

Machine Learning Algorithms

A Visual Framework · Classification · Domains · Use Cases · Explanations

Dennis Kankam Hagan · AIML501-01AMV · Indiana Wesleyan University · Dr. Eskinder Belete

APPLICATION DOMAINS:

Tabular

Computer Vision

NLP

Generative AI

◆ SUPERVISED LEARNING

6 algorithms

Learns from labeled training data · predicts known output categories

Linear Regression

Tabular

Fits a straight line through training points by minimizing the sum of squared errors (MSE) between predicted and actual values via gradient descent or closed-form solution.

Price prediction · Sales forecasting · Trend analysis

Logistic Regression

Tabular

NLP

Applies a sigmoid function to a linear combination of features to output a class probability; decision boundary is a hyperplane. Trained by maximizing log-likelihood (cross-entropy loss).

Spam detection · Disease diagnosis · Sentiment classification

Decision Tree

Tabular

Computer Vision

Recursively partitions the feature space by choosing the split that minimizes Gini impurity or entropy. Leaf nodes return the majority class; depth is controlled via pruning.

Classification · Credit scoring · Medical diagnosis

Random Forest

Tabular

Computer Vision

Trains an ensemble of decision trees on bootstrap samples with random feature subsets. Final prediction is the majority vote (classification) or mean (regression) across all trees.

Image detection · Remote sensing · Feature importance ranking

Support Vector Machine

Tabular

NLP

Computer Vision

Finds the maximum-margin hyperplane separating classes. Kernel functions (RBF, polynomial) implicitly map inputs to higher-dimensional spaces to handle non-linear decision boundaries.

Image classification · Text categorization · Bioinformatics

Neural Network

Computer Vision

NLP

Generative AI

Stacked layers of weighted neurons apply non-linear activation functions (ReLU, Sigmoid). Backpropagation and Adam optimizer iteratively reduce loss across thousands of training iterations.

Image recognition · Language translation · Speech recognition

◆ UNSUPERVISED LEARNING

4 algorithms

No labels required · discovers hidden structure and natural groupings

K-Means Clustering

Tabular

Randomly initializes k centroids, assigns each point to the nearest centroid by Euclidean distance, then recomputes centroids. Iterates until assignments no longer change (convergence).

Image segmentation · Document grouping · Market analysis

Principal Component Analysis

Tabular

Computer Vision

Computes eigenvectors of the covariance matrix; projects data onto the top-k orthogonal axes of maximum variance. Reduces dimensionality while preserving the most information.

Face recognition · Gene expression · Noise reduction

DBSCAN

Tabular

Labels a point as a core point if its ϵ -neighborhood contains at least MinPts neighbors; expands clusters by density-reachability. Points in sparse regions are marked as noise (outliers).

Image segmentation · Network intrusion detection · Social graphs

Generative Adversarial Network

Computer Vision

Generative AI

A generator network synthesizes fake samples from random noise; a discriminator classifies real vs. fake. Adversarial training drives the generator to produce increasingly realistic outputs.

Image generation · Text generation · Style transfer

Both rely on
Feature Engineering
& Data Preprocessing

Shared domains:
Tabular · Computer Vision
NLP · Generative AI

REFLECTION ON LEARNING

Classifying machine learning algorithms reveals a fundamental insight: the availability of labeled data shapes every design decision. Supervised algorithms—from Linear Regression to Neural Networks—learn mappings between inputs and known outputs, making them precise but dependent on annotation effort. Unsupervised algorithms instead uncover latent structure—clusters, principal axes, density regions—without any ground truth. Application domain imposes a second constraint: convolutional networks excel at spatial image patterns; transformers capture long-range sequential dependencies in text; GANs model full data distributions for generation. Building this framework has deepened my appreciation that algorithm selection is as much an art of matching inductive bias to data geometry as it is about maximizing a performance metric.

Supervised · Linear Regression · Logistic Regression · Decision Tree · Random Forest · SVM · Neural Network

Unsupervised · K-Means Clustering · Principal Component Analysis · DBSCAN · Generative Adversarial Network