Data Science Topics

07 April 2021 21:13

Machine Learning

1. Maximum Likelihood Estimation
    a. Outputs drawn from Gaussian Distribution (MSE)
    b. Outputs drawn from Binomial Distribution (BCE)
2. Bayesian Parameter Estimation
    a. Outputs drawn from Gaussian Distribution (MSE)
    b. Outputs drawn from Binomial Distribution (BCE)
3. Linear Regression
4. Gaussian Regression
5. Logistic Regression
6. Eigen Value Decomposition
7. Principal Component Analysis
8. Singular Value Decomposition
9. Naïve Bayes
10. K Means
11. K Nearest Neighbors
12. Support Vector Machines
13. Trees
    a. Decision Trees
    b. Regression Trees
    c. Random Forests
    d. GBDT
        i. LambdaRank
        ii. LambdaMART
14. Gaussian Mixture Models
15. Expectation Maximization
16. Hidden Markov Models
17. Conditional Random Fields
18. Markov Random Fields
19. Latent Dirichlet Allocation
20. Gibbs Sampling (Monte Carlo Markov Chain)
21. Thompsons Sampling
22. Hopfield Networks
23. Boltzmann Machines
24. Deep Belief Networks
25. Mixture Density Networks
26. Kernel Methods

Deep Learning

1. Perceptron
2. Convolutional Neural Networks
3. Graph Neural Networks
    a. GCN
    b. Graph Sage
    c. Node2Vec
4. Generative Adversarial Networks
5. BERT
6. Transformers / Attention
7. Knowledge Distillation
8. Transfer Learning
9. Auto Encoders
10. Variational Auto Encoders
11. Recurrent Neural Networks
12. Long-Short Term Models
13. Gated Recurrent Units
14. Auto Regressive Models

Optimizers

1. SGD
2. Momentum
3. AdaGrad
4. RMSProp
5. Adam

Regularization

1. L1
2. L2
3. Batch Normalization
4. Dropout

Important Topics

1. Discriminative v/s Generative Models
2. Overfitting v/s Underfitting
3. Bias and Variance
4. Maximum Likelihood v/s Bayesian
5. Graphical Models
6. Variational Inference
7. Approximate Inference
8. Bayesian Inference
9. Graph Spectral Theory
10. Sufficient Statistics
11. PageRank
12. Topic Models
13. Jacobian / Hessian
14. Backpropagation
15. Second Order Methods
16. Imbalanced Data
    a. Preprocessing
    b. Metrics
    c. Models
17. Ensemble Models
    a. Bagging / Bootstrapping
    b. Boosting
    c. Adaboost

Loss Functions

1. Logistic Loss - Binary Cross Entropy
2. Log Loss
3. Hinge Loss
4. Mean Squared Error
5. Mean Absolute Error
6. Bayesian Page Rank
7. Pairwise Logistic Loss with NDCG/MRR

Metrics

1. AUC
2. Precision/Recall
3. NDCG
4. Mean Reciprocal Rank
5. MicroAUC@k
6. pAp@k
7. CTR
8. SSR
9. PSP
10. R-Squared / Adjusted R-Squared

Probability Distributions

1. Uniform
2. Normal
3. Binomial
4. Poisson
5. Bernoulli
6. Beta
7. Dirichlet
8. T-distribution