MACHINE LEARNING ALGORITHMS

Short Notes By Shivani Bhatia

Linear Regression

- Purpose: Predict continuous outcomes.
- Concept: Finds the line of best fit for data points to predict values.
- Equation: $y = b0 + b1x + \varepsilon$, where b0 is the intercept, b1 is the slope.
- Assumptions:
- – Linearity: Relationship between features and target is linear.
- – Independence of errors.
- – Homoscedasticity: Constant variance of errors.
- – Normal distribution of errors.
- Evaluation: R-squared, Adjusted R-squared, Mean Squared Error (MSE), Mean Absolute Error (MAE).
- Regularization Techniques: Lasso (L1), Ridge (L2) regression to prevent overfitting by penalizing large coefficients.

Logistic Regression

- Purpose: Classify binary outcomes (0 or 1).
- Concept: Uses a sigmoid function to map predictions to probabilities.
- Equation: $\log(p/(1 - p)) = b0 + b1x$, where p is the probability of class 1.
- Assumptions:

- Binary or ordinal outcome variable.
- Linear relationship between features and the logit of the outcome.

- Evaluation: Accuracy, Precision, Recall, F1 Score, ROC-AUC.
- Applications: Classification problems like spam detection, fraud detection.

Decision Trees

- Purpose: Handle both classification and regression tasks.
- Concept: Recursively splits data into branches to form tree-like structures.
- Impurity Measures: Gini Index (for classification), Entropy (Information Gain).
- Advantages: Easy to interpret and visualize, Handles non-linear data well.
- Disadvantages: Prone to overfitting, especially with deep trees.
- Evaluation: Classification accuracy, R-squared (for regression).
- Pruning: Reduces tree complexity to avoid overfitting.

Random Forest

- Purpose: Improve decision tree performance through ensemble learning.
- Concept: Builds multiple decision trees on random subsets and combines their outputs.
- Parameters: Number of trees, Max depth, Min samples per leaf.
- Advantages: Reduces overfitting (less variance than single trees), Works well on high-dimensional datasets.
- Evaluation: Accuracy, Precision, Recall, F1 Score, Feature Importance.
- Applications: Image classification, healthcare diagnostics, finance.

Support Vector Machine (SVM)

- Purpose: Separate classes with the widest possible margin.
- Concept: Finds the optimal hyperplane that maximizes the margin between classes.
- Kernel Trick: Maps data to higher dimensions to make classes linearly separable (Linear, Polynomial, RBF).
- Hyperparameters: C (regularization), Gamma (for RBF kernel).
- Advantages: Effective for high-dimensional data, Works well for binary classification.
- Evaluation: Accuracy, Precision, Recall, F1 Score, ROC-AUC.

K-Nearest Neighbors (KNN)

- Purpose: Classify data based on similarity.
- Concept: Classifies data points based on the majority class of its k-nearest neighbors.
- Hyperparameters: Number of neighbors (k), Distance metric (Euclidean, Manhattan).
- Advantages: Simple, easy to implement, Non-parametric (no assumption about data distribution).
- Disadvantages: Computationally expensive for large datasets, Sensitive to irrelevant features and scaling.
- Evaluation: Accuracy, Confusion Matrix, ROC-AUC.

Naive Bayes

- Purpose: Classify data based on Bayes' theorem.
- Concept: Assumes feature independence and uses the likelihood of features to make predictions.
- Equation: $P(y|X) = (P(X|y)P(y)) / P(X)$.
- Types: Gaussian Naive Bayes (continuous data), Multinomial Naive Bayes (text classification), Bernoulli Naive Bayes (binary features).
- Advantages: Works well with small data and text classification, Fast and requires less training data.
- Disadvantages: Assumes feature independence, which is often not true.
- Evaluation: Accuracy, Precision, Recall, F1 Score.

K-Means Clustering

- Purpose: Cluster data into groups based on similarity.

- Concept: Partitions data into k clusters by minimizing variance within each cluster.

- Algorithm:

1. Initialize centroids randomly
2. Assign points to nearest centroid,
3. Recalculate centroids.

- Hyperparameter: Number of clusters (k).

- Evaluation: Elbow Method for selecting k, Silhouette Score for clustering quality.

- Applications: Customer segmentation, image compression.

AdaBoost

- Purpose: Improve weak learners to achieve a stronger ensemble classifier.

- Concept: Sequentially trains weak classifiers (usually decision trees) on weighted data, adjusting weights for misclassified points to focus on difficult cases in subsequent rounds.

- Algorithm:

1. Initialize weights for all samples.
2. Train a weak classifier and calculate its error.
3. Increase weights for misclassified samples and decrease for correctly classified ones.
4. Combine weak classifiers' predictions based on their accuracy.

- Parameters: No.of estimators (number of weak learners), Learning rate (controls contribution of each weak learner).

- Advantages: Simple to implement, Works well on both classification and regression, Focuses on difficult cases.

- Evaluation: Accuracy, Precision, Recall, F1 Score, AUC-ROC.

Gradient Boosting

- Purpose: Ensemble method for both classification and regression to reduce errors iteratively.
- Concept: Uses a sequence of weak learners to correct errors of the previous learners. Each new model is trained on the residuals (errors) of the previous model, gradually improving performance.
- Algorithm:

1. Start with an initial prediction (mean value for regression).
2. Calculate residuals.
3. Train a weak learner on residuals and update predictions by adding weak learner output.
4. Repeat until the desired performance or number of iterations.

- Parameters: n_estimators, learning_rate, max_depth
- Advantages: Effective for both classification and regression, High predictive accuracy.
- Disadvantages: Computationally intensive, Prone to overfitting without careful tuning.
- Evaluation: Accuracy, RMSE, MAE, ROC-AUC.
- Applications: Predictive modeling in finance, healthcare, marketing, and insurance.

PCA (Principal Component Analysis)

- Purpose: Dimensionality reduction while preserving maximum variance.
- Concept: Transforms data onto a new set of uncorrelated components (principal components).
- Steps:

1. Standardize data
2. Compute covariance matrix
3. Find eigenvectors
4. Select top components

- Advantages: Reduces overfitting and computational cost.
- Applications: Image processing, data visualization, feature extraction.

XGBoost (Extreme Gradient Boosting)

- Purpose: Optimized gradient boosting for large and complex datasets.

- Concept: Improves gradient boosting with regularization, handling of missing values, parallel processing, and optimized tree structure.

- Key Parameters: n_estimators, learning_rate, max_depth, gamma, subsample

- Advantages:
  -High accuracy and efficiency.
  -Reduces overfitting with regularization.

- Applications: Used widely in data science competitions, finance, marketing, and customer churn prediction.

THANK YOU

I hope you can get useful knowledge from this presentation. Good luck !