Machine Learning Fundamentals – Refined Summary 1. Learning Paradigms **Supervised Learning:** Learns from labeled data (input–output pairs). Examples: Regression, Classification.
Unsupervised Learning: Finds hidden patterns in unlabeled data. Examples: Clustering, Dimensionality Reduction.
Reinforcement Learning: Learns by interacting with environment and receiving rewards/penalties. Examples: Game AI, Robotics.
2. Model Performance and Generalization **Overfitting:** Fits training data too closely; fails to generalize.
Underfitting: Model too simple to capture data patterns.
Bias–Variance Tradeoff: High bias = underfitting; high variance = overfitting.
Cross-Validation: Splits data into folds to evaluate generalization (e.g., K-Fold).
Regularization: Adds penalty to loss function to discourage complexity.
L1 (Lasso) vs L2 (Ridge): L1 zeros out coefficients; L2 shrinks large coefficients.
Curse of Dimensionality: Too many features can degrade performance.
Feature Scaling: Normalizes features to ensure equal contribution.
Parameters vs Hyperparameters: Parameters learned; Hyperparameters set externally.
3. Regression and Classification **Multicollinearity:** High correlation among independent variables in regression.
Logistic Regression: Uses sigmoid function for binary output (0 or 1).
ROC Curve: Plots TPR vs FPR; AUC measures separation power.
Precision, Recall, F1: Balance accuracy for imbalanced data.
Confusion Matrix: Summarizes predicted vs actual classes.
Imbalanced Data & SMOTE: Synthetic oversampling of minority class.
Regression Metrics: MAE, MSE, RMSE, R^2 measure prediction errors.
4. Decision Trees & Ensemble Methods **Splitting:** Chooses features maximizing information gain (Gini or Entropy).
Pruning: Removes weak branches to avoid overfitting.
Bagging vs Boosting: Bagging reduces variance; Boosting reduces bias.
Random Forest: Ensemble of multiple trees for stability.
Feature Importance: Quantifies feature contribution.
Boosting Variants: AdaBoost, XGBoost, LightGBM, CatBoost.
5. Clustering & Dimensionality Reduction **K-Means:** Clusters data into K groups minimizing SSE.
Elbow Method: Optimal K where SSE decrease flattens.
Hierarchical Clustering: Agglomerative (bottom-up) or Divisive (top-down).
PCA: Orthogonal transformation maximizing variance across components.
Eigenvalues/Vectors: Define principal axes and magnitude of variance.
6. Ensemble Learning Overview Bagging (parallel models), Boosting (sequential improvements), Stacking (meta-model). Normalization helps prevent overfitting.
7. Generative AI & LLMs **Transformers:** Use self-attention and positional encoding (no recurrence).
Self-Attention: Captures token relationships in parallel.
Positional Encoding: Adds sequence order for context.
GPT vs BERT: GPT = Decoder (generation); BERT = Encoder (understanding).
Embeddings: Vector representations of semantic meaning.
Pre-training vs Fine-tuning: Pre-training = general learning; Fine-tuning = task-specific adaptation (e.g., LoRA).
8. Advanced Model Optimization **RAG:** Combines retrieval with generation for factual outputs.
Quantization: Reduces precision for faster inference.
Model Distillation: Trains smaller “student” model to mimic a larger “teacher.”
9. Evaluation, Deployment & Ethics **Evaluation Metrics:** AUC/F1 (classification), RMSE/MAPE (regression).
A/B Testing: Compare model variants statistically.
Explainability: LIME (local), SHAP (game-theoretic), PDPs for feature effects.
Ethics: Focus on fairness, transparency, and responsible AI use.