

Machine Learning

| Classification – Regression

Prof. Dr. Mostafa Elhosseini

Professor of Smart Systems Engineering

<https://youtube.com/drmelhosseini>



| Agenda

- ML Classification
- Supervised Learning
 - Classification
 - Regression
- Examples

| Types of ML

Machine learning systems can be classified into several broad categories based on specific criteria:

- Supervision during Training: This includes:
 - **Supervised Learning:** Where models are trained with labeled data.
 - **Unsupervised Learning:** Where models work with unlabeled data.
 - **Semi-Supervised Learning:** A mix of labeled and unlabeled data is used.
 - **Self-Supervised Learning:** Where the model generates its own labels from the input data.
 - **Reinforcement learning:** where the model learns through rewards and penalties.

| Supervised Learning

- The training data you feed to the algorithm includes the desired solutions, called labels
 - Classification
 - Regression
- **Classification:** to predict which category, or class, an input belongs to. It involves training a model on a dataset containing inputs that are already tagged with correct output labels, and the model learns to assign labels to new, unseen inputs based on this training.

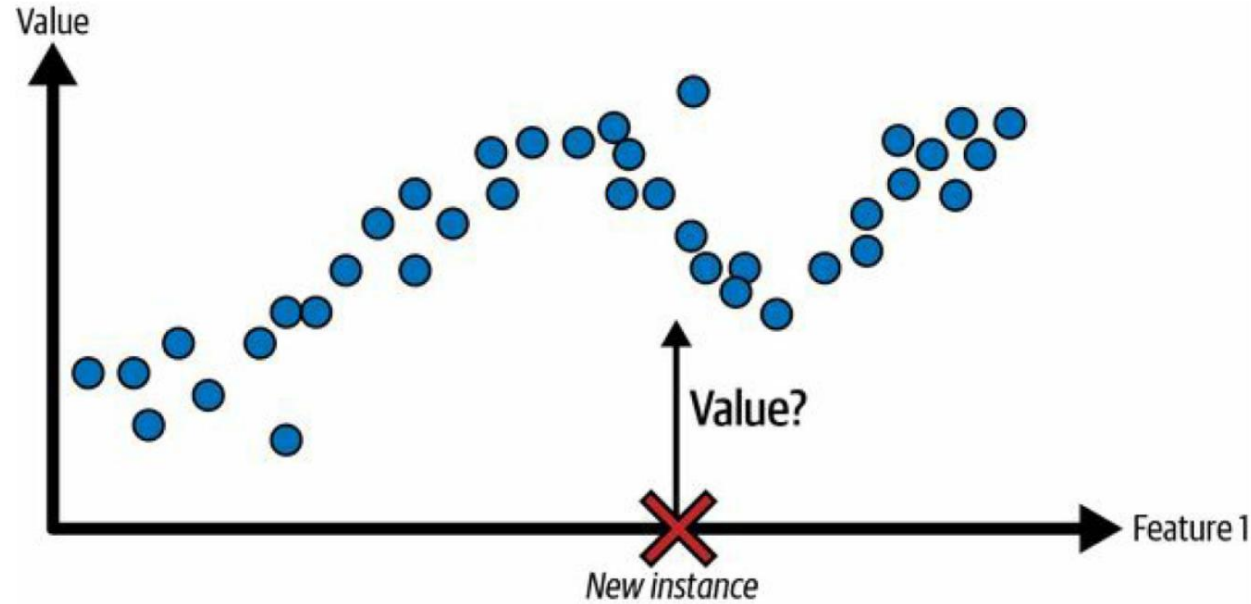
| Classification

- Email Spam Filter: Classifies incoming emails as 'Spam' or 'Not Spam'
- Loan Approval System: Approved / Rejected
- Disease Diagnosis: classifying X-ray images into categories like 'Normal', 'Pneumonia', or 'Tuberculosis'.
- Handwriting Recognition: recognizing handwritten numbers on postal envelopes for automatic sorting
- classifying text data like product reviews into sentiments such as 'Positive', 'Negative', or 'Neutral'

| Classification

- Facial Recognition: Faces are classified into identities based on facial features.
- Credit Card Fraud Detection: Classifying transactions as 'Fraudulent' or 'Legitimate' based on transaction history, user behavior, and other relevant features.
- Image Classification: In computer vision, classifying images into various categories like 'Cat', 'Dog', 'Car', etc., based on the content of the image.
- Customer Segmentation: Classifying customers into different groups or segments based on purchasing behavior, demographics, and other factors for targeted marketing.
- Voice Assistant Understanding: When you interact with voice assistants like Siri or Alexa, they classify your speech into intents or commands to provide appropriate responses.

| Regression



- Predicting a continuous outcome variable (or target) based on one or more predictor variables.
- The goal of regression is to model the relationship between the predictors and the target and use this model to predict the target variable for new, unseen data.

| Regression

- **Real Estate Pricing:** Predicting the market price of houses based on features like location, size, number of bedrooms, and age of the property. This is typically done using linear regression models.
- **Stock Price Forecasting:** Using historical stock market data to predict future stock prices. Time series analysis and regression models can be employed for such predictions.

| Regression

- **Weather Forecasting:** Predicting temperature, humidity, or precipitation levels based on historical weather data and atmospheric conditions using regression analysis.
- **Sales Forecasting:** Estimating future sales for businesses based on past sales data, market trends, and other economic indicators.
- **Salary Estimation:** Predicting an employee's salary based on factors like experience, education level, and industry.

| Regression

- **Life Expectancy Prediction:** Estimating life expectancy or health outcomes based on variables such as diet, exercise, lifestyle choices, and medical history.
- **Demand Forecasting:** Predicting product demand in retail or manufacturing based on factors like seasonality, price, and consumer trends.
- **Energy Consumption Analysis:** Estimating energy usage of households or buildings based on factors like size, location, weather conditions, and occupancy.

| Regression

- **Vehicle Fuel Efficiency Prediction:** Predicting a vehicle's miles per gallon (MPG) based on attributes like engine size, weight, horsepower, and manufacturer.
- **Loan Amount Determination:** Determining the loan amount a person is eligible for based on their credit score, income, employment history, and past loan repayment history.
- **Biometric Analysis:** In sports, predicting athletic performance metrics such as running speed or jump height based on biometric data.
- **Health Risk Assessment:** Estimating the risk of developing certain medical conditions like diabetes or heart disease based on health indicators and lifestyle factors..

| Regression

- **Vehicle Fuel Efficiency Prediction:** Predicting a vehicle's miles per gallon (MPG) based on attributes like engine size, weight, horsepower, and manufacturer.
- **Loan Amount Determination:** Determining the loan amount a person is eligible for based on their credit score, income, employment history, and past loan repayment history.
- **Biometric Analysis:** In sports, predicting athletic performance metrics such as running speed or jump height based on biometric data.
- **Health Risk Assessment:** Estimating the risk of developing certain medical conditions like diabetes or heart disease based on health indicators and lifestyle factors.

| Note:

- Some regression algorithms can be used for classification as well
- Logistic Regression is commonly used for classification, as it can output a value that corresponds to the probability of belonging to a given class (e.g., 20% chance of being spam).
 - classifying inputs based on a threshold on this probability

| Note:

- Imagine a bank wants to decide whether to grant loans to customers based on their credit history. The bank can use logistic regression for this classification task.
- Training: The model is trained on historical customer data, where the inputs include features like income, credit score, employment history, and the output variable is whether they defaulted on a loan (1 for default, 0 for non-default).
- Modeling: Logistic regression is used to estimate the probability of a customer defaulting based on their features.

| Note:

- Decision Making: If the model predicts a probability greater than 0.5 (or some other threshold set by the bank), the customer is classified as high risk (likely to default), and the loan application is rejected. If the probability is lower, the customer is deemed low risk, and the loan is approved.
- This approach is powerful because it not only classifies customers but also quantifies the risk (as a probability) associated with each decision. Logistic regression provides a clear decision-making framework that balances risk against the potential for profit, a crucial aspect in the finance industry.