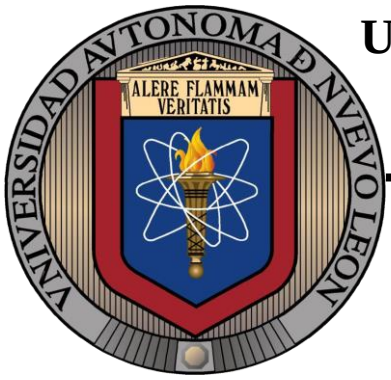


UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

Facultad de Ingeniería Mecánica y Eléctrica

Artificial Intelligence

Assignment 3

Machine Learning mind map

Student ID: 2109461

Name: Jocelyn Benítez Ramírez

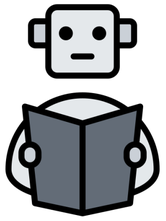
Career: Biomedical Engineering

Educator: Daniel Isaías López Páez
Group: 002
Day: Thursday
Hour: N4
Semester: January – June 2024

April 12, 2024

San Nicolás de los Garza, Nuevo León

MACHINE LEARNING



Is the science and art of programming computers so they can learn from data.

Field of study that gives computers the ability to learn without being explicitly programmed.

WHAT IS?

APPLICATIONS

Healthcare

Medical image analysis for detecting abnormalities in X-rays, MRI scans, and CT scans.

Finance

Algorithmic trading and stock price prediction

Fraud detection in banking and financial transactions.

Natural Language Processing

Chatbots and virtual assistants for customer service.

TESTING

Process of evaluating the performance of a trained model on a dataset that was not used during the training phase.

- The data is split into two sets: the training set and the test set.
  The model is train using the training set, and the it will be test it using the test set.
- The error rate on new cases is called the generalization error.
  This value tells you how well your model will perform on instances it has never seen before.

VALIDATING

Process of assessing the performance of a model during the training phase.

- Holdout validation: hold out part of the training set to evaluate several candidate models and select the best one.
- The new heldout set is called the validation set.

The training set is used to train the model, while the validation set is used to tune hyperparameters and monitor the model's performance during training.

TYPES OF MACHINE LEARNING

Supervised/Unsupervised Learning

Whether or not they are trained with human supervision. There are four major categories

Supervised learning
The training data you feed to the algorithm includes the desired solutions, called labels.

Unsupervised learning
The training data is unlabeled. The system tries to learn without a teacher.

Semisupervised learning
It can deal with partially labeled training data, usually a lot of unlabele data and a little bit of labeled data.

Reinforcement Learning
The learning system, called an agent in this context, can observe the environment, select and perform actions, and get rewards in return

Batch and Online Learning

Whether or not they can learn incrementally on the fly.

Batch learning
It is incapable of learning incrementally: it must be trained using all the available data.
This will generally take a lot of time and computing resources, so it is typically done offline.

Online learning
They train the system incrementally by feeding it data instances sequentially, either individually or by small groups called mini-batches.
Each learning step is fast and cheap, so the system can learn about new data on the fly, as it arrives.

Instance-Based Versus Model-Based Learning

Whether they work by simply comparing new data points to known data points, or instead detect patterns in the training data and build a predictive model, much like scientists do

Instance-based learning
The system learns the examples by heart, then generalizes to new cases by comparing them to the learned examples using a similarity measure.

Model-based learning
Another way to generalize from a set of examples is to build a model of these examples, then use that model to make predictions.

MAIN CHALLENGES

⚠ Insufficient Quantity of Training Data
Even for very simple problems you typically need thousands of examples, and for complex problems such as image or speech recognition you may need millions of examples

⚠ Overfitting the Training Data
The model performs well on the training data, but it does not generalize well.

⚠ Underfitting the Training Data
The model is too simple to learn the underlying structure of the data

⚠ Nonrepresentative Training Data
Use a training set that is representative of the cases you want to generalize to.
If the sample is too small, you will have sampling noise but even very large samples can be nonrepresentative if the sampling method is flawed. This is called sampling bias.

⚠ Irrelevant Features
Garbage in, garbage out.
Your system will only be capable of learning if the training data contains enough relevant features and not too many irrelevant ones.

⚠ Poor-Quality Data
Training with full of errors, outliers, and noise will make it harder for the system to detect the underlying patterns, so your system is less likely to perform well.

References

Géron, A. (2022). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. " O'Reilly Media, Inc.”