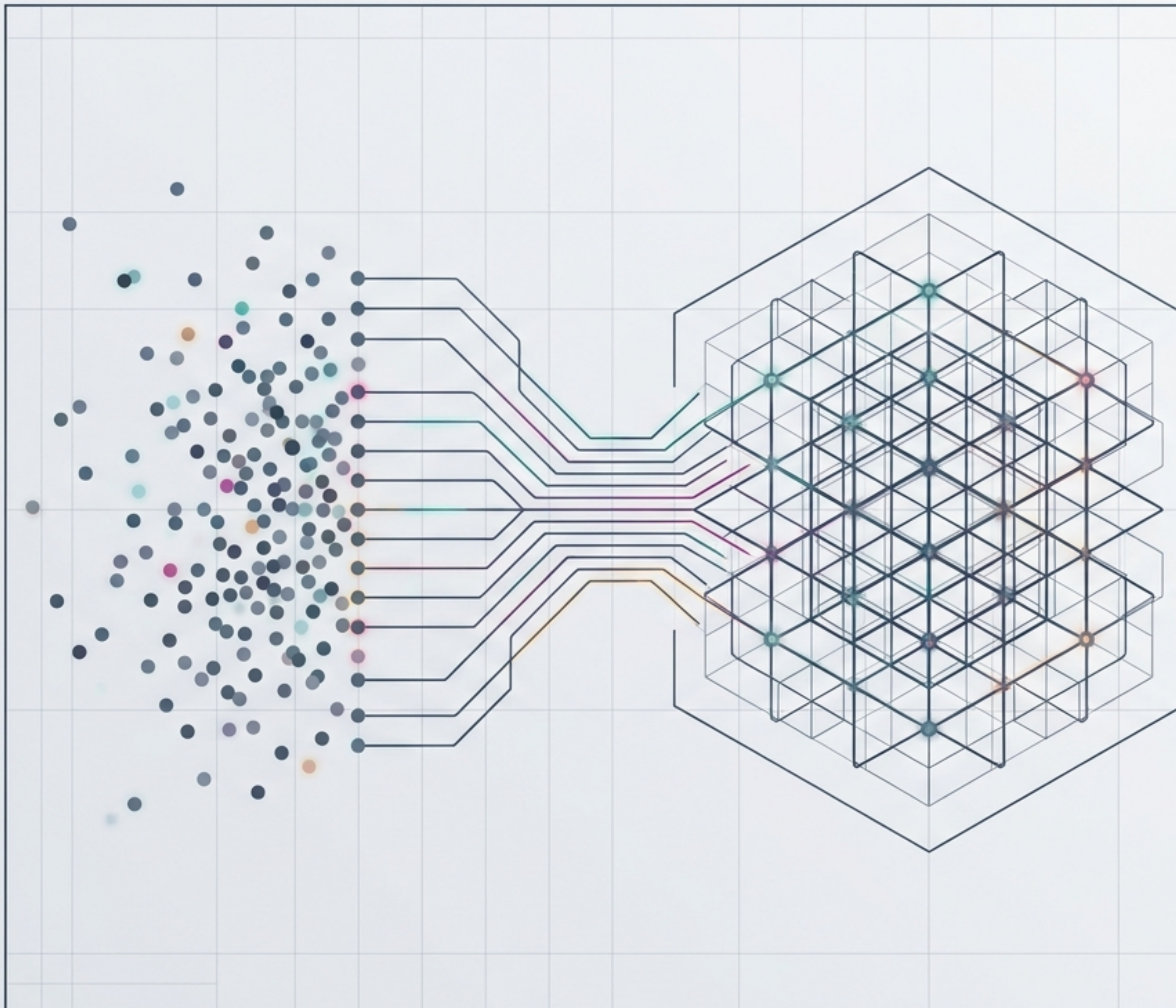
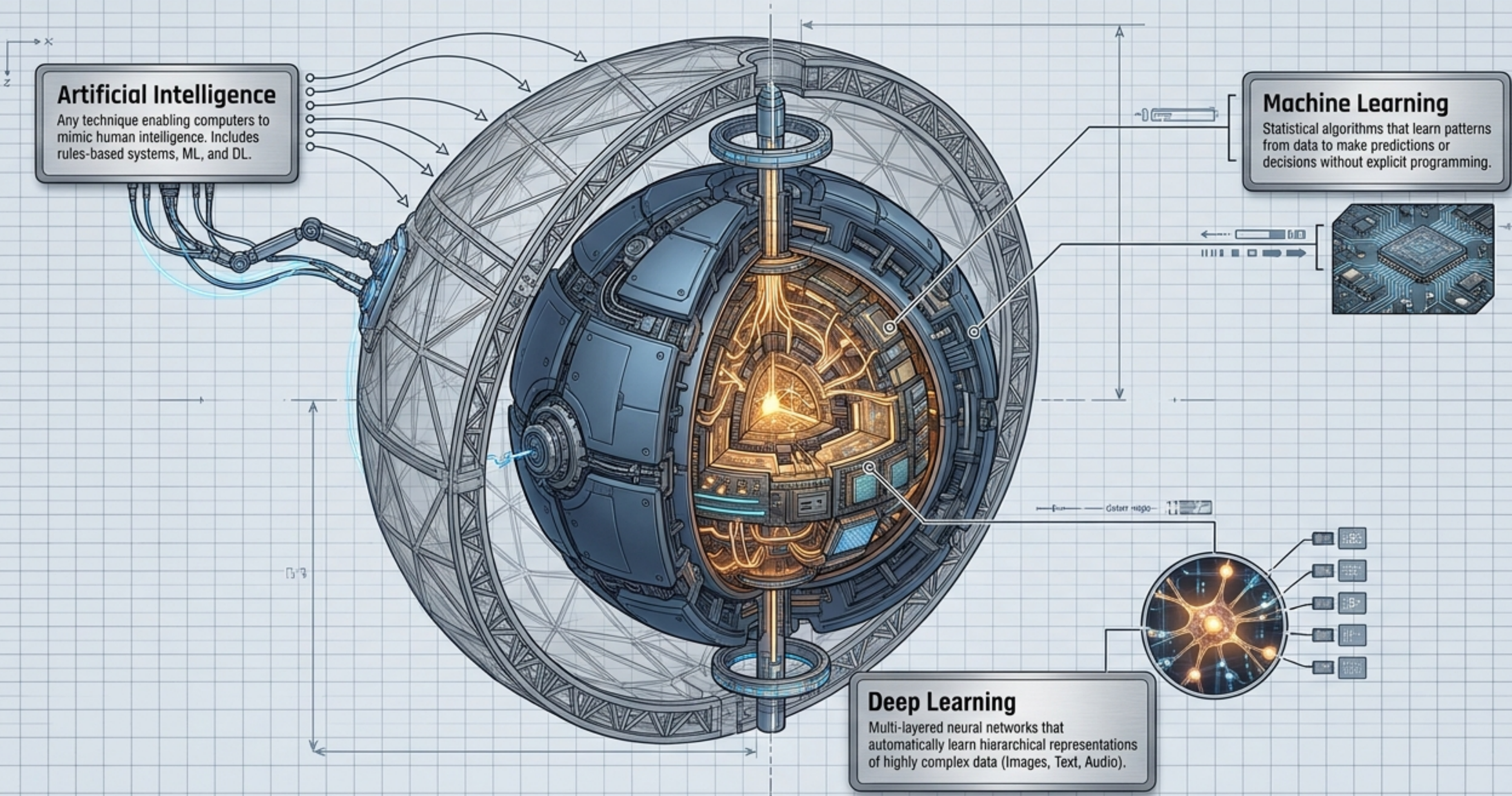




The Machine Learning Algorithm Atlas

A visual guide to the models, architectures, and paradigms powering modern AI.



SUPERVISED LEARNING



Mechanism
Learns from labeled data (input-output pairs).
Direct feedback loop.

Goal
Predict specific outcomes (Classification & Regression).

UNSUPERVISED LEARNING



Mechanism
Learns from unlabeled data.
No explicit feedback or predefined answers.

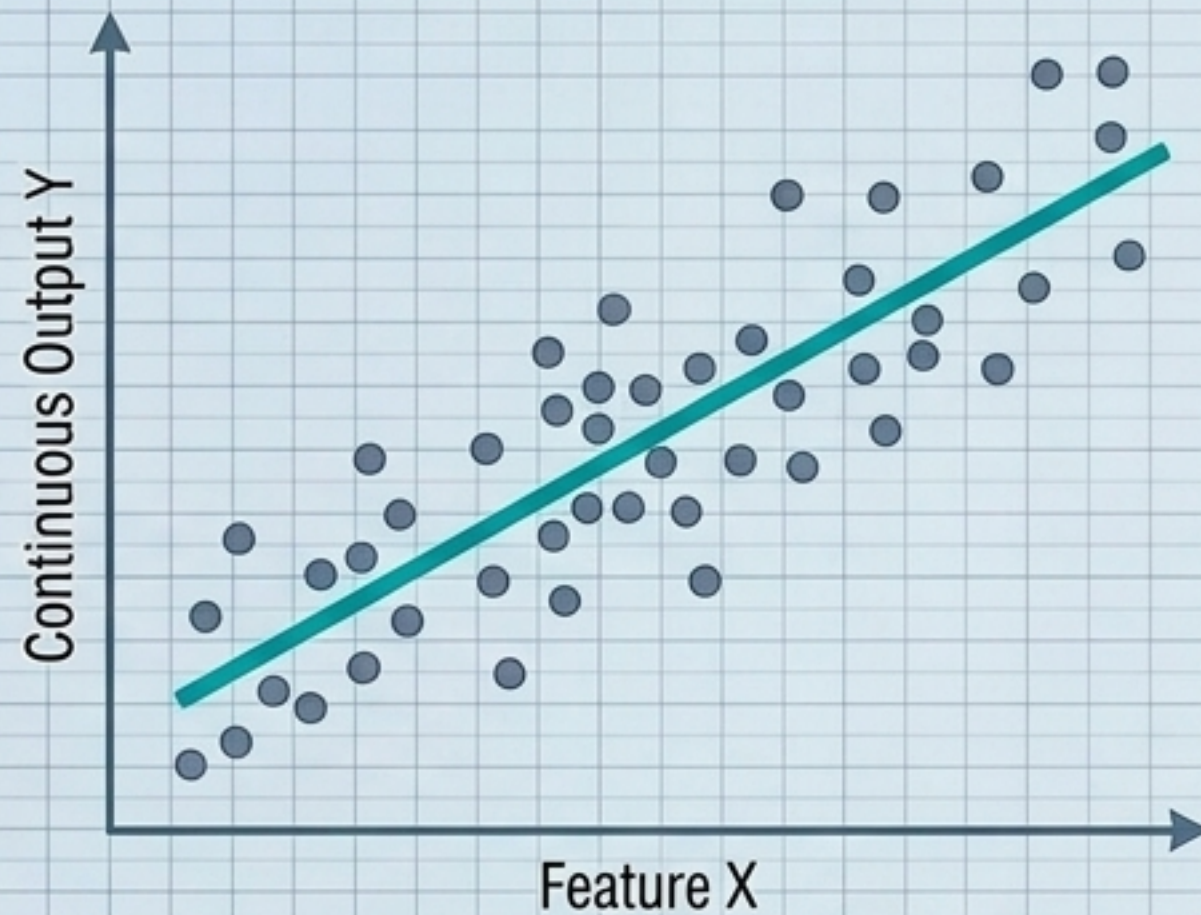
Goal
Discover hidden patterns and underlying structures (Clustering & Anomaly Detection).

REINFORCEMENT LEARNING



Mechanism
Learns via environment interactions. Delayed feedback through rewards/penalties.

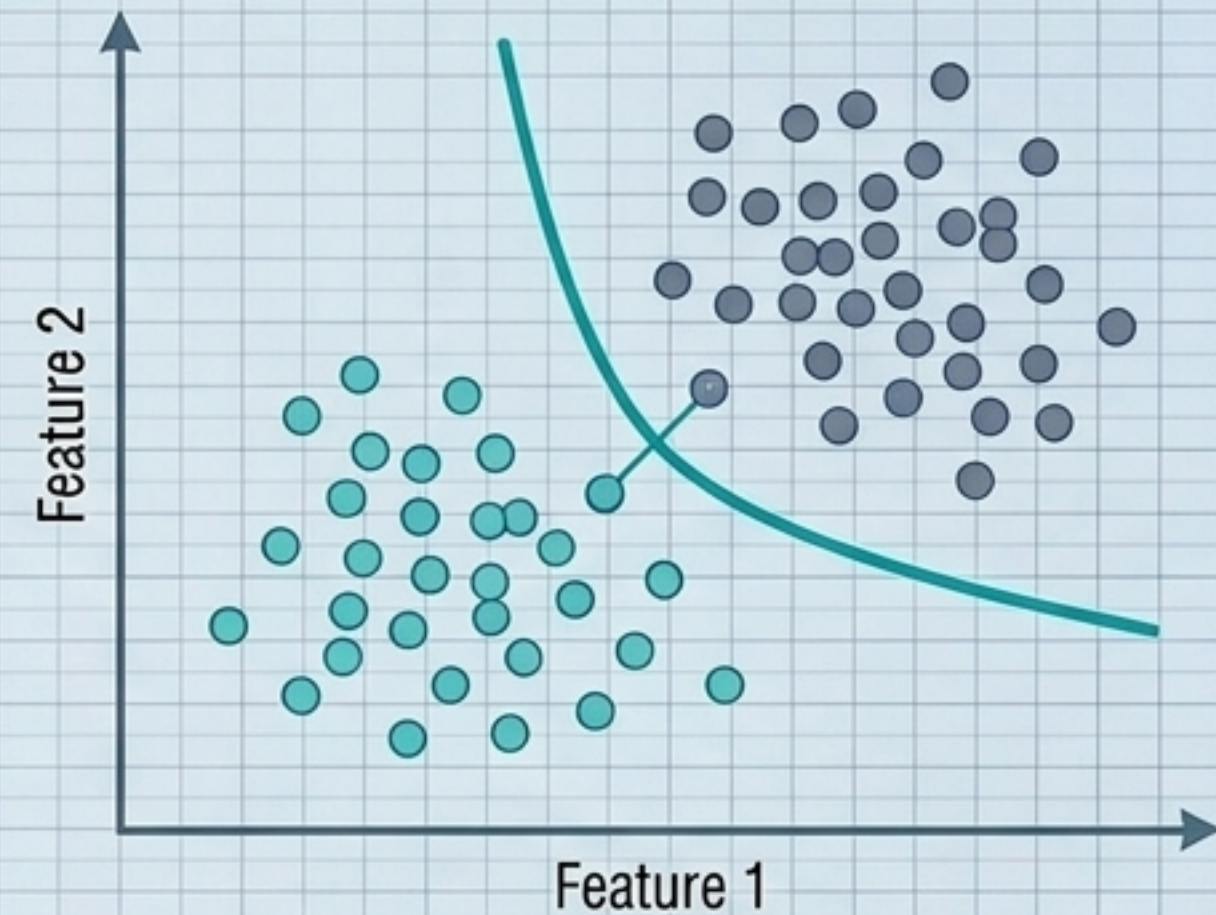
Goal
Formulate a sequence of optimal decisions (Game AI, Robotics).

REGRESSION (Predicting Quantities)**DEFINITION**

Outputs a continuous numerical value.

APPLICATION

Forecasting temperature, predicting house prices based on square footage, projecting future sales.

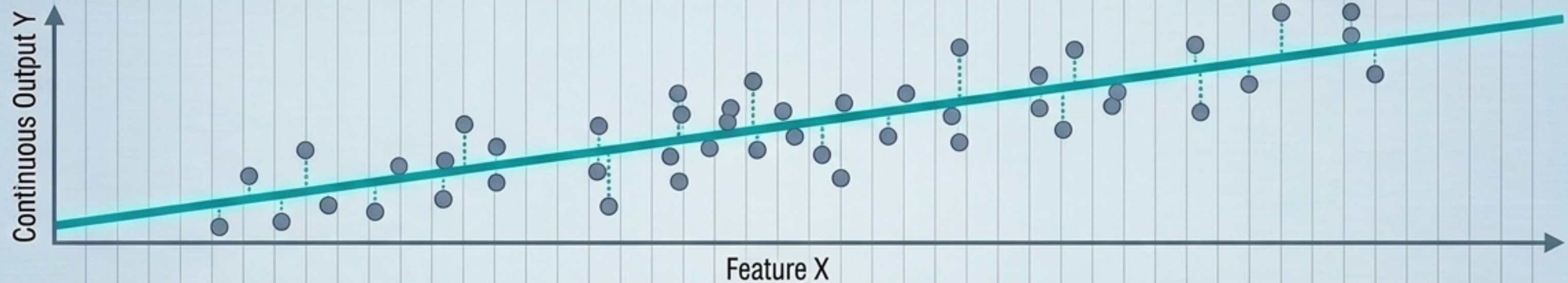
CLASSIFICATION (Predicting Categories)**DEFINITION**

Outputs a discrete categorical label.

APPLICATION

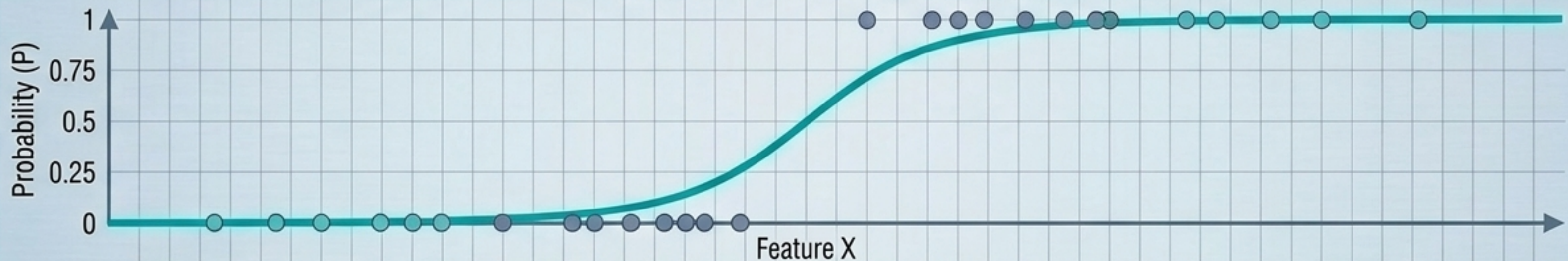
Diagnosing disease vs. no disease, flagging spam vs. not spam, image recognition (cat vs. dog).

LINEAR REGRESSION

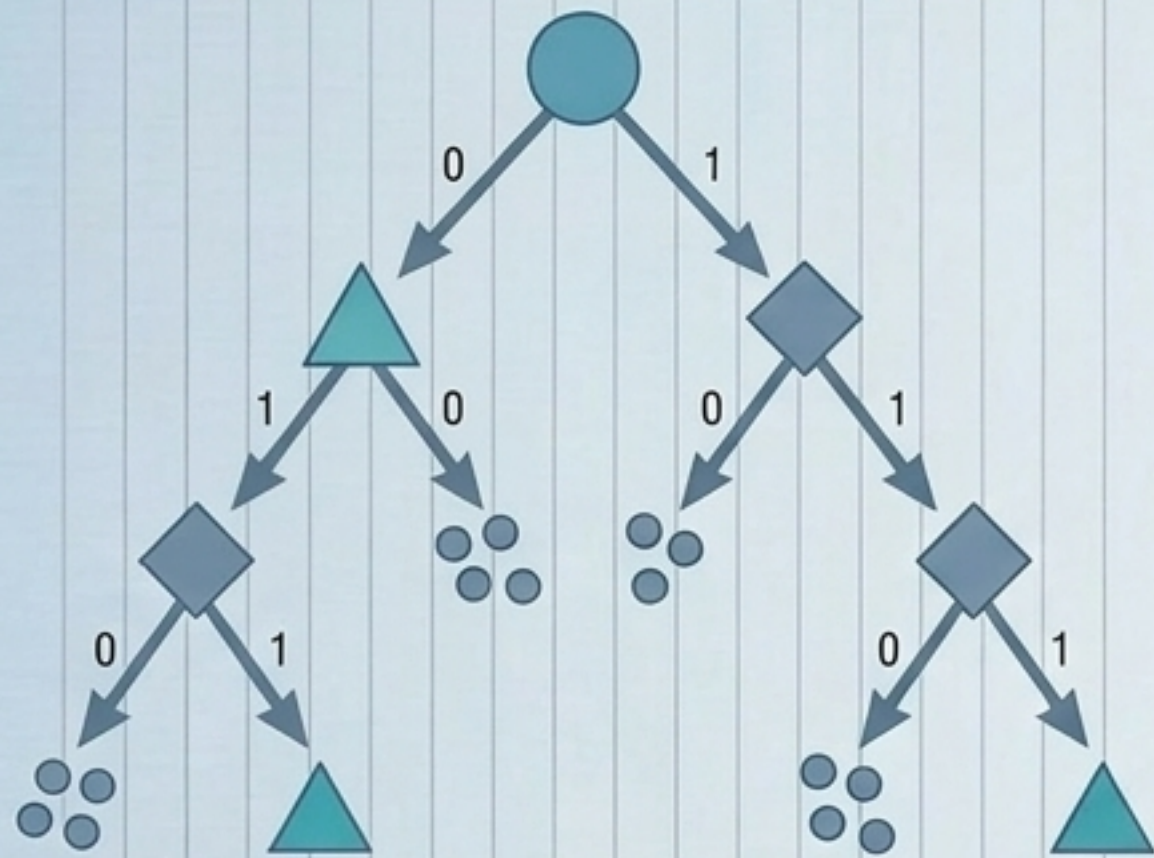


Fits a linear equation to the data. Assumes a constant straight-line relationship between inputs and outputs. Fast, simple, but fails with non-linear data.

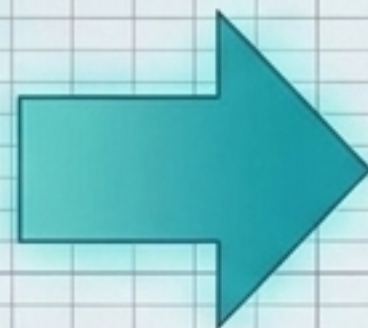
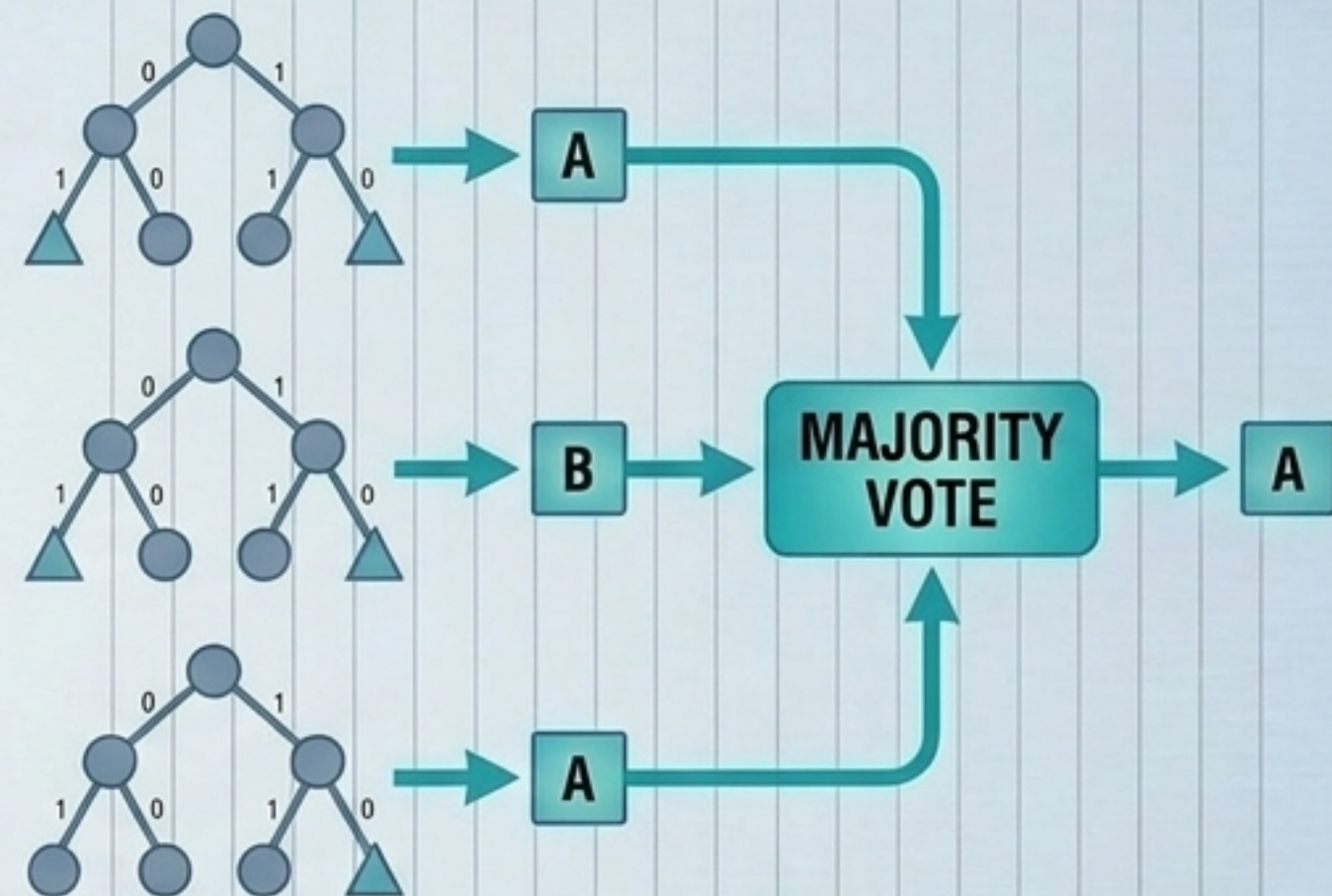
LOGISTIC REGRESSION



Estimates the log-odds of a binary outcome. Highly interpretable, fast scoring, ideal for binary choices (e.g., Spam vs. Inbox).

DECISION TREES

Recursively splits data based on feature values to maximize separation. Trade-off: Requires minimal data prep, but highly prone to overfitting.

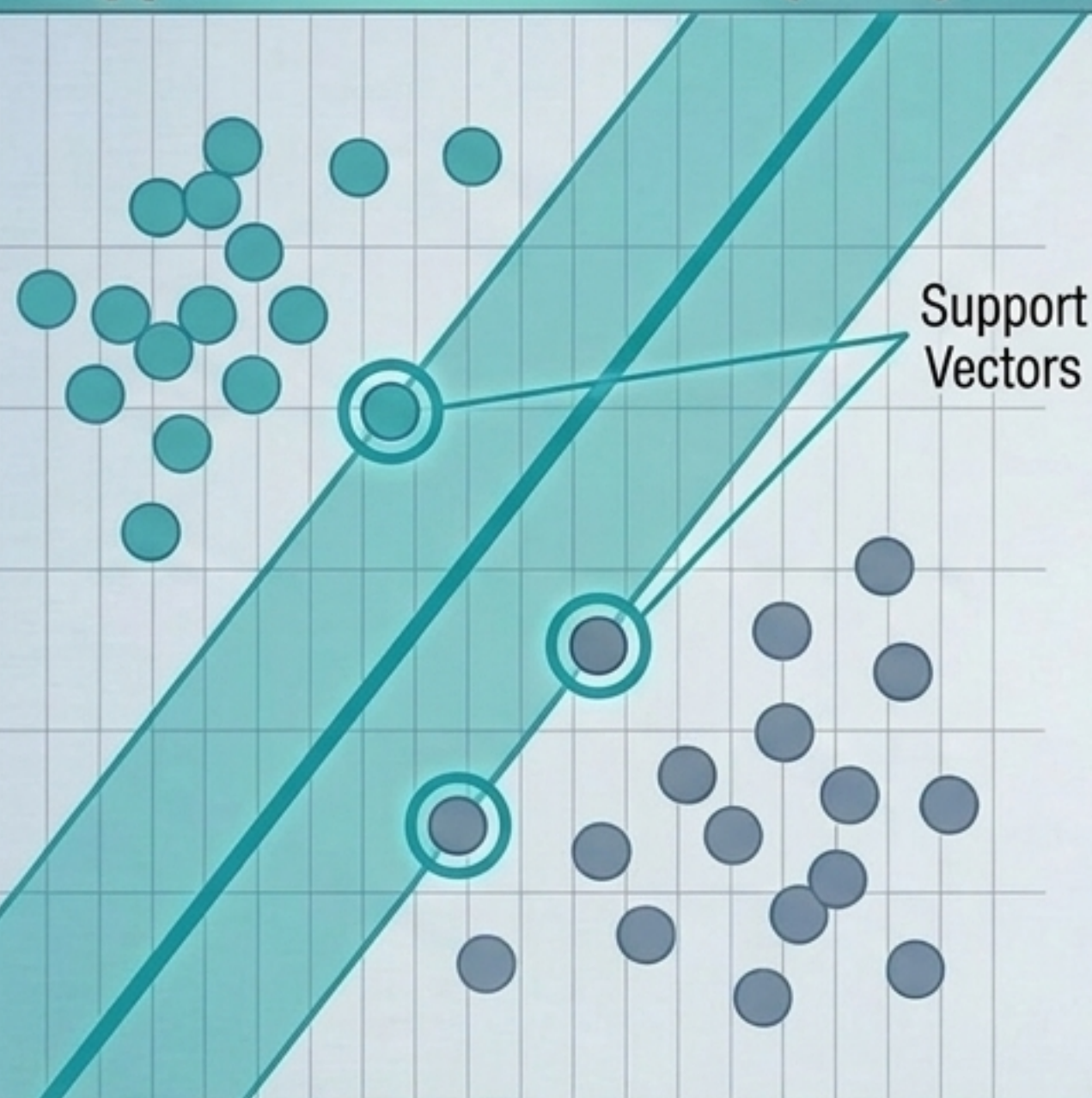
**RANDOM FORESTS (Bagged Ensemble)**

Builds multiple independent trees on random data subsets. Final output is averaged or voted on. Trade-off: High accuracy, but creates a "black box".

GRADIENT BOOSTING (Sequential Ensemble)

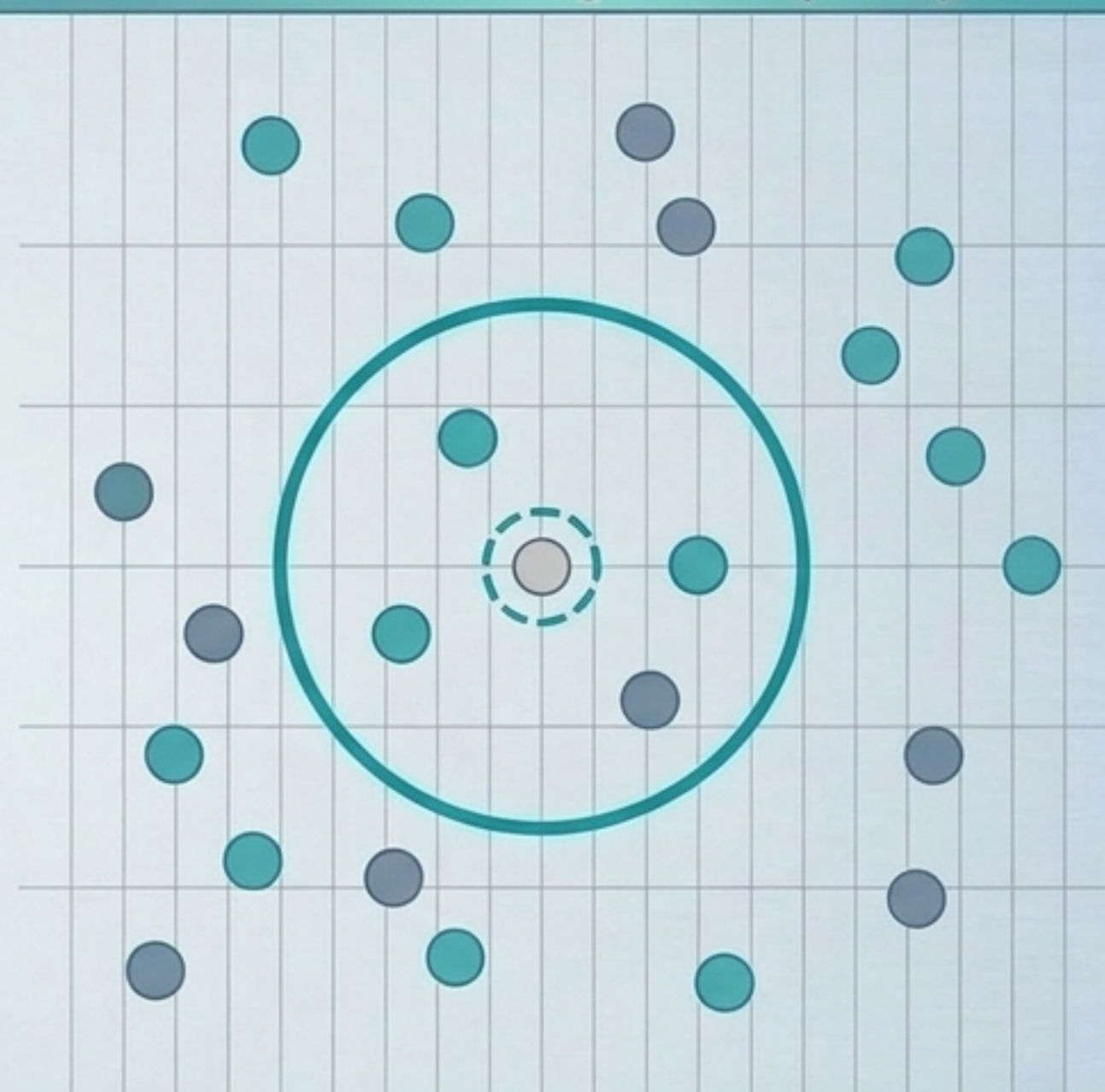
Builds a series of weak trees sequentially, where each new tree attempts to correct the errors of the previous one to minimize loss.

Support Vector Machines (SVM)



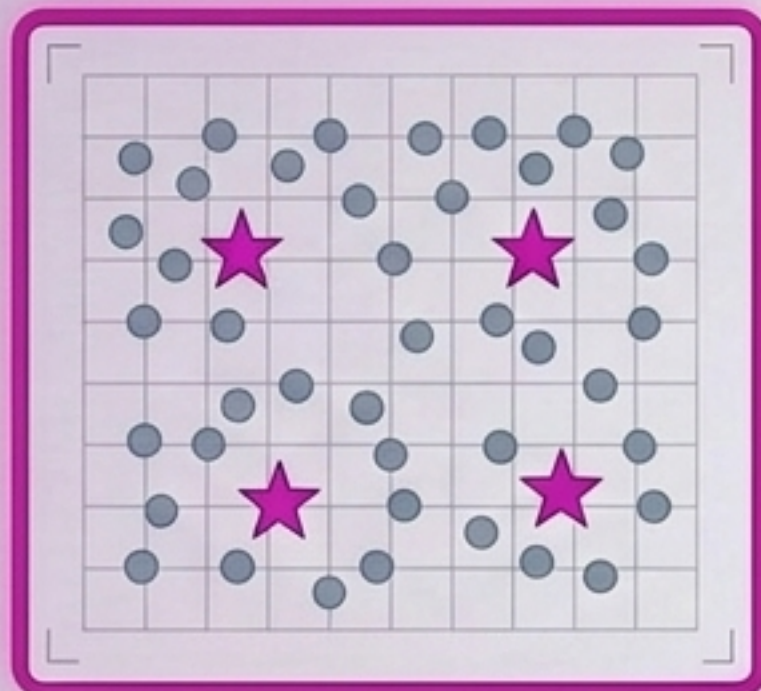
Maximizes the margin between classes using support vectors. Can map data into high-dimensional space for non-linear separation via the 'kernel trick'.

K-Nearest Neighbors (KNN)

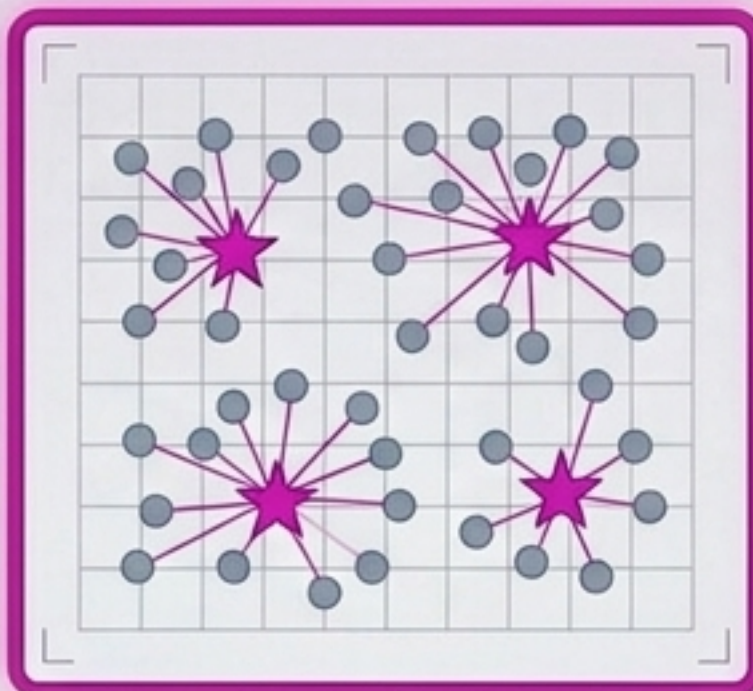
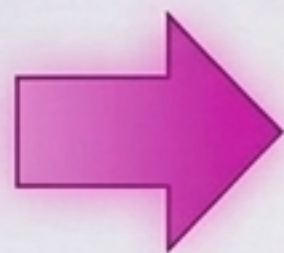


A distance-based, instance-learning algorithm. Classifies new points based on the majority label of its 'k' closest neighbors. Simple, but computationally heavy at inference.

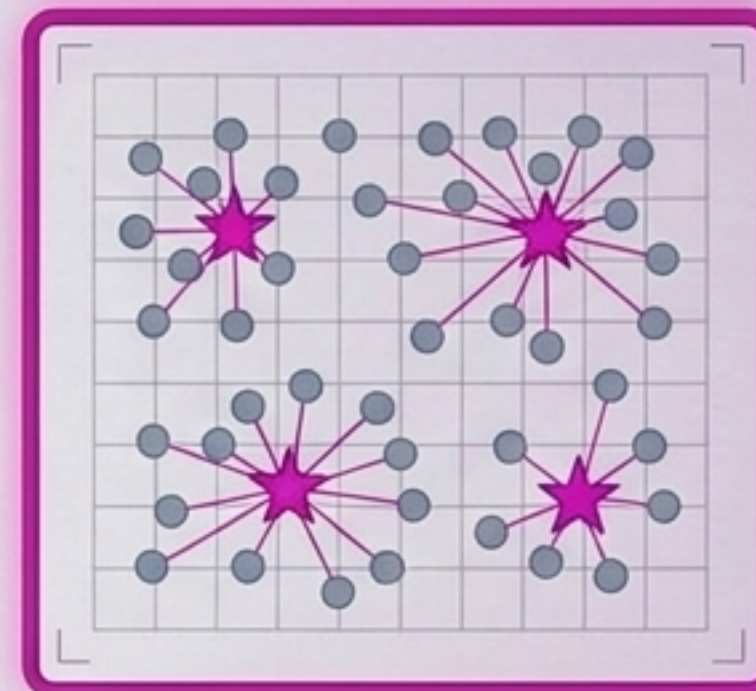
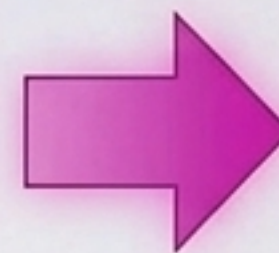
THE K-MEANS FLOW



Initialize: Randomly place k centroids.



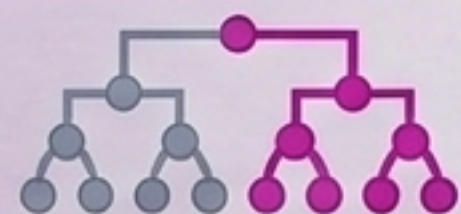
Assign: Associate each data point to its nearest centroid.



Update: Recalculate centroids to be the mean of their assigned points.

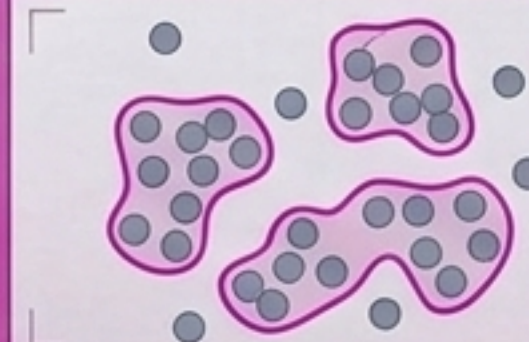
K-Means: Partitions data into k predefined clusters by minimizing intra-cluster distance to the centroid.

ALTERNATIVE ARCHITECTURES

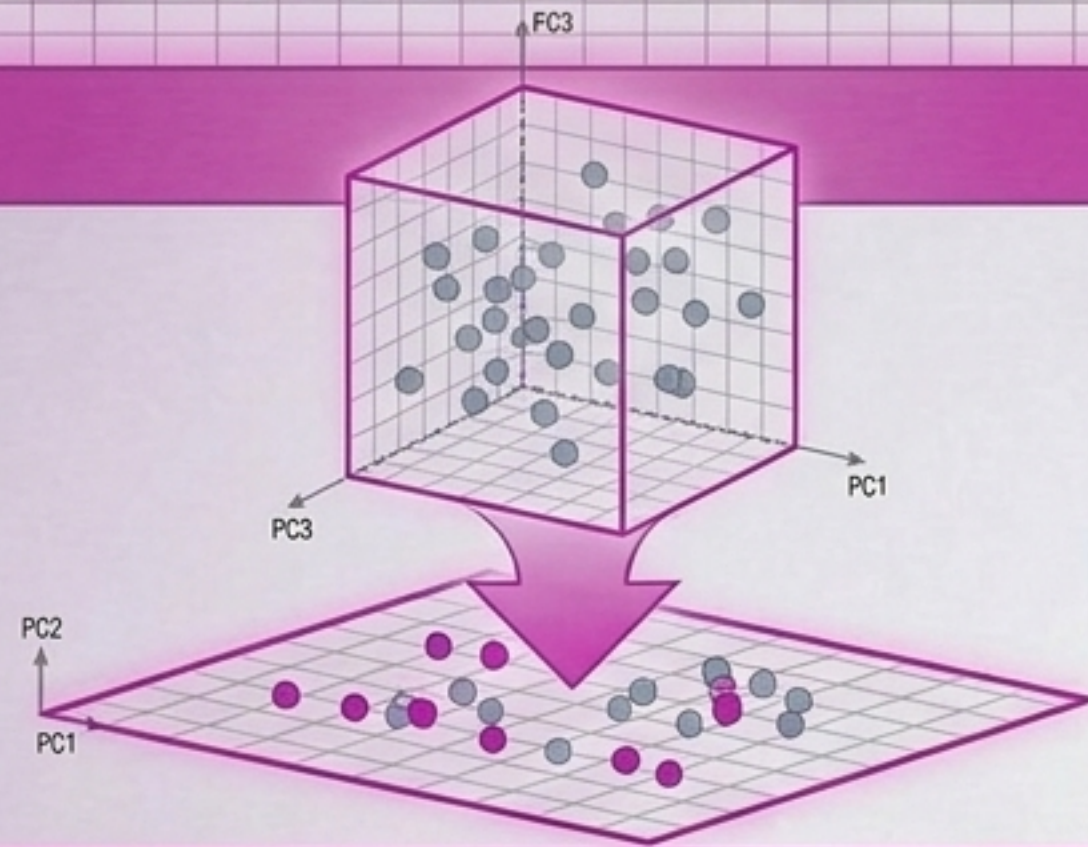
**Hierarchical Clustering**

Builds a nested tree (dendrogram) by iteratively merging the closest clusters. No need to pre-define k .

ALTERNATIVE ARCHITECTURES

**DBSCAN**

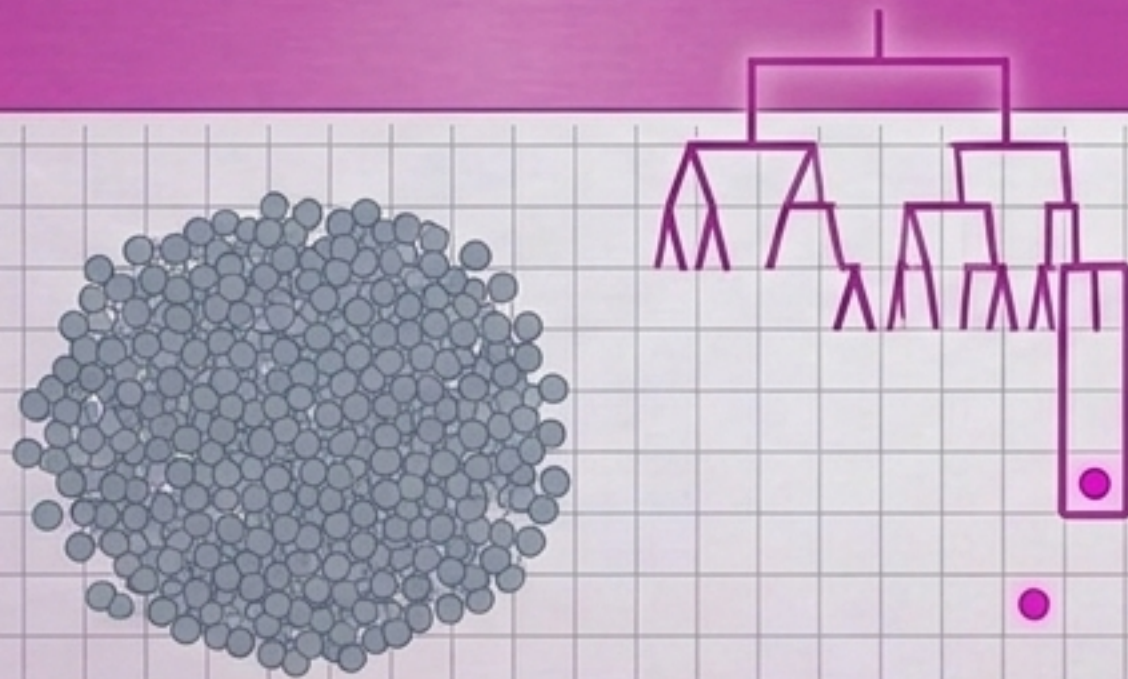
Density-based clustering. Finds irregular shapes and isolates sparse data points as noise.



Principal Component Analysis (PCA)

Transforms data into a new coordinate system. Reduces feature dimensionality by finding the axes (principal components) that explain the maximum variance.

Use Case: Noise reduction, image compression.

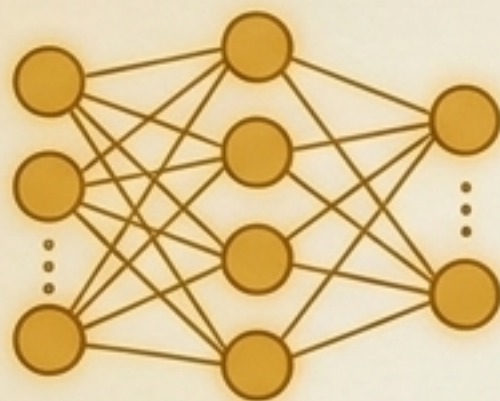


Isolation Forests

A tree-based algorithm that isolates outliers by randomly partitioning data. Anomalies exist in sparse regions, meaning they require significantly fewer splits to isolate than normal points.

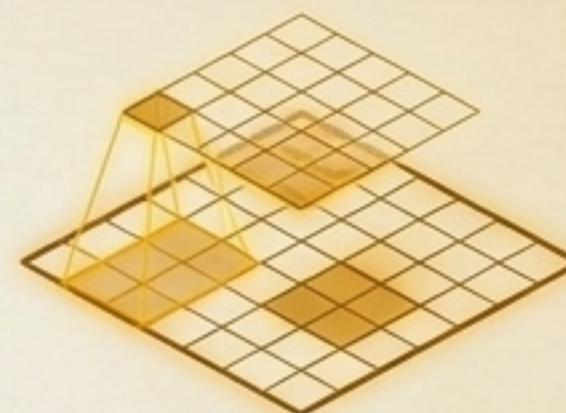
Use Case: Fraud detection.

Multi-Layer Perceptrons (MLPs)



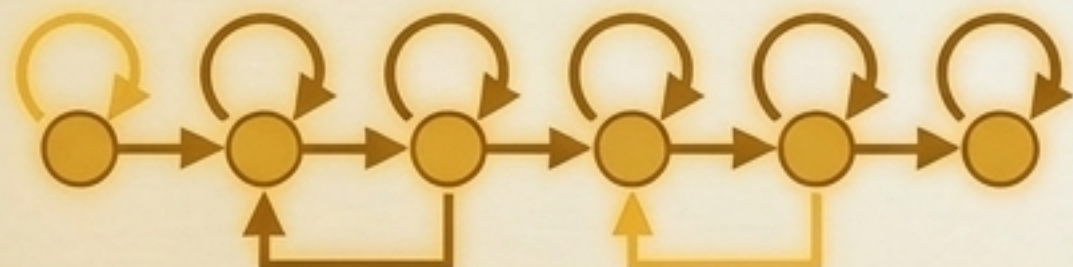
The foundation. Uses activation functions (ReLU, Sigmoid) and error backpropagation.

Convolutional Neural Networks (CNNs)



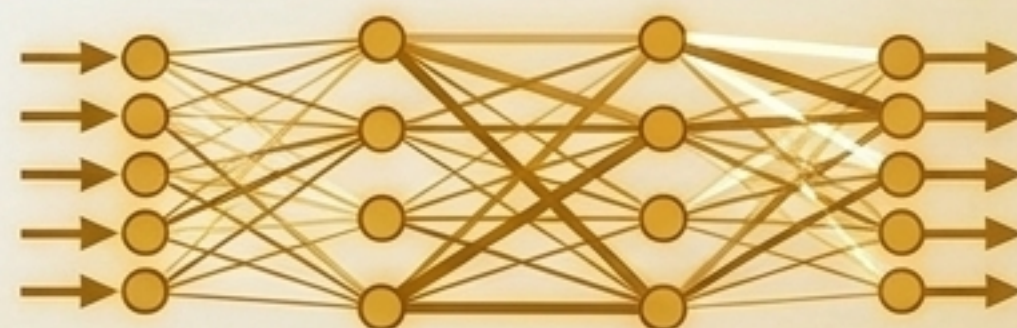
Processes grid-like spatial data via convolution.
Best for: Image & Video Classification.

Recurrent Neural Networks (RNNs/LSTMs)



Processes sequential data by maintaining a hidden state (memory) across time steps.
Best for: Time-Series, Speech Recognition.

Transformers



Replaced sequential processing with parallel self-attention mechanisms.
Best for: Generative AI, LLMs, Complex NLP.

Tokenization

The model learns

[The] [mod] [##e] [learns]

Breaking text into manageable units (Words, Subwords, or Characters). This bridges the gap between human language and machine limits.

Embeddings

[The]

[mod]

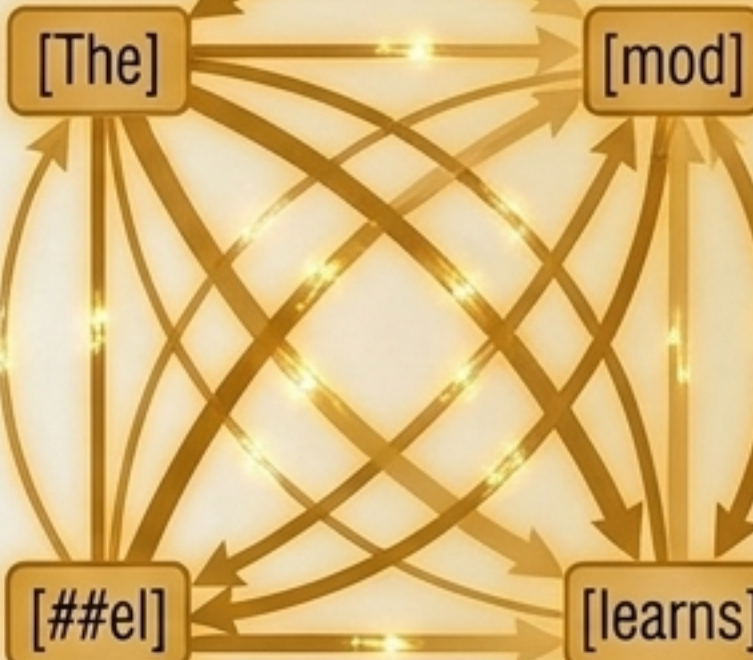
[##e]

[learns]

0.0023	6.09	0.21	0.0000	0.90	...	0.3700	0.36
0.0168			0.0000			0.0107	
0.0055			0.0000			0.0000	
0.0007			0.0000			0.0000	
0.0016	6.01	0.25	0.0000	6.29	...	0.0000	0.25
0.0011			0.0000			0.0000	
0.0000			0.0000			0.0000	
0.0000			0.0000			0.0000	
0.0000	6.31	0.30	0.0000	6.33	...	0.0000	0.47
0.0002			0.0000			0.0000	
0.0000			0.0000			0.0000	
0.0000	0.35	0.29	0.0000	0.18	...	0.0000	0.28
0.0000			0.0000			0.0000	
0.0000			0.0000			0.0000	
0.0027			0.0000			0.0000	
0.0000	0.46	0.29	0.0000	6.50	...	0.0000	0.29
0.0000			0.0000			0.0000	
0.0000			0.0000			0.0000	
0.0002	0.18	0.20	0.0022	0.87	...	0.0000	0.32
0.0000			0.0000			0.0000	
0.0000			0.0000			0.0000	
0.0000	6.31	0.20	0.0000	6.50	...	0.0000	0.23
0.0000			0.0000			0.0000	
0.0000			0.0000			0.0000	
0.0000	0.91	0.53	0.4300	0.56	...	0.0000	0.74
0.0000			0.0000			0.0000	
0.0000			0.0000			0.0000	
0.0000	6.31	0.36	0.0016	0.93	...	0.0000	0.54
0.0000			0.0000			0.0000	
0.0000			0.0000			0.0000	
0.0000	0.31	0.23	0.0000	0.35	...	0.0000	0.72
0.0000			0.0000			0.0000	
0.0000			0.0000			0.0000	
0.0000	6.01	0.20	0.0000	0.39	...	0.0000	0.22
0.0000			0.0000			0.0000	
0.0000			0.0000			0.0000	
0.0000	0.46	0.59	0.0000	0.93	...	0.0000	0.54
0.0000			0.0000			0.0000	
0.0000			0.0000			0.0000	
0.0000	8.11	0.00	0.0000	0.00	...	0.0000	0.11

Mapping tokens to unique numerical indices that capture semantic meaning. Words with similar meanings cluster together mathematically.

Self-Attention Mechanism



The core of the Transformer. The model computes scores to determine the importance of every other word in a sequence relative to the target word, capturing deep context.

THE HOLISTIC ML LIFECYCLE

16:9





Classification Metrics (Categorical)



- **Accuracy:** Overall correctness.



- **Precision & Recall:** Balancing false positives and false negatives.

- **Confusion Matrix:** Visual grid of True vs. Predicted classes.

	T	P
T	True	False Ps
N	False Ps	Neg.



Regression Metrics (Continuous)

- **Mean Squared Error (MSE):** Measures average squared difference between estimated and actual values.



The Golden Rule: Optimize for Generalization



Overfitting

Model memorizes training data but fails on new data



Generalization

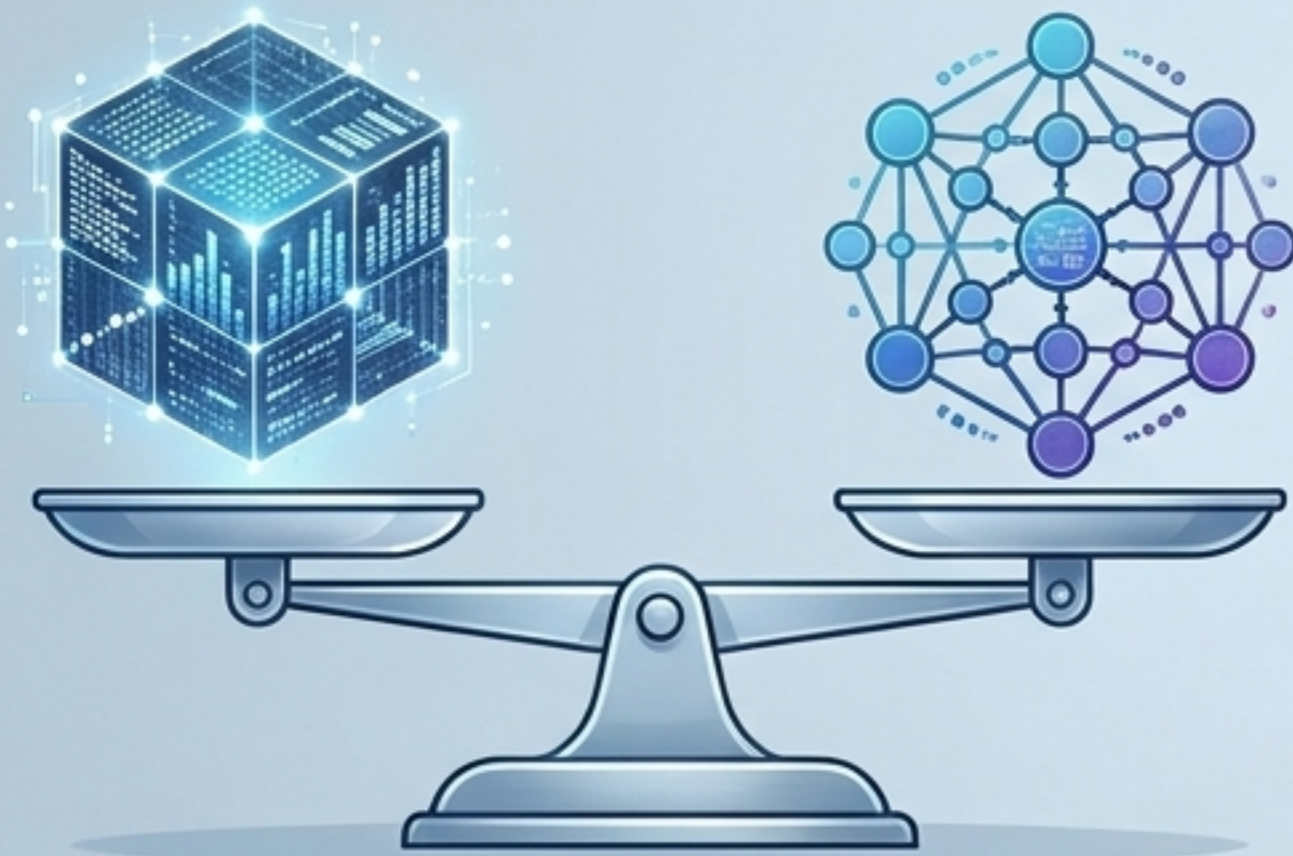
Captures underlying patterns accurately for new, unseen data

Avoid Underfitting (too simple) and Overfitting (too complex).

ALGORITHM CHEAT SHEET: DECISION MATRIX

Algorithm	Type	Best Use Case	Core Pro	Core Con
Linear Regression	Supervised	Predict continuous	Simple	Sensitive to outliers
Logistic Regression	Supervised	Binary classification	Probabilistic	Weak on non-linear
Random Forest	Supervised	Ensemble accuracy	Robust	Slower, less interpretable
SVM	Supervised	Max-margin classification	High dimensions	Slow on large data
KNN	Supervised	Few-shot classification	No training phase	Slow inference
K-Means	Unsupervised	Customer segmentation	Fast	Sensitive to scale
Transformers	Deep Learning	NLP tasks	Long context	Heavy compute

THE BOTTOM LINE



AI success is not about finding a magic bullet; it relies on matching the right mathematical algorithm to the right business data structure.

AI ASSISTANT TOOLKIT

LEVERAGING GENAI ASSISTANTS

When coding models or cleaning data with tools like Copilot, use structured prompting frameworks to ensure precision:

RTFC (Best for Data/Coding)

Role, Task, Format, Constraints.

Highly direct, eliminates conversational filler.

APE (Best for Strategy/Logic)

Action, Purpose, Expectation.

Aligns AI reasoning with broader business objectives.