ML Cheatsheet

✅ 1. Supervised Learning
• Regression
o Linear Regression
o Logistic Regression
o Polynomial Regression
o Ridge Regression
o Lasso Regression
o ElasticNet
o Support Vector Machines (SVM)
o Decision Trees
o Random Forest
• Classification
o Logistic Regression
o K-Nearest Neighbors (KNN)
o Support Vector Machines (SVM)
o Decision Trees
o Random Forest
o Naive Bayes
o Confusion Matrix
o Stochastic Gradient Descent
o Gradient Boosting
o AdaBoost
o XGBoost
o LightGBM
o CatBoost

____

2. Unsupervised Learning
• Clustering
1. Centroid-Based Clustering
• K-Means
• K-Medoids
• Mean-Shift

____

2. Density-Based Clustering
• DBSCAN
• OPTICS
• HDBSCAN

____

3. Hierarchical Clustering
• Agglomerative Clustering
• BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)
• Affinity Propagation

____

4. Distribution-Based Clustering
• Gaussian Mixture Models (GMM)
• Dimensionality Reduction
o PCA (Principal Component Analysis)
o t-SNE
o UMAP
o ICA (Independent Component Analysis)
o LDA (Linear Discriminant Analysis)

____

3. Semi-Supervised Learning
• Self-Training
• Label Propagation
• Label Spreading

____

4. Reinforcement Learning
• Q-Learning
• Deep Q-Networks (DQN)
• SARSA
• Policy Gradient Methods
• Actor-Critic
• Proximal Policy Optimization (PPO)
• Deep Deterministic Policy Gradient (DDPG)

____

5. Deep Learning Algorithms
1. Feedforward Networks (FNN)
• Multilayer Perceptron (MLP)
• Deep Neural Networks (DNN)

____

2. Convolutional Neural Networks (CNN)
• LeNet
• AlexNet
• VGGNet
• GoogLeNet (Inception)
• ResNet
• DenseNet
• EfficientNet
• MobileNet
• SqueezeNet

____

3. Recurrent Neural Networks (RNN)
• Vanilla RNN
• Long Short-Term Memory (LSTM)
• Gated Recurrent Unit (GRU)

• Bidirectional RNN
• Deep RNNs
• Echo State Networks (ESN)

---

 4. Attention-Based Models / Transformers
• Transformer
• BERT
• GPT (GPT-1, GPT-2, GPT-3, GPT-4)
• RoBERTa
• ALBERT
• XLNet
• T5
• DistilBERT
• Vision Transformer (ViT)
• Swin Transformer
• DeiT
• Performer
• Longformer

---

 5. Autoencoders
• Vanilla Autoencoder
• Sparse Autoencoder
• Denoising Autoencoder
• Contractive Autoencoder
• Variational Autoencoder (VAE)

---

 6. Generative Adversarial Networks (GANs)
• Vanilla GAN
• Deep Convolutional GAN (DCGAN)
• Conditional GAN (cGAN)
• CycleGAN
• StyleGAN
• Pix2Pix
• BigGAN
• StarGAN
• WGAN (Wasserstein GAN)
• WGAN-GP

---

 7. Reinforcement Learning (Deep RL)
• Deep Q-Network (DQN)
• Double DQN
• Dueling DQN
• Policy Gradient
• REINFORCE
• Actor-Critic
• A3C (Asynchronous Advantage Actor-Critic)
• PPO (Proximal Policy Optimization)
• DDPG (Deep Deterministic Policy Gradient)
• TD3 (Twin Delayed DDPG)
• SAC (Soft Actor-Critic)

Supervised Learning

Evaluation Metrics

Confusion Matrix: A table that describes the performance of a classification model.

- **Key Metrics:** True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN).

Accuracy: (TP + TN) / Total

- **Limitation:** Can be misleading with imbalanced datasets.

Precision: TP / (TP + FP)

- **Meaning:** What proportion of positive identifications was actually correct?

Recall (Sensitivity): TP / (TP + FN)

- **Meaning:** What proportion of actual positives was identified correctly?

F1-Score: 2 * (Precision * Recall) / (Precision + Recall)

- **Meaning:** Harmonic mean of precision and recall.

AUC-ROC: Area Under the Receiver Operating Characteristic curve.

- **Meaning:** Measures the ability of a classifier to distinguish between classes.

Regression Algorithms

Linear Regression: Models the relationship between variables by fitting a linear equation to observed data.

- **Use Case:** Predicting housing prices based on size and location.

Logistic Regression: Predicts the probability of a categorical outcome.

- **Use Case:** Predicting whether an email is spam or not.

Polynomial Regression: Models non-linear relationships using polynomial functions.

- **Use Case:** Modeling growth rates where the increase accelerates over time.

Ridge Regression: Linear regression with L2 regularization to prevent overfitting.

- **Use Case:** When dealing with multicollinearity.

Lasso Regression: Linear regression with L1 regularization, performs feature selection.

- **Use Case:** When many features are irrelevant.

Elastic Net: Combines L1 and L2 regularization.

- **Use Case:** Combines the benefits of both Ridge and Lasso.

Support Vector Regression (SVR),Random Forest,Decision Trees

Classification Algorithms

Logistic Regression: (Also used for classification) Predicts probability of a binary outcome.

- **Use Case:** Predicting disease presence.

K-Nearest Neighbors (KNN): Classifies based on the majority class among its k nearest neighbors.

- **Use Case:** Recommendation systems.

Support Vector Machines (SVM): Finds an optimal hyperplane to separate classes.

- **Use Case:** Image classification.

Decision Trees: (Also used for classification) Splits data into subsets based on feature values.

- **Use Case:** Risk assessment.

Random Forest: (Also used for classification) Ensemble of decision trees for improved accuracy.

- **Use Case:** Fraud detection.

Naive Bayes: Applies Bayes' theorem with strong (naive) independence assumptions between features.

- **Use Case:** Text classification.

Stochastic Gradient Descent (SGD): Optimization algorithm used to train linear classifiers under convex loss functions.

- **Use Case:** Large-scale learning problems.

Gradient Boosting: Builds an ensemble of weak learners sequentially, where each learner corrects the errors of its predecessors.

- **Use Case:** Predictive analytics.

AdaBoost: Adaptive Boosting, focuses on correcting mistakes of previous classifiers.

- **Use Case:** Boosting weak classifiers.

XGBoost: Optimized Gradient Boosting implementation.

- **Use Case:** Winning Kaggle competitions.

LightGBM: Gradient Boosting framework that uses tree based learning algorithms.

- **Use Case:** Efficient for large datasets.

CatBoost: Gradient Boosting algorithm that handles categorical features natively.

- **Use Case:** Datasets with many categorical features.

Unsupervised Learning & Dimensionality Reduction

Clustering Algorithms

K-Means: Partitions n observations into k clusters, each with the nearest mean.
- **Use Case:** Customer segmentation.

K-Medoids: Similar to K-Means but chooses data points as cluster centers (medoids).
- **Use Case:** More robust to outliers than K-Means.

Mean-Shift: Locates the maxima of a density function.
- **Use Case:** Image segmentation, object tracking.

DBSCAN: Density-Based Spatial Clustering of Applications with Noise; identifies clusters based on density.
- **Use Case:** Anomaly detection.

OPTICS: Ordering Points To Identify the Clustering Structure; extends DBSCAN to handle varying densities.
- **Use Case:** When clusters have different densities.

HDBSCAN: Hierarchical DBSCAN; combines hierarchical clustering with DBSCAN.
- **Use Case:** Discovering clusters of varying densities and sizes.

Agglomerative Clustering: Bottom-up approach where each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.
- **Use Case:** Document clustering.

BIRCH: Balanced Iterative Reducing and Clustering using Hierarchies; designed for large datasets.
- **Use Case:** Clustering large transactional data.

Affinity Propagation: Creates clusters by sending messages between pairs of samples until convergence.
- **Use Case:** Gene expression data analysis.

Gaussian Mixture Models (GMM): Assumes data is generated from a mixture of Gaussian distributions.
- **Use Case:** Soft clustering where each point belongs to multiple clusters with different probabilities.

Dimensionality Reduction

PCA (Principal Component Analysis): Reduces dimensionality by projecting data onto principal components.
- **Use Case:** Image compression.

t-SNE: Reduces dimensionality while keeping similar instances close and dissimilar instances apart.
- **Use Case:** Visualizing high-dimensional data.

UMAP: Uniform Manifold Approximation and Projection; similar to t-SNE but faster and can preserve more global structure.
- **Use Case:** Exploring relationships in data.

ICA (Independent Component Analysis): Separates multivariate signals into additive subcomponents that are statistically independent.
- **Use Case:** Blind source separation (e.g., separating audio sources).

LDA (Linear Discriminant Analysis): Finds a linear combination of features that characterizes or separates two or more classes of objects or events.
- **Use Case:** Feature extraction for classification.

Semi-Supervised Learning

Self-Training: Train a model on labeled data, predict labels for unlabeled data, add high-confidence predictions to the labeled set, and retrain.
- **Use Case:** When labeled data is scarce.

Label Propagation: Assigns labels to unlabeled data points based on the labels of their neighbors.
- **Use Case:** Recommending articles based on user reading history.

Label Spreading: Similar to label propagation but uses a graph-based approach to spread labels.
- **Use Case:** Image segmentation.

Semi-Supervised & Reinforcement Learning

Reinforcement Learning

Q-Learning: Model-free RL algorithm that learns a Q-function representing the expected reward for taking an action in a state.
- **Use Case:** Game playing.

Deep Q-Networks (DQN): Uses a neural network to approximate the Q-function.
- **Use Case:** Playing Atari games.

SARSA: On-policy RL algorithm that updates the Q-function based on the action actually taken.
- **Use Case:** Robot navigation.

Policy Gradient Methods: Directly optimize the policy function.
- **Use Case:** Continuous control tasks.

Actor-Critic: Combines policy gradient and value-based methods.
- **Use Case:** Autonomous driving.

Proximal Policy Optimization (PPO): Policy gradient method that constrains policy updates.
- **Use Case:** Training robots to walk.

Deep Deterministic Policy Gradient (DDPG): Actor-critic method that handles continuous action spaces.
- **Use Case:** Robotics and control systems.

Convolutional Neural Networks (CNN)

LeNet: Early CNN architecture for handwritten digit recognition.
- **Use Case:** Historical significance in CNN development.

AlexNet: Deeper CNN architecture that won the 2012 ImageNet competition.
- **Use Case:** Image classification.

VGGNet: CNN architecture with very deep layers and small convolutional filters.
- **Use Case:** Image classification and object detection.

GoogLeNet (Inception): Uses inception modules to reduce computational cost and improve performance.
- **Use Case:** Image recognition.

ResNet: Introduces residual connections to train very deep networks.
- **Use Case:** Image classification and object detection.

DenseNet: Connects each layer to every other layer in a feed-forward fashion.
- **Use Case:** Image recognition.

EfficientNet: Balances network depth, width, and resolution.
- **Use Case:** Efficient image recognition.

MobileNet: Designed for mobile and embedded devices.
- **Use Case:** Mobile vision applications.

SqueezeNet: Achieves AlexNet-level accuracy with fewer parameters.
- **Use Case:** Low-power devices.

Deep Learning Algorithms

Feedforward Networks

Deep Neural Networks (DNN): Neural networks with multiple hidden layers.
- **Use Case:** Complex pattern recognition.

Multilayer Perceptron (MLP): Basic feedforward neural network with one or more hidden layers.
- **Use Case:** Tabular data classification.

Recurrent Neural Networks (RNN)

Vanilla RNN: Basic RNN architecture for sequential data.
- **Use Case:** Language modeling.

Long Short-Term Memory (LSTM): RNN architecture with memory cells to capture long-range dependencies.
- **Use Case:** Machine translation.

Gated Recurrent Unit (GRU): Simplified version of LSTM.
- **Use Case:** Time series prediction.

Bidirectional RNN: Processes sequences in both forward and backward directions.
- **Use Case:** Sentiment analysis.

Deep RNNs: RNNs with multiple layers.
- **Use Case:** Complex sequence modeling tasks.

Echo State Networks (ESN): RNN with a sparsely connected hidden layer.
- **Use Case:** Time series forecasting.