"Bumper Sticker" Supplement to "A Comprehensive Approach to DS, ML & AI"

Paul F. Roysdon, Ph.D.

1 Unsupervised Learning - Dimension Reduction

- **Principle Component Analysis**: "Principle Component Analysis (PCA) is a statistical technique that transforms high-dimensional data into a lower-dimensional subspace by identifying the directions of maximum variance."
- **Singular Value Decomposition**: "Singular Value Decomposition (SVD) is a simple two-step method to perform dimension reduction of large datasets, and reconstruct them with only the most important information."
- **Manifold Learning**: "Manifold learning is a dimensionality reduction technique that discovers low-dimensional structures embedded in high-dimensional data by preserving local neighborhood relationships."

2 Unsupervised Learning - Clustering

- **k-Means**: "K-means is a clustering algorithm that partitions data into k clusters by minimizing the sum of squared distances between data points and their respective cluster centroids."
- **Kernel Density Estimation**: "Kernel Density Estimation is a non-parametric method for estimating the probability density function of a random variable by averaging over localized kernel functions centered at the data points."
- **DBSCAN**: "DBSCAN is a density-based clustering algorithm that groups together points in high-density regions while labeling isolated points as noise."
- **Expectation Maximization**: "The Expectation Maximization algorithm is an iterative method for computing maximum likelihood estimates in models with latent variables by alternately computing expected values (E-step) and maximizing the likelihood (M-step)."
- **Gaussian Mixture Models**: "Gaussian Mixture Models represent a probabilistic framework that models complex data distributions as a weighted sum of multiple Gaussian components, typically estimated via the Expectation Maximization algorithm."

3 Supervised Learning - Classification

- **Decision Trees**: "Decision Trees are supervised learning models that recursively partition the feature space into regions with homogeneous responses by applying a sequence of decision rules."

- **Random Forest**: "Random Forests are an ensemble learning method that constructs multiple decision trees using bootstrapped data and random feature subsets, aggregating their predictions to achieve robust and accurate classification or regression."
- **Gradient Boosting Trees**: "Gradient Boosting Trees are an ensemble learning method that sequentially builds decision trees to correct the errors of previous models by performing gradient descent in function space."
- **AdaBoost**: "AdaBoost is an ensemble learning algorithm that iteratively combines weak classifiers by reweighting training examples, ultimately forming a strong classifier that minimizes the exponential loss."
- **XGBoost**: "XGBoost is an advanced gradient boosting algorithm that builds an ensemble of decision trees using second-order optimization and regularization to achieve high predictive accuracy."
- **K-Nearest Neighbors**: "K-Nearest Neighbors is a non-parametric algorithm that classifies or predicts a new data point based on the majority label (or average value) of its kk closest training examples."
- **Naive Bayes**: "Naive Bayes is a probabilistic classification algorithm that applies Bayes' Theorem with the assumption of feature independence to predict class membership based on observed features."
- **Logistic Regression**: "Logistic Regression is a binary classification algorithm that maps a linear combination of input features to a probability through the sigmoid function."
- **Support Vector Machines**: "Support Vector Machines are supervised learning models that determine the optimal hyperplane separating classes by maximizing the margin between data points."
- **Linear Discriminant Analysis**: "Linear Discriminant Analysis is a supervised dimensionality reduction technique that identifies the linear combinations of features which best separate multiple classes."

4 Supervised Learning - Regression

- **Linear Regression**: "Linear Regression is a statistical method that models the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data."
- **Graph Optimization**: "The network flow problem seeks the maximum feasible flow from a source to a sink in a directed graph, subject to capacity constraints on the edges, and is solved optimally using algorithms like Ford-Fulkerson."
- **Linear Programs**: "Linear Programming is an optimization technique that maximizes or minimizes a linear objective function subject to linear constraints, with optimal solutions typically found at the vertices of a convex feasible region."
- **Mixed-Integer Linear Programs**: "Mixed Integer Linear Programming is an optimization framework that seeks to maximize or minimize a linear objective function subject to linear constraints, where some decision variables are restricted to integer values."

5 Filtering & Estimation

- **Maximum Likelihood Estimation**: "Maximum Likelihood Estimation is a statistical method for estimating model parameters by maximizing the likelihood that the observed data were generated by the model."
- **Extended Kalman Filter**: "The Extended Kalman Filter is a recursive state estimator that linearizes nonlinear system and measurement models about the current estimate, enabling real-time fusion of predictions with noisy measurements."

- **Contemplative Real-Time Estimator**: "The Contemplative Real-Time Estimator is a real-time parameter estimation algorithm that combines optimal Bayesian estimation with a sliding-window approach and soft-thresholding to yield robust and sparse estimates."

6 Deep Learning - Foundations

- **Vanilla Neural Networks**: "Artificial Neural Networks, also known as Vanilla Neural Networks or Multi-Layer Perceptron, are a basic, fully-connected feed-forward model that learns complex functions by iteratively adjusting its weights via backpropagation to minimize prediction error."

7 Deep Learning - Sequence Models

- **Recurrent Neural Networks**: "A Recurrent Neural Network is a neural network architecture that processes sequential data by maintaining and updating a hidden state across time steps to capture temporal dependencies."
- **Long Short-Term Memory Networks**: "A Long Short-Term
Memory network is a recurrent neural network architecture that uses input, forget, and output gates to maintain and update a cell state, enabling it to capture long-term dependencies in sequential data."
- **Transformers**: "The Transformer architecture is a deep learning model that employs self-attention mechanisms to process sequences in parallel, effectively capturing global dependencies without relying on recurrence."

8 Deep Learning - Graphical Models

- **Convolutional Neural Networks**: "A Convolutional Neural Network is a deep learning model that employs convolutional and pooling layers to automatically extract and learn hierarchical features from grid-structured data, such as images."
- **Graph Convolutional Networks**: "Graph Convolutional Neural Networks are deep learning models that extend traditional convolution operations to graph-structured data, enabling effective representation learning and prediction on complex networks."

9 Deep Learning - Generative Models

- **Deep Belief Networks**: "Deep Belief Networks are a class of deep generative models that learn hierarchical representations by stacking Restricted Boltzmann Machines and fine-tuning with back-propagation."
- **Latent Dirichlet Allocation**: "Latent Dirichlet Allocation (LaDA) is a generative probabilistic model that discovers latent topics in a collection of documents by representing each document as a mixture of topics and each topic as a distribution over words."
- **Auto-Encoders**:"Autoencoder Networks are unsupervised deep learning models that learn efficient, low-dimensional representations of data by minimizing the reconstruction error between the input and its decoded output."

- **Variational Auto-Encoders**: "A Variational Autoencoder is a generative model that learns to encode input data into a structured latent space and decodes samples from this space to reconstruct the data, all while optimizing a variational lower bound on the likelihood."
- **Generative Adversarial Networks**: "Generative Adversarial Networks are a class of deep learning models in which a generator and a discriminator are trained simultaneously in an adversarial framework to produce realistic synthetic data."
- **Diffusion Networks**: "Diffusion Models are generative models that progressively corrupt data with noise in a forward process and learn to reverse this process to generate realistic samples."
- **Adversarial Machine Learning**: "Adversarial Machine Learning techniques generate carefully crafted input perturbations that exploit vulnerabilities in machine learning models, thereby revealing weaknesses and inspiring more robust defenses."

10 Reinforcement Learning

- **Q-learning**: "Q-Learning is a model-free reinforcement learning algorithm that iteratively updates the value of state-action pairs to learn an optimal policy based on the Bellman equation."
- **Policy Gradient**: "Policy Gradient is a model-free reinforcement learning algorithm that optimizes the policy directly by ascending the gradient of the expected return with respect to the policy parameters."
- **Deep Deterministic Policy Gradient**: "Deep Deterministic Policy Gradient is a model-free, off-policy actor-critic algorithm that learns deterministic policies for continuous action spaces by leveraging deep function approximators and stabilized training techniques."
- **Proximal Policy Optimization**: "Proximal Policy Optimization is a reinforcement learning algorithm that employs a clipped surrogate objective to ensure stable, incremental policy updates, thereby balancing exploration and exploitation."
- **Group Relative Policy Optimization**: "Group Relative Policy Optimization is a reinforcement learning approach that compares an individual agent's advantage to a group average, enabling more balanced and cooperative policy updates in multi-agent or multi-task environments."

11 Notable Concepts & Advanced Techniques

- **Numerical Differentiation**: "Numerical differentiation approximates the derivative of a function using finite difference methods and is widely used for gradient checking in back-propagation."
- **ADAM**: "Adaptive Moment Estimation (ADAM) is an adaptive
stochastic optimization algorithm that leverages first and second moment estimates of gradients to achieve fast and robust convergence in training deep learning models."
- **Word2Vec**: "Word2Vec is a neural network-based algorithm that learns dense vector representations of words by predicting a target word from its surrounding context (or vice versa), capturing rich semantic and syntactic relationships."
- **Supervised Fine-Tuning**: "Supervised Fine-Tuning is a technique that adapts pre-trained models to new tasks by further training on labeled data to optimize task-specific performance."
- **LoRA**: "Low-Rank Adaptation (LoRA) is a fine-tuning technique that adapts pre-trained models by injecting trainable, low-rank update matrices, enabling efficient parameter adaptation with minimal computational cost."

- **Reinforcement Learning with Human Feedback**: "Reinforcement Learning with Human Feedback is an algorithm that augments traditional RL by incorporating human-provided reward signals to guide the learning process, thereby aligning agent behavior with human preferences."
- **Mixture of Experts**: "The Mixture of Experts algorithm is a model combination framework that uses a gating network to weight the outputs of specialized expert models, yielding improved performance on complex tasks."
- **Markov Chain Monte Carlo**: "Markov Chain Monte Carlo is a set of algorithms that generate samples from a target probability distribution by constructing a Markov chain whose stationary distribution is the target distribution."
- **Monte Carlo Tree Search**: "Monte Carlo Tree Search is a decision-making algorithm that builds a search tree using random simulations and the Upper Confidence Bound for Trees (UCT) criterion to balance exploration and exploitation in complex domains."
- **Model distillation**: "Model distillation is an algorithm that transfers the knowledge of a large teacher model to a smaller student model by training the student to mimic the teacher's soft predictions."
- **RAG**: "The Retrieval Augmented Generation (RAG) algorithm is a hybrid approach that integrates document retrieval with generative modeling to produce contextually enriched, knowledge-driven outputs."