

- 672, 674, 677
- Ridge regression, *see* weight decay
- Risk, 273
- RNN-RBM, 679
- Saddle points, 283
- Sample mean, 124
- Scalar, xi, xii, 30
- Score matching, 509, 613
- Second derivative, 85
- Second derivative test, 88
- Self-information, 72
- Semantic hashing, 521
- Semi-supervised learning, 241
- Separable convolution, 359
- Separation (probabilistic modeling), 568
- Set, xii
- SGD, *see* stochastic gradient descent
- Shannon entropy, xiii, 73
- Shortlist, 462
- Sigmoid, xiv, *see* logistic sigmoid
- Sigmoid belief network, 26
- Simple cell, 362
- Singular value, *see* singular value decomposition
- Singular value decomposition, 43, 146, 475
- Singular vector, *see* singular value decomposition
- Slow feature analysis, 489
- SML, *see* stochastic maximum likelihood
- Softmax, 182, 415, 446
- Softplus, xiv, 67, 195
- Spam detection, 3
- Sparse coding, 319, 353, 492, 626, 686
- Sparse initialization, 302, 403
- Sparse representation, 145, 224, 251, 501, 552
- Spearmint, 433
- Spectral radius, 401
- Speech recognition, *see* automatic speech recognition
- Sphering, *see* whitening
- Spike and slab restricted Boltzmann machine, 674
- SPN, *see* sum-product network
- Square matrix, 37
- ssRBM, *see* spike and slab restricted Boltzmann machine
- Standard deviation, 60
- Standard error, 126
- Standard error of the mean, 126, 276
- Statistic, 121
- Statistical learning theory, 109
- Steepest descent, *see* gradient descent
- Stochastic back-propagation, *see* reparametrization trick
- Stochastic gradient descent, 15, 149, 277, 292, 666
- Stochastic maximum likelihood, 608, 666
- Stochastic pooling, 263
- Structure learning, 578
- Structured output, 100, 679
- Structured probabilistic model, 76, 554
- Sum rule of probability, 57
- Sum-product network, 549
- Supervised fine-tuning, 525, 656
- Supervised learning, 104
- Support vector machine, 139
- Surrogate loss function, 274
- SVD, *see* singular value decomposition
- Symmetric matrix, 40, 42
- Tangent distance, 267
- Tangent plane, 511
- Tangent prop, 267
- TDNN, *see* time-delay neural network
- Teacher forcing, 379, 380
- Tempering, 599
- Template matching, 140
- Tensor, xi, xii, 32
- Test set, 109
- Tikhonov regularization, *see* weight decay
- Tiled convolution, 349
- Time-delay neural network, 364, 371
- Toeplitz matrix, 330
- Topographic ICA, 489
- Trace operator, 45
- Training error, 109
- Transcription, 100
- Transfer learning, 532