

# Index

- accuracy, 82, 103, 178
- activation function, 137, 140, 146, 178, 202
- adjustable basis, *see* basis
- automated medical diagnosis, 11, 186
- automatic feature design, *see* feature
  
- back-propagation, 146, 178, 202
- backtracking line search, 226, 236
- bag of words (BoW), 109
- basis
  - adjustable basis, 136, 141
  - discrete cosine transform basis, 254
  - fixed basis, 135, 141, 174
  - Fourier basis, 136
  - neural network basis, 136, 140, 202
  - orthogonal basis, 158
  - polynomial basis, 135
- bounded curvature, *see* curvature
- buffer zone, *see* margin perceptron
  
- chain rule, 202, 269
- classification
  - classification accuracy, *see* accuracy
  - classification boundary, 178
  - ideal classification scenario, 166
  - linear classification, 73
  - multiclass classification, 95, 179
  - nonlinear classification, 166
  - real classification scenario, 171
- clustering, *see* *K*-means
- convex function, 24
- convolutional neural network, 188
- cost function, 15
- cross-validation
  - hold out cross-validation, 149, 182
  - k*-fold cross-validation, 152, 184
  - leave-one-out cross-validation, 155
  - 12 regularized cross-validation, 209
- counting cost, 83
- curse of dimensionality, 134
- curvature, 22, 241
  
- data types
  - audio data, 115
  - image data, 112
  - text data, 109
- data generating function, 53, 107, 131, 166
- dimension reduction techniques, 245
- discrete cosine transform (DCT), *see* basis
  
- edge detection, 14, 113
- eigenvalue, 25, 276
  
- face detection, 10, 186
- features
  - knowledge-driven feature design, 51, 104
  - automatic feature design/learning, 131, 166
  - feature transformation/map, 51, 104
  - histogram features, 107
- first order condition for optimality, 22
- fixed basis, *see* basis
- Fourier basis, *see* basis
- function approximation
  - continuous function approximation, 134
  - discrete function approximation, 133
  - indicator function approximation, 170
  - piecewise continuous function approximation, 166
  - universal approximator, 165
- fusion rule, 99, 179
  
- Galileo's ramp experiment, 12, 53, 155
- genome-wide association (GWA), 8
- gradient descent
  - gradient descent with adaptive step length, 226
  - gradient descent with fixed step length, 29, 219
  - gradient descent with momentum, 206
  - stochastic gradient descent, 229
- gross domestic product (GDP), 47
  
- handwritten digit recognition, 96
- Hessian matrix, 22, 268
- Hilbert matrix, 160
- hinge cost, 75
- histogram of oriented gradients (HoG), 115
- hold out cross-validation, *see* cross-validation
- hyperbolic tangent function, 137

- image compression, 254
- indicator function, 168
- $k$ -fold cross-validation, *see* cross-validation
- $K$ -means clustering, 246
- $K$ -nearest neighbors, 192
- Kepler's law, 67
- kernel
  - kernelization cost functions, 197
  - polynomial kernel, 199
  - Fourier kernel, 200
  - radial basis function (RBF) kernel, 201
- Kleiber's law, 62
- $\ell_2$  norm, 59
- learning rate, *see* step length
- least squares, 47
- leave-one-out, *see* cross-validation
- linear classification, *see* classification
- linear regression, *see* regression
- Lipschitz constant, 223, 241
- logistic regression, 56, 86
- logistic sigmoid function, 56
- loss, *see* cost function
- matrix completion, 257
- matrix factorization, 245, 274
- mean squared error (MSE), 50
- minimum
  - global minimum, 24
  - local minimum, 24
- Moore's law, 65
- multiclass classification, *see* classification
- multiclass softmax classifier, 99
- neural network, *see* basis
- Newton's method, 33
- nonlinear classification, *see* classification
- nonlinear regression, *see* regression
- numerical optimization, *see* optimization
- object detection, 10, 114
- Occam's razor, 148
- Ohm's law, 66
- one-versus-all (OvA) classifier, 96, 179
- optimization, 15, 21, 219
- outliers, 90
- overfitting, 147, 149, 181, 209
- principal component analysis (PCA), 250
- perceptron
  - basic perceptron, 73
  - margin perceptron, 78
  - squared margin perceptron, 81
- polynomial basis, *see* basis
- population growth modeling, 56
- positive (semi)-definite matrix, 25
- radial basis function, *see* kernel
- recommender system, 256
- rectified linear unit, 137
- regression
  - ideal regression scenario, 131
  - linear regression, 45
  - nonlinear regression, 56, 131
  - real regression scenario, 141
- regularization
  - $\ell_2$  regularization, 56, 95, 208
- revenue forecasting, 8
- saddle point, 26
- second order definition of convexity, 25
- sentiment analysis, 11, 110
- similarity matrix, 201
- singular value decomposition (SVD), 274
- spam detection, 111
- speech recognition, 117
- softmax cost, 76, 101
- softmax function, 75
- spectrogram, 117
- step function, 87
- step length
  - fixed step length rule, 219
  - adaptive step length rule, 226
  - step length for stochastic gradient, 233
- steepest descent, *see* gradient descent
- stochastic gradient descent, *see* gradient descent
- stopping condition, 27
- student loan debt, 7, 49
- stationary point, 23
- support vector machine (SVM)
  - hard-margin SVM, 93
  - log-loss SVM, 95
  - soft-margin SVM, 93
- Taylor series approximation, 21
- testing phase
  - testing error, 152
  - testing set, 4, 151
- training phase
  - training error, 152
  - training set, 1, 151
- transformed feature space, 52, 55, 107
- underfitting, 147, 149, 184
- universal approximator, *see* function approximation