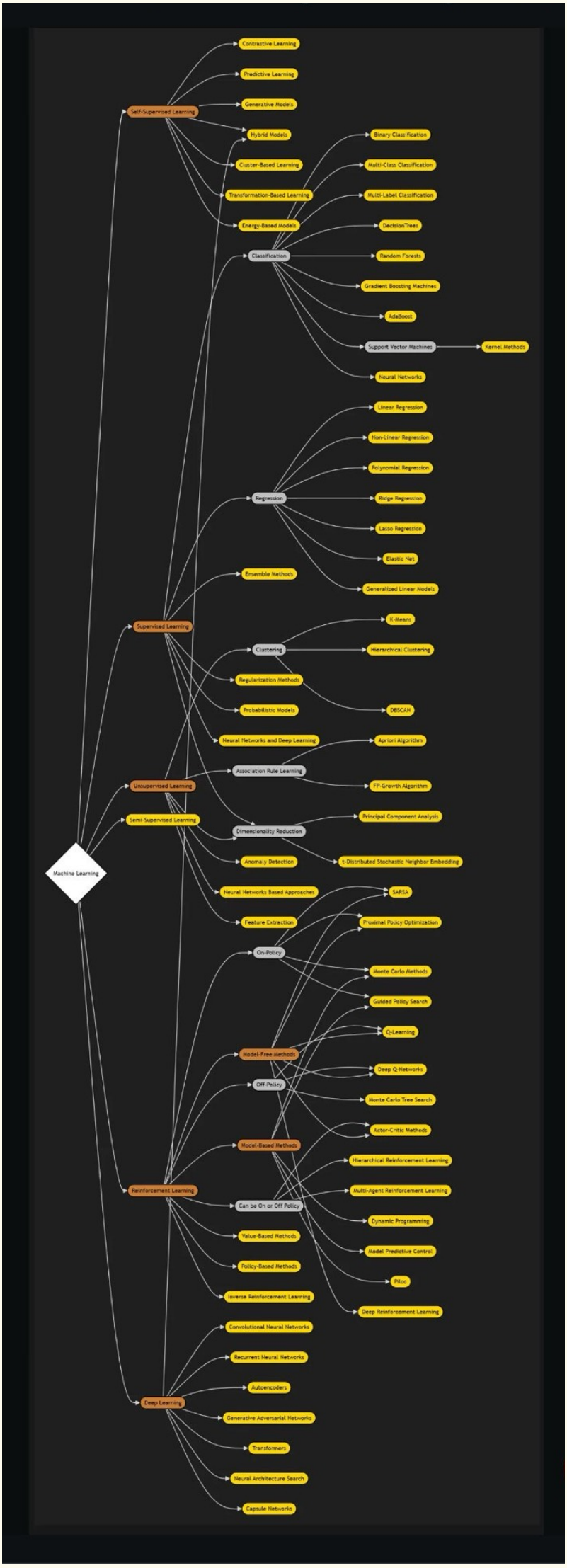

Machine Learning
Self-Supervised Learning
Contrastive Learning
Predictive Learning
Generative Models
Hybrid Models
Supervised Learning
Cluster-Based Learning
Transformation-Based Learning
Energy-Based Models
Classification
Binary Classification
Multi-Class Classification
Multi-Label Classification
DecisionTrees
Random Forests
Gradient Boosting Machines
AdaBoost
Support Vector Machines
Kernel Methods
Neural Networks
Regression
Linear Regression
Non-Linear Regression
Polynomial Regression
Ridge Regression
Lasso Regression
Elastic Net
Generalized Linear Models
Ensemble Methods
Unsupervised Learning
K-Means
Clustering
Hierarchical Clustering
Regularization Methods
DBSCAN
Probabilistic Models
Neural Networks and Deep Learning
Apriori Algorithm
Association Rule Learning
FP-Growth Algorithm
Semi-Supervised Learning
Principal Component Analysis
Dimensionality Reduction
t-Distributed Stochastic Neighbor Embedding
Anomaly Detection
Neural Networks Based Approaches
Feature Extraction
Reinforcement Learning
SARSA
Proximal Policy Optimization
On-Policy
Monte Carlo Methods
Guided Policy Search
Q-Learning
Model-Free Methods
Deep Q-Networks
Off-Policy
Monte Carlo Tree Search
Actor-Critic Methods
Model-Based Methods
Hierarchical Reinforcement Learning
Multi-Agent Reinforcement Learning
Can be On or Off Policy
Dynamic Programming
Value-Based Methods
Model Predictive Control
Policy-Based Methods
Pilco
Inverse Reinforcement Learning
Deep Reinforcement Learning
Deep Learning
Convolutional Neural Networks
Recurrent Neural Networks
Autoencoders
Generative Adversarial Networks
Transformers
Neural Architecture Search
Capsule Networks

25 Most important Mathematical Definitions in Data Science

blog.DailyDoseofDS.com

1) Gradient Descent

$$\theta_{j+1} = \theta_j - \alpha \nabla J(\theta_j)$$

2) Normal distribution

$$f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

3) Z-score

$$z = \frac{x - \mu}{\sigma}$$

4) Sigmoid

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

5) Correlation

$$\text{Correlation} = \frac{\text{Cov}(X, Y)}{\text{Std}(X) \cdot \text{Std}(Y)}$$

6) Cosine Similarity

$$\text{similarity} = \frac{A \cdot B}{\|A\| \|B\|}$$

7) Naive Bayes

$$P(y|x_1, \ldots, x_n) = \frac{P(y) \prod_{i=1}^{n} P(x_i|y)}{P(x_1, \ldots, x_n)}$$

8) MLE

$$\text{argmax}_\theta \prod_{i=1}^{n} P(x_i|\theta)$$

9) OLS

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

10) F1 Score

$$\frac{2 \cdot P \cdot R}{P + R}$$

11) ReLU

$$\max(0, x)$$

12) Softmax

$$P(y = j|x) = \frac{e^{x^T w_j}}{\sum_{k=1}^{K} e^{x^T w_k}}$$

13) R2 score

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

14) MSE

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

15) MSE + L2 Reg

$$\text{MSE}_{\text{regularized}} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

16) Eigen vectors

$$Av = \lambda v$$

17) Entropy

$$\text{Entropy} = -\sum_i p_i \log_2(p_i)$$

18) KMeans

$$\text{argmin}_S \sum_{i=1}^{k} \sum_{x \in S_i} \|x - \mu_i\|^2$$

19) KL Divergence

$$D_{\text{KL}}(P\|Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

20) Log-loss

$$-\frac{1}{N} \sum_{i=1}^{N} \left(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)\right)$$

21) SVM

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \max(0, 1 - y_i(w \cdot x_i - b))$$

22) Linear regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \epsilon$$

23) SVD

$$A = U \Sigma V^T$$

24) Lagrange multiplier

$$\max f(x) \; ; \; g(x) = 0$$

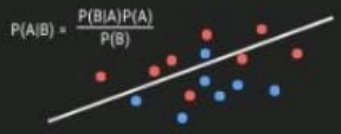
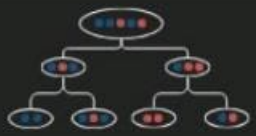
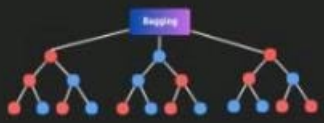
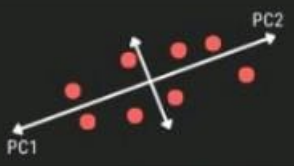
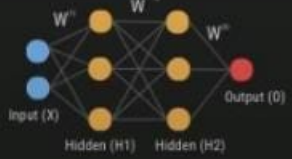
$$L(x, \lambda) = f(x) - \lambda * g(x)$$

25) What will you add?

• • •

Machine Learning Hyperparameters

DataInterview.com

ML Algorithms	Hyperparameters
Linear Regression 	• L1/L2 Penalty • Fit Intercept • Solver
Logistic Regression 	• L1/L2 Penalty • Class Weight • Solver
Naive Bayes 	• Alpha • Binarize • Fit Prior
Decision Tree 	• Criterion • Min Sample Split • Max Depth
Random Forest 	• Criterion • N Estimators • Max Depth • Max Features
Gradient Boosted Trees 	• Criterion • Min Sample Split • Max Depth • Learning Rate • N Estimators
Principal Component 	• N Component • Iterated Power • SVD Solver
K-Nearest Neighbor 	• N Neighbors • Algorithm ('kd_tree', 'brute') • Weights
K-Means 	• N Clusters • Max Iter • Init
Dense Neural Networks 	• Hidden Layer Sizes • Solver • Activation • Alpha • Droput • Learning rate