**AGENTIC TRAINING**

SESSION-1

**GitHub Basics**

* GitHub is a cloud-based platform, it helps to store and manage data, collaborate with others efficiently.
* Repository is like a project folder.
* Branch is a copy of the project where we can make changes without affecting the main code.
* Pull Requests: When you finish working on a feature in a branch, you open a pull request. It is done to propose changes to the main branch, get feedback from team members, merge changes if approved.

git push: sends local changes to the to the current repo.

git pull: pulls the latest changes from the remote repo into your local branch.

Pull request: team members can review, comment, suggest and approve the requests.

Merge conflicts: In general, conflicts arise when two people edit the same line in the same file.

**Introduction to Machine Learning**

**Artificial Intelligence (AI):** The broad field focused on creating machines or software that can perform tasks that typically require human intelligence.

**Machine Learning (ML):** A subset of AI that enables computers to learn from data and improve their performance on a task without being explicitly programmed.

Data Collection --- Data Processing --- Training --- Evaluation --- Tuning **Supervised Learning:** The most common type, where the model learns from labeled data. You have input data (features) and corresponding correct output (labels). The goal is to learn a mapping from inputs to outputs.

**Unsupervised Learning:** The model learns patterns from unlabeled data, where no explicit output is provided. The goal is to find hidden structures or patterns.  
**Reinforcement Learning:** The model learns by interacting with an environment, receiving rewards or penalties based on actions taken.

**Linear Regression:**

Linear regression is a **supervised learning algorithm** used for **predicting a continuous output variable** (like price, age, salary) based on one or more input features.

* Simple: involves 1 dependent and 1 independent variable.
* Multiple: involves 2 or more input features.
* Learning rate is a small positive value that controls how fast the model learns.
* **Ordinary Least Squares (OLS):** Method to minimize the sum of squared differences between observed and predicted values.
* **Loss Function:** Quantifies the error of the model’s predictions (e.g., Mean Squared Error).
* **Mean Square Error (MSE):** Average of squared differences between predicted and actual values.
* **Updating Intercepts and Coefficients:** Using gradient descent to iteratively adjust parameters to minimize loss.

**Logistic Regression:**

Logistic regression is a **supervised learning algorithm** used for **classification.**

* Uses logistic function (sigmoid) to output probabilities.
* Decision Boundary: Threshold to decide class labels (commonly 0.5).
* Cost Function (Log Loss): Measures difference between predicted probabilities and actual labels.
* Gradient Descent for Optimization: Algorithm to minimize cost function by updating parameters.

**Decision Trees:** It is a supervised ML algorithm used for classification and regression tasks.

**Ensemble Learning:** It is a technique in machine learning where **multiple models (weak learners)** are combined to produce a **stronger, more accurate model**.

* Goals of Ensemble Learning are reducing bias (under fitting), variance (over fitting) and improving accuracy, stability of predictions.
* Bagging category: use different random subsets of training data to train multiple models, it reduces variance.
* Boosting category: models are trained sequentially, each trying to correct the errors of the previous models.
* Stacking category: combines different types of models.

**Random Forest:** An ensemble of decision trees using bagging and random feature selection to reduce over fitting and improve accuracy.

**Grid Search:** Hyper parameter tuning technique to exhaustively search over specified parameter values. **Cross Validation:** Technique to evaluate model performance by splitting data into training and validation sets multiple times (e.g., k-fold cross-validation).