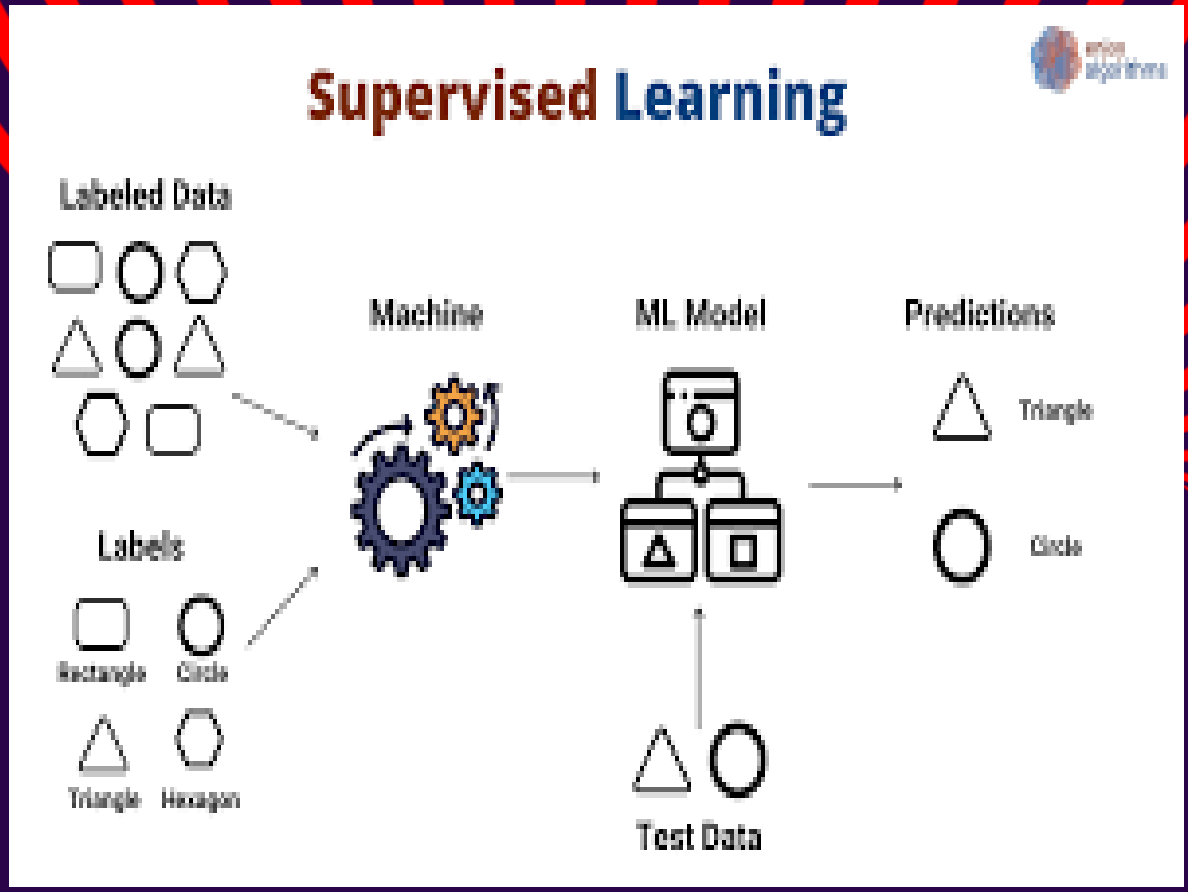
MACHINE LEARNING

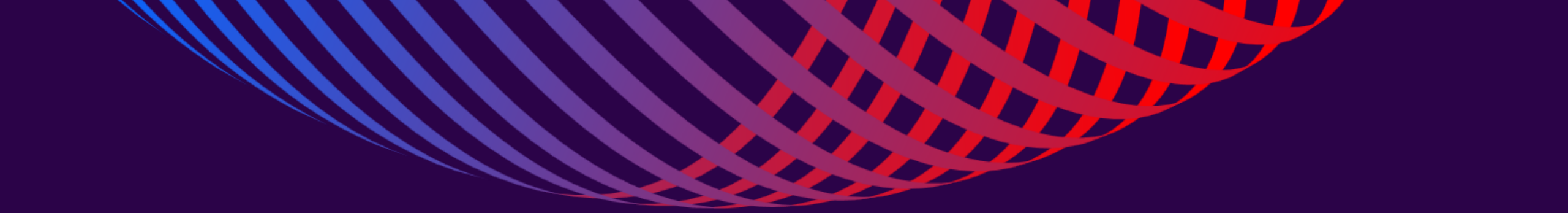
Trainer:

Kaleem U Allah

SUPERVISED MACHINE LEARNING

Supervised machine learning is a type of machine learning where the model is trained on a labeled dataset, meaning that each input data point is associated with a corresponding target label or output. Supervised learning aims to learn a mapping from inputs to outputs, based on the examples provided during training.





TYPES OF SUPERVISED LEARNING

CLASSIFICATION

In classification tasks, the goal is to predict a categorical label for each input. For example, given an image of a handwritten digit, the task might be to classify it as one of the digits from 0 to 9. Common algorithms for classification include decision trees, random forests, support vector machines (SVM), and neural networks.

REGRESSION

In regression tasks, the goal is to predict a continuous value for each input. For example, predicting house prices based on features such as square footage, number of bedrooms, etc. Common algorithms for regression include linear regression, decision trees, random forests, and neural networks.

REGRESSION

- In regression problems, the goal is to predict a continuous numerical value based on input features. The target variable in regression can take on an infinite number of possible values within a given range. Examples of regression problems include predicting house prices, stock prices, temperature, sales revenue, etc.

Characteristics:

- The target variable is continuous.
- The output is a quantity that can vary over a range.
- Evaluation metrics commonly used for regression include Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), etc.

Examples:

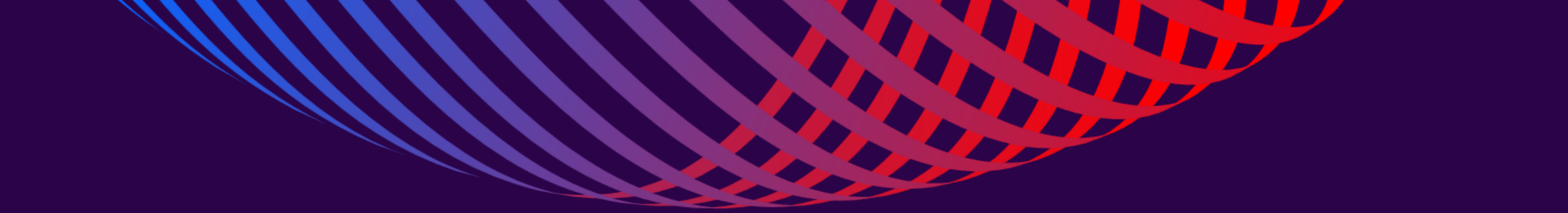
- Predicting house prices based on features like square footage, number of bedrooms, location, etc.
- Forecasting stock prices based on historical data and market indicators.
- Estimating the amount of rainfall based on weather conditions.

CLASSIFICATION

- In classification problems, the goal is to predict a categorical label or class for each input instance. The target variable in classification is discrete and represents different classes or categories. Examples of classification problems include spam email detection, sentiment analysis, image classification, etc.
- **Characteristics**:
- The target variable is categorical.
- The output is a class label indicating the category to which an instance belongs.
- Evaluation metrics for classification include accuracy, precision, recall, F1-score, etc.

Examples:

- Classifying emails as spam or not spam based on their content and metadata.
- Predicting whether a customer will churn or not churn based on their behavior and interactions with a service.
- Identifying handwritten digits in images (e.g., classifying digits from 0 to 9).



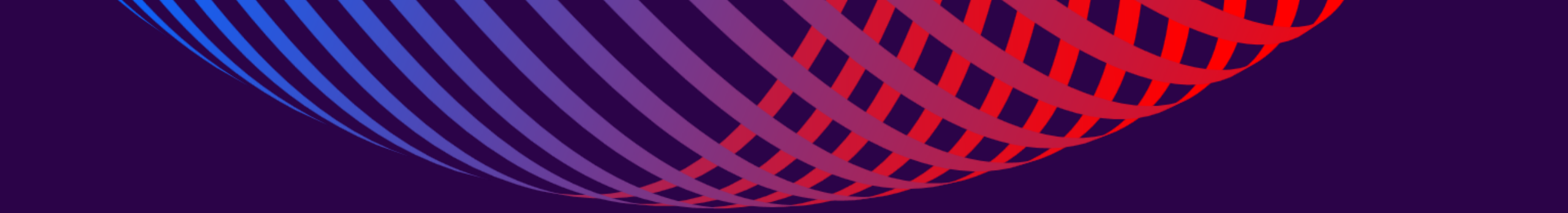
COMPONENTS OF SUPERVISED LEARNING

LABEL DATA

Labeled data consists of input-output pairs, where each input is associated with a corresponding output or target label. This data is used to train the model by showing it examples of inputs and their corresponding correct outputs.

HYPOTHESIS

The hypothesis, also known as the model, is a function that maps inputs to outputs. It represents the learned relationship between the input features and the target labels. The hypothesis is parameterized by a set of parameters (weights and biases) that are adjusted during training to minimize the difference between predicted outputs and actual outputs.



COMPONENTS OF SUPERVISED LEARNING

COST FUNCTION

The cost function, also known as the loss function or objective function, measures the difference between the model's predictions and the actual target labels. It quantifies how well the model is performing on the training data. The goal during training is to minimize this cost function. Common cost functions include Mean Squared Error (MSE) for regression problems and Cross-Entropy Loss for classification problems.

OPTIMIZER

The optimizer is an algorithm used to update the parameters of the model (weights and biases) during training to minimize the cost function. It adjusts the parameters based on the gradients of the cost function concerning the model parameters. Gradient Descent is a widely used optimization algorithm, and variations such as Stochastic Gradient Descent (SGD), Mini-batch Gradient Descent, and Adam are commonly employed to efficiently train models.

LINEAR REGRESSION

Linear regression is a statistical method used to model the relationship between a dependent variable (target) and one or more independent variables (features). It assumes a linear relationship between the independent variables and the dependent variable. The simplest form of linear regression is known as simple linear regression, which involves only one independent variable. In more complex scenarios with multiple independent variables, it's called multiple linear regression.

Mathematically, the equation for simple linear regression can be expressed as

$$Y = \theta_1 + \theta_2 x$$