Machine Learning

Supervised Learning

Learning Objectives

By the end of this lesson, you will be able to:

- Learn about supervised learning and its applications
- Identify the appropriate way to prepare the data
- Discuss overfitting and underfitting and how to detect and prevent them
- Explain the concept of regularization

Business Scenario

A manufacturing company is using supervised learning techniques to inspect and classify the quality of their products. They have implemented a machine learning model that uses linear regression, decision trees and support vector machines to identify defects and classify the products into different grades based on the level of damage.

The training data used to create the model includes information about the severity of dents, scratches and other factors. The model helps the company automate and optimize their production process, reduce defects and increase the overall quality of their products.

The model also helps the company save time and resources by eliminating the need for manual inspection and grading. The company can use this model to make data-driven decisions, improve their production efficiency and maintain customer satisfaction by delivering high-quality products.

Supervised Learning

Discussion: Supervised Learning

Duration: 10 minutes

- What is supervised learning?
- What are the different types of supervised learning?
- Give an example of supervised learning.

Supervised Learning

In supervised learning, machines are trained with labeled data as input and expected output.

The machine learning model must:

Identify patterns and methods

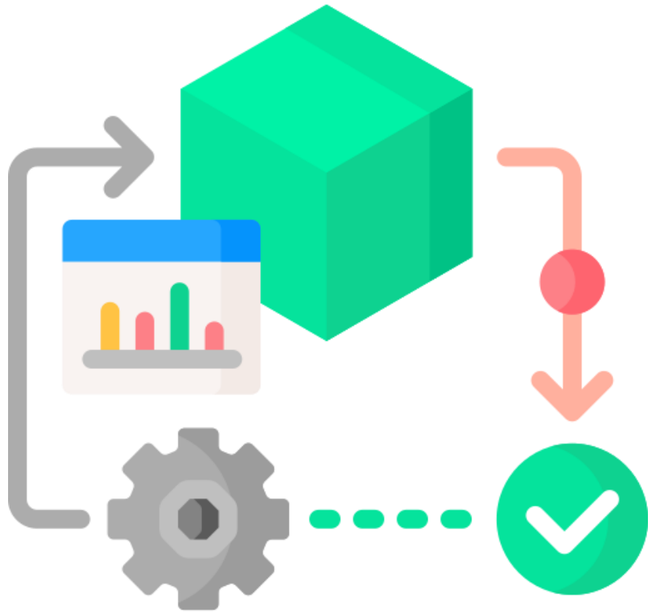
Learn from them

Predict the output for new data

It is a type of machine learning where the input data as well as the correct output data are provided.

Supervised Learning

The model learns from the training dataset by generating predictions on the data iteratively and adjusts for the right output.

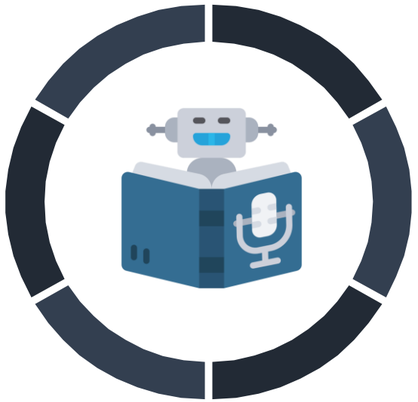


The process continues until the algorithm achieves the highest accuracy possible.

Supervised Learning

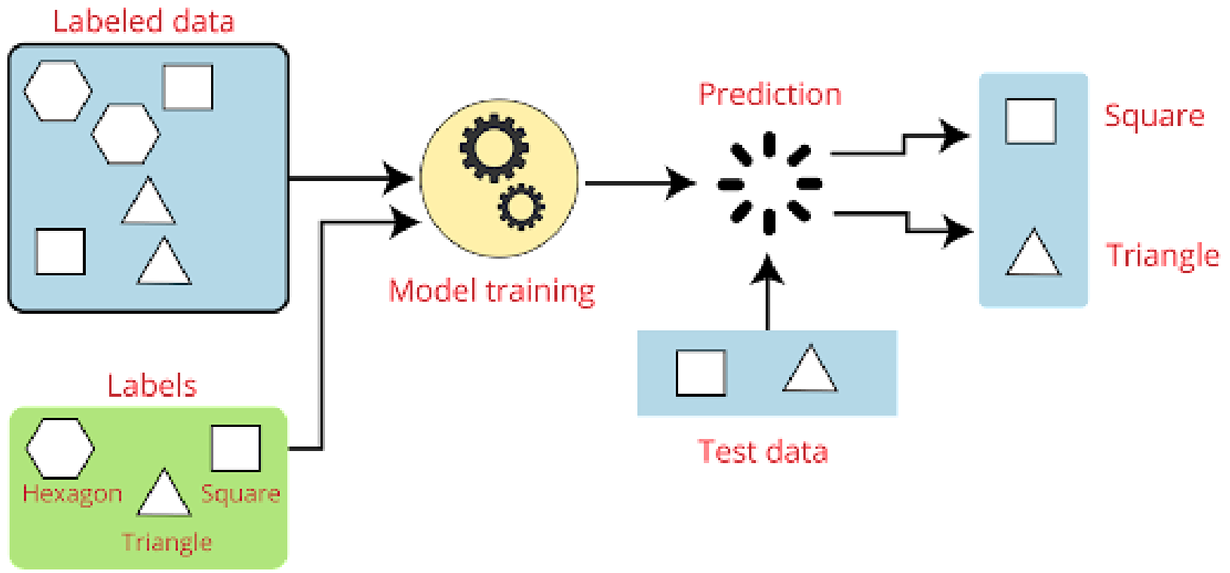
The supervised machine learning algorithm includes:

- Linear and logistic regression
- Multi-class classification
- Decision trees
- Support vector machines



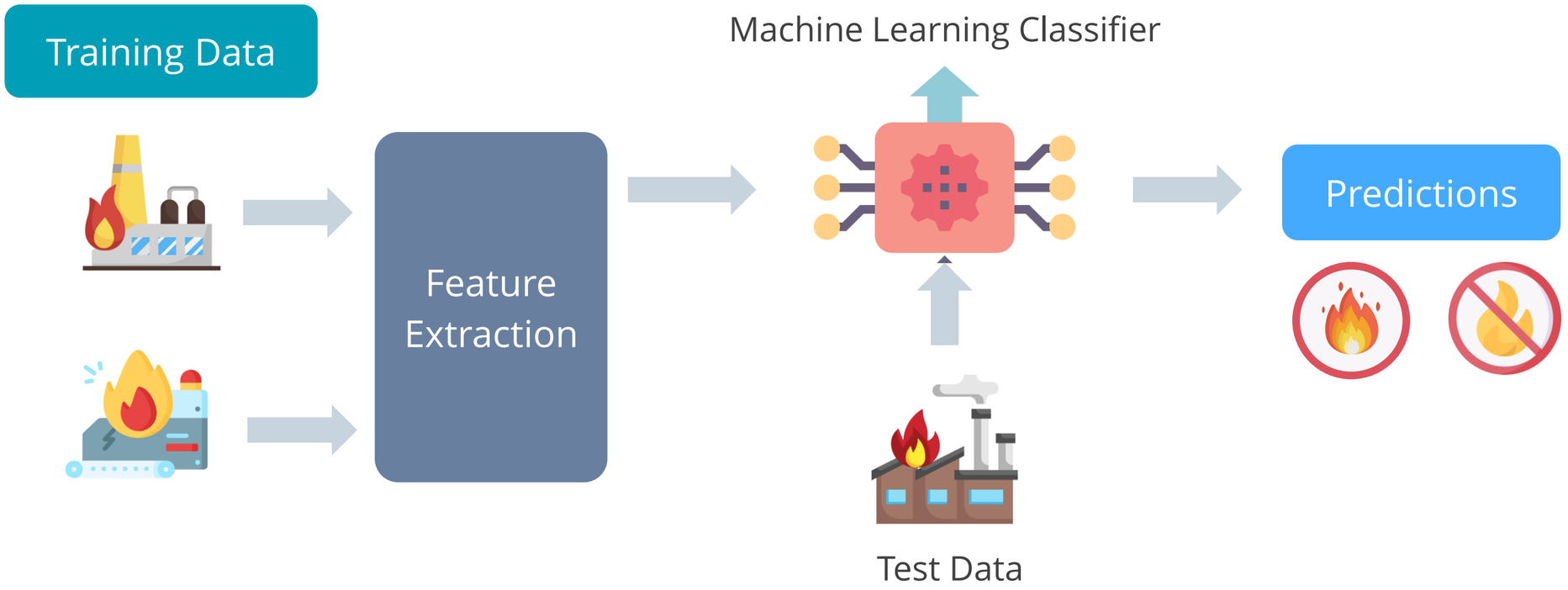
Supervised Learning: Examples

The following is an example of a supervised machine learning model:



Supervised Learning: Examples

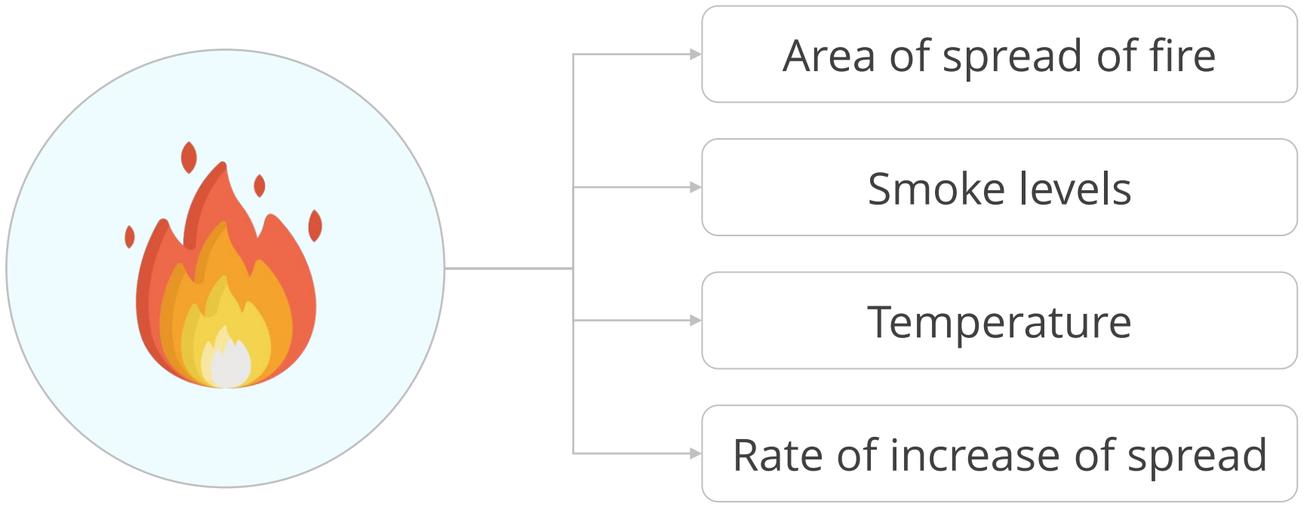
Example 1: Industrial illustration of a model to determine if an incident is a fire incident



It helps detect forest fires, oil and gas tragedies, shipping fires and building fires.

Training Data

The training data to create a classification model for fire incident detection includes:



The output data is binary (1: fire incident, 0: no fire incident).

Training Data

This training data helps:

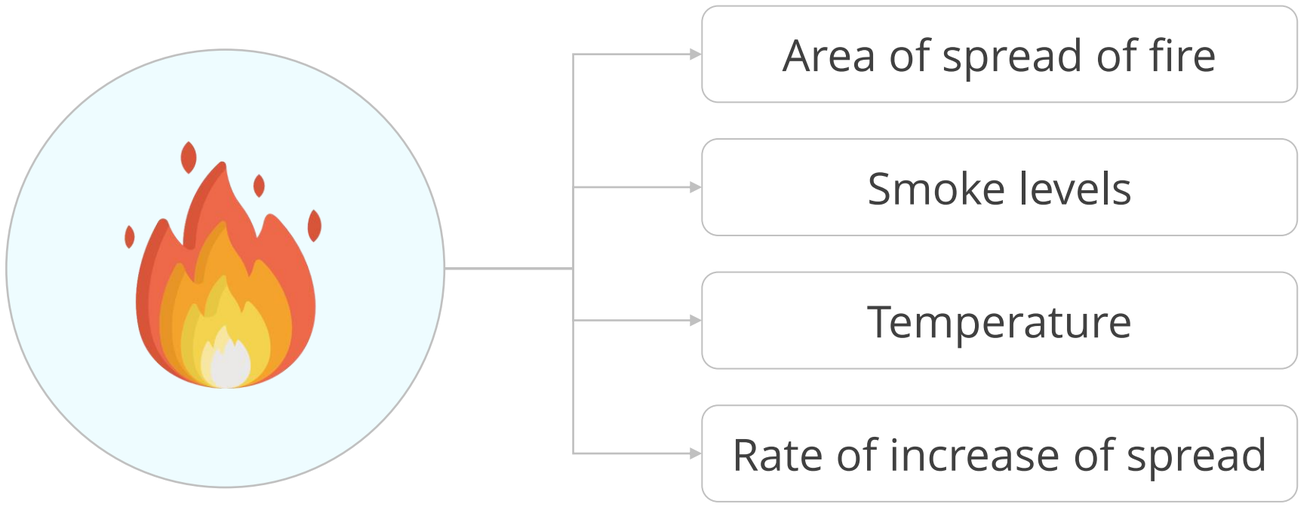
Eliminate false positives like a cooking fire or a campfire being misconstrued as a fire incident

Classify levels of fire incidents

The fire department recognize and classify real life situations and dispatch sufficient resources

Testing Data

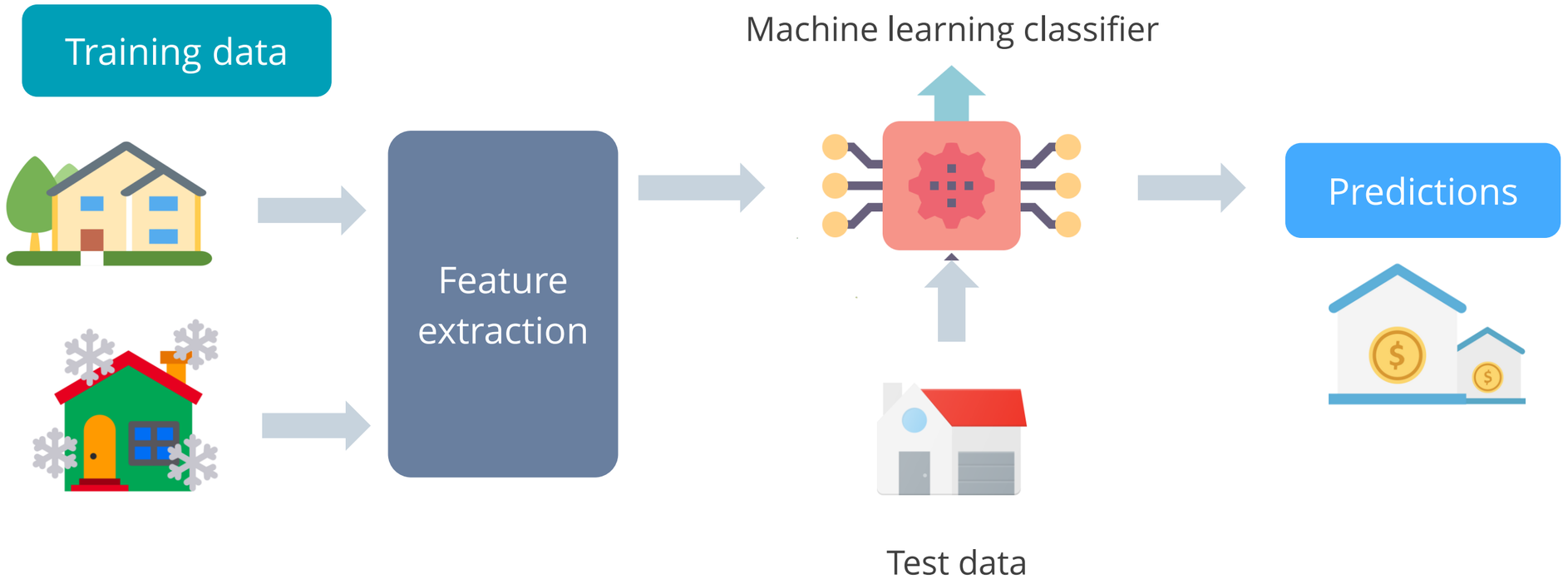
Once the machine learning model finishes training (using training data), it needs to be evaluated using testing data.



The output is labeled as a fire incident or not a fire incident based on the factors given above.

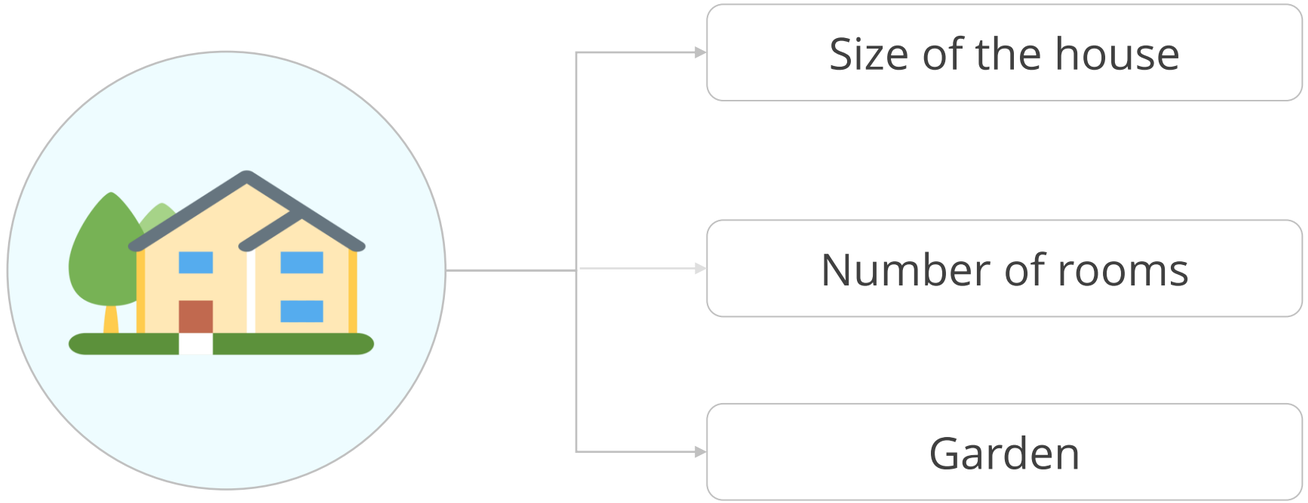
Supervised Learning: Examples

Example 2: predicting the house prices



Training and Testing Data

Here, the training data creates the classification model for house price prediction which includes:



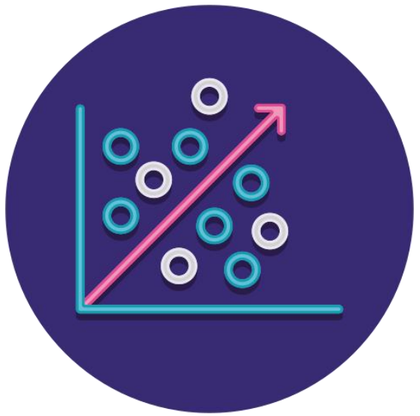
The output predicts the price of a house based on the factors given above.

Once the model is built, testing data is used to test the accuracy of the model.

Supervised Learning Algorithms

There are two types of supervised learning algorithms:

Classification



Regression

Classification Algorithm

It segregates data into two or more categories, given one or more inputs.

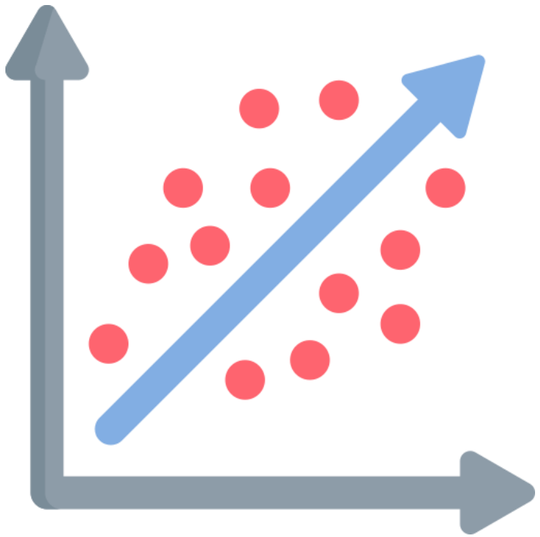
Example

To segregate the emails into spam or ham, the model first examines the data to find patterns and methods.

A classification model predicts the value of one or more outcomes.

Regression Algorithm

It attempts to establish a connection between the input and output variables and is appropriate for circumstances in which the output variable has a real or continuous value.



Example

A regression algorithm estimates a car’s price based on features like car size, brand, and mileage.

A regression algorithm is ideal for forecasting, time series analysis, and finding the cause-and-effect relationship between variables.

Discussion: Supervised Learning

Duration: 10 minutes

- What is supervised learning?

In supervised learning, the machine learns under supervision. It contains a model that is able to predict with the help of a labeled dataset. In a labeled dataset, the target answer is already known.

- What are the different types of supervised learning?

There are two types of supervised learning: regression and classification.

- Give an example of supervised learning.

Following are the examples of supervised learning,

  - Predicting the price of a house.
  - Predicting the price of gold.
  - Predicting whether a person is suffering from cancer or not.
  - Predicting whether a mail is spam or ham.

Applications of Supervised Learning

Discussion: Applications of Supervised Learning

Duration: 5 minutes

- What are the different applications of supervised learning?

Supervised Learning

Supervised learning can optimize and automate processes across industries.

It enhances decision-making, reduces manual effort and enables more accurate predictions.

Applications of Supervised Learning

1. HR operations: It is used to find the right candidates for job vacancies.

Supervised learning can help a machine to shortlist resumes based on accepted or rejected candidate profiles, improving its ability to recommend similar profiles in the future.

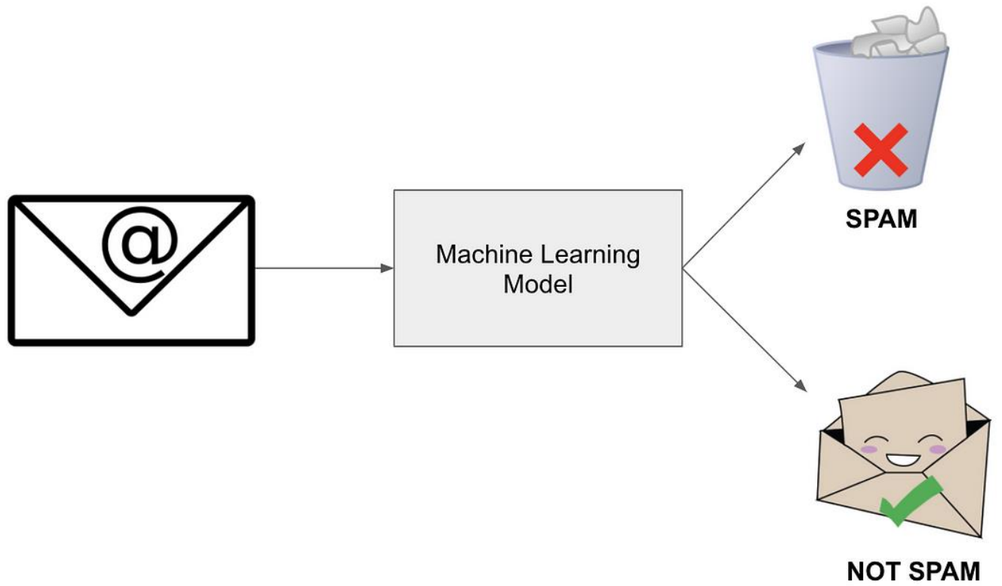
Applications of Supervised Learning

2. Finance: It is used to segregate good and bad loans.

Supervised learning can automate loan approval and improve loan risk assessment by using historical loan data to train a model.

Applications of Supervised Learning

3. Email: Spam filters for emails



Supervised learning helps by training a machine learning model using labeled examples of spam and non-spam emails to make accurate predictions on new, unseen emails.

Applications of Supervised Learning

Quality inspection in manufacturing

Forecasting for the maritime industry

Fraud protection measures

Waste management

Quality Inspection in Manufacturing

Supervised learning is used in manufacturing and production engineering to:

- Inspect quality
- Classify the manufactured products into grades

Quality Inspection in Manufacturing

Supervised learning can also be used to detect the level of damage in products.

Example

Grading of shrimps as A, B and C based on disease and discoloration

Example

Classifying manufacturing defects in an aircraft based on severity of dents, scratches and other factors

These can be automated with the help of supervised learning.

Forecasting in the Maritime Industry

The maritime industry is critical for global supply chains, and any disruption to it can have significant impacts on global trade.

The Suez Canal being blocked in 2021 impacted global trade.

Forecasting in the Maritime Industry

Forecasting is done based on a combination of historical events and weather conditions, which helps develop precautionary incident management.

For example, a regression model developed with supervised learning can:

- Use past data to predict tidal currents
- Forecast demand and supply based on past consumer records, thus reducing inventory losses.

Fraud Protection Measures

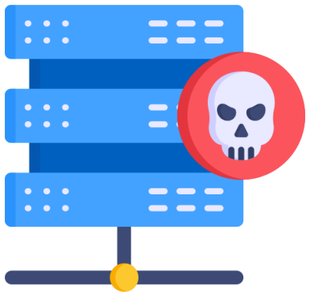
Supervised learning models help detect and prevent fraud.

Email clients mark certain mails as spam with the help of a learning model.

Fraud Protection Measures

To predict fraud, the model learns from data such as:

Financial transactions with fraud patterns



Cyber attacks on servers

Fake social media profiles

Waste Management

Supervised learning is used to improve waste management practices.

It helps:

1. Classify waste into dry, wet and e-waste
2. Improve composting activities and optimize e-waste recycling
3. Categorize customers based on environmental consciousness
4. Educate potential consumers based on their e-commerce patterns

Discussion: Applications of Supervised Learning

Duration: 5 minutes

- What are the different applications of supervised learning?

  Supervised learning can be used in the following domains:
  - HR operations
  - Finance
  - Maritime
  - Manufacturing

Assisted Practices

Let’s understand the topic below using Jupyter Notebook.

- 3.04_Preparing and Shaping Data

Note: Please download the pdf files for each topic mentioned above from the Reference Material section.

ASSISTED PRACTICE

Overfitting and Underfitting

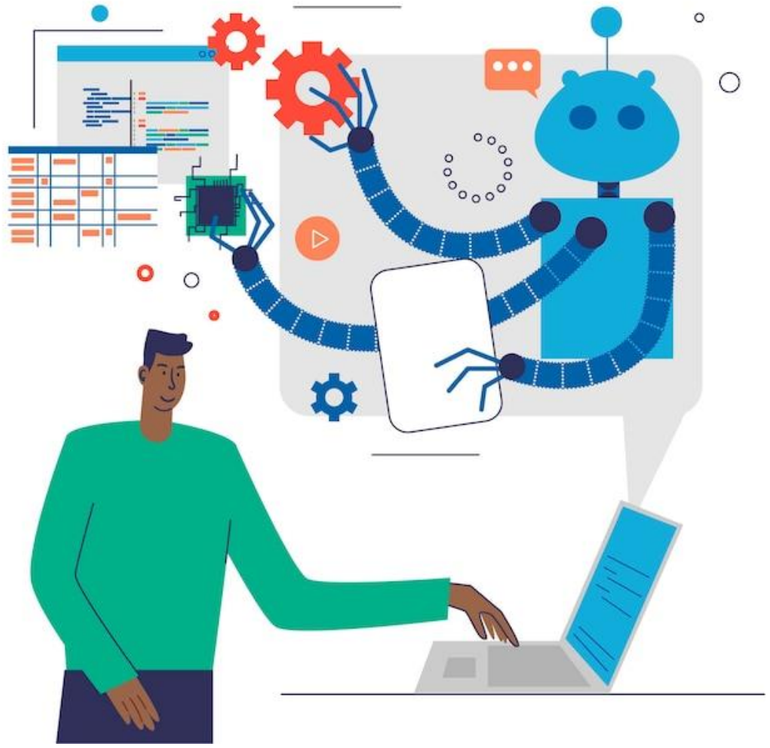
Discussion: Overfitting and Underfitting

Duration: 5 minutes

- What is overfitting and underfitting?
- What is bias and variance?

Overfitting and Underfitting

Overfitting and underfitting define how well machine learning models are learning and applying what they learned.



Bias

Bias is an error introduced in the model.

High bias means that there is a big difference between the actual and predicted values.

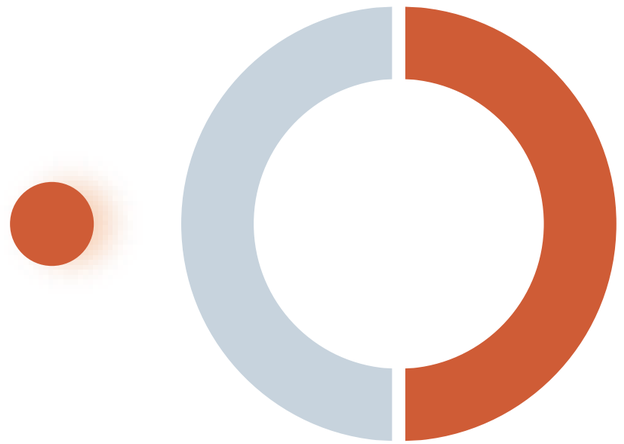
Low bias indicates that the difference between the actual and predicted values is low.

High bias is not good for the model.

Variance

Variance indicates how scattered the data is.

High variance indicates more scattered data.



Low variance indicates less scattered data.

Overfitting

The data is overfitted when there is low bias and high variance.

Example

If you train a machine learning model that predicts climate change to consider only -2°F as cold temperature, it is an overfit and the model will not consider any other temperature as cold and will not make accurate predictions.

Overfitting happens when a model focuses on too many details in the training data set to the extent that it negatively impacts the performance of the model on a new data set.

Underfitting

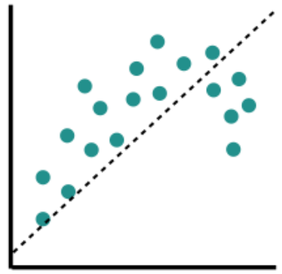
The data is said to be underfitted when there is high bias and high variance.

Example

If a model to predict climate change is trained with limited features like temperature and wind speed, it is an underfit. If the model encounters other climate conditions like rain or snow, it won't be able to detect patterns.

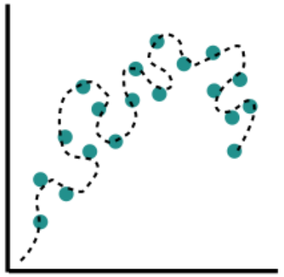
Underfitting is easily detectable as it will exhibit poor performance even on the training data set.

Overfitting and Underfitting



If a model performs well on training data as well as on testing data, it is said to be a good model.

If the model performs well with training data, but not with testing data, it is an overfit.



If the model does not perform well on both training data and testing data, it is an underfit.

Discussion: Overfitting and Underfitting

Duration: 5 minutes

- What is overfitting and underfitting?

Overfitting is a situation when the model performs well on training data and not on testing data. Underfitting is a situation when the model does not perform well on training data and testing data.

- What is bias and variance?

Bias is a training error, and variance is a testing error.

Assisted Practices

Let’s understand the topic below using Jupyter Notebook.

- 3.06_Detecting and Preventing: Overfitting and Underfitting

Note: Please download the pdf files for each topic mentioned above from the Reference Material section.

ASSISTED PRACTICE

Regularization

Discussion: Regularization

Duration: 5 minutes

- What is regularization?
- What are the different ways for regularization in machine learning?

Regularization

It is a form of regression that shrinks the coefficient toward zero to reduce the complexity of the data.

It reduces the variance of the model without an increase in bias.

It discourages more complex models and prevents overfitting.

Example: In autonomous vehicles, neural networks uses regularization for image classification to prevent overfitting, enhance accuracy and recognize objects in new and unseen images captured by the sensors.

Regularization

Fitting involves a loss function known as the residual sum of squares (RSS).

RSS = Actual value - Predicted value

$$RSS = \sum_{i=1}^{n} (yi - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2$$

In this equation:

- y is the independent variable
- β is the regression coefficient value
- x is the dependent variable

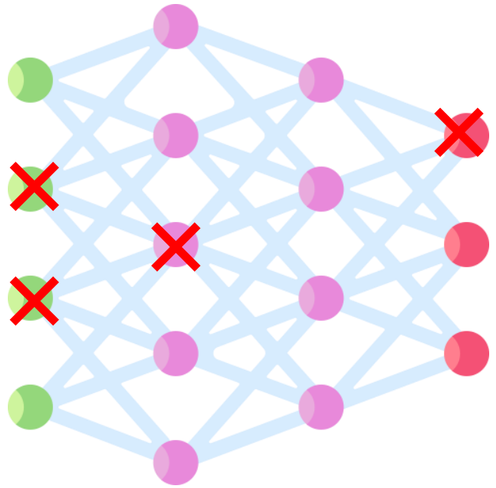
Types of Regularizations

The different types of regularizations are:

1. Dropout regularization
2. Early stopping
3. Coadaptation
4. Lasso regression
5. Ridge regression

Types of Regularizations: Dropout Regularization

Dropout regularization works by removing a random selection.

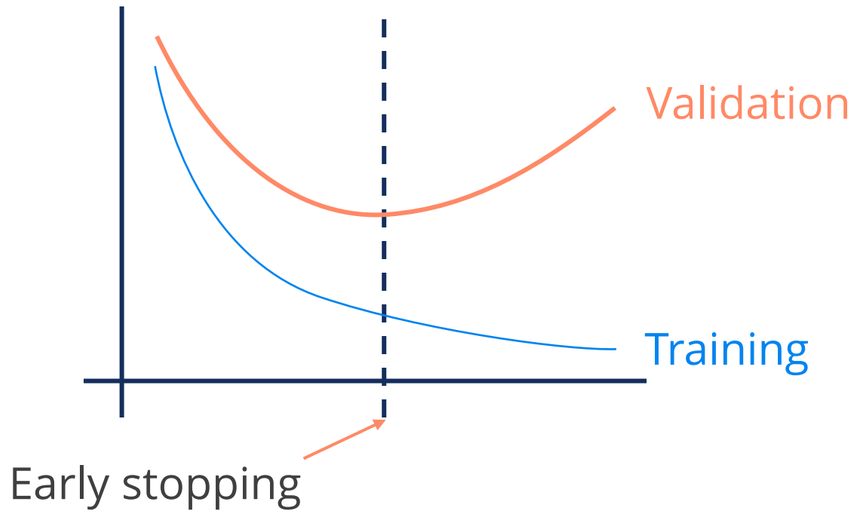


The more units dropped, the stronger the regularization.

This type of regularization is good for training neural networks.

Types of Regularizations: Early Stopping

Early stopping is a second type of regularization where the user uses a large number of epochs and plots a validation loss graph.

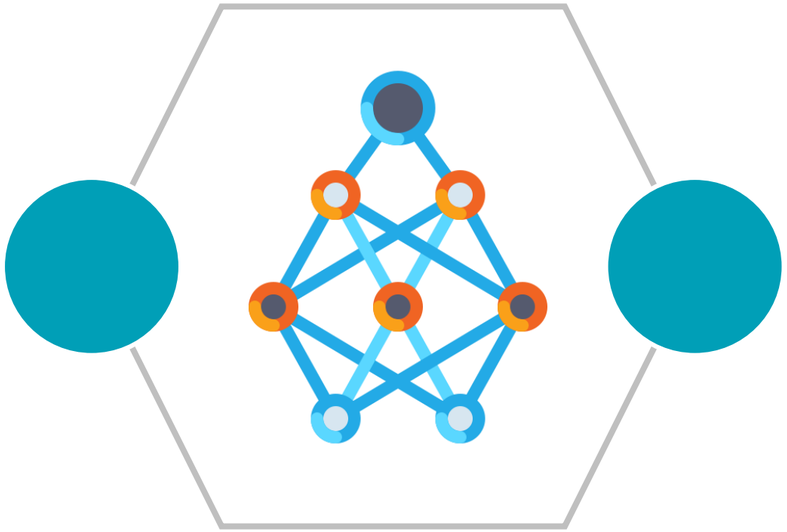


The training stops and the model is saved when the validation loss moves from decreasing to increasing.

Types of Regularizations: Coadaptation

In coadaptation, neurons predict patterns in the training data by:

Relying almost exclusively on the output of specific neurons

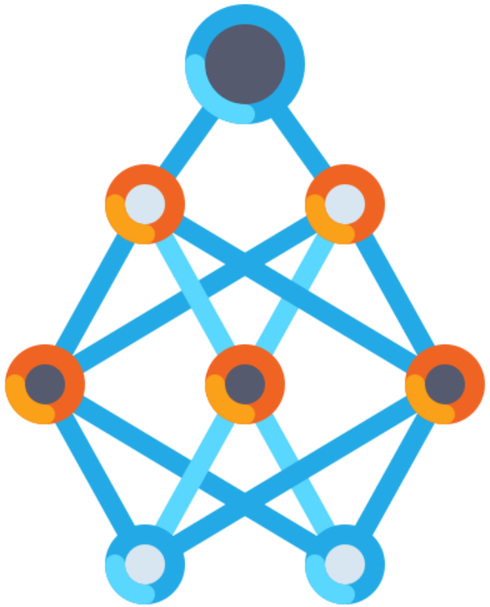


Not relying on the overall network behavior

If the validation data does not contain the patterns that cause coadaptation, it could cause overfitting.

Types of Regularizations: Coadaptation

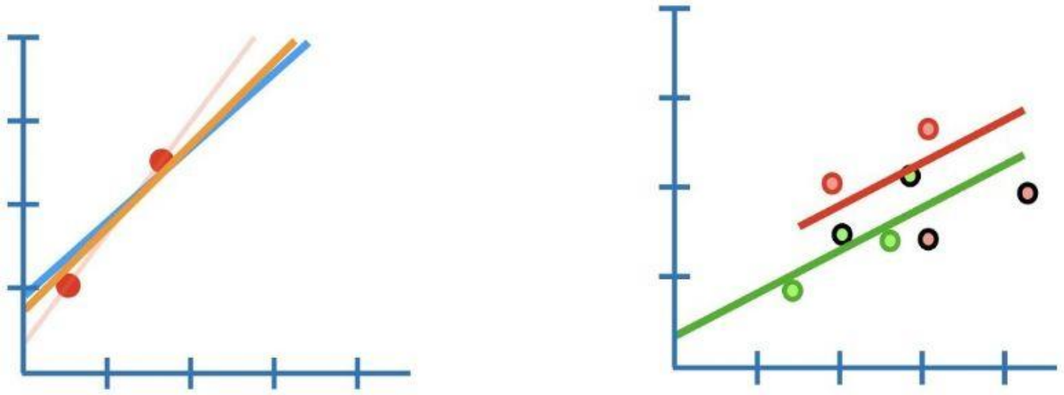
Dropout regularization reduces coadaptation.



It ensures neurons cannot rely solely on other neurons.

Types of Regularizations: Lasso Regression

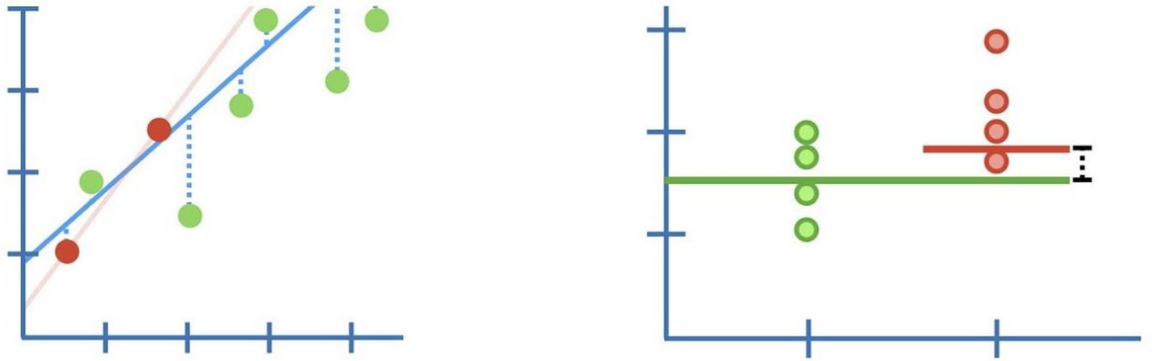
Lasso regression (L1 regularization) penalizes weights in proportion to the sum of the absolute values of the weights.



It helps drive the weights of irrelevant or barely relevant features to exactly zero, which removes those features from the model.

Types of Regularizations: Ridge Regression

Ridge regression (L2) regularization shrinks coefficients close to zero for those predictors that are unimportant.



The final model will include all the predictors.

Discussion: Regularization

Duration: 5 minutes

- What is regularization?

It is a technique used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.

- What are the different ways for regularization in machine learning?

Following are the different ways for regularization

  - Ridge Regression
  - Lasso Regression
  - Elastic Net Regression

Key Takeaways

- Supervised learning is a machine learning model that identifies patterns and methods and learns from them.
- Supervised learning gives predictions on a new set of data.
- The two types of supervised learning algorithms are classification and regression.
- Overfitting and underfitting define how well machine learning models are learning and applying what they learnt.

Knowledge Check

Knowledge Check

1

What is supervised learning?

A. A type of machine learning where input data is not labeled

B. A type of machine learning where input data is labeled

C. A type of machine learning where output data is labeled

D. A type of machine learning where neither input nor output data is labeled

Knowledge Check

1

What is supervised learning?

A. A type of machine learning where input data is not labeled

B. A type of machine learning where input data is labeled

C. A type of machine learning where output data is labeled

D. A type of machine learning where neither input nor output data is labeled

The correct answer is **B**

Supervised learning is a type of machine learning where input data is labeled.

Knowledge Check

2

What are the two types of supervised learning algorithms?

A. Linear and logistic regression

B. Multi-class classification and decision trees

C. Classification and regression

D. Support vector machines and deep learning

Knowledge Check 2

What are the two types of supervised learning algorithms?

A. Linear and logistic regression

B. Multi-class classification and decision trees

C. Classification and regression

D. Support vector machines and deep learning

The correct answer is **C**

The two types of supervised learning algorithms are classification and regression.

Knowledge Check

3

What is overfitting?

A. A situation where the model focuses on too many details in the training data set

B. A situation where the model negatively impacts the performance on a new data set

C. A situation where there is high bias and high variance in the data

D. A situation where the model performs well on training data but not on testing data

Knowledge Check

3

What is overfitting?

A. A situation where the model focuses on too many details in the training data set

B. A situation where the model negatively impacts the performance on a new data set

C. A situation where there is high bias and high variance in the data

D. A situation where the model performs well on training data but not on testing data

The correct answer is **D**

Overfitting is a situation where the model performs well on training data but not on testing data.

Thank You