Supervised Machine Learning Algorithms Cheatsheet

Algorithm	Definition	Uses	Advantages	Disadvantages
1. Linear Regression	A linear approach to modeling the relationship between a dependent variable and one or more independent variables.	- Predicting numerical values (regression)	- Simple and interpretable - Works well when the relationship between variables is approximately linear	- Assumes a linear relationship - Sensitive to outliers and multicollinearity
2. Logistic Regression	A statistical model that predicts the probability of a binary outcome based on one or more predictor variables.	- Binary classification - Probability estimation	- Simple and interpretable - Provides probability estimates	- Assumes a linear relationship - May not perform well with complex data
3. Decision Trees	A tree-like model of decisions and their consequences; it breaks down a dataset into smaller subsets based on different conditions.	- Classification - Regression - Feature selection	- Easy to understand and visualize - Can handle both numerical and categorical data	- Prone to overfitting - Instability (small changes in data can lead to different trees)
4. Random Forest	An ensemble learning method that builds multiple decision trees and combines their predictions.	- Classification - Regression - Feature importance	- Reduces overfitting - Handles high-dimensional data well - Robust to outliers	- More complex than a single decision tree - Slower training time compared to decision trees
5. Support Vector Machines (SVM)	A classification algorithm that finds the hyperplane that best separates classes in a high-dimensional space.	- Classification - Regression - Anomaly detection	- Effective in high-dimensional spaces - Works well with clear margin of separation	- Computationally intensive for large datasets - Sensitivity to kernel choice

Algorithm	Definition	Uses	Advantages	Disadvantages
6. k-Nearest Neighbors (KNN)	A lazy-learning algorithm that classifies data points based on the majority class among their k-nearest neighbors.	- Classification - Regression	- Simple to understand and implement - No training phase	- Slower prediction for large datasets - Sensitive to the choice of k
7. Naive Bayes	A probabilistic classification algorithm based on Bayes' theorem with the assumption of independence among features.	- Text classification - Spam detection - Sentiment analysis	- Simple and efficient - Works well with high-dimensional data	- Assumes independence of features (naive assumption)
8. Neural Networks	A complex network of interconnected nodes (neurons) that can learn and make predictions from data.	- Image recognition - Natural language processing - Speech recognition	- Can model complex relationships - Deep learning can learn hierarchical features	- Requires large amounts of data - Computationally intensive - Prone to overfitting
9. Gradient Boosting (e.g., XGBoost, LightGBM)	An ensemble method that builds multiple weak models sequentially and combines their predictions.	- Classification - Regression - Anomaly detection	- High predictive accuracy - Handles missing data well - Feature importance	- Can be computationally expensive - Sensitive to hyperparameters
10. Linear Discriminant Analysis (LDA)	A dimensionality reduction technique that finds the linear combinations of features that best separate different classes.	- Classification - Feature extraction	- Reduces dimensionality - Preserves class information	- Assumes normally distributed data

Algorithm	Definition	Uses	Advantages	Disadvantages
11. Ridge Regression	A linear regression variant that adds L2 regularization to prevent overfitting.	- Regression	- Mitigates multicollinearity - Reduces overfitting	- Cannot perform feature selection
12. Lasso Regression	A linear regression variant that adds L1 regularization to perform feature selection.	- Regression	- Performs automatic feature selection - Reduces overfitting	- May select too few variables if regularization is too strong
Algorithm	Definition	Uses	Advantages	Disadvantages
AdaBoost	Ensemble method that adjusts the weights of misclassified instances iteratively to focus on difficult cases.	Classification and regression tasks.	1. Can improve the accuracy of weak learners.	1. Sensitive to noisy data and outliers.
			2. Automatically handles feature selection.	2. Can be computationally expensive.
			3. Reduces overfitting.	
			4. Works well with a variety of base models.	
Gradient Boosting	Builds an ensemble of decision trees, where each new tree corrects errors made by the previous ones.	Regression and classification tasks.	1. Excellent predictive performance.	1. Can be prone to overfitting, needs tuning.

Algorithm	Definition	Uses	Advantages	Disadvantages
			2. Handles both numeric and categorical data.	2. Sensitive to outliers.
			3. Feature importance estimation.	3. Computationally intensive.
			4. Works well with missing data.	
XGBoost	Extreme Gradient Boosting, a regularized version of gradient boosting with tree-based models.	Classification and regression tasks.	1. Improved regularization for preventing overfitting.	1. Requires parameter tuning.
			2. Efficient and scalable.	2. Sensitive to noisy data.
			3. Handles missing data.	3. May require feature engineering.
LightGBM	Gradient boosting framework that uses a histogram-based learning method.	Classification and regression tasks.	1. Fast and memory-efficient.	1. May not perform well with small datasets.
			2. Handles large datasets well.	2. Sensitive to overfitting if not tuned.
			3. Supports categorical features without one-hot encoding.	3. Limited interpretability.

Algorithm	Definition	Uses	Advantages	Disadvantages
CatBoost	Gradient boosting algorithm designed for categorical feature support.	Classification and regression tasks.	1. Handles categorical features natively.	1. Slower training compared to some others.
			2. Automatically deals with missing values.	2. Requires careful parameter tuning.
			3. Robust to overfitting.	3. Limited interpretability.