Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

The Machine Learning Landscape

A Basic Introduction to Machine Learning

Dr. Muhammad Usman Ghani
Associate Professor, Computer Science & Engineering

University of Engineering and Technology, Lahore, Pakistan

Overview

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- What Is Machine Learning?
- Why Use Machine Learning?
- Machine Learning Systems
- Machine Learning Challenges
- Testing and Validation

What Is Machine Learning?

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- **Machine Learning** is the field of study that gives computers the ability to learn without being explicitly programmed.
- **Training set:** The examples that the system uses to learn are called the training set.
- **Training instance:** Each training example is called a training instance (or sample).
- For example: A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.
  - Experience 'E' is the training data
  - particular performance measure is called **Accuracy**.

What Is Machine Learning?

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- Problems for which existing solutions require a lot of hand-tuning or long lists of rules: one Machine Learning algorithm can often simplify code and perform better.
- Complex problems for which there is no good solution at all using a traditional approach: the best Machine Learning techniques can find a solution.
- **Fluctuating environments:**A Machine Learning system can adapt to new data.
- Getting insights about complex problems and large amounts of data.

Why Use Machine Learning?

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

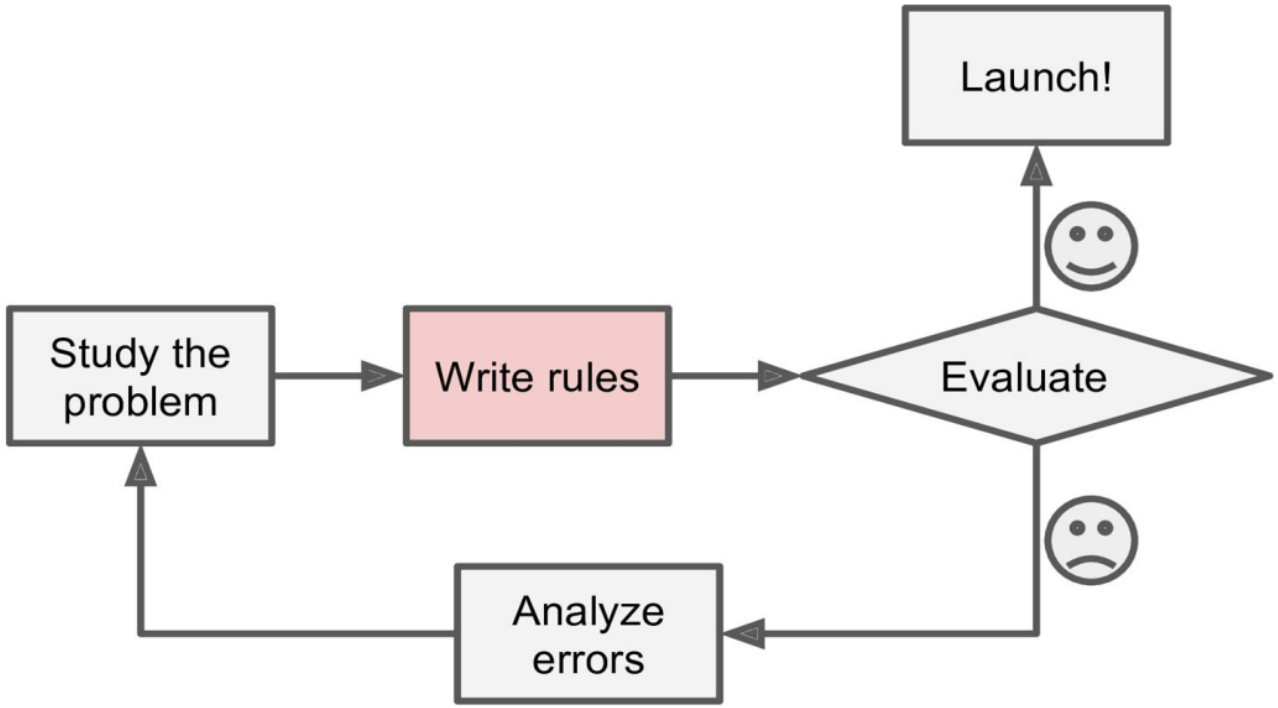


Figure 1.1 The traditional approach

Why Use Machine Learning?

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

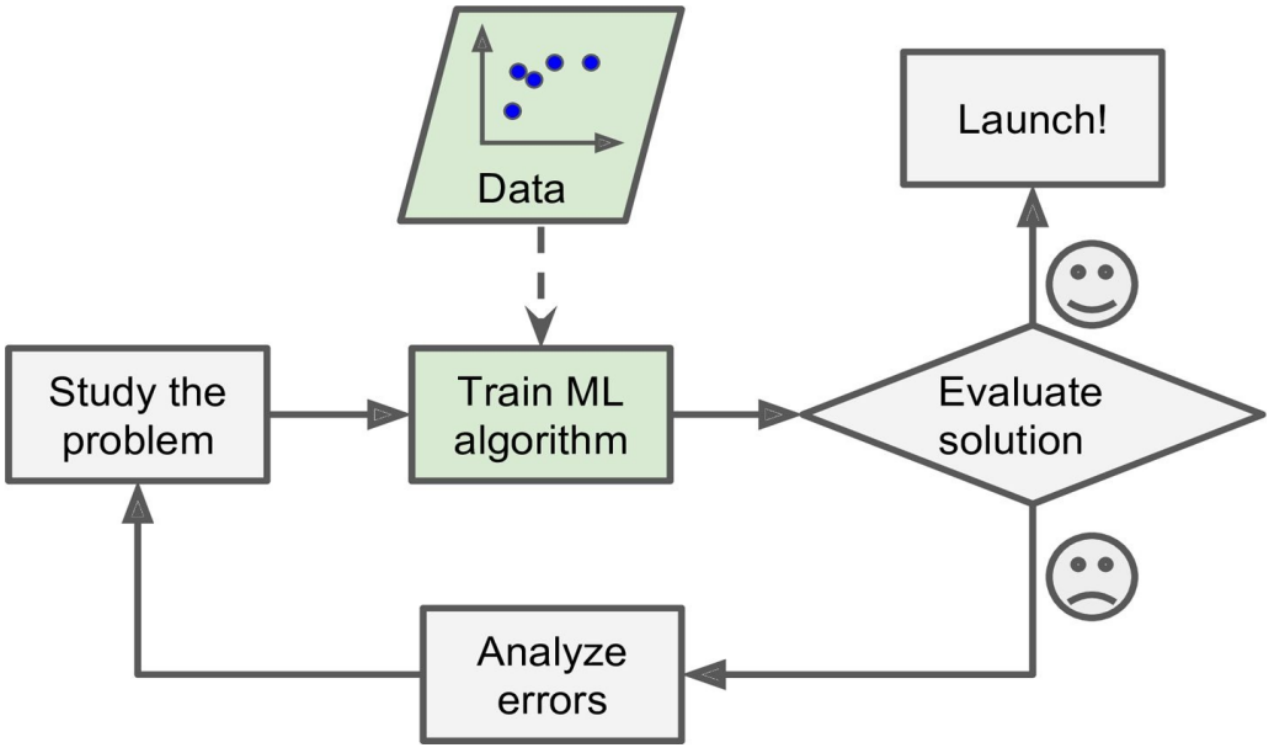


Figure 1.2. Machine Learning approach

Why Use Machine Learning?

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

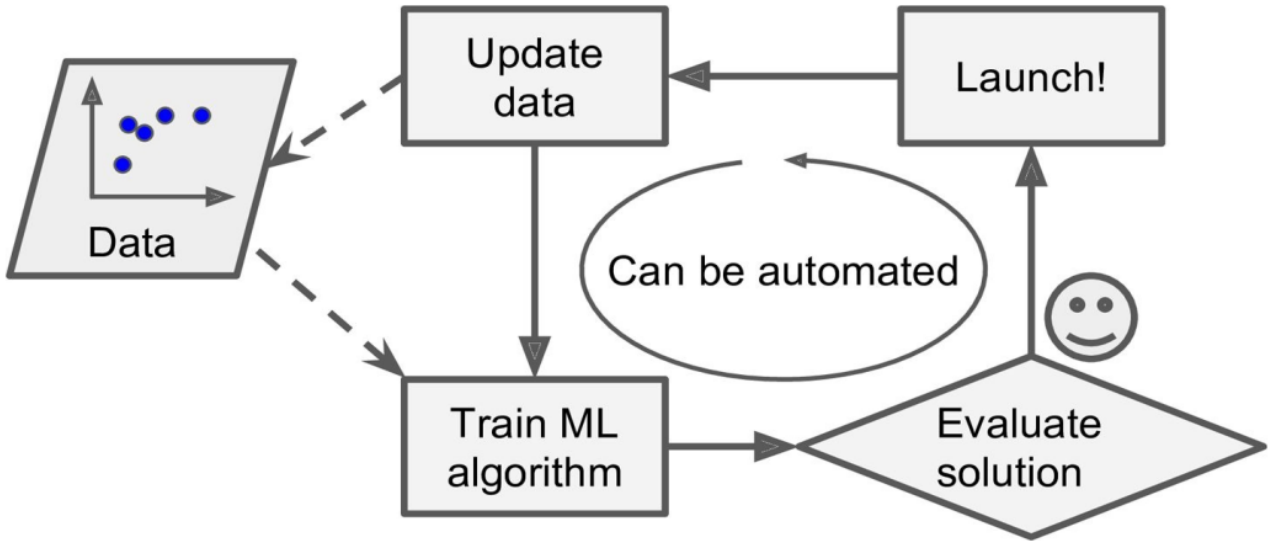


Figure 1.3. Automatically adapting to change

Why Use Machine Learning?

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

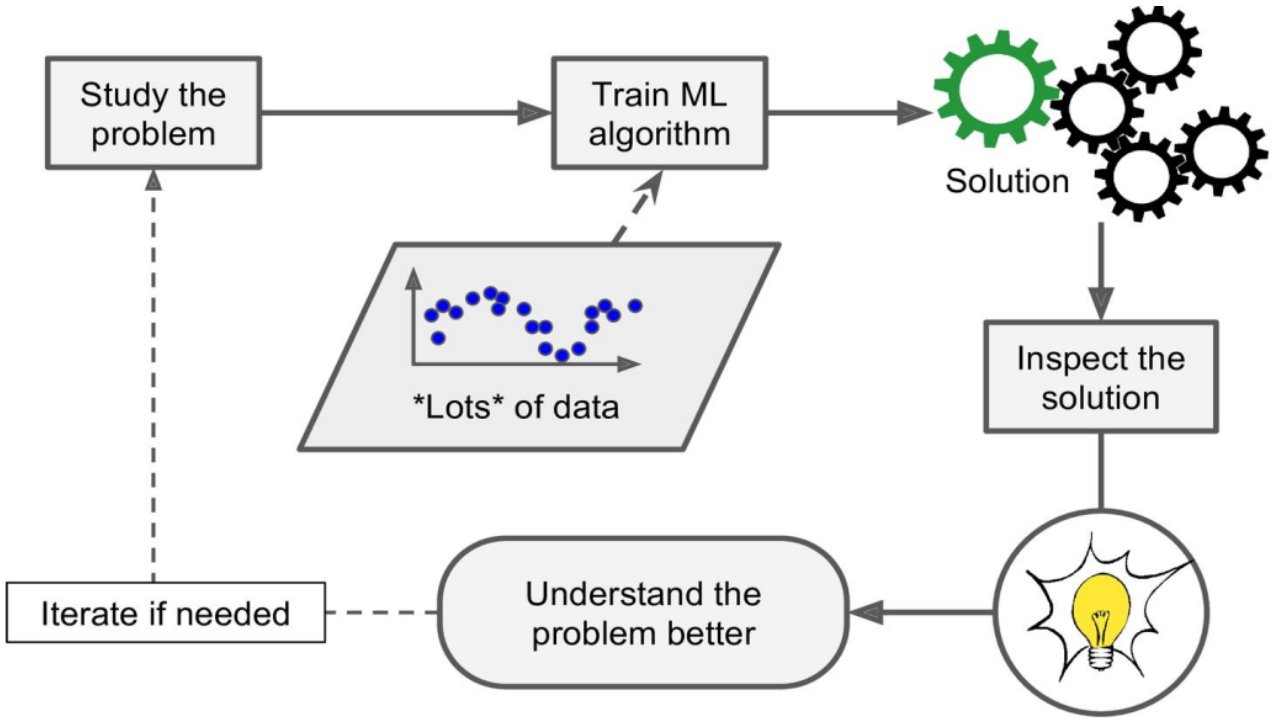


Figure 1.4. Machine Learning can help humans learn

What Machine Learning?

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- Machine learning is needed for tasks that are too complex for humans to code directly.
- Some tasks are so complex that are impractical and impossible for humans to code explicitly.
  - Sentiment analysis data in the form of 200 documents, 200 words per doc
  - Spam and non-spam 200 emails of size 10 lines & 2 images
  - Looking for activities in a CCTV footage.

Why Use Machine Learning?

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

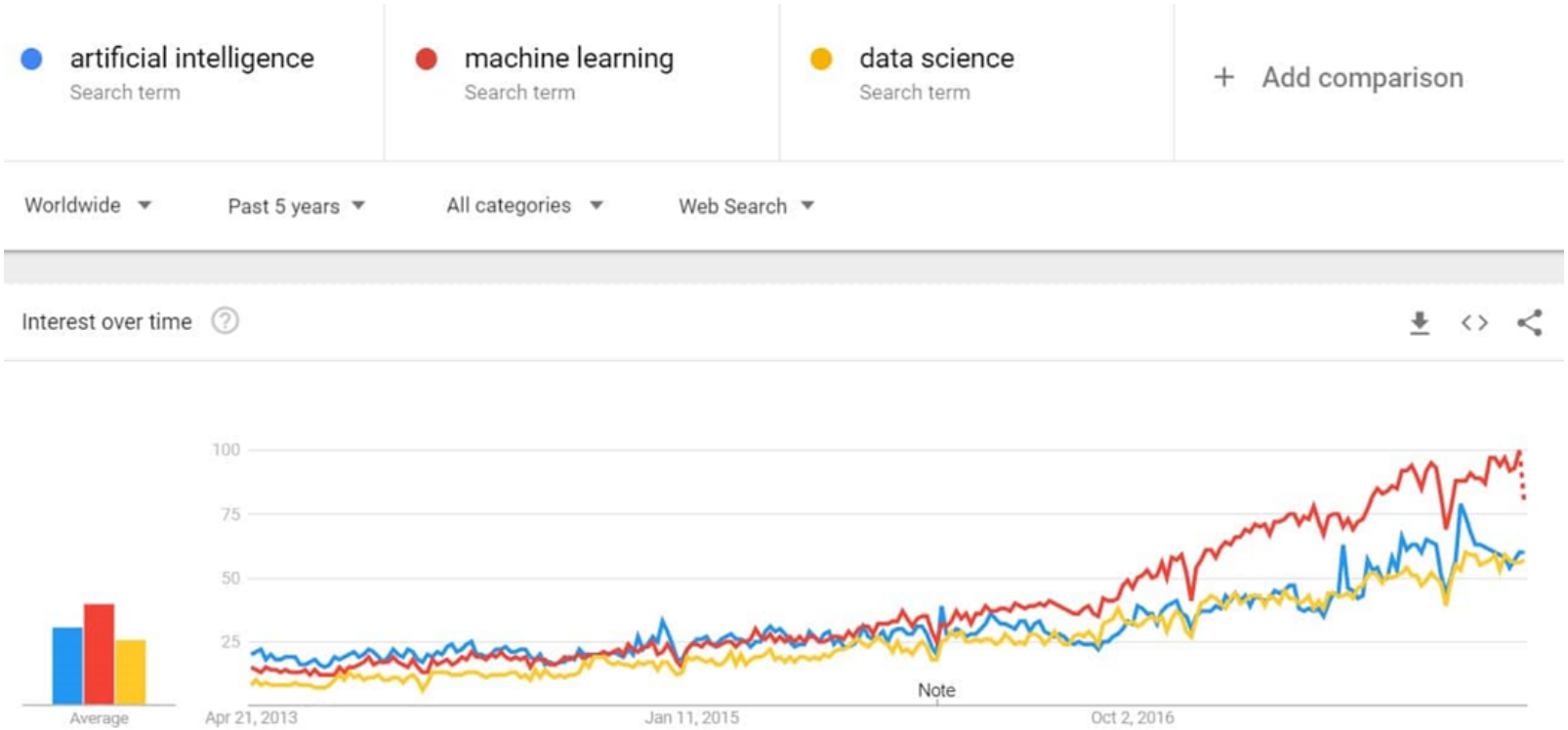


Figure 1.4. Machine Learning can help humans learn

Applications of Machine Learning?

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- Virtual Personal Assistants (Siri, Alexa, Google)
- Videos Surveillance (ASSVS)
- Social Media Services
- People You May Know
- Face Recognition
- Email Spam and Malware Filtering
- Product Recommendations (Amazon, Alibaba)
- Online Fraud Detection (Paypal using ML for protection against money laundering)

Machine Learning Systems

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- Exists different types of Machine Learning systems; it is useful to classify them in broad categories.
- Three major types are:
  - Supervised / Unsupervised Learning ()
  - Online versus batch learning (when incrementally on fly)
  - Instance-based versus model-based learning

Supervised / Unsupervised Learning

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- Machine Learning systems can be classified according to the amount and type of supervision they get during training.
- There are four major categories:
  - Supervised Learning
  - Unsupervised Learning
  - Semi supervised Learning
  - Reinforcement

Supervised learning

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- In supervised learning, the training data feed to the algorithm includes the labels.
- A typical supervised learning task is classification.
- Spam email classification is the example of Supervised learning.

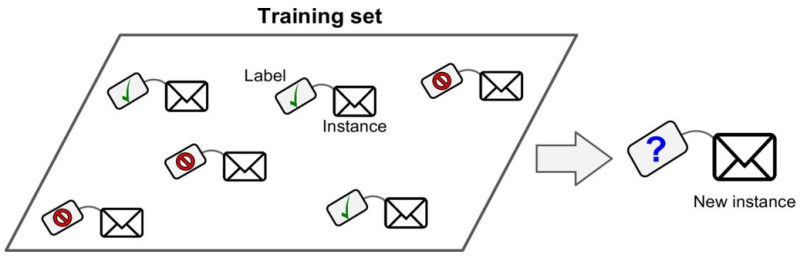


Figure 1.5. A labeled training set for supervised learning (e.g., spam classification)

Supervised Learning

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- Supervised learning is also used in Regression problem classification.
- Different algorithms for Supervised Learning are:
  - k-Nearest Neighbors
  - Linear Regression
  - Logistic Regression
  - Support Vector Machines (SVMs)
  - Decision Trees and Random Forests
  - Neural networks

Unsupervised Learning

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- Training data is unlabeled.
- Most common algorithms are
  - Clustering
  - Visualization and dimensionality reduction
  - Associate rule learning

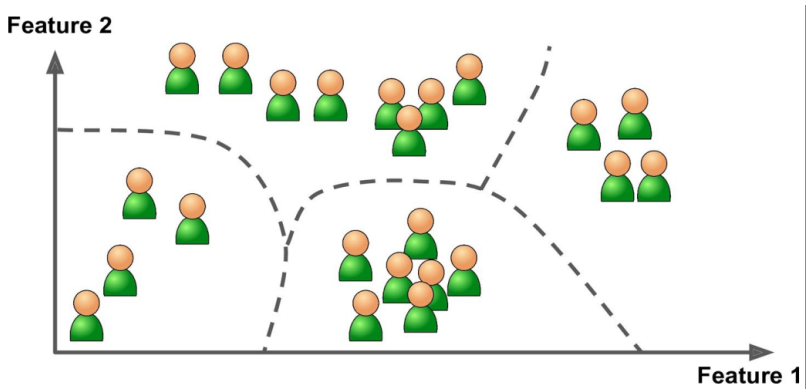


Figure 1.6. Clustering through Unsupervised learning

Semisupervised Learning

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- Lot of unlabeled data and a little bit of labeled data
- For example in many pictures just 1 label per person and able to label every person in every photo.
- Most semisupervised learning algorithms are combinations of unsupervised and supervised algorithms.
- For example RBMs are trained sequentially in an unsupervised manner, and then the whole system is fine-tuned using supervised learning techniques

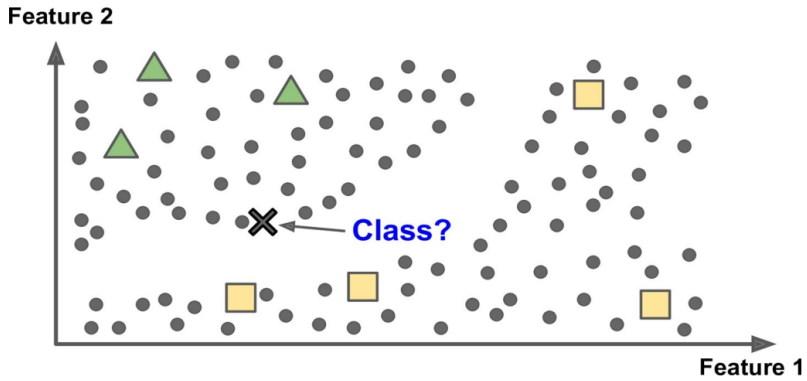


Figure 1.7. Semisupervised learning

Reinforcement Learning

- It is learn by itself what is the best strategy.

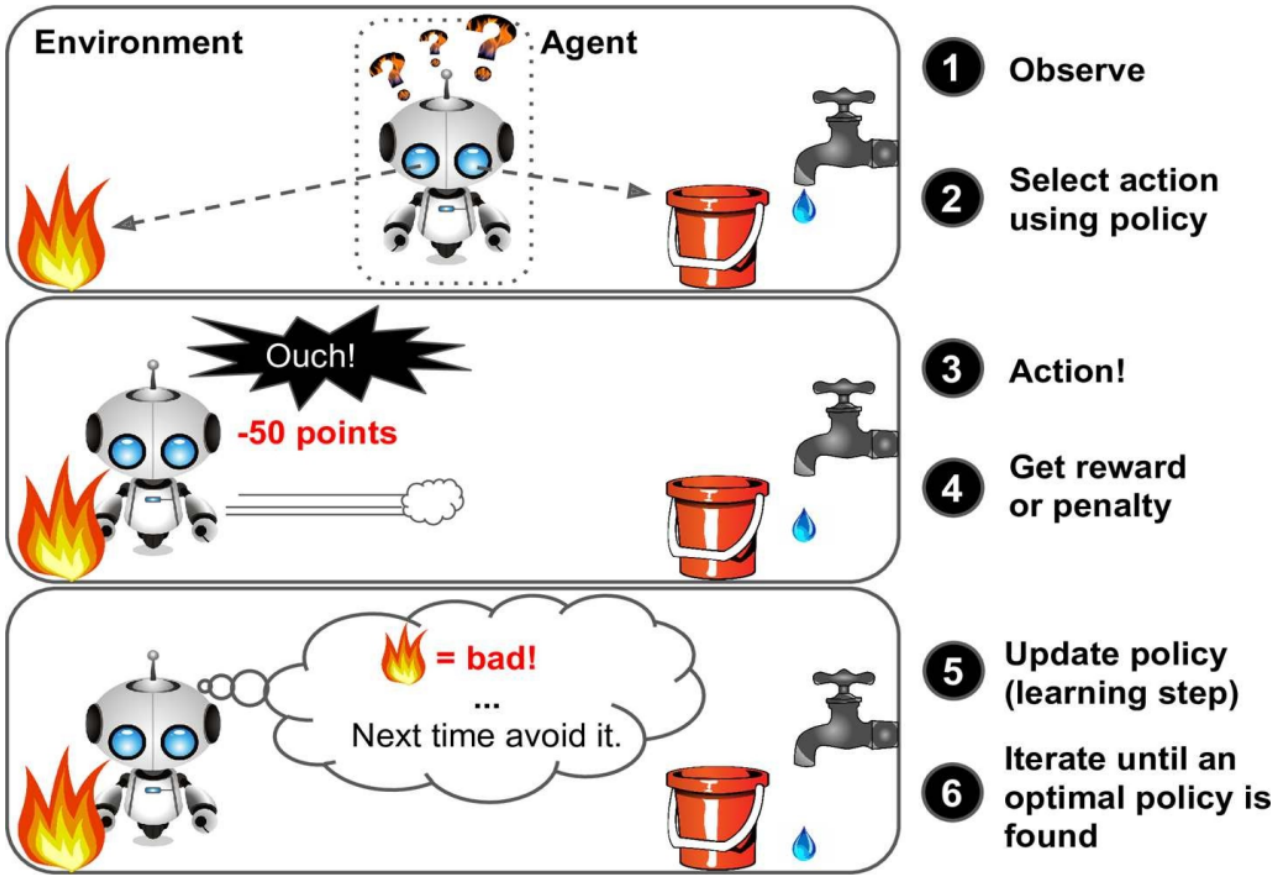


Figure 1.8. Reinforcement learning

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

Batch Learning

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- The system is incapable of learning incrementally.
- It must be trained using all the available data.
- Take a lot of time and computing resources.
- It typically done offline.
- It just applies what it has learned. This is also called offline learning.
- **Problem:** If we want a batch learning system to know about new data than we need to train a new version of the system from scratch on the full dataset. Means stop the old system and replace it with the new one.

Online Learning

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

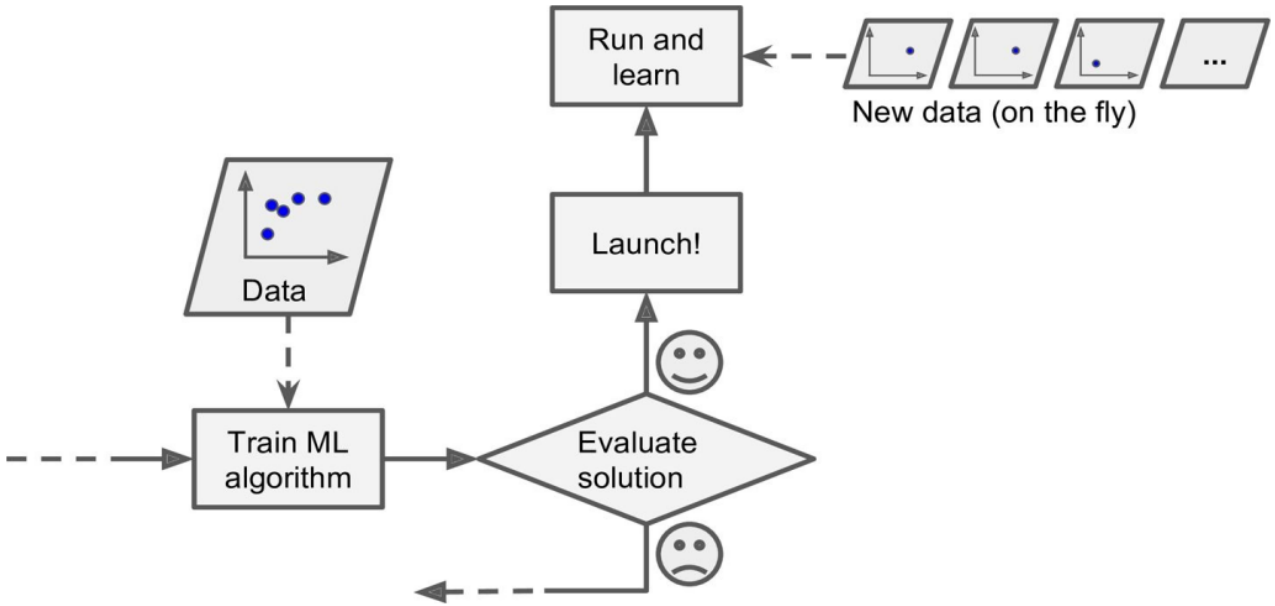


Figure 1.9. Online Learning

Online Learning for huge dataset

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

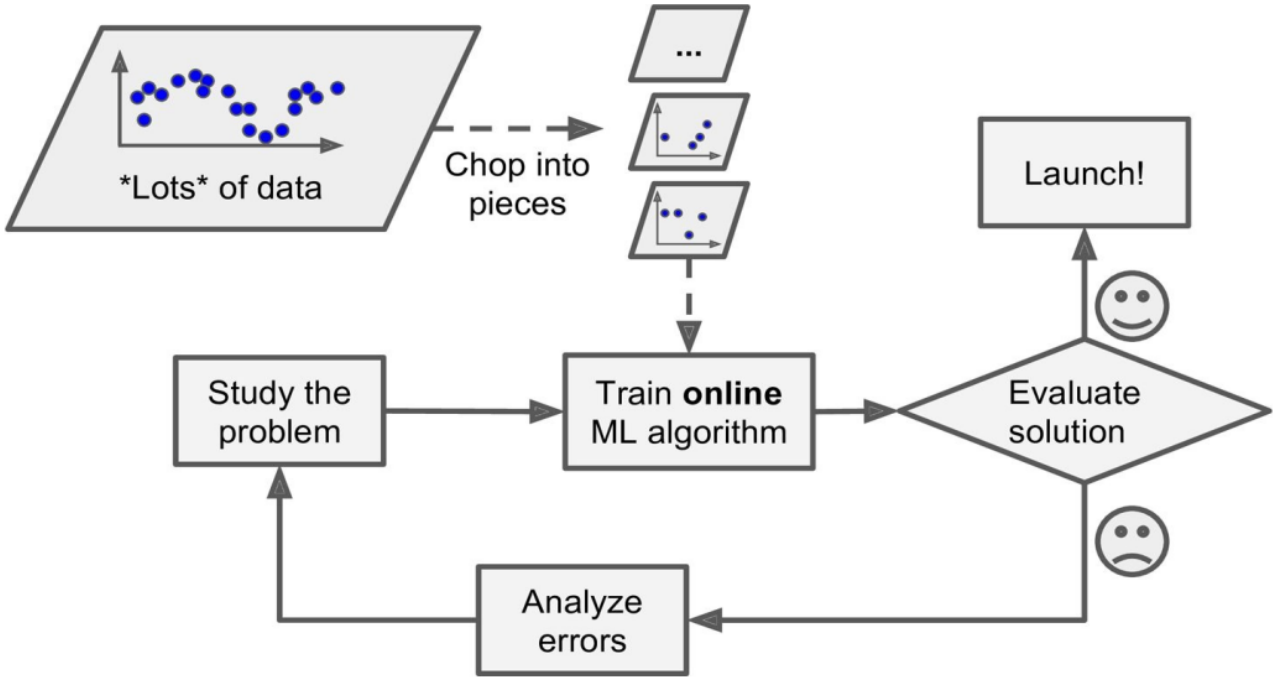


Figure 1.10. Online Learning for huge dataset

Instance-Based Versus Model-Based Learning

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- Another way to categorize Machine Learning systems is by how they generalize.
- Good performance measure on the training data is good, but insufficient.
- **Goal:** True goal is to perform well on new instances.
- There are two main approaches to generalize:
  - Instance based Learning
  - Model based Learning

Instance based Learning

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- This system learn the examples by heart.
- They generalizes to new cases using a similarity measure.
- For example, in spam email it use the email flags to measure the similarity between two mails.
- A (very basic) similarity measure between two emails could be to count the number of words they have in common.
- The system would flag an email as spam if it has many words in common with a known spam email.

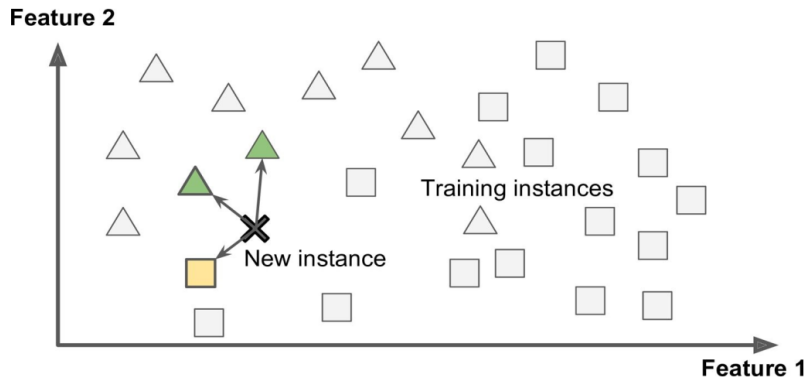


Figure 1.11. Instance based learning

Model based Learning

- Firstly Studied the data.
- Select the model.
- Trained the model on the training data (i.e., the learning algorithm searched for the model parameter values that minimize a cost function).
- Finally, applied the model to make predictions on new cases (this is called inference), hoping that this model will generalize well.

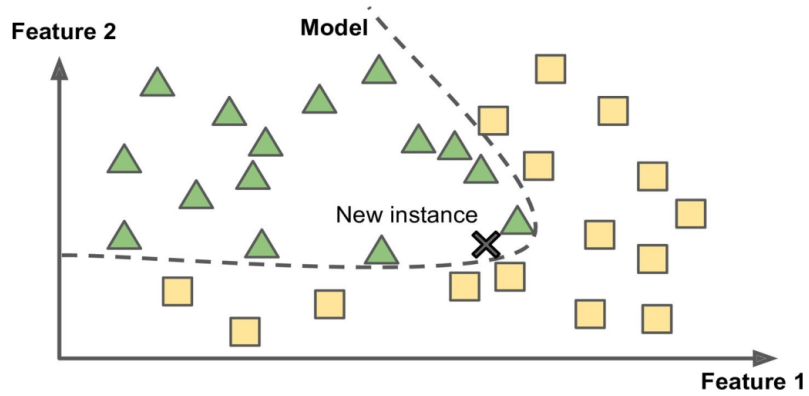


Figure 1.12. Model based learning

Chapter 1
Machine Learning
Supervised / Unsupervised Learning
Batch and Online Learning
Instance-Based Versus Model-Based Learning
ML Development Life Cycle
Main Challenges of Machine Learning

ML Development Life Cycle

Chapter 1

Machine Learning

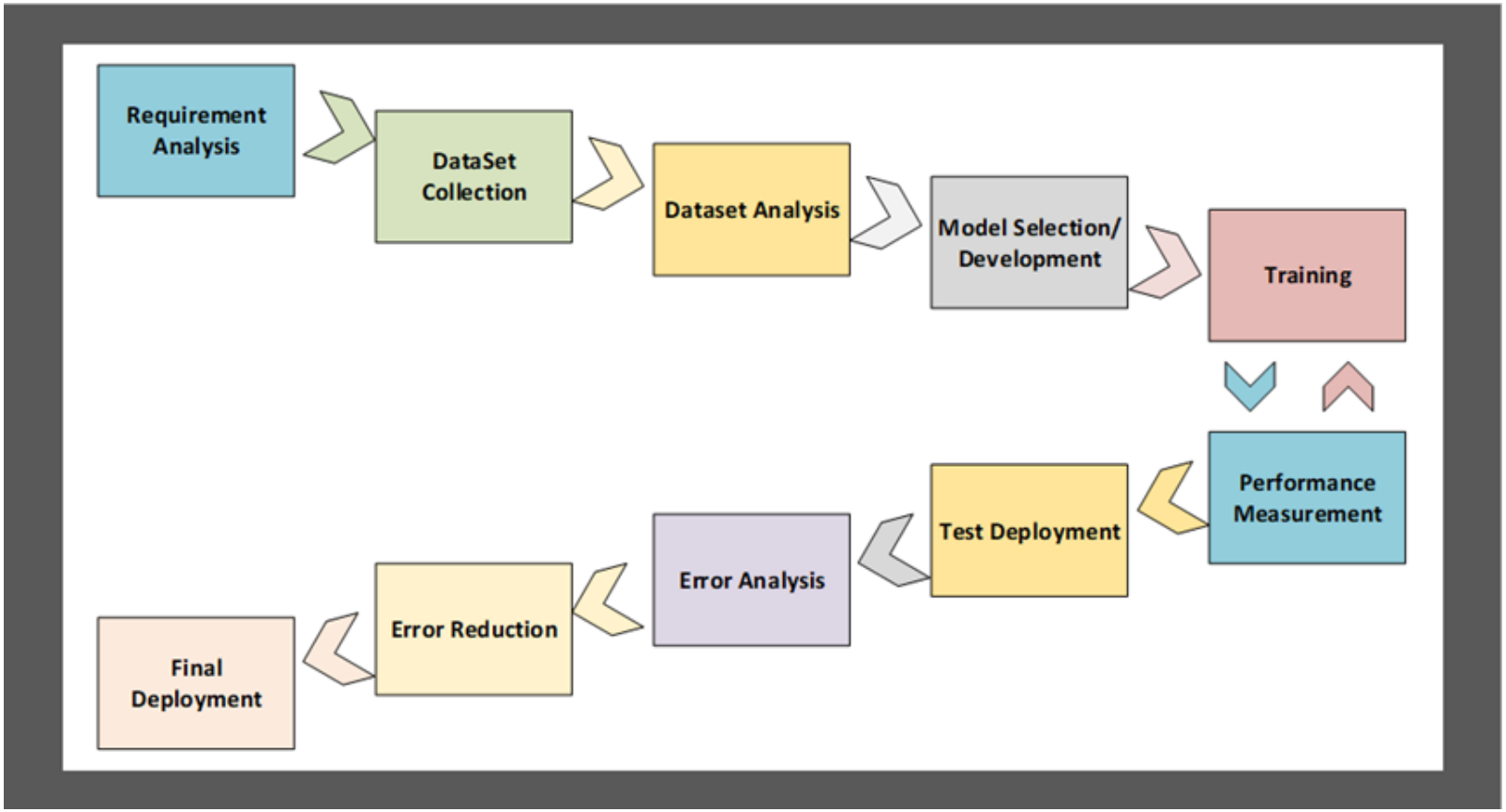
Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning



Requirement Analysis

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- Purpose of the automation in the desired area
- Type of dataset
  - Image/Video
  - Text
  - Tabular etc
- Number of target classes
- Most important : Domain Knowledge
- Deployment Details

Dataset Analysis

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- Involves Preprocessing
  - Remove unwanted data, reduce noise etc
  - Can be done by acquiring domain knowledge
- Features Engineering (Optional)
  - It depends upon the data type
  - Often it is required in case of large data such as
    - High resolution images
    - Textual data
    - Tabular data with large number of instances and attributes

ML Models

Chapter 1

Machine Learning

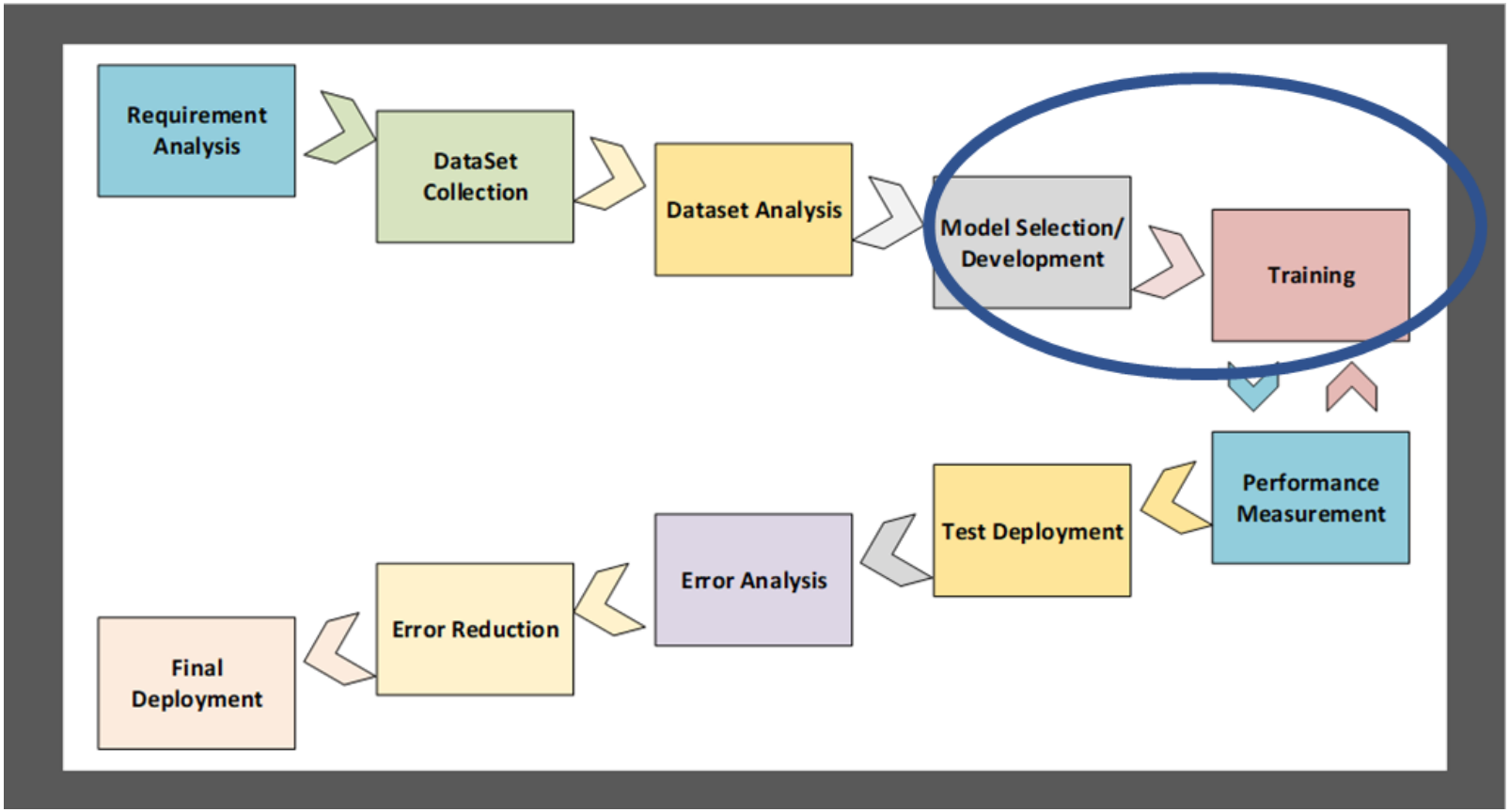
Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning



Some ML Models

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- Most of the models are available in prebuilt API's.
- We choose the model based on the data we have
- Recall list of already studied models
- Some examples given here are
  - MLP
  - DT
  - SVM

Neural Networks

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

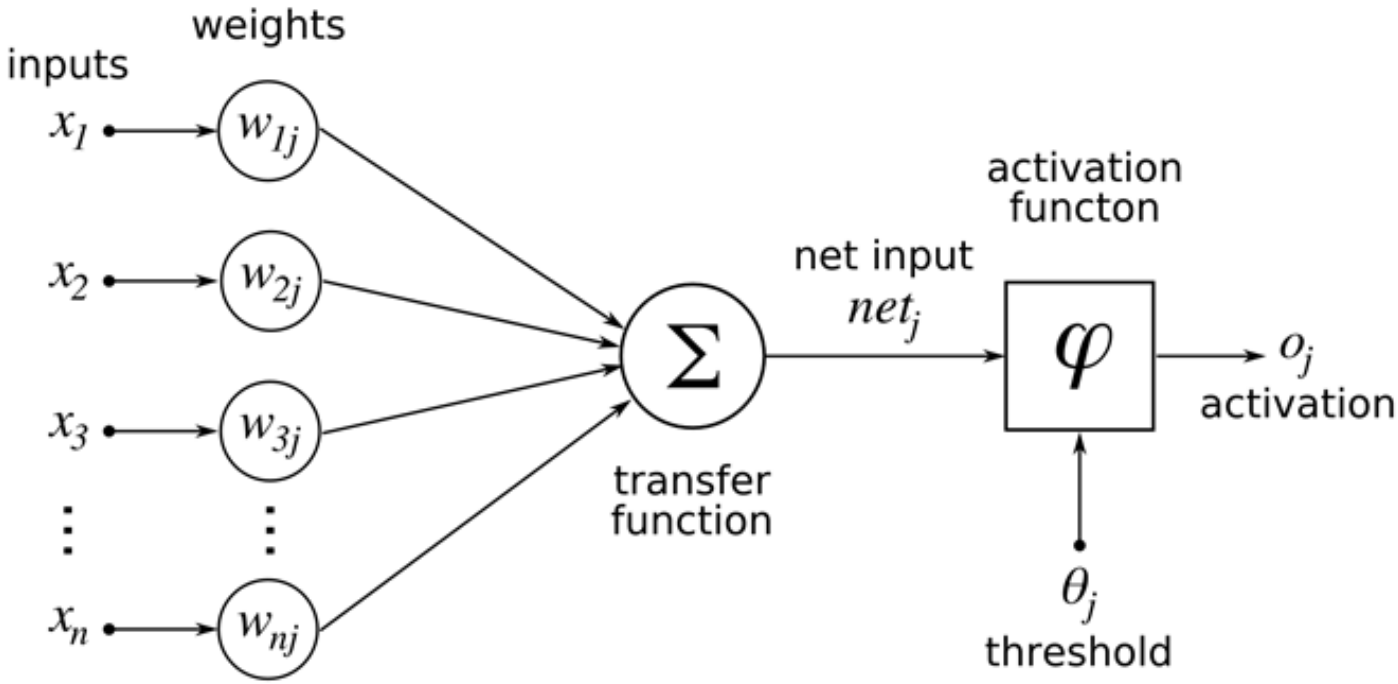
Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- Neural networks are a set of units, modeled loosely after the human brain, that are designed to recognize patterns.
- Every unit contains weights, transfer function and activation function.
- The weights are learned by passing and updating the whole dataset several times.



Neural Networks

Chapter 1

Machine Learning

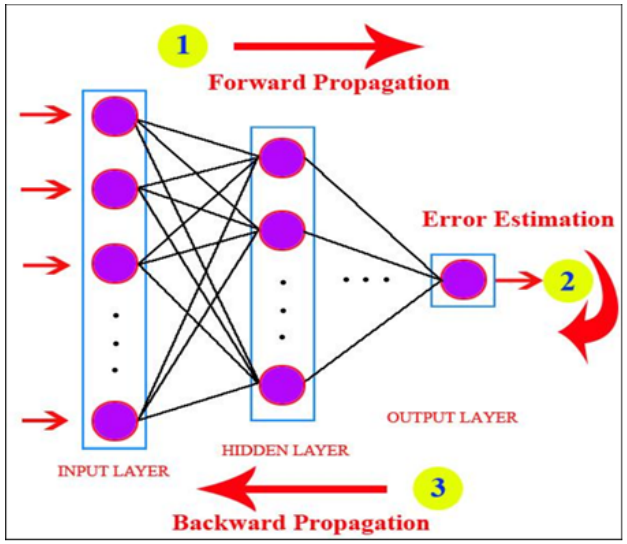
Supervised / Unsupervised Learning

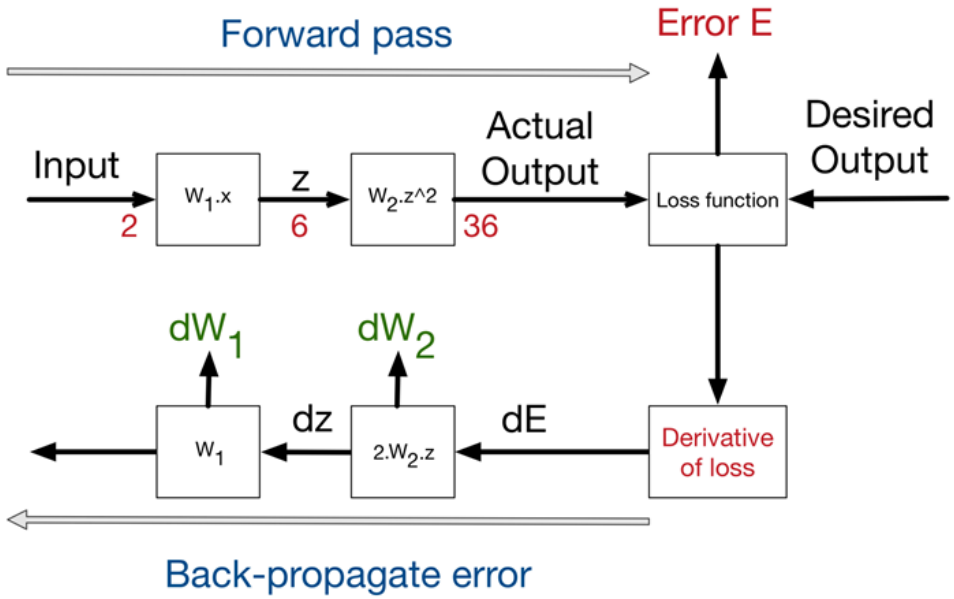
Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning





Decision Trees

Chapter 1

Machine Learning

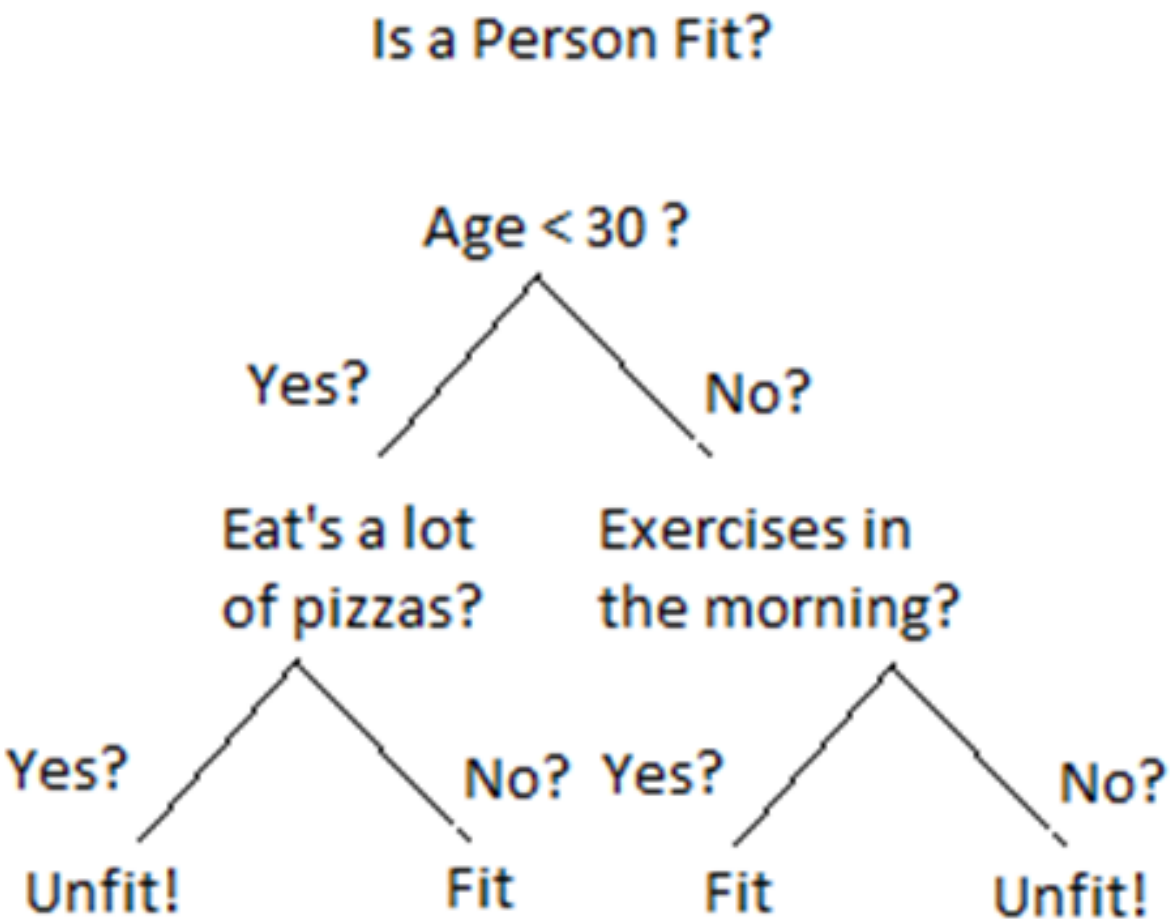
Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning



Decision Trees

Chapter 1

Machine Learning

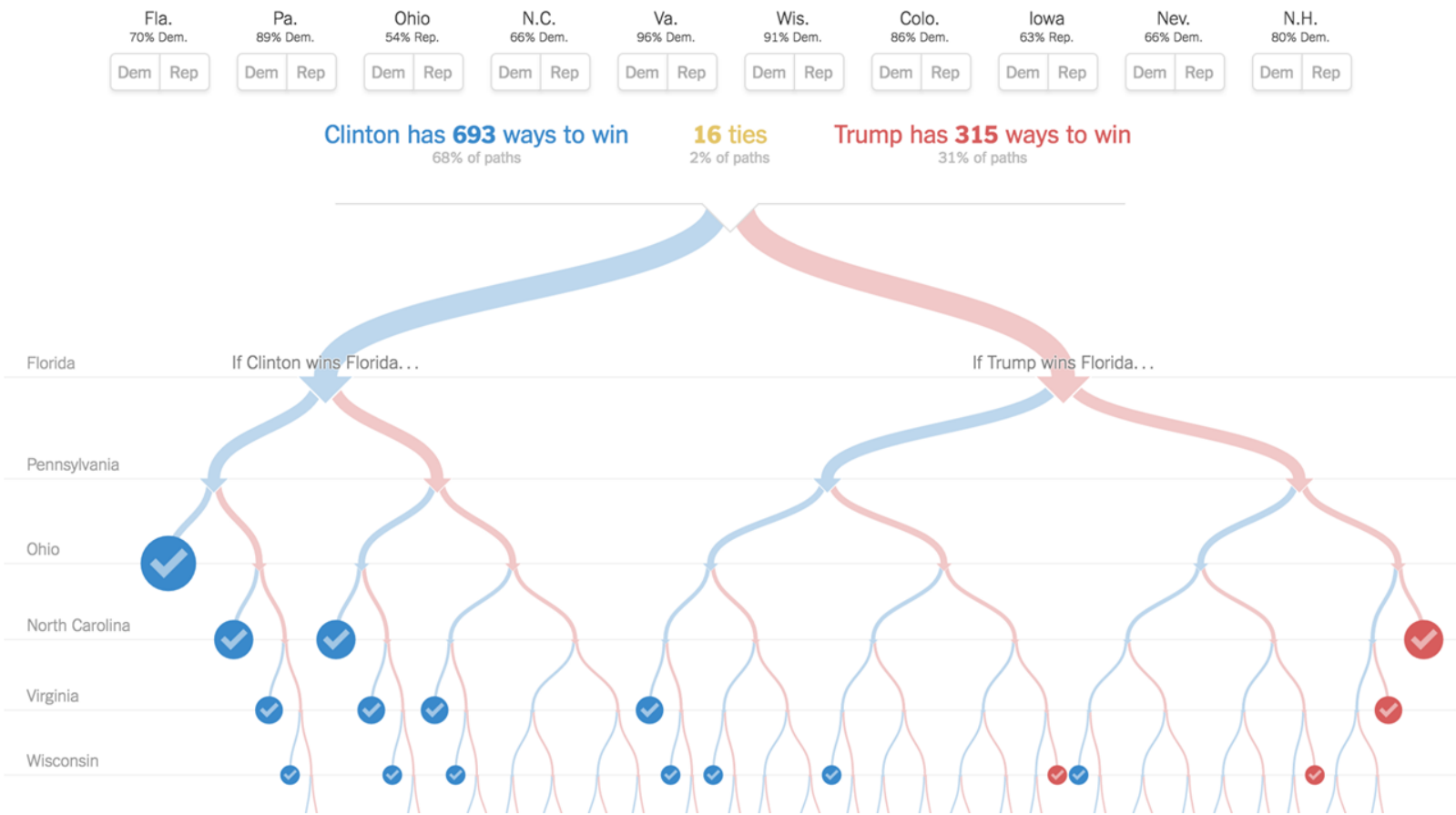
Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning



Support Vector Machines

Chapter 1

Machine Learning

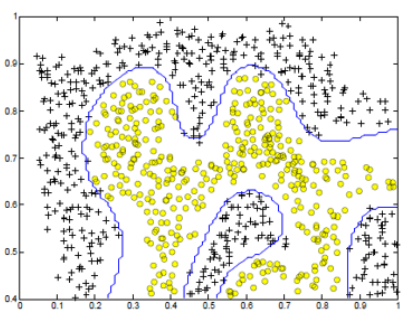
Supervised / Unsupervised Learning

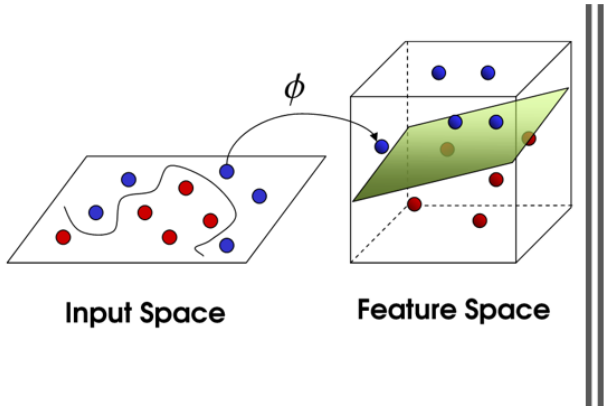
Batch and Online Learning

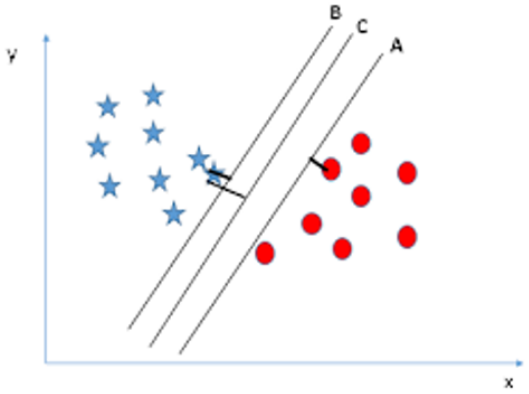
Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning







Clustering

Chapter 1

Machine Learning

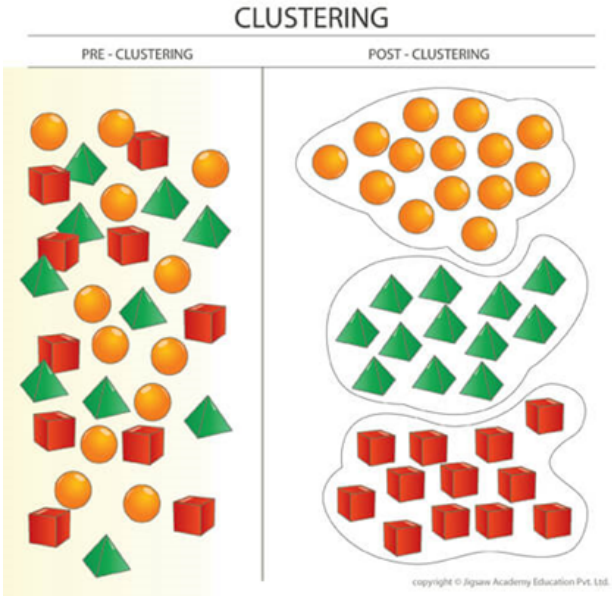
Supervised / Unsupervised Learning

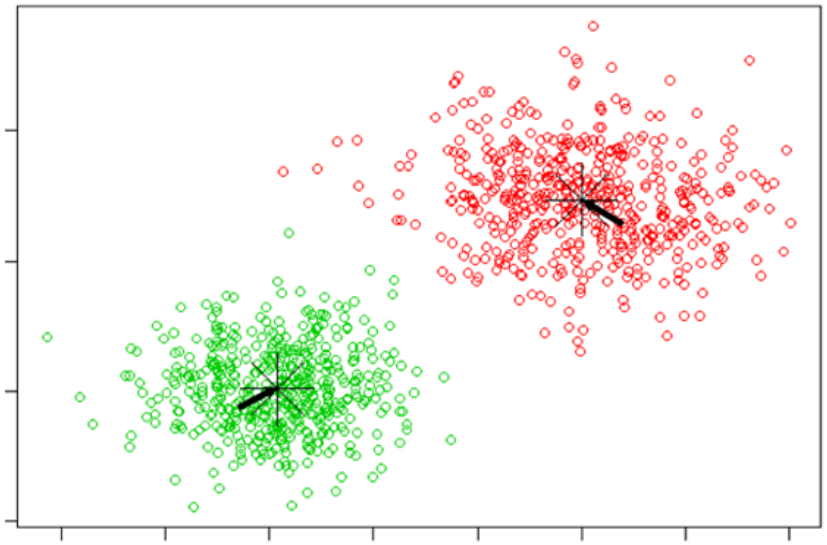
Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning





Test Deployment

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- Choose validation and test sets to reflect data you expect to get in the future and want to do well on.
- It is necessary to check model performance in real time scenarios
- Deployed where model will be used for predictions
- Regressive testing by bombarding various types of data instances

Error Analysis and Error Reduction

Chapter 1

Machine Learning

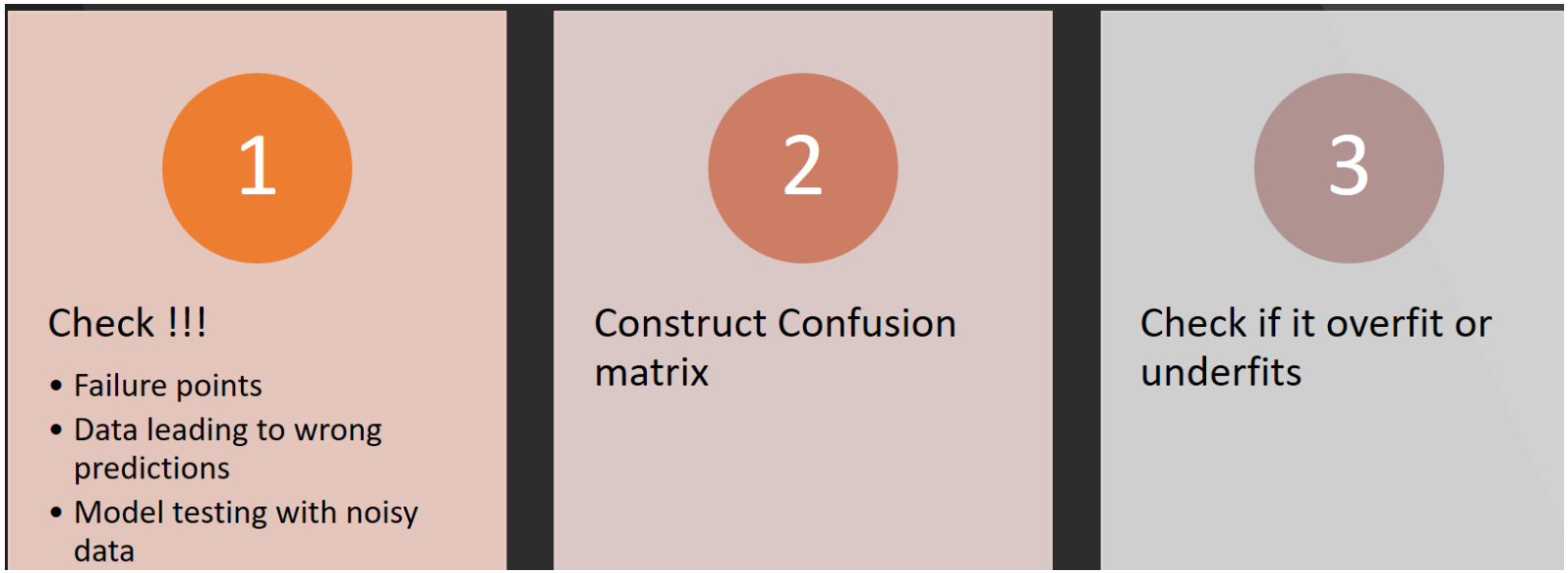
Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning



Error Analysis and Error Reduction

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

True Positive (TP): • Reality: A wolf threatened. • Shepherd said: "Wolf." • Outcome: Shepherd is a hero.	**False Positive (FP):** • Reality: No wolf threatened. • Shepherd said: "Wolf." • Outcome: Villagers are angry at shepherd fo up.
False Negative (FN): • Reality: A wolf threatened. • Shepherd said: "No wolf." • Outcome: The wolf ate all the sheep.	**True Negative (TN):** • Reality: No wolf threatened. • Shepherd said: "No wolf." • Outcome: Everyone is fine.

Error Analysis and Error Reduction

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- Reduce the wrong predictions by taking following steps
- Change the parameters of the model
- Increase the training and validation data
- Change the ML model if the problem persists
- Investigate data, if some instances or attributes are irrelevant
- If it overfits, reduce error by introducing regularization
- If it underfits , increase data

Training and Performance Measurement

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- For efficient training our dataset is divided in three folds
  - Training set
  - Validation set
  - Test set
- Training data is used to train the classifier.
- Validation data is used to measure how well the classifier has learned unseen data.
  - Used during training of the data for tuning the parameters
  - Validation and test sets are to direct you towards most important changes to your machine learning model.
- Test set is used to measure the classification accuracy of trained model.

Main Challenges of Machine Learning

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- Insufficient Quantity of Training Data
- Nonrepresentative Training Data
- Poor-Quality Data (full of errors, noise, outliers)
- Irrelevant Features
- Overfitting the Training Data
- Underfitting the Training Data
- Stepping Back:
  - Learning from data instead of explicit rules
  - Too many different types of ML systems: supervised or not, batch or online, instance-based or model-based
  - Size of training data

Testing and Validating

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

- **Problem:** Measured the generalization error multiple times on the test set, and adapted the model and hyperparameters to produce the best model for that set.
- **Solution:** Have the second holdout Validation set with training data, thus select the model and hyperparameters that perform best on the validation set.
- Finally single final test against the test set to get an estimate of the generalization error.

The End

Chapter 1

Machine Learning

Supervised / Unsupervised Learning

Batch and Online Learning

Instance-Based Versus Model-Based Learning

ML Development Life Cycle

Main Challenges of Machine Learning

Any Questions?