Machine Learning Notes

1. Introduction to Machine Learning

• Definition: Machine Learning (ML) is a subfield of Artificial Intelligence that enables systems to learn and improve from experience without being explicitly programmed.
• Types of learning: supervised, unsupervised, semi-supervised, reinforcement.
• Applications: recommendation systems, fraud detection, natural language processing, computer vision.

2. Supervised Learning

• Definition: Learning a mapping from inputs to outputs based on labeled training data.
• Key algorithms: Linear Regression, Logistic Regression, Decision Trees, Random Forests, Support Vector Machines, Neural Networks.
• Evaluation metrics: accuracy, precision, recall, F1-score, ROC-AUC.
• Overfitting & Underfitting: causes, bias-variance tradeoff, regularization (L1, L2).

3. Unsupervised Learning

• Definition: Learning patterns from unlabeled data.
• Clustering: K-means, Hierarchical, DBSCAN.
• Dimensionality reduction: PCA, t-SNE, UMAP.
• Applications: customer segmentation, anomaly detection, data compression.

4. Reinforcement Learning

• Definition: Learning by interacting with an environment to maximize cumulative reward.
• Key elements: agent, environment, state, action, reward, policy, value function.
• Algorithms: Q-Learning, SARSA, Deep Q-Networks (DQN), Policy Gradient methods.
• Applications: robotics, game playing (AlphaGo), resource optimization.

5. Neural Networks & Deep Learning

• Artificial Neural Networks (ANN): inspired by biological neurons, consist of input, hidden, and output layers.
• Activation functions: Sigmoid, Tanh, ReLU, Leaky ReLU, Softmax.
• Deep architectures: Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Transformers.
• Training: backpropagation, gradient descent variants (SGD, Adam).
• Regularization: dropout, batch normalization, data augmentation.

6. Feature Engineering & Data Preprocessing

• Handling missing values: deletion, imputation.
• Feature scaling: normalization, standardization.
• Encoding categorical variables: one-hot encoding, label encoding, embeddings.
• Feature selection: filter, wrapper, embedded methods.

7. Model Evaluation & Validation

- Train-test split, k-fold cross-validation.
- Confusion matrix, precision-recall curves.
- Hyperparameter tuning: grid search, random search, Bayesian optimization.
- Model interpretability: SHAP, LIME, feature importance.

8. Advanced Topics

- Ensemble learning: bagging, boosting (AdaBoost, XGBoost, LightGBM, CatBoost).
- Semi-supervised learning: self-training, pseudo-labeling.
- Transfer learning: using pre-trained models for new tasks.
- Federated learning: distributed model training without data centralization.
- Ethics in ML: fairness, bias, accountability, transparency.

9. Tools & Frameworks

- Python libraries: scikit-learn, TensorFlow, PyTorch, Keras, XGBoost, LightGBM.
- ML workflow platforms: MLflow, Kubeflow, TensorBoard.
- Data manipulation & visualization: NumPy, pandas, Matplotlib, Seaborn.

10. Future of Machine Learning

- Integration with edge devices and IoT.
- Advances in self-supervised learning and generative AI.
- Interdisciplinary applications: healthcare, finance, climate science.
- Trends: AutoML, Explainable AI (XAI), Responsible AI.