

- Gradient clipping, 287, 411
- Gradient descent, 82, 84
- Graph, xii
- Graphical model, *see* structured probabilistic model
- Graphics processing unit, 441
- Greedy algorithm, 321
- Greedy layer-wise unsupervised pretraining, 524
- Greedy supervised pretraining, 321
- Grid search, 429

- Hadamard product, xii, 33
- Hard tanh, 195
- Harmonium, *see* restricted Boltzmann machine
- Harmony theory, 567
- Helmholtz free energy, *see* evidence lower bound
- Hessian, 221
- Hessian matrix, xiii, 86
- Heteroscedastic, 186
- Hidden layer, 6, 166
- Hill climbing, 85
- Hyperparameter optimization, 429
- Hyperparameters, 119, 427
- Hypothesis space, 111, 117

- i.i.d. assumptions, 110, 121, 265
- Identity matrix, 35
- ILSVRC, *see* ImageNet Large Scale Visual Recognition Challenge
- ImageNet Large Scale Visual Recognition Challenge, 22
- Immorality, 573
- Importance sampling, 588, 620, 691
- Importance weighted autoencoder, 691
- Independence, xiii, 59
- Independent and identically distributed, *see* i.i.d. assumptions
- Independent component analysis, 487
- Independent subspace analysis, 489
- Inequality constraint, 93
- Inference, 558, 579, 626, 628, 630, 633, 643, 646
- Information retrieval, 520
- Initialization, 298
- Integral, xiii
- Invariance, 339
- Isotropic, 64

- Jacobian matrix, xiii, 71, 85
- Joint probability, 56

- k -means, 361, 542
- k -nearest neighbors, 141, 544
- Karush-Kuhn-Tucker conditions, 94, 235
- Karush-Kuhn-Tucker, 93
- Kernel (convolution), 328, 329
- Kernel machine, 544
- Kernel trick, 139
- KKT, *see* Karush-Kuhn-Tucker
- KKT conditions, *see* Karush-Kuhn-Tucker conditions
- KL divergence, *see* Kullback-Leibler divergence
- Knowledge base, 2, 479
- Krylov methods, 222
- Kullback-Leibler divergence, xiii, 73

- Label smoothing, 241
- Lagrange multipliers, 93, 641
- Lagrangian, *see* generalized Lagrangian
- LAPGAN, 695
- Laplace distribution, 64, 492
- Latent variable, 66
- Layer (neural network), 166
- LCN, *see* local contrast normalization
- Leaky ReLU, 191
- Leaky units, 404
- Learning rate, 84
- Line search, 84, 85, 92
- Linear combination, 36
- Linear dependence, 37
- Linear factor models, 485
- Linear regression, 106, 109, 138
- Link prediction, 480
- Lipschitz constant, 91
- Lipschitz continuous, 91
- Liquid state machine, 401