Machine Learning Syllabus & Contents

A Machine Learning (ML) syllabus typically covers a broad range of concepts and techniques that enable the development of models capable of making predictions or decisions without being explicitly programmed for each task. The following is a detailed syllabus for a Machine Learning course, covering fundamental to advanced topics.

1. Introduction to Machine Learning

What is Machine Learning

- Overview of ML and its applications in various industries.
- Types of ML: Supervised Learning, Unsupervised Learning, Semi-supervised Learning, and Reinforcement Learning.

Data Preprocessing

- Data cleaning and preparation.
- Handling missing data, scaling, encoding categorical variables.

Introduction to Python for Machine Learning

- Libraries: NumPy, Pandas, Matplotlib, Scikit-learn, TensorFlow, and Keras.

2. Supervised Learning Algorithms

Linear Regression

- Simple Linear Regression and Multiple Linear Regression.
- Model evaluation using Mean Squared Error (MSE).

Logistic Regression

- Binary and Multinomial Logistic Regression.
- Understanding of Sigmoid function, odds, and probability.

Support Vector Machines (SVM)

- Linear and non-linear classification using kernels.
- Margin maximization, soft margin, and regularization.

Decision Trees and Random Forests

- Decision tree construction (ID3, CART algorithms).
- Overfitting and pruning.
- Random Forests and Bagging techniques.

K-Nearest Neighbors (KNN)

- Instance-based learning.

- Distance metrics (Euclidean, Manhattan).
- KNN for classification and regression.

Naive Bayes Classifier

- Bayes' Theorem and its application to classification.
- Assumptions in Naive Bayes (independence of features).

3. Model Evaluation and Tuning

Cross-Validation

- K-fold cross-validation.
- Bias-Variance tradeoff.

Metrics for Classification

- Accuracy, Precision, Recall, F1-score, ROC-AUC curve.

Metrics for Regression

- Mean Absolute Error (MAE), Mean Squared Error (MSE), R^2 (Coefficient of Determination).

Hyperparameter Tuning

- Grid Search and Random Search for hyperparameter optimization.

Overfitting and Underfitting

- Techniques to avoid overfitting (regularization, dropout, cross-validation).

4. Unsupervised Learning Algorithms

Clustering Algorithms

- K-Means Clustering, Hierarchical Clustering, DBSCAN.
- Understanding centroid-based, density-based, and agglomerative methods.

Dimensionality Reduction

- Principal Component Analysis (PCA).
- t-SNE (t-Distributed Stochastic Neighbor Embedding).
- Feature selection techniques.

Association Rule Learning

- Apriori Algorithm for Market Basket Analysis.
- FP-growth Algorithm.

5. Advanced Topics in Machine Learning

Ensemble Methods

- Bagging (Bootstrap Aggregating).
- Boosting (AdaBoost, Gradient Boosting).
- Stacking.

Deep Learning

- Introduction to Neural Networks: Perceptrons, Backpropagation.
- Deep Neural Networks, Activation functions, and Gradient Descent.
- Convolutional Neural Networks (CNNs) for image classification.
- Recurrent Neural Networks (RNNs) for sequential data (LSTMs, GRUs).
- Generative Adversarial Networks (GANs).

Reinforcement Learning

- Overview of RL, Markov Decision Processes (MDP).
- Exploration vs. Exploitation dilemma.
- Q-Learning and Policy Gradient methods.

6. Model Deployment and Real-World Applications

Model Deployment Techniques

- Saving and loading models (Pickle, Joblib).
- Web deployment using Flask or FastAPI.
- Cloud-based deployment using AWS, GCP, or Azure.

Real-World Applications

- Predictive analytics (e.g., stock market prediction, fraud detection).
- NLP (Natural Language Processing) techniques, such as Sentiment Analysis and Text Classification.
- Computer Vision (e.g., object detection, facial recognition).
- Recommender Systems (e.g., collaborative filtering).

7. Ethical Considerations in Machine Learning

Bias in Machine Learning Models

- Sources of bias in training data and models.
- Techniques for mitigating bias.

Interpretability and Explainability

- Importance of model transparency.
- Techniques for explaining complex models (LIME, SHAP values).

8. Tools and Libraries in Machine Learning

Python Libraries:

- NumPy for numerical computation.
- Pandas for data manipulation and analysis.
- Matplotlib, Seaborn, Plotly for data visualization.

- Scikit-learn for implementing classical machine learning algorithms.
- TensorFlow and Keras for building and training deep learning models.
- XGBoost and LightGBM for gradient boosting techniques.

Jupyter Notebooks:

- Setting up Jupyter Notebooks for interactive coding and experimentation.