|  |  |  |
| --- | --- | --- |
| **Linear Regression**   * **Use Case:** Predicting continuous numerical outputs (e.g., housing prices). * **Input:** Features (independent variables), target (dependent variable). * **Parameters:**   + Learning rate (if using gradient descent)   + Regularization strength (L1 or L2). * **Output:** A continuous value. * **Metrics:**   + Mean Squared Error (MSE)   + Root Mean Squared Error (RMSE)   + Mean Absolute Error (MAE)   + R-squared | **Logistic Regression**   * **Use Case:** Binary or multiclass classification (e.g., spam email detection). * **Input:** Features, class labels. * **Parameters:**   + Regularization strength (L1 or L2).   + Multi-class options (e.g., one-vs-rest). * **Output:** Probability of class membership. * **Metrics:**   + Accuracy   + Precision, Recall   + F1 Score   + ROC-AUC Score | **Decision Trees**   * **Use Case:** Both classification and regression tasks (e.g., customer segmentation). * **Input:** Features, target values. * **Parameters:**   + Maximum depth   + Minimum samples per split   + Criterion (e.g., Gini impurity, entropy). * **Output:** Predicted class or value. * **Metrics:**   + Classification: Accuracy, F1 Score, Precision, Recall   + Regression: MSE, RMSE, MAE |
| **Random Forest**   * **Use Case:** Improves on decision trees by reducing overfitting (e.g., feature importance analysis). * **Input:** Features, target values. * **Parameters:**   + Number of trees   + Maximum depth   + Minimum samples per leaf. * **Output:** Predicted class or value. * **Metrics:**   + Classification: Accuracy, F1 Score, Precision, Recall   + Regression: MSE, RMSE, MAE | **Support Vector Machines (SVM)**   * **Use Case:** Classification and regression, especially with small datasets (e.g., image recognition). * **Input:** Features, labels. * **Parameters:**   + Kernel (linear, polynomial, RBF)   + Regularization parameter (C)   + Gamma (for RBF and polynomial kernels). * **Output:** Predicted class or value. * **Metrics:**   + Accuracy   + Precision, Recall   + F1 Score   + ROC-AUC Score | **K-Nearest Neighbors (KNN)**   * **Use Case:** Classification or regression with non-linear boundaries (e.g., recommendation systems). * **Input:** Features, labels. * **Parameters:**   + Number of neighbors (k)   + Distance metric (e.g., Euclidean, Manhattan). * **Output:** Predicted class or value. * **Metrics:**   + Classification: Accuracy, F1 Score   + Regression: MSE, MAE. |
|  |  |  |
| **Naïve Bayes**   * **Use Case:** Text classification, spam detection, sentiment analysis. * **Input:** Features, labels. * **Parameters:**   + Smoothing parameter (e.g., Laplace smoothing). * **Output:** Predicted class probabilities. * **Metrics:**   + Accuracy   + Precision, Recall   + F1 Score   + ROC-AUC Score | **Neural Networks**   * **Use Case:** Complex and large-scale tasks (e.g., image recognition, NLP). * **Input:** Features, target values. * **Parameters:**   + Number of layers and neurons   + Learning rate   + Activation functions   + Regularization (dropout, weight decay). * **Output:** Predicted values or class probabilities. * **Metrics:**   + Classification: Accuracy, F1 Score, ROC-AUC   + Regression: MSE, RMSE, MAE | **Gradient Boosting (e.g., XGBoost, LightGBM, CatBoost)**   * **Use Case:** Both classification and regression (e.g., predictive analytics). * **Input:** Features, target values. * **Parameters:**   + Learning rate   + Number of estimators   + Maximum depth. * **Output:** Predicted class or value. * **Metrics:**   + Classification: Accuracy, F1 Score, Precision, Recall   + Regression: MSE, RMSE, MAE |
| **Clustering Algorithms (e.g., K-Means, DBSCAN)**   * **Use Case:** Unsupervised learning for grouping data points (e.g., customer segmentation). * **Input:** Features. * **Parameters:**   + Number of clusters (for K-Means)   + Epsilon and minimum samples (for DBSCAN). * **Output:** Cluster assignments. * **Metrics:**   + Silhouette Score   + Davies-Bouldin Index   + Inertia (for K-Means) |  | **General Metrics to Gauge Model Performance**   * **Classification Metrics:**   + Confusion Matrix   + Precision-Recall Curve   + ROC Curve   + Log Loss. * **Regression Metrics:**   + Explained Variance   + Mean Bias Deviation. * **Unsupervised Learning Metrics:**   + Silhouette Coefficient   + Dunn Index. |
|  |  |  |
| **Type 1 Error (Alpha, α):**   * **Definition:** Rejecting the null hypothesis when it is actually true. * **Example:** Convicting an innocent person. * **Impact:** False positive. * **Control:** By setting a low significance level (α, e.g., 0.05).   **Type 2 Error (Beta, β):**   * **Definition:** Failing to reject the null hypothesis when it is actually false. * **Example:** Letting a guilty person go free. * **Impact:** False negative. * **Control:** Increasing sample size or power of the test. |  | **Confusion Matrix:**   |  |  |  | | --- | --- | --- | | **Actual \ Predicted** | **Positive (P)** | **Negative (N)** | | Positive (P) | True Positive (TP) | False Negative (FN) | | Negative (N) | False Positive (FP) | True Negative (TN) |   **Key Terms:**   * **True Positive (TP):** The model correctly predicts the positive class. * **False Positive (FP):** The model incorrectly predicts the positive class (Type 1 Error). * **True Negative (TN):** The model correctly predicts the negative class. * **False Negative (FN):** The model incorrectly predicts the negative class (Type 2 Error). |
| **Accuracy:**   * + **Definition:** The proportion of total correct predictions out of all predictions made.   + **Formula:**   + **Insight:** High accuracy can be misleading in imbalanced datasets.   **Precision:**   * + **Definition:** The proportion of true positive predictions out of all positive predictions made.   + **Formula:**   + **Insight:** Focuses on how many predicted positives are actually correct. | **Recall (Sensitivity):**   * **Definition:** The proportion of true positive predictions out of all actual positives. * **Formula:** * **Insight:** Indicates the model's ability to detect positive cases. | **F1 Score:**   * + **Definition:** The harmonic mean of precision and recall.   + **Formula:**   + **Insight:** A balanced measure for imbalanced datasets.   **Specificity:**   * + **Definition:** The proportion of true negative predictions out of all actual negatives.   + **Formula:**   + **Insight:** Highlights the model's ability to identify negative cases accurately. |