Supervised Learning

1. What is Supervised Learning?

Supervised learning is a machine learning paradigm where a model learns a mapping from **input features (X)** to **known output labels (y)** using **labeled training data**.

Goal: Learn a function

F(X)->y

📌 Key idea: *The correct answer is already known during training.*

2. Types of Supervised Learning

1️⃣ Classification

Predicts **discrete / categorical outputs**.

Examples:

- Spam vs Not Spam
- Disease vs No Disease
- Fraud vs Non-fraud

Common algorithms:

- Logistic Regression
- K-Nearest Neighbors (KNN)
- Naive Bayes
- Support Vector Machine (SVM)
- Decision Tree
- Random Forest
- Gradient Boosting (XGBoost, LightGBM)

- Neural Networks

2️⃣ Regression

Predicts **continuous numerical values**.

Examples:

- House price prediction
- Temperature forecasting
- Salary prediction

Common algorithms:

- Linear Regression
- Polynomial Regression
- Ridge / Lasso / ElasticNet
- Decision Tree Regressor
- Random Forest Regressor
- Gradient Boosting Regressor
- Neural Networks

3. Key Components of Supervised Learning

1️⃣ Dataset

A labeled dataset consists of:

- **Features (X)** → independent variables
- **Target (y)** → dependent variable (label)

Example:

Size (sqft)	Bedrooms	Price
1200	2	300k

2️⃣ Training Set, Validation Set, Test Set

- **Training set** → learns parameters
- **Validation set** → hyperparameter tuning
- **Test set** → final unbiased evaluation

Typical split:

- 70% Train
- 15% Validation
- 15% Test

3️⃣ Model

A mathematical function with learnable parameters.

Example (Linear Regression):

$y = wx + b$

4️⃣ Loss Function

Measures **how wrong** the model's predictions are.

Task	Common Loss Functions
Regression	MSE, MAE, RMSE, Huber
Classification	Cross-Entropy, Log Loss, Hinge Loss

5️⃣ Optimization Algorithm

Minimizes the loss function.

Examples:

- Gradient Descent
- Stochastic Gradient Descent (SGD)
- Adam, RMSProp

4. Common Supervised Learning Algorithms (Interview Focus)

◇ Linear Regression

- Assumes linear relationship
- Minimizes Mean Squared Error
- Sensitive to outliers

📌 Assumptions:

- Linearity
- Independence
- Homoscedasticity
- No multicollinearity

◇ Logistic Regression

- Used for classification
- Uses **sigmoid function**
- Outputs probability (0–1)

[

\sigma(z) = \frac{1}{1+e^{-z}}

]

◇ K-Nearest Neighbors (KNN)

- Instance-based, lazy learner
- No training phase
- Sensitive to scale and choice of K

◇ Naive Bayes

- Probabilistic classifier
- Assumes feature independence
- Fast and works well on text data

◇ Support Vector Machine (SVM)

- Maximizes margin
- Uses kernels for non-linear separation
- Memory-intensive for large datasets

◇ Decision Tree

- Rule-based model
- Easy to interpret
- Prone to overfitting

◇ Random Forest

- Ensemble of decision trees
- Reduces variance
- Robust and widely used

◇ Gradient Boosting (XGBoost, LightGBM)

- Sequential learning
- Corrects previous errors
- High performance but needs tuning

◇ Neural Networks

- Multi-layer nonlinear models
- Requires large data
- Used in deep learning

5. Evaluation Metrics

📊 Regression Metrics

- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)
- R^2 Score

📊 Classification Metrics

- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC
- Confusion Matrix

📌 Tip:

- Use **Precision** when false positives are costly
- Use **Recall** when false negatives are costly

6. Bias–Variance Tradeoff

- **High Bias** → Underfitting
- **High Variance** → Overfitting

Problem	Solution
Underfitting	Add features, complex model
Overfitting	Regularization, more data, pruning

7. Regularization Techniques

Used to prevent overfitting.

- **L1 (Lasso)** → Feature selection
- **L2 (Ridge)** → Shrinks weights
- **ElasticNet** → Combination of L1 + L2

8. Feature Engineering (Very Important)

- Handling missing values
- Encoding categorical variables
- Feature scaling (StandardScaler, MinMax)
- Polynomial features
- Feature selection

9. Data Leakage

Occurs when information from the test set leaks into training.

Example:

- Scaling entire dataset before splitting

📌 Always fit preprocessors **only on training data**.

10. Supervised vs Unsupervised (Common Question)

Supervised	Unsupervised
Labeled data	No labels
Prediction tasks	Pattern discovery
Classification, Regression	Clustering, PCA

11. Real-World Supervised Learning Pipeline

1. Data collection
2. Data cleaning
3. Feature engineering
4. Train-test split
5. Model selection
6. Training
7. Hyperparameter tuning
8. Evaluation
9. Deployment
10. Monitoring