte 581  
 Hessien, 232  
*Hill climbing*, 104  
 Hyper-paramètre, 137, 428  
 Hypothèse i.i.d., 127, 139, 275  
 Hypothèse de variété, 175  
  
 Identifiabilité du modèle, 291  
 ILSVRC, ImageNet Large Scale Visual Recognition Challenge 50  
 ImageNet Large Scale Visual Recognition Challenge, 50  
 Images couleur, 362  
 Immoralité, 572  
*Importance sampling*, 587  
 Inégalité triangulaire, 62  
 Indépendance, xxiii, 81  
 Indépendance conditionnelle, xxiii, 81  
 Indépendance spécifique au contexte, 569  
 Inférence, 559, 578, 624, 627, 629, 638, 641  
 Inférence approximée (ou approximative), 578  
 Information propre, 93  
 Initialisation, 306  
 Initialisation de paramètres, 306, 405  
 Initialisation normalisée, 308  
 Initialisation parcimonieuse, 309, 405  
 Intégrale, xxiii  
 Intelligence artificielle, 27  
 Invariance, 344  
 Isotrope, 86  
  
 Jeu d'échecs, 28  
 Jeu de données, 121  
  
*k*-means, 365, 544  
*k*-plus proches voisins, 158, 544  
 Karush-Kuhn-Tucker, 111  
 Kernel (convolution), 336  
 KKT, Karush-Kuhn-Tucker 111  
  
 Lagrangien, Lagrangien généralisé 111  
 LAPRGA, 689  
 LCN, Normalisation de contraste local 454  
 Fléau de la dimension, 170  
*Leaky unit*, Unité avec fuite 406  
 Liaison de paramètre, Partage de paramètre 260  
 Lipschitz-continue, 110  
 Lissage d'étiquettes, 251  
 Liste courte, 463  
 Logit, 90  
 Loi multinoulli, 84  
 Loi catégorielle, Loi multinoulli 84  
 Loi de Bernoulli, 83  
 Loi de Laplace, 86  
 Loi de mélange, 87  
 Loi de probabilité, 78  
 Loi empirique, 86  
 Loi exponentielle, 86  
 Loi gaussienne, Loi normale 84

Loi multinomiale, 84  
 Loi normale, 84, 142  
 Loi uniforme, 79  
 $L^p$  norme, 62  
  
 Mélange gaussien, 201  
 Méthode de dégradation des pondérations, 190  
 Méthode de la plus grande pente, Descente de gradient 104  
 Méthode de Newton, 108, 316  
 Méthode de continuation, 330  
 Méthodes de Krylov, 233  
 Méthode de quasi-Newton, 320  
 Méthode ensembliste, 263  
 Mélange d'experts, 447, 544  
 Mélange gaussien, 87  
 Mémoire à court et long terme (MCLT), 43, 50, 407  
 Méthode de dégradation des pondérations, 135  
 Méthode de Monte-Carlo par chaîne de Markov (MCMC), 590  
 Machine à état liquide, 403  
 Machine à noyau, 544  
 Machine à vecteur de support, 156  
 Machine Boltzmann restreinte, 447  
 Machine de Boltzmann, 566, 645  
 Machine de Boltzmann *spike-and-slab*, 670  
 Machine de Boltzmann profonde, 48, 528, 642, 648  
 Machine de Boltzmann prévisionnelle, 663  
 Machine de Boltzmann restreinte, 358, 476, 581, 623, 647, 670  
 Machine de Boltzmann restreinte, 647, 666, 668  
 Machine de Turing neuronale, 415  
 Machine de Boltzmann profonde, 623, 653  
 Approximation MAP, 504  
 Marge, 352  
 Matrice, xxi, 56  
 Matrice carrée, 62  
 Matrice de covariance, 83  
 Matrice de conception, 123  
 Matrice de Toeplitz, 339  
 Matrice diagonale, 64  
 Matrice doublement circulante par blocs, 339  
 Matrice hessienne, xxiii, 105  
 Matrice identité, 59  
 Matrice inverse, 60  
 Matrice jacobienne, xxiii, 92, 105  
 Matrice orthogonale, 65  
  
 Matrice symétrique, 64, 66  
*Max pooling* probabiliste, 672  
*Max pooling*, 344  
 Maximum de vraisemblance, 147  
 Maximum de vraisemblance stochastique, 605, 662  
*Maxout*, 205, 423  
 MBR, Machine de Boltzmann restreinte 647  
 MBR conditionnelle, 674  
 MBR discriminantes, 675  
 MCLT, 285, Mémoire à court et long terme 407  
 MCMC, Méthodes de Monte-Carlo par chaîne de Markov 590  
  
 Mesure nulle, 91  
 Minimisation du risque empirique, 283  
 MLN, Modèle de langage neuronal 461  
 MNIST, 46, 662  
 Modèle de langage neuronal, 473  
 Modèle de mélange, 508, 201  
 Modèle de reconnaissance, 685  
 Modèle fondé sur l'énergie, 590  
 Modèle graphique dirigé, 506  
 Modèle graphique non dirigé, 505  
 Modèle de langage par classe, 461  
 Modèles factoriels linéaires, 487  
 Modèle fondé sur l'énergie, 566, 645, 653  
 Modèle graphique, Modèle probabiliste structuré 555  
 Modèle graphique dirigé, 97, 560, 680  
 Modèle graphique non dirigé, 97  
 Modèle non-paramétrique, 131  
 Modèle probabiliste structuré, 96  
 Modèle de langage neuronal, 461  
 Modèle non dirigé, 562  
 Modèle probabiliste structuré, 555  
*Model compression*, 446  
 Momentum, 302  
 Pseudo-inverse de Moore-Penrose, 248  
 Moyennage de modèle, 263  
 Moyenne de l'échantillon, 142  
 MRF, modèles non dirigé 562  
 Multiplicateurs de Lagrange, 111, 637  
 MVS, Maximum de vraisemblance stochastique 605  
  
*n*-gramme, 459  
 NADE, 696  
 Nat, 93  
 Natural image, 556

Neocognitron, 42, 48, 49, 369  
 Nesterov Momentum, Accélération de Nesterov 305  
 Grand Prix Netflix , 265, 476  
 Neuroscience, 41  
 NLP, Traitement du langage naturel 458  
 Nombre de condition, 286  
 Normalisation de contraste global, 452  
 Normalisation de contraste local, 454  
 Normalisation par lot, 275, 423  
 Norme, xxiv, 62  
 Norme de Frobenius, 69  
 Norme euclidienne, 63  
 Norme uniforme, 63  
 Norme unitaire, 64  
 Noyau (convolution), 337  
 Noyau gaussien, 157  
  
 OMP- $k$ , Poursuite adaptative orthogonale 263  
 Opérateur trace, 69  
 Opération, 215  
 Optimisation, 99, 101  
 Optimisation bayésienne des hyper-paramètres, 433  
 Optimisation convexe, 157  
 Optimisation des hyper-paramètres, 430  
 Optimisation sous contrainte, 245  
 Optimisation sous contraintes, 110  
 Orthogonalité, 64  
  
 Période de chauffe, 592  
 Paire de quadrature, 371  
 Parallélisation des données, 445  
 Parallélisation du modèle, 445  
 Paramètre de biais, 126  
 Partage de paramètre, 260  
 Partage de paramètre, 340  
 Partage de paramètres, 375, 388  
 Pas, 351  
 Pas d'apprentissage, 104  
 Perceptron, 41, 49  
 Perceptron multicouche, 31  
 Perte 0-1, 120  
 Phase négative, 467, 600  
 Phase positive, 467, 600, 647, 658  
 Plan tangent, 514  
 Plongement, 515  
 Plongement de mot, 461  
 Plongement lexical, Plongement de mot 461  
  
 Plus grande pente, Descente de gradient 104  
 Poids, 40, 124  
 Point-selle, 292  
 Politique, 477  
*Pooling*, 335, 344, 672  
 Porte *forget* 311  
 Porte d'oubli, 383  
 Potentiel de clique, Facteur (modèle graphique) 563  
 Poursuite adaptative orthogonale, 263  
 Poursuite adaptative orthogonale , 23  
 Pré-entraînement non supervisé, 457, 527  
 Pré-entraînement, 527  
 Pré-entraînement couche par couche glouton et non-supervisé, 527  
 Pré-traitement, 451  
 Précision, 421  
 Précision (d'une loi normale), 84, 86  
 Prévission de liens, 481  
 Presque partout, 91  
 Probabilité bayésienne, 77  
 Probabilité conditionnelle, 80  
 Probabilité fréquentiste, 77  
 Probabilité jointe, 79  
 Probabilité marginale, 80  
 Processus générateur de données, 127  
 Produit d'experts, 566  
 Produit de Hadamard, xxii  
 Produit matriciel, 58  
 Produit matriciel de Hadamard, 58  
 Produit matriciel terme-à -terme, 58  
 Produit scalaire, 58, 156  
 Propagation de croyance par boucle, 580  
 Propagation tangentielle, 277  
 Propagation vers l'avant, 214  
 Pseudo-inverse de Moore-Penrose, 68  
 Pseudo-vraisemblance, 608  
  
 Réglage fin, 327  
 Régression *ridge*, *weight decay* 240  
 Régularisation, 190, 237  
 Régularisation de Tikhonov, *weight decay* 240  
 Réseau d'inférence, 685  
 Réseau de neurones convolutif, 261  
 Réseau MCLT (*Mémoire à court et long terme*), 44  
 Réseaux à mélanges de densités, 201  
 Réseaux de neurones à propagation avant, 181  
 Rétropropagation, 214

Réglage fin supervisé, 528  
Régession, 118  
Régession linéaire, 124, 126, 155  
Régession logistique, 29, 156  
Régularisation, 136, 428  
Réseau à mémoire, 414  
Réseau bayésien complètement intelligible, 693  
Réseau bayésien, modèle graphique dirigé 560  
Réseau convolutif, 42  
Réseau de croyance, modèle graphique dirigé 550  
Réseau de croyances profond, 49, 528, 648, 681  
Réseau de Markov, modèles non dirigés 562  
Réseau de neurones, 39  
Réseau de neurones à retard temporel, 376  
Réseau de neurones artificiels, 39  
Réseau de neurones convolutif, 335, 423, 457  
Réseau de neurones profond acyclique, 423  
Réseau de neurones récurrents, 379  
Réseau de neurones récurrents à mémoire court-terme et long terme, 423  
Réseau dynamique en écho, 48  
Réseau génératif, 682  
Réseau génératif par antagonisme, 688  
Réseau récurrent, 49  
Réseau somme-produit, 550  
Réseau de croyance profond, 623  
Réseau de neurones à retard temporel, 369  
Réseau génératif à correspondance de moment, 691  
Réseau génératif par antagonisme, 541, 678  
Rétropropagation stochastique, 677  
Rétropropagation temporelle, 385  
Règle de Bayes, 90  
Règle de la chaîne de probabilité, 81  
Règle de sommation des probabilités, 80  
Règle du produit, Règle de la chaîne de probabilité 81  
Rappel, 421  
Rayon spectral, 403  
RBCI, Réseau bayésien complètement intelligible 693  
RBF, 208  
Recherche aléatoire, 432  
Recherche d'information, 522  
Recherche linéaire, 104, 111  
Recherche sur grille, 430  
Recirculation, 502  
Reconnaissance d'objet, 450  
Reconnaissance de la parole, 455  
Recouvrement, 422  
Rectification en valeur absolue, 205  
Recuit par échantillonnage d'importance, 618, 658  
REINFORCE, 678  
Relation binaire, 479  
Relations, 479  
Représentation parcimonieuse, 235, 262  
Représentation distribuée, 43, 166, 543  
Représentation parcimonieuse, 161, 543, 553  
*Retroprop*, Rétropropagation 214  
RGAs, Réseau génératif par antagonisme 688  
Risque, 282  
Risque empirique, 283  
RNA, Réseau de neurones artificiels 39  
RNC, Réseau de neurones convolutif 261  
RNR-MBR, 675  
Sélection des fonctionnalités, 245  
Séparation (modélisation probabiliste), 567  
Sac de mots, 468  
Sans biais, 141  
Scalaire, xxi, 55  
Semi-supervised learning, 251  
Seuillage du gradient, 295, 412  
*Short list*, Liste courte 463  
Sigmoïde, xxiv  
Sigmoïde, Fonction logistique sigmoïde 89  
*Sigmoid belief network*, 49  
*Softmax*, 196, 415, 447  
*Softplus*, xxiv, 89, 208  
Sommeil paradoxal, 602, 642  
Sortie structurée, 118, 674  
*Sparse coding*, Codage parcimonieux 325  
*Spearmint*, 433  
*Sphering*, *Whitening* 454  
SPN, Réseau somme-produit 550  
ssMBR, Machine de Boltzmann *spike-and-slab* restreinte 670  
Statistique, 139  
Statistique bayésienne, 151  
Statistique fréquentielle, Statistique fréquentiste 151  
*Stride*, Pas 351  
Structure dynamique, 446  
SVD, Décomposition en valeurs singulières 67  
Symétrie de l'espace de poids, 291

Système de recommandation fondé sur le contenu, 476  
 Système dynamique classique, 377  
 Système de recommandation, 474  
  
 TDNN, Réseaux de neurones à retard temporel 369  
*Tempering*, tempérage, 597  
*Template matching*, 157  
 Tenseur, xxi, 57  
 Test de la dérivée seconde, 107  
 Théorème d'approximation universelle, 209  
 Théorème central limite, 85  
 Théorie de l'apprentissage statistique, 127  
 Théorie de l'harmonie, 567  
 Théorie de la mesure, 91  
*Tiled convolution*, Convolution pavée 355  
 Traduction automatique, 118  
 Traitement distribué parallèle, 42  
 Traitement du langage naturel, 458  
 Transcription, 118  
 Transformée de Fourier, 362, 364  
 Transposée, xxii  
 Transposition, 57  
 Tri-gramme, 459  
  
 Uni-gramme, 459  
 Unité avec fuite, 406  
 Unité de rectification linéaire, 185, 204, 506  
 Unité de rectification linéaire paramétrique ReLU, 205  
 Unité linéaire rectifiée, 423  
 Unité récurrente fermée, 423  
 Approximateur universel, 550  
  
*V-structure*, Structure en V, Explication par un autre facteur 570  
 V1, 366  
 Valeur espérée, Espérance 82  
 Valeur propre, 65  
 Valeur singulière, Décomposition en valeurs singulières 67  
 Validation croisée, 139  
*Vanishing gradient*, 296  
 Variété, 497  
 Variété, 174  
 Variable aléatoire, 78  
 Variable latente, 87  
 Variance, xxiii, 82, 238  
  
 Vecteur, xxi, 56  
 Vecteur propre, 65  
 Vecteur singulier, Décomposition en valeurs singulières 67  
 Vecteur unitaire, 64  
 Veille-sommeil, 641, 652  
 Vision par ordinateur, 450  
  
*Weight decay*, Dégradation des pondérations 429  
*Weight scaling*, Échelonnage des poids 270  
*Whitening*, 454  
**Wikibase**, 480  
*Word embedding*, Plongement de mot 461  
**WordNet**, 480  
  
*Zero padding*, Marge 352