Table of Content

Part i: Preliminary Concepts

Chapter 1: Introduction and Basic Concepts 3

Ultra Short History of AI and Machine Learning 3

What is Data Science, Machine Learning, Data Mining, and Artificial Intelligence? 4

- Why is data science important? 5
- What is an algorithm? 6

Evaluating a Machine Learning Algorithm 7

- Computational Complexity and Efficiency 7
  - Algorithms Runtime 9
  - Accuracy 10
  - The Balance between Accuracy and Efficiency 12

Types of Input Dataset 13

- Tabular Dataset 13
- Temporal Datasets 14
- Data Streams 15
- Graph Datasets 15
- Text 17
- Image and Video 17
- Audio 18
- Numerical Data Types 18

Tasks that Machine Learning Can Perform 21

- Unsupervised Learning 22
  - Clustering 22
  - Association Rule and Sequence Mining 22
  - Dimensionality Reduction 23
  - Deviation and Anomaly Detection 24
- Supervised Learning 21
  - Regression and Correlation 25
  - Classification 26
- Self-Supervised Learning 27
- Generative AI 27
- Reinforcement Learning 28

Training and Evaluation in Supervised Learning 30

- Ground Truth Dataset 30
- k-fold Cross Validation 32

Working with Data Step by Step 33

Summary 38
Chapter 2: Visualization 39
Background and History 39
Basic Visualizations 41
Pie chart 41
Bar chart and Histogram 41
Line chart 42
Scatter plot 42
Broken axis chart 43
More than One Dimensional Data Visualizations 43
Surface plot 44
Contour plot 44
Area plot 45
Radar chart 46
Heatmap 47
Calendar Plot 48
Timeline plot 48
Box plot and Violin plot 49
Hierarchical Visualizations 50
Dendrogram 50
TreeMap 51
Sunburst 52
Graph and Network Visualizations 52
Node-link Diagram 52
Arc Diagram 53
Chord Diagram (Circular Link) 53
Deviation and Change Visualizations 54
Divergent bar plot 54
Dumbbell plot 55
Waterfall 56
Beeswarm 57
3D Data Visualizations 57
Geographical Map Visualizations 59
Choropleth 59
Bubble Map 60
Customized and Other Multidimensional Visualizations 61
Multi-Dimensional Scaling 62
Summary 63
Further Reading and Watching 64

Chapter 3: Statistics & Probability 67
Concepts and Definitions 68
Variable and Value 68
Random or Stochastic Variable 68
Continuous and Discrete Variable 68
Different Data Types 69
Dependent, Independent, and Control Variables 69
Independent versus Dependent Trials 70
First Insight on the Data & Basic Statistical Concepts 70
Mean 71
Arithmetic Mean 71
Geometric Mean 71
Harmonic Mean 72
Median 73
Mode 73
Variance, Standard Deviation, and Covariance 74
Range and Quartile 75
Whisker plot or Box plot 76
Degree of Freedom 78
Probability 78
Basic Probability Concepts 79
Joint Probability 79
Bayes Rule 80
Probability Density/Mass Functions and Cumulative Distribution Functions 81
Statistical Distribution 82
Normal (Gaussian) Distribution 84
Uniform Distribution 85
Beta Distribution 85
Dirichlet Distribution 87
Binomial Distribution 88
Bernoulli Distribution 90
Geometric Distribution 90
Poisson Distribution 92
Weibull Distribution 94
Power Law (Long Tail), Exponential, and ZipfLaw Distributions 95
Zipf law 96
Pareto principle 96
Chi-Square Distribution 97
Boltzmann Distribution 98

Distribution Check with P-P Plot and Q-Q Plot 100
Expected Value and Expectation of a Function 102
Normalization 102
z-score 104
How Much Data is Enough? 105
Central Limit Theorem (CLT) 106
Law of Large Numbers (LLN) 106
Bias in Sampling 106
Confidence Interval 107
Hypothesis and Significance Tests 110
Hypothesis Error 113
A/B Testing and Significance Test 113
Parametric Significance Tests 114
t-test 115
One-sample t-test 115
Independent t-test 116
Paired t-test 116
ANOVA, MANOVA and ANCOVA 117
One-Way ANOVA 117
Repeated Measure (Dependent) ANOVA 118
Factorial ANOVA 118
MANOVA 118
ANCOVA 119
Non-Parametric Significance Tests 120
Chi-Square Test 120
Test of independence 121
Goodness-of-Fit 123
Kolmogorov-Smirnov 124
Kruskal-Wallis Test 124
Mann-Whitney-U Test 126
Significance Test Error Correction 128
Bonferroni correction 128
Tukey correction 129
Effect size 130
Cohen’s d Test 131
Odds Ratio 131
Correlation Coefficients 132
Entropy & Information Gain 134
Entropy 134

Measuring Distances between Distributions 137
KL-Divergence 137
Cross-entropy 138
Jensen Shanon Divergence (JS-Divergence) 138
Probability Estimations 140
Maximum Likelihood Estimation (MLE) Approach 141
Expectation Maximization (EM) 142
Summary 143
Further Readings and Watching 145
Part ii: Unsupervised Learning
Chapter 4: Clustering 149
Similarity and Dissimilarity 149
Similarity Measurement Methods 152
Euclidean Distance 152
Manhattan Distance 153
Mahalanobis Distance 154
Hamming Distance 156
Levenshtein (Edit) Distance 157
Longest Common Subsequences (LCS) Distance 157
Cosine Distance 157
Jaccard distance or Jaccard Index 159
Dynamic Time Warping (DTW) 160
Graph Similarities 162
Clustering Algorithms 164
Partitioning (k-representative) Methods 164
k-mean 166
k-median 166
k-medoid 156
Density Based Methods 168
Density-Based Spatial Clustering Applications with Noise (DBSCAN) 169
OPTICS 171
Hierarchical Methods 175
Single-Linkage (SLINK) 177
DIANA 179
Large Scale Hierarchical Clustering 182
BIRCH 183
CURE 187
Probabilistic and Fuzzy Clustering 189
Gaussian Mixture Model (GMM) 191

Fuzzy C-Mean Clustering (FCM) 194
Other Types of Clustering 194
Clustering Result Evaluation 197
Intrinsic Cluster Evaluation Methods 198
Elbow Method 198
Within Cluster Sum of Squares Errors (WSS) 199
Silhouette Index 199
Dunn Index 200
Davies-Bouldin Index 200
Extrinsic Cluster Evaluation Methods 201
Purity (Entropy) 201
Rand Index 202
Summary 203
Further Reading and Watching 205
Chapter 5: Frequent Itemset, Sequence Mining and Information Retrieval 207
Basic Concepts 208
Information Retrieval Methods 211
Concepts 212
Hash Structures 213
Hash Table 213
Minwise Independent Permutations Hashing (MinHash) 214
Tree Data Structure 217
Binary Tree 218
2-3 Search Tree 219
B-Tree and B+Tree 220
Red-black Tree 220
Trie and Radix Tree 221
Tree Search Methods 222
BFS and DFS 222
Beam Search 222
Monte Carlo Tree Search 224
Bloom Filters 226
Sliding Windows 228
Sliding Window Types 228
SkipList 230
Frequent Pattern Mining Algorithms 232
Apriori 232
FP-Growth 234
ECLAT 239

Sequence Mining Algorithms 241
Generalized Sequential Pattern (GSP) 242
SPADE 244
FreeSpan 247
PrefixSpan 250
Sequential Data Prediction with Hidden Markov Model 254
Probabilistic Graphical Model 254
Hidden Markov Model Concept 255
Markov Model for Prediction 257
Hidden Markov Model for Prediction 259
Likelihood (Forward Algorithm) 261
Decoding (Viterbi Algorithm) 263
Training (Baum-Welch) Algorithm 264
Summary 266
Further Reading and Watching 267
Chapter 6: Feature Engineering 269
Basic Concepts of Feature Engineering 270
Feature Selection 270
Filtering methods 271
Wrapper methods 272
Sequential Feature Selection and Sequential Backward Selection 272
Genetic Algorithm 274
Embedded methods 276
Feature Generation 276
Feature Engineering for Numerical Data 277
Scaling Data 278
Min-Max scaling 279
Standardization and z-score Standardization 279
L1 and L2 norms 280
The Magic Power of Transformation 280
Logarithmic Transformation 282
Box-Cox Transformation 282
Feature Engineering for Categorical Data 284
One-hot Encoding 284
Dummy Coding 285
Effect Coding 285
Feature Hashing 285
Bin Counting 286
Target Encoding 287

Feature Engineering for Textual Data 288
Bag-of-words 291
Subword Tokenization 291
N-Grams 291
Part of Speech Tagging 292
Word Embeddings and Language Models 292
Training a Word Embedding 294
Word Embedding Models 294
Word2Vec 295
GloVe 296
FastText 297
Theme and Keyword Extraction 297
TF/IDF 297
TextRank 297
Rapid Automatic Keyword Extraction (RAKE) 298
Yet Another Keyword Extractor (YAKE) 298
LLM Keyword Extraction 299
Feature Engineering for Image and Video Data 300
Image Processing Concepts and Components 300
Features Extraction from Image Data 303
Image Feature Extraction Algorithms 305
Harris-Stephens (HS) corner detection 305
Maximally Stable Extremal Regions (MSER) 305
Histogram Oriented Gradient (HOG) 307
Scale Invariant Feature Transform (SIFT) 308
Watershed Transformation 312
Video Feature Extraction 314
Motion Vectors 314
Optical Flow 314
3D Convolutional Neural Network (3D CNN) 315
Graph Convolutional Networks 316
Feature Engineering for Signals and Time Series 316
Time series Features 317
Stationarity and non-stationarity time series 317
Periodicity and Seasonal Behaviors 317
Trends 318
Motif 318
Lag 319
Step, Burst and Change Point 319

Signal Features 320
Signal Concepts 320
Types of Signals 320
Signal and Time Series Smoothing 322
Summary 323
Further Reading or Watching 326
Chapter 7: Dimensionality Reduction and Data Decomposition 327
Dimensionality Reduction Methods 328
The Curse of Dimensionality 329
Linear Dimensionality Reduction Methods 330
Principal Component Analysis (PCA) 330
Incremental PCA 334
Linear Discriminant Analysis 335
Fisher Linear Discriminant (FLD) 335
Non-linear Dimensionality Reduction Methods 338
Locally Linear Embedding (LLE) 339
t-distributed Stochastic Neighbor Embedding (tSNE) 341
Uniform Manifold Approximation and Projection (UMAP) 344
Signal and Time Series Decomposition 347
Signal and Time Series Concepts 347
Fourier Transform 349
Wavelet Transform 350
Continuous and Discrete Wavelet Transform 352
Approximate Aggregation Methods (Time Series) 354
Matrix Decomposition 355
Cholesky Decomposition 356
Non-negative Matrix Factorization (NMF) 357
Singular Value Decomposition (SVD) 357
Topic Modeling (Clustering with Matrix Decompositions) 358
Latent Semantic Indexing (LSI) 359
Latent Dirichlet Allocation (LDA) 361
Tensor Decompositions 363
Tensor Concepts 364
Tensor and Matrix Operations 366
Mode-n unfolding 367
n-mode product (multiplication) 368
Tensor Decomposition Methods 369
Canonical or Parallel Factor Decomposition (CP) 369
Tucker Decomposition (TD) 370

Tensor Train Decomposition (TT) 371
Summary 372
Further Reading and Watching 374
Part iii: Supervised Learning
Chapter 8: Regression, Regularization, and Optimization 377
Objective, Cost, and Loss 378
Linear Regressions 379
Univariate Linear Regression 379
Model Parameters (Coefficients) Estimation 381
Multiple Linear Regression 383
Deciding About Model Variables? 384
Linear Regression Challenges and Resolutions 385
Polynomial Regression 386
Model Parameters (Degrees and Coefficients) Estimation 387
Coefficients 387
Degrees 388
Piecewise, Segmented, or Non-Additive Regression 388
Evaluating the fitness of training sets in linear models. 390
Residual Standard Error (RSE) 391
R^2 391
Root Mean Square Error (RMSE) 392
AutoRegressive Integrated Moving Average (ARIMA) 393
Extrapolation and Interpolation 394
ARIMA Model 394
ARIMA Parameters Estimation 395
Determining the best order of differencing (d) 398
Determining the best order of Auto Regression (p) 398
Determining the best order of Moving Average (q) 398
Logistic Regression 399
Model Parameters Estimation 401
Softmax Regression (Classifier) 403
Model Parameter Estimation 406
Evaluating Regression Models Fitness 407
k-fold cross validation 408
Learning Curve 408
ROC Curve 409
Pseudo R^2 410
Wald Test 410
Information Criterion 411

Likelihood Ratio Test 412
Devils of Model Building 414
Overfitting and Underfitting 414
Bias-Variance Tradeoff 415
Regularization 418
Ridge 419
LASSO 421
Elastic Net 422
Non-Negative Garrote 422
Optimization Algorithms 424
Mathematical Concepts Required for Optimization 424
Derivative 425
Second-Order Derivative 428
Partial Derivative 428
Gradient 429
Jacobian 430
Hessian 430
Integral 430
Taylor Series 432
What is Optimization in Machine Learning? 432
Gradient Descent 434
Types of Gradient Descent 437
Batch Gradient Descent (BGD) 437
Stochastic Gradient Descent (SGD) 438
Mini Batch Gradient Descent (miniBGD) 439
Newton Method 440
Early Stopping 443
Summary 444
Further Readings and Watching 446
Chapter 9: Classification 449
Rule Based Classifier 450
Naive Bayes 451
Bayes Theorem 451
Prediction with Naive Bayes 453
Gaussian Naive Bayes 455
Naive Bayes Prediction Example 456
k Nearest Neighbor (kNN) 459
Voronoi Tessellation 460
KD-Tree 461

Locality Sensitive Hashing (LSH) 463
Support Vector Machine (SVM) 466
Handling Non-linear Data with Maximum Margin Classifier 470
Kernel Trick 471
Kernel Functions 472
Multi-label classification for SVM 473
How Does the SVM Perform the Prediction? 474
SVM Computational Complexity 474
Decision Trees 475
Iterative Dichotomiser 3 (ID3) 6 477
Chi-square Automatic Interaction Detector (CHAID) 479
C4.5 481
Classification and Regression Trees (CART) 484
Decision Tree Challenges and the Need for Pruning. 485
Decision Trees Computational Complexity 487
Ensemble Learning Methods 489
Bootstrap Aggregating (Bagging) 489
Boosting 491
Stacking 492
Random Forest 493
Adaptive Boosting (AdaBoost) 496
Gradient Boosting Decision Tree 501
Regression with GBDT 502
Classification with GBDT 505
eXtreme Gradient Boosting (XGBoost) 510
LightGBM 518
CatBoost 519
How to Select the Best Classification Model? 522
Summary 523
Further Reading or Watching 524
Part iv: Deep Learning
Chapter 10: Neural Networks and Deep Learning 531
Universal Approximation Theory 533
Biological Neural Network 533
Artificial Neural Network 535
Perceptron Algorithm 539
Multilayer Perceptron 541
Activation Functions 544
Neural Network Cost Functions 547

Quadratic Cost (Mean Squared Error) 547
Cross Entropy 548
Kullback-Leibler Divergence 548
Hellinger (Bhattacharyya) Distance 549
Neural Network Optimizers 551
Stochastic Gradient Descent with Momentum 552
Exponentially Weighted Average 553
SGD with Momentum 554
Nesterov Momentum 555
Adagrad 557
RMSprop 559
Adam 560
Optimizers Summary 561
Backpropagation 562
Forward and Backward Pass Example 567
Regularization in Neural Network 568
Vanishing and Exploding Gradients 568
Weight Initialization 569
Batch Normalization 570
Gradient Clipping 571
Other Normalization Techniques 571
Dropout 572
Early Stopping 573
Convolutional Neural Network (CNN) 574
Convolution and Cross correlation 576
CNN Architecture 577
Different Types of Convolutions 584
Recurrent Neural Network (RNN) 587
Long-Term Short-Term Memory (LSTM) 590
Gated Recurrent Unit (GRU) 593
Bidirectional RNN 594
Deep RNN 596
RNN Examples 596
Summary 599
Further Reading or Watching 601
Chapter 11: Self-Supervised Neural Networks 603
Representation Learning Concepts 605
Generative vs. Discriminative Model 605
Deterministic vs. Stochastic Model 606

Self Organizing Maps (SOM) 607
Boltzmann Machines 611
Restricted Boltzmann Machine (RBM) 612
The Intractable Problem of Z in RBM 613
RBM Training 615
Deep Belief Network and Deep Boltzmann Machine 617
Autoencoders 621
How do Autoencoders work? 622
Autoencoders can Cheat 623
Autoencoder Types 623
Sparse Autoencoder 624
Denoising Autoencoder 626
Contractive Autoencoder 627
Stacked Autoencoder 628
Variational Auto-Encoder (VAE) 630
The Magic of Generative Models in Data Reconstruction 630
How does VAE work? 631
VAE Cost Function 633
U-Net 633
Generative Adversarial Network 637
Training GAN 639
GAN Cost Function 640
GAN Challenges 641
Loss Oscillation 641
Slow Convergence 641
Mode Collapse 641
Uninformative Loss 642
Evaluating GAN Result 642
Inception score (IS) 642
Fréchet Inception Distance (FID) 643
GAN Architectures 644
Conditional GAN (CGAN) 644
Deep Convolutional GAN (DCGAN) 645
Wasserstein GAN (WGAN) 646
Wasserstein or Earth Mover Distance 647
Continuity Condition 649
WGAN Cost Function 649
Weight Clipping 650
WGAN with Gradient Penalty (WGAN-GP) 650

Pix2Pix 651
CycleGAN 653
CycleGAN Architecture 653
CycleGAN Cost Functions 655
Generator and Discriminator Networks 655
StyleGAN Models 656
Noise Sources and AdaIN 657
Progressive Growing 658
Style Mixing 659
StyleGAN Successors 660
Contrastive Representation Learning 663
Contrastive Learning Loss Functions 664
Contrastive Loss 665
Triplet Loss 665
Siamese Network 666
Training Siamese Network 666
Data Preparation for Training 667
Testing Siamese Network 667
Text-to-Image Models 668
Text-to-Image Related Concepts 668
Zero-shot Learning 668
Autoregressive Models 669
Diffusion Models 669
Inpainting and Outpainting 671
CLIP (Contrastive Language–Image Pre-training) 671
Contrastive Pretraining 672
Training 673
Experimenting Zero-Shot learning 673
VQ-GAN (Vector Quantized GAN) 674
VQGAN-CLIP 677
DALL-E Models 677
DALL-E v1 678
DALL-E v2 678
DALL-E v3 680
Imagen 681
Parti 682
Stable Diffusion Models 684
Latent Diffusion Model 684
Training LDM 685

Stable Diffusion XL 686
Stable Diffusion 3 687
Other models that we didn't explain 689
Summary 690
Further Reading and Watching 691
Chapter 12: Deep Learning Models and Applications (Text, Vision, and Audio) 693
Attention and Transformer Neural Network 694
Sequence-to-Sequence 694
Seq2Seq with Attention 696
Attention Models 699
Dot Product Attention 699
Bahdanau (Additive) Attention 699
Luong Attention 700
Self-Attention 700
Cross-Attention 704
Transformer Network 705
Skip Connection and Positional Encoding 706
Skip (Residual) Connection 706
Positional Encoding 706
Multi Head Attention 708
Transformer Architecture 710
Encoder 710
Decoder 710
Masked Multi-Head Attention 712
Structured State Space Sequence Models (S4) 713
State Space Model 713
Linear Time Invariants SSM 714
Structured State Space for Sequence Modeling (S4) 714
Mamba Architecture 714
Large Language Models 717
BERT 718
BERT Architecture 719
BERT Input 719
Pre-training BERT 720
Fine-tuning BERT 721
Derivations of BERT 722
RoBERTa 723
DistilBERT 723
XLM 724

Transfer Learning with a Unified Text to Text Transformer (T5) 724
Generative Pre-trained Transformer (GPT) Models 725
GPT-1 725
GPT-2 726
GPT-3 727
Instruct GPT (Commercialized as ChatGPT) 729
Post ChatGPT LLMs 730
LLama Models 731
LLama2 732
LLama 3 733
Successors of Llama 733
Mistral Model Series 735
Mistral 7B 735
Mixtral-8x7B 736
Mixtral-8x22B 738
Fine-Tuning of LLM 738
Reinforcement Learning from Human Feedback (RLHF) 738
Direct Preference Optimization (DPO) 739
Task-specific and Domain-specific Fine-Tuning 740
Adversarial Fine-Tuning 740
Low-Rank Adaptation (LoRA) 741
Natural Language Model Evaluations 741
LLM Benchmarks 746
Summary of LLMs 747
Computer Vision 749
Image Classification with Convolutional Neural Networks 749
LeNet 750
AlexNet 750
VGG 751
GoogleLeNet / Inception Net 753
Residual Networks (ResNet) 754
Inception-v4 and Inception-ResNet 757
Vision Transformers 758
Object Detection 761
Object Detection Concepts 762
R-CNN Models 763
R-CNN 763
Fast R-CNN 764
Faster R-CNN 765

Single Shot MultiBox Detection (SSD) 767
YOLO Models 771
YOLO v1 771
YOLO v2 772
YOLO 9000 773
YOLO v3 773
Semantic Segmentation and Instance Segmentation 775
U-Net 776
Fully Convolutional Network (FCN) 776
Mask R-CNN 778
DeepLab Models 779
DeepLab v1 779
DeepLab v2 781
DeepLab v3 781
DeepLab v3+ 782
Segment Anything Model (SAM) 783
SAM v1 783
SAM v2 787
3D View Synthesis 788
Concepts 789
Neural Radiance Field (NeRF) Model 792
3D Gaussian Splatting 793
Other computer vision models that we didn’t explain 797
Audio 800
Audio Signal Concepts and Features 801
Sound Modeling 804
WaveNet Models 805
WaveNet 1 805
Parallel WaveNet (WaveNet 2) 808
Tacotron Models 809
Tacotron 1 809
Tacotron 2 811
wav2vec Models 811
wav2vec v1 811
wav2vec v2 813
Whisper 814
Training Whisper 814
Whisper Architecture 815
Summary 816

Further Reading and Watching 816

Part v: Reinforcement Learning

Chapter 13: Reinforcement Learning 821
Reinforcement Learning Concepts 822
Exploitation versus Exploration 824
Markov Decision Process 825
Grid-world 826
Deterministic, Nondeterministic, and Stochastic Models 829
Policy 830
Expected Return (Goal) 830
Value, Value function, and Action-value Function 831
State Transition Diagram 832
Bellman Equation 834
Bellman Equation Examples 835
Multi-Armed Bandits Problem 839
Epsilon Greedy (-Greedy) 840
Upper Confidence Bound (UCB) 841
Thompson Sampling 843
Real-world examples of MAB problem 845
Optimal Policy and Optimal Value Function 847
Control and Prediction Problems 848
Tabular Methods 850
Monte Carlo Method (Policy Evaluation and Policy Improvement) 852
Monte Carlo Challenges 852
Monte Carlo Policy Evaluation and Policy Improvement 853
When to use MC method 854
Dynamic Programming (Policy Iteration and Value Iteration) 854
How does a DP algorithm work? 855
Converting Bellman equation into interactive updates 856
When to use DP method 858
Temporal Differences (TD) Learning 860
SARSA and Q-Learning 861
SARSA 862
Q-Learning 863
Dyna-Q 864
Approximation Methods and Deep Reinforcement Learning 867
n-Step Methods and TD(λ) 867
TD(λ) 868
Eligibility trace 870

Continuous Space and Function Approximation 872
Deep Q-Networks (DQN) 873
DQN cost function 874
Two problems of Deep RL (i.d.d. samples and non-stationary target) 875
Experience Replay 876
DQN's target network 877
Convolutional Layers of DQN to play Atari games 878
Double Deep Q-Networks (DDQN) 879
Dueling Networks 881
Prioritized Experience Replay (PER) 882
Policy Gradient Methods 884
Policy Gradient Advantages 884
Policy Gradient Objective Function 884
REINFORCE 885
REINFORCE with Baseline 886
Trust Region Policy Optimization (TRPO) 888
Trust Region 888
Minorization-Maximization (MM) algorithm 889
Surrogate Loss 890
Fisher Information 891
Natural Policy Gradient 891
TRPO Algorithm 895
Proximal Policy Optimization (PPO) 895
Penalty Coefficient 896
Clipped Objective 896
PPO as Actor-Critic Model 897
Actor-Critic Methods 898
Asynchronous Advantage Actor-Critic (A3C) 899
Deep Deterministic Policy Gradient (DDPG) 901
Twin Delayed Deep Deterministic Policy Gradient (TD3) 903
Soft Actor-Critic (SAC) 905
Entropy Regularization (Entropy Bonus) 906
Soft Q-value 906
Critics and Target Critics 907
Actor 907
Policy Update and Reparameterization Trick 908
SAC Architecture 909
Dreamer Models 910
Straight-Through Gradient 910

Dreamer v1 911
Dreamer v2 913
Dreamer v3 914
Summary 916
Further Reading or Watching 918

Part vi: Other Concepts and Algorithms

Chapter 14: Making Lighter Neural Networks and Machine Learning Models 923
Data Compression Algorithms 924
Byte Pair Encoding (BPE) 925
Bitmap index 926
Huffman Coding 927
Lempel–Ziv–Welch (LZW) Encoding 928
Sparse Coding 930
Quantization 933
Vector Quantization and Cluster Quantization 933
Non-uniform Quantization (Signal Companding) 935
μ-Law 936
A-Law 936
Quantization for Neural Networks 937
Fixed Point versus Floating Point Forms 937
How to Quantize? 938
When to Quantize a Neural Network? 939
Dynamic versus Static Quantizations 939
Post-Training Quantization (PTQ) 939
Pre-training Quantization and Quantization-Aware Training (QAT) 940
Channel-wise Quantization 940
Pruning and Sparsification in Neural Network 942
What to prune/sparsify? 943
When to prune? 944
Structured versus Unstructured pruning 944
How to identify pruning candidates? 945
Optimal Brain Damage (OBD) and Optimal Brain Surgeon (OBS) 945
Magnitude-based Pruning 947
Activation-based Pruning 947
Taylor Expansion based Pruning 948
Low-Rank Adaptation of Large Language Models (LoRA) 950
Lighter Self-Attention Mechanisms 951
FlashAttention Models 951
Multi Query Attention and Group Query Attention 953

Multi Query Attention (MQA) 953
Group Query Attention (GQA) 955
Sliding Window and Dilated Sliding Window Attentions 955
Memory Efficient Gradient Descent 956
Gradient Accumulation 956
Gradient Checkpointing 957
Conjugate Gradient 958
Computing the next step in CGD 959
Neural Network Training Paradigms 959
Transfer Learning 960
Multitask Learning 961
Meta-learning 963
Model-based Meta-Learning (MML) 964
Metric-based Meta-Learning 964
Optimization-based Meta-Learning 965
Curriculum Learning 966
Federated Learning 967
Knowledge Distillation 968
Knowledge Distillation Concepts 969
Soft and hard targets 969
Softmax temperature 971
Transfer set 971
Balancing factor 971
Knowledge Distillation Training 971
Common Knowledge Distillation Architectures 972
Automatic Machine Learning 974
Hyperparameter Tuning 974
Random Search 974
Grid Search 975
Neural Architectural Search 975
Summary 976
Further Readings and Watching 976
Chapter 15: Graph Mining Algorithms 979
Basic Concepts 980
Minimum Spanning Tree 981
Prim 982
Kruskal 984
Shortest Path Finding 985
Dijkstra 985

A* Search 987
Matching Algorithm 990
Hungarian Algorithm 990
Hopcroft-Karp Algorithm 992
Centrality Measurement Algorithms 995
Degree Centrality 995
Closeness Centrality 996
Betweenness centrality 997
Eigenvector centrality 998
Link Analysis Algorithms 998
PageRank 999
Hyperlinked Induced Topic Search (HITS) 1002
Community Detection in Graphs 1004
Spectral Clustering 1005
Louvain 1007
Leiden 1010
Graph Neural Network 1011
Challenges of working with Graphs in Neural Network 1012
Message Passing Layer (Propagation) 1013
Graph Pooling 1014
Spectral versus Spatial Representation 1015
Graph Convolutional Network (GCN) 1015
Graph Attention Network 1001
Graph SAGE 1017
High-Dimensional Search with Graphs 1020
Hierarchical Navigable Small World 1021
Summary 1023
Further Reading and Watching 1024
Chapter 16: Challenges of Working with Data 1025
Problem Complexity 1025
Sampling from Complex Data Structure 1028
Stratified Sampling and Clustered Sampling 1028
Monte Carlo (MC) Sampling 1029
Monte Carlo Markov Chain (MCMC) Methods 1031
Metropolis-Hastings 1032
Gibbs Sampling 1033
Importance Sampling 1036
Noise 1037
Types of Noises 1021

Noises Based on Statistical Distributions 1038
White noise 1038
Gaussian noise 1038
Other types of noises 1039
Salt-and-Pepper noise 1039
Gradient noise 1039
Speckle noise 1040
Noise Reduction 1040
Machine Learning based noise reduction 1041
Signal-based Noise Filtering 1041
Butterworth filter 1043
Wiener filter 1043
Kalman filter 1044
Imbalance Data Problem and Data Augmentation 1049
Data Augmentation Methods 1050
Tabular Data 1050
Image 1051
Text 1053
Signal and Time Series 1054
Reconstructing Missing Data 1055
Imputation 1056
Mean/Median or Mode 1056
K-Nearest Neighbors (kNN) for Imputation 1056
Hot deck imputation 1056
Cold deck imputation 1056
Regression Imputation 1057
Griffin-Lim Algorithm (GLA) 1057
Customized methods 1058
Interpolation 1058
Linear and Bilinear interpolation 1058
Polynomial interpolation 1060
Spline interpolation 1060
Kriging interpolation 1060
Radial Basis Function Interpolation (RBFI) 1062
Anomaly and Outliers Detection 1062
Unsupervised Methods 1063
Isolation-based Methods 1063
Isolation Forest (iForest) 1063
One-Class SVM 1065

Local Outlier Factor (LOF) 1065
Random Sample Consensus (RANSAC) Algorithm 1066
Data and Model Changes over Time 1067
Concept, Model and Data Drift 1068
Tackling Drift Challenges 1068
Cold Start Problem 1069
Tackling Cold Start Problem 1070
Raters Agreement Methods 1071
Percentage Agreement 1071
Cohen's Kappa 1071
Fleiss' Kappa 1072
Krippendorff’s Alpha 1073
Utilizing LLMs 1075
Retrieval Augmented Generation (RAG) 1075
Fine Tuning the LLM with External Data 1075
Indexing the External Data 1075
Store External Data in an Embedding 1076
Prompt Engineering 1077
Zero-shot and Few-shot Prompting 1077
Chain-of-Thought (CoT) 1078
Self-Consistency 1079
Tree-of-Thought (ToT) 1081
Fairness, Bias, and Transparency 1083
Concepts and Terms 1084
Fairness Parity Metrics 1085
Bias and Fairness Mitigation Methods 1086
Interpretability and Transparency 1088
SHapley Additive exPlanations (SHAP) 1088
Local Interpretable Model-agnostic Explanations (LIME) 1091
Summary 1092
Further Reading and Watching 1093
Bibliography 1095
Index 1123