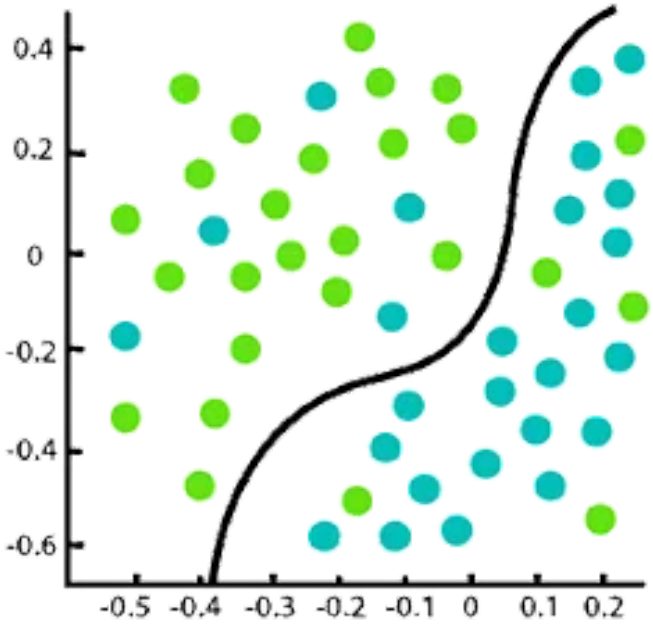
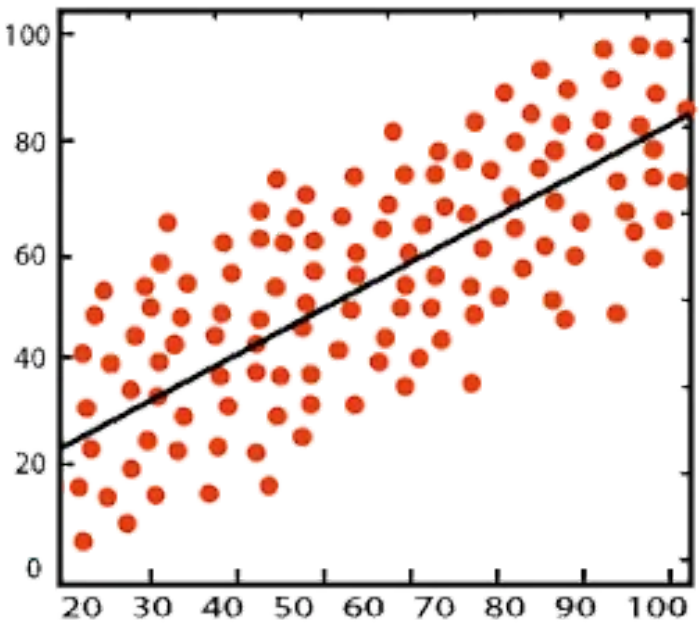
# REGRESSION
*vs*
# CLASSIFICATION

## in Machine Learning Explained



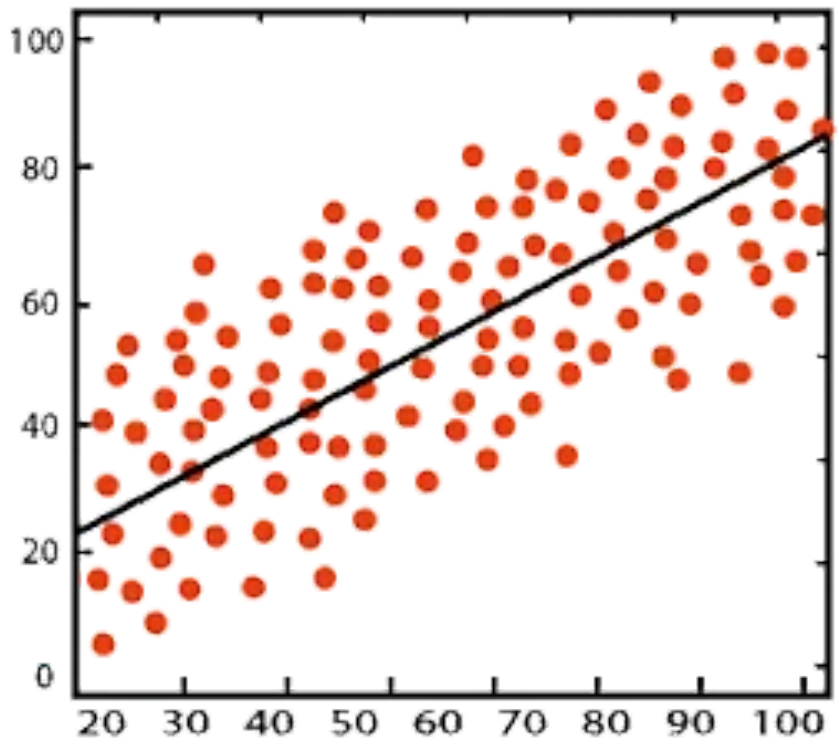

Classification




Regression

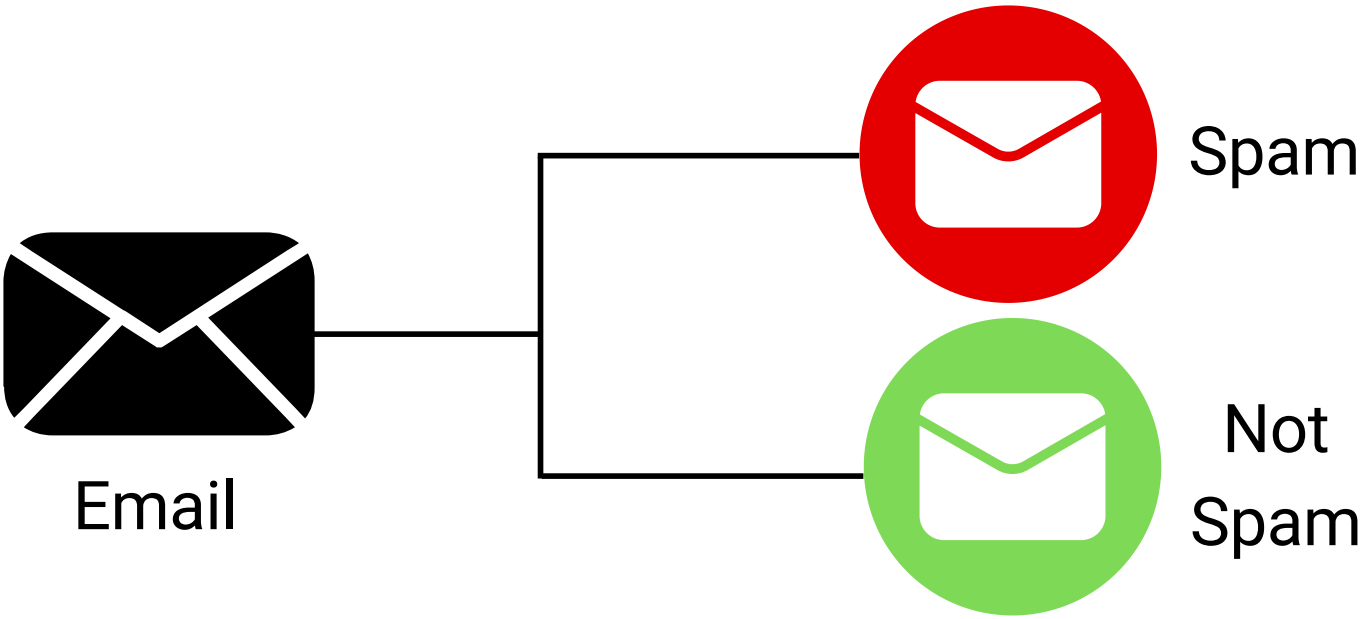
# What is Regression

Regression algorithms predict continuous values from input data, making them ideal for supervised learning tasks. ML engineers and data scientists commonly use them to map estimations with labeled datasets.



# What is Classification?

Classification is a method that categorizes data into distinct classes based on independent features. It uses If-Then rules to derive a mapping function, enabling the classification or prediction of values such as spam/not spam, yes/no, and true/false. For instance, it can predict the likelihood of an actor visiting a mall for a promotion based on historical events, resulting in labels like Yes or No.



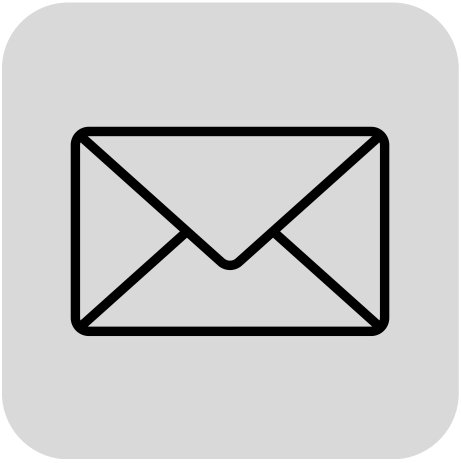
# Application of Regression

Predicting Stock Prices

Sales Forecasting

Real Estate Valuation

# Application of Classification



Email Spam Filtering

Credit Scoring

Image Recognition

# Advantages of Regression

1. **Valuable Insights**: Helps to analyze the relationships between distinct variables and comprehensively understand the data.

2. **Prediction Power**: Prediction of dependent variable values with high accuracy using independent variables.

3. **Flexibility**: Regression is a flexible algorithm used to find or predict models of a wide range, including logistic, linear, polynomial, and many more.

# Disadvantage of Regression

1

**False Assumptions**: There are many assumptions underlying regression algorithms, leading to false assumptions. It includes normality of errors, linearity, and independence.

2

**Overfitting:** Inadequate performance can be applied to new and unseen data because the regression models are overly customized for the training data.

3

**Outliers**: Regression models are sensitive to outliers, thus, can have a significant effect on analyzed prediction results.

# Advantages of Classification

1. **Accuracy in Prediction**: With fitting training, the classification algorithm achieves high accuracy in the model prediction.

2. **Flexible**: Classification algorithms have many applications, like spam filtering, speech, and image recognition.

3. **Scalable Datasets**: Easy to apply in real-time applications that can scale up huge datasets easily.

# Disdvantages of Classification

1

**Bias**: If the training data does not represent the complete dataset, the classification algorithm may get biased with certain trained data.

2

**Imbalanced Data**: If the classes of the datasets are not determined equally, the classification algorithm will read the majority and leave the minority class.

3

**Selection of Features:** Features must be defined in the classification algorithms, else the prediction of data is challenging with multiple or undefined features.

# Difference between Regression and Classification

| Features | Regression | Classification |
| --- | --- | --- |
| Main goal | Predicts continuous values like salary and age. | Predicts discrete values like stock and forecasts. |
| Input and output variables | Input: Either categorical or continuousOutput: Only continuous | Input: Either categorical or continuousOutput: Only categorial |
| Types of algorithm | Linear regressionPolynomial regressionLasso regressionRidge regression | Decision treesRandom forestsLogistic regressionNeural networksSupport vector machines |
| Evaluation metric | R2 scoreMean squared errorMean absolute errorAbsolute percentage error (MAPE) | Receiver operating characteristic curveRecallAccuracyPrecisionF1 score |