ML Cheatsheet

A comprehensive cheat sheet covering various machine learning algorithms, including supervised, unsupervised, semi-supervised, and reinforcement learning, along with deep learning architectures.

Supervised Learning

Regression Algorithms

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

Example: Predicting house prices based on square footage.

Logistic Regression: Predicts the probability of a binary outcome.

Example: Predicting whether an email is spam or not.

Polynomial Regression: Models non-linear relationships by fitting a polynomial equation to the data.

Example: Modeling growth rates that increase over time.

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

Use Case: When multicollinearity is present among the features.

Lasso Regression: Linear regression with L1 regularization, which can perform feature selection.

Use Case: Situations with many features, some of which are irrelevant.

ElasticNet: Combines L1 and L2 regularization.

Use Case: When you need both regularization and feature selection.

Support Vector Machines (SVM): Effective in high dimensional spaces.

Example: Regression tasks with complex, non-linear relationships.

Decision Trees: Tree-like model that makes decisions based on features.

Example: Predicting customer churn based on various attributes.

Random Forest: Ensemble of decision trees.

Example: Improving prediction accuracy and reducing overfitting in regression tasks.

Classification Algorithms

Logistic Regression: Predicts the probability of a binary outcome, classifying data points into one of two categories.

Example: Classifying patients as having a disease or not.

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

Example: Image recognition tasks.

Support Vector Machines (SVM): Finds the optimal hyperplane to separate classes in high-dimensional space.

Example: Text categorization.

Decision Trees: Classifies data by recursively splitting the data based on feature values.

Example: Credit risk assessment.

Random Forest: Ensemble of decision trees to improve classification accuracy and reduce overfitting.

Example: Complex classification tasks with many features.

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

Example: Spam filtering.

Stochastic Gradient Descent: Optimization algorithm used to train linear classifiers.

Example: Large-scale classification problems.

Gradient Boosting: Builds an ensemble of weak learners (usually decision trees) sequentially, with each tree correcting errors of the previous ones.

Example: Fraud detection.

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

Example: Face detection.

XGBoost, LightGBM, CatBoost: Advanced gradient boosting algorithms known for their efficiency and accuracy.

Example: Used widely in competitive machine learning.

Unsupervised Learning

Clustering Algorithms

K-Means: Partitions data into k clusters based on distance to centroids.

Example: Customer segmentation.

K-Medoids: Similar to K-Means but uses medoids (actual data points) as cluster centers.

Example: Robust to outliers compared to K-Means.

Mean-Shift: Discovers clusters by shifting points towards the mode of the data distribution.

Example: Image segmentation and object tracking.

DBSCAN: Density-Based Spatial Clustering of Applications with Noise; identifies clusters based on data point density.

Example: Anomaly detection.

OPTICS: Ordering Points To Identify the Clustering Structure; an extension of DBSCAN that creates a cluster ordering.

Example: Identifying hierarchical cluster structures.

HDBSCAN: Hierarchical DBSCAN; combines hierarchical clustering with DBSCAN to find clusters of varying densities.

Example: More robust to parameter selection than DBSCAN.

Agglomerative Clustering: Bottom-up hierarchical clustering; each data point starts as a cluster, and clusters are merged iteratively.

Example: Document clustering.

BIRCH: Balanced Iterative Reducing and Clustering using Hierarchies; builds a CF-tree to summarize cluster information.

Example: Large datasets where memory is limited.

Affinity Propagation: Clusters data points based on message passing between pairs of data points.

Example: Identifying exemplars in a dataset.

Gaussian Mixture Models (GMM): Models data as a mixture of Gaussian distributions.

Example: Soft clustering and density estimation.

Dimensionality Reduction

PCA (Principal Component Analysis): Reduces dimensionality by projecting data onto principal components.

Example: Noise reduction and feature extraction.

t-SNE: Visualizes high-dimensional data by reducing it to a lower-dimensional space while preserving local similarities.

Example: Visualizing clusters in gene expression data.

UMAP: Uniform Manifold Approximation and Projection; similar to t-SNE but faster and preserves more of the global structure.

Example: Visualizing and exploring high-dimensional datasets.

ICA (Independent Component Analysis): Separates mixed signals into independent components.

Example: Blind source separation.

LDA (Linear Discriminant Analysis): Supervised dimensionality reduction technique to find the best linear combination of features that separates classes.

Example: Face recognition.

Semi-Supervised Learning

Self-Training: Iteratively trains a model on labeled data and then uses the model to predict labels for unlabeled data, adding high-confidence predictions to the labeled set.

Example: Document classification with limited labeled data.

Label Propagation: Assigns labels to unlabeled data points based on the labels of their neighbors in a graph.

Example: Image segmentation.

Label Spreading: Similar to label propagation but uses a different algorithm to propagate labels through the graph.

Example: Community detection in social networks.

Reinforcement Learning

Reinforcement Learning

Q-Learning: An off-policy RL algorithm that learns the optimal Q-value for each state-action pair.

Example: Training an agent to play a game.

Deep Q-Networks (DQN): Uses a deep neural network to approximate the Q-function.

Example: Playing Atari games.

SARSA: An on-policy RL algorithm that updates the Q-value based on the action taken in the current state.

Example: Robot navigation.

Policy Gradient Methods: Directly optimizes the policy without using a value function.

Example: Training a robot to walk.

Actor-Critic: Combines policy gradient and value-based methods.

Example: Continuous control tasks.

Proximal Policy Optimization (PPO): A policy gradient method that constrains policy updates to ensure stable learning.

Example: Complex control tasks with high-dimensional state spaces.

Deep Deterministic Policy Gradient (DDPG): An actor-critic algorithm for continuous action spaces.

Example: Robotics and autonomous vehicles.

Deep Learning

CNN

LeNet: An early CNN architecture for digit recognition.

Example: Handwritten digit recognition.

AlexNet: A deeper CNN architecture that won the ImageNet competition in 2012.

Example: Image classification.

VGGNet: A CNN architecture with very deep layers.

Example: Image classification and object detection.

GoogLeNet (Inception): A CNN architecture that uses inception modules to capture features at different scales.

Example: Image classification.

ResNet: A CNN architecture that uses residual connections to train very deep networks.

Example: Image classification and object detection.

DenseNet: A CNN architecture that connects each layer to every other layer in a feedforward fashion.

Example: Image classification.

EfficientNet: A CNN architecture that scales all dimensions of the network (width, depth, and resolution) in a principled way.

Example: Image classification with high efficiency.

MobileNet: A CNN architecture designed for mobile devices with limited resources.

Example: Mobile vision applications.

SqueezeNet: A CNN architecture that uses fire modules to reduce the number of parameters.

Example: Image classification with a small model size.

Deep Learning Algorithms

Recurrent Neural Networks (RNN)

Vanilla RNN: A basic recurrent neural network.

Example: Sequence modeling.

Long Short-Term Memory (LSTM): A type of RNN that is designed to handle the vanishing gradient problem.

Example: Natural language processing.

Gated Recurrent Unit (GRU): A simplified version of LSTM.

Example: Machine translation.

Bidirectional RNN: Processes the input sequence in both directions.

Example: Text classification.

Deep RNNs: RNNs with multiple layers.

Example: Speech recognition.

Echo State Networks (ESN): A type of RNN with a randomly connected reservoir.

Example: Time series prediction.

Feedforward Networks

Multilayer Perceptron (MLP): A basic feedforward neural network with one or more hidden layers.

Example: Classification and regression tasks.

Deep Neural Networks (DNN): Neural networks with multiple hidden layers.

Example: Complex pattern recognition tasks.