SEMINAR PRESENTATION ON MACHINE LEARNING

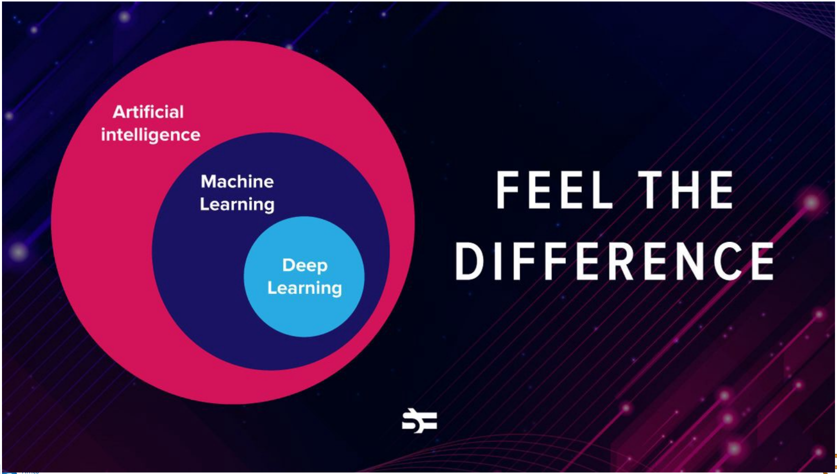
By:
Ishaya Jeremiah Ayock

ECOLE POLYTECHNIQUE DE THIES

September 12, 2020

Outline

Machine Learning vs Deep Learning vs AI.



Introduction

What is Learning?

Introduction

What is Learning?

Learning is any process by which a system improves performance from experience.

Introduction

What is Learning?

Learning is any process by which a system improves performance from experience.
"Herbert Simon"

Machine Learning vs Deep Learning vs AI ?

Artificial Intelligence

Machine Learning vs Deep Learning vs AI ?

Artificial Intelligence

Artificial intelligence is a **technique** which enables machines to mimic human behavior.

Machine Learning vs Deep Learning vs AI ?

Artificial Intelligence

Artificial intelligence is a **technique** which enables machines to mimic human behavior.

Machine Learning

Machine Learning vs Deep Learning vs AI ?

Artificial Intelligence

Artificial intelligence is a **technique** which enables machines to mimic human behavior.

Machine Learning

Machine learning is a subset of **AI technique** which uses a statistical methods to enable machines to improve with experience.

Machine Learning vs Deep Learning vs AI ?

Artificial Intelligence

Artificial intelligence is a **technique** which enables machines to mimic human behavior.

Machine Learning

Machine learning is a subset of **AI technique** which uses a statistical methods to enable machines to improve with experience.

Deep Learning

Machine Learning vs Deep Learning vs AI ?

Artificial Intelligence

Artificial intelligence is a **technique** which enables machines to mimic human behavior.

Machine Learning

Machine learning is a subset of **AI technique** which uses a statistical methods to enable machines to improve with experience.

Deep Learning

Deep learning is a subset of **Machine Learning** which make the computation of multi-layer neutral network feasible.

Machine Learning vs Deep Learning vs AI ?

Artificial Intelligence

Artificial intelligence is a **technique** which enables machines to mimic human behavior.

Machine Learning

Machine learning is a subset of **AI technique** which uses a statistical methods to enable machines to improve with experience.

Deep Learning

Deep learning is a subset of **Machine Learning** which make the computation of multi-layer neutral network feasible.

Machine Learning

Definition

Tom Mitchell (1998) defines Machine Learning as the study of algorithms that;

Machine Learning

Definition

Tom Mitchell (1998) defines Machine Learning as the study of algorithms that;

- improve their performance P

Machine Learning

Definition

Tom Mitchell (1998) defines Machine Learning as the study of algorithms that;

- improve their performance P
- at some task T

Machine Learning

Definition

Tom Mitchell (1998) defines Machine Learning as the study of algorithms that;

- improve their performance P
- at some task T
- with experience E.

Machine Learning

Definition

Tom Mitchell (1998) defines Machine Learning as the study of algorithms that;

- improve their performance P
- at some task T
- with experience E.

Goal

Machine Learning

Definition

Tom Mitchell (1998) defines Machine Learning as the study of algorithms that;

- improve their performance P
- at some task T
- with experience E.

Goal

The goal of machine learning generally is to understand the structure of data and fit that data into models that can be understood and utilized by people.

Africa Digital Native

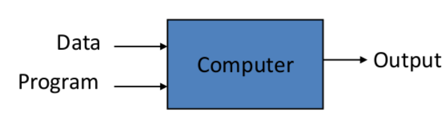
Sagesse Devoir Ecole Polytechnique Thiès

Historically

Traditional Programming

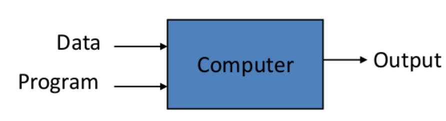
Historically

Traditional Programming



Historically

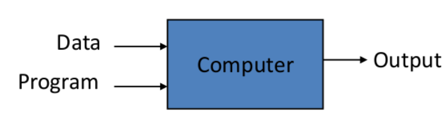
Traditional Programming



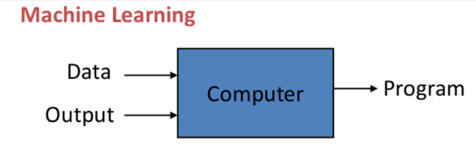
Machine Learning

Historically

Traditional Programming



Machine Learning



When Do We Use Machine Learning?

When?

When Do We Use Machine Learning?

When?

Machine Learning is used when;

- Human expertise does not exist (Navigating on Mars)

When Do We Use Machine Learning?

When?

Machine Learning is used when;

- Human expertise does not exist (Navigating on Mars)
- Humans can't explain their expertise (Speech recognition)

When Do We Use Machine Learning?

When?

Machine Learning is used when;

- Human expertise does not exist (Navigating on Mars)
- Humans can't explain their expertise (Speech recognition)
- Models must be customized (Personalized medicine)

When Do We Use Machine Learning?

When?

Machine Learning is used when;

- Human expertise does not exist (Navigating on Mars)
- Humans can't explain their expertise (Speech recognition)
- Models must be customized (Personalized medicine)
- Models are based on huge amounts of data (Genomics)

When Do We Use Machine Learning?

When?

Machine Learning is used when;

- Human expertise does not exist (Navigating on Mars)
- Humans can't explain their expertise (Speech recognition)
- Models must be customized (Personalized medicine)
- Models are based on huge amounts of data (Genomics)

Application of Machine Learning

Health Care

Application of Machine Learning

Health Care

- Identification of Diseases and Diagnosis

Application of Machine Learning

Health Care

- Identification of Diseases and Diagnosis
- Drug Discovery and Manufacturing

Application of Machine Learning

Health Care

- Identification of Diseases and Diagnosis
- Drug Discovery and Manufacturing
- Medical Imaging

Application of Machine Learning

Health Care

- Identification of Diseases and Diagnosis
- Drug Discovery and Manufacturing
- Medical Imaging
- Personalized Medicine/Treatment

Application of Machine Learning

Health Care

- Identification of Diseases and Diagnosis
- Drug Discovery and Manufacturing
- Medical Imaging
- Personalized Medicine/Treatment
- Predicting Diseases

Application of Machine Learning

Health Care

- Identification of Diseases and Diagnosis
- Drug Discovery and Manufacturing
- Medical Imaging
- Personalized Medicine/Treatment
- Predicting Diseases

Retail

Retail

- Market Basket Analysis

Retail

- Market Basket Analysis
- Customer Relation Management(CRM)

Retail

- Market Basket Analysis
- Customer Relation Management(CRM)

Finance

Retail

- Market Basket Analysis
- Customer Relation Management(CRM)

Finance

- Credit Scoring

Retail

- Market Basket Analysis
- Customer Relation Management(CRM)

Finance

- Credit Scoring
- Process Automation

Retail

- Market Basket Analysis
- Customer Relation Management(CRM)

Finance

- Credit Scoring
- Process Automation
- Fraud Detection and Prevention

Retail

- Market Basket Analysis
- Customer Relation Management(CRM)

Finance

- Credit Scoring
- Process Automation
- Fraud Detection and Prevention

Manufacturing

Manufacturing

- Optimization

Manufacturing

- Optimization
- Troubleshooting

Manufacturing

- Optimization
- Troubleshooting

Bioinformatics

Manufacturing

- Optimization
- Troubleshooting

Bioinformatics

- Motifs

Manufacturing

- Optimization
- Troubleshooting

Bioinformatics

- Motifs
- Alignment

Manufacturing

- Optimization
- Troubleshooting

Bioinformatics

- Motifs
- Alignment

Web Mining

Manufacturing

- Optimization
- Troubleshooting

Bioinformatics

- Motifs
- Alignment

Web Mining

- Search Engines

Manufacturing

- Optimization
- Troubleshooting

Bioinformatics

- Motifs
- Alignment

Web Mining

- Search Engines

Education

Education

- Accurately Grading Assignments

Education

- Accurately Grading Assignments
- Learning Analytic

Education

- Accurately Grading Assignments
- Learning Analytic
- Adaptive Learning

Education

- Accurately Grading Assignments
- Learning Analytic
- Adaptive Learning
- Predictive Analytics

Agriculture

Education

- Accurately Grading Assignments
- Learning Analytic
- Adaptive Learning
- Predictive Analytics

Agriculture

- Species Recognition

Education

- Accurately Grading Assignments
- Learning Analytic
- Adaptive Learning
- Predictive Analytics

Agriculture

- Species Recognition
- Yield Prediction

Education

- Accurately Grading Assignments
- Learning Analytic
- Adaptive Learning
- Predictive Analytics

Agriculture

- Species Recognition
- Yield Prediction
- Disease Detection

Education

- Accurately Grading Assignments
- Learning Analytic
- Adaptive Learning
- Predictive Analytics

Agriculture

- Species Recognition
- Yield Prediction
- Disease Detection
- Weed Detection

Education

- Accurately Grading Assignments
- Learning Analytic
- Adaptive Learning
- Predictive Analytics

Agriculture

- Species Recognition
- Yield Prediction
- Disease Detection
- Weed Detection

Telecommunication

Education

- Accurately Grading Assignments
- Learning Analytic
- Adaptive Learning
- Predictive Analytics

Agriculture

- Species Recognition
- Yield Prediction
- Disease Detection
- Weed Detection

Telecommunication

- Quality of Service Optimization.

Education

- Accurately Grading Assignments
- Learning Analytic
- Adaptive Learning
- Predictive Analytics

Agriculture

- Species Recognition
- Yield Prediction
- Disease Detection
- Weed Detection

Telecommunication

- Quality of Service Optimization.

Machine Learning Groupings

Machine Learning Groupings

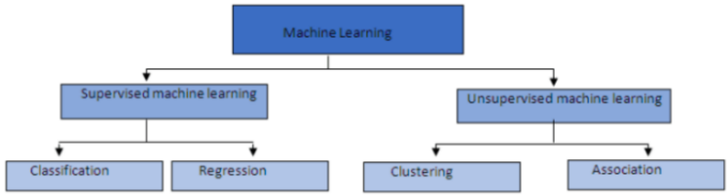


Figure: Machine Learning Groupings

Supervised Learning

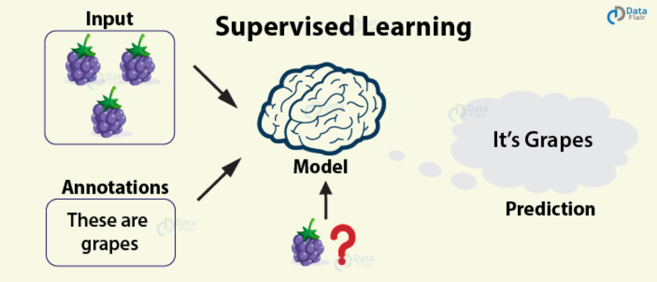


Figure: Supervised learning.

Supervised Learning

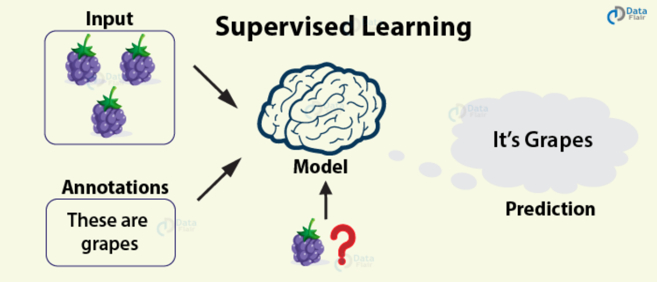


Figure: Supervised learning.

Supervised Learning is is a type of learning where the input data is labeled/already known.

Classification

Classification

Classifier:
An algorithm that maps the input data to a specific category.

Classification

Classifier:
An algorithm that maps the input data to a specific category.

- Classification is a technique where we categorize data into a given number of classes.

Classification

Classifier:
An algorithm that maps the input data to a specific category.

- Classification is a technique where we categorize data into a given number of classes.
- The main goal of a classification problem is to identify the category/class to which a new data will fall under.

Classification

Classifier:
An algorithm that maps the input data to a specific category.

- Classification is a technique where we categorize data into a given number of classes.
- The main goal of a classification problem is to identify the category/class to which a new data will fall under.

Types of Classifiction

Classification

Classifier:
An algorithm that maps the input data to a specific category.

- Classification is a technique where we categorize data into a given number of classes.
- The main goal of a classification problem is to identify the category/class to which a new data will fall under.

Types of Classifiction

- **Binary:** Classification task with two possible outcomes.

Classification

Classifier:
An algorithm that maps the input data to a specific category.

- Classification is a technique where we categorize data into a given number of classes.
- The main goal of a classification problem is to identify the category/class to which a new data will fall under.

Types of Classifiction

- **Binary:** Classification task with two possible outcomes.
- **Multi-class:** Classification with more than two classes.

Classification

Classifier:
An algorithm that maps the input data to a specific category.

- Classification is a technique where we categorize data into a given number of classes.
- The main goal of a classification problem is to identify the category/class to which a new data will fall under.

Types of Classifiction

- **Binary:** Classification task with two possible outcomes.
- **Multi-class:** Classification with more than two classes.
- **Multi-label:** Classification task where each sample is mapped to a set of target labels (more than one class).

Classification

Classifier:
An algorithm that maps the input data to a specific category.

- Classification is a technique where we categorize data into a given number of classes.
- The main goal of a classification problem is to identify the category/class to which a new data will fall under.

Types of Classifiction

- **Binary:** Classification task with two possible outcomes.
- **Multi-class:** Classification with more than two classes.
- **Multi-label:** Classification task where each sample is mapped to a set of target labels (more than one class).

Examples of Classification Algorithms

Examples of Classification Algorithms

- Logistics regression

Examples of Classification Algorithms

- Logistics regression
- Support vector machine(SVM)

Examples of Classification Algorithms

- Logistics regression
- Support vector machine(SVM)
- Random Forest

Examples of Classification Algorithms

- Logistics regression
- Support vector machine(SVM)
- Random Forest
- Nearest Neighbor(KNN)

Examples of Classification Algorithms

- Logistics regression
- Support vector machine(SVM)
- Random Forest
- Nearest Neighbor(KNN)
- Naive Bayes

Examples of Classification Algorithms

- Logistics regression
- Support vector machine(SVM)
- Random Forest
- Nearest Neighbor(KNN)
- Naive Bayes

Implementation

Examples of Classification Algorithms

- Logistics regression
- Support vector machine(SVM)
- Random Forest
- Nearest Neighbor(KNN)
- Naive Bayes

Implementation

- **Initialize** The classifier to be used.

Examples of Classification Algorithms

- Logistics regression
- Support vector machine(SVM)
- Random Forest
- Nearest Neighbor(KNN)
- Naive Bayes

Implementation

- **Initialize** The classifier to be used.
- **Train the classifier**. *All scikit-learn uses fit(X, y).*

Examples of Classification Algorithms

- Logistics regression
- Support vector machine(SVM)
- Random Forest
- Nearest Neighbor(KNN)
- Naive Bayes

Implementation

- **Initialize** The classifier to be used.
- **Train the classifier**. *All scikit-learn uses fit(X, y).*
- **Predict the target.**

Examples of Classification Algorithms

- Logistics regression
- Support vector machine(SVM)
- Random Forest
- Nearest Neighbor(KNN)
- Naive Bayes

Implementation

- **Initialize** The classifier to be used.
- **Train the classifier**. *All scikit-learn uses fit(X, y).*
- **Predict the target.**
- **Evaluate the classifier model.**

Examples of Classification Algorithms

- Logistics regression
- Support vector machine(SVM)
- Random Forest
- Nearest Neighbor(KNN)
- Naive Bayes

Implementation

- **Initialize** The classifier to be used.
- **Train the classifier**. *All scikit-learn uses fit(X, y).*
- **Predict the target.**
- **Evaluate the classifier model.**

Examples of Classification Problems

Examples

Examples of Classification Problems

Examples

- Fraud Detection

Examples of Classification Problems

Examples

- Fraud Detection
- Email Spam Detection

Examples of Classification Problems

Examples

- Fraud Detection
- Email Spam Detection
- Medical Diagnostics

Examples of Classification Problems

Examples

- Fraud Detection
- Email Spam Detection
- Medical Diagnostics
- Image Classification

Examples of Classification Problems

Examples

- Fraud Detection
- Email Spam Detection
- Medical Diagnostics
- Image Classification

Regression

Regression

Regression

Regression

It is a supervised machine learning algorithm that predict or forecast numerical values.

Examples of Regression Algorithms

Regression Algorithms

Examples of Regression Algorithms

Regression Algorithms

- Linear regression

Examples of Regression Algorithms

Regression Algorithms

- Linear regression
- Multiple Linear regression

Examples of Regression Algorithms

Regression Algorithms

- Linear regression
- Multiple Linear regression
- GLM

Examples of Regression Algorithms

Regression Algorithms

- Linear regression
- Multiple Linear regression
- GLM
- Neural Network

Examples of Regression Algorithms

Regression Algorithms

- Linear regression
- Multiple Linear regression
- GLM
- Neural Network
- Support Vector Regression(SVR)

Examples of Regression Algorithms

Regression Algorithms

- Linear regression
- Multiple Linear regression
- GLM
- Neural Network
- Support Vector Regression(SVR)
- Elastic Net Regression

Examples of Regression Algorithms

Regression Algorithms

- Linear regression
- Multiple Linear regression
- GLM
- Neural Network
- Support Vector Regression(SVR)
- Elastic Net Regression

Examples of Regression Problems

Examples

Examples of Regression Problems

Examples

- Risk Assessment

Examples of Regression Problems

Examples

- Risk Assessment
- Score Prediction

Examples of Regression Problems

Examples

- Risk Assessment
- Score Prediction
- Sales Predictions

Examples of Regression Problems

Examples

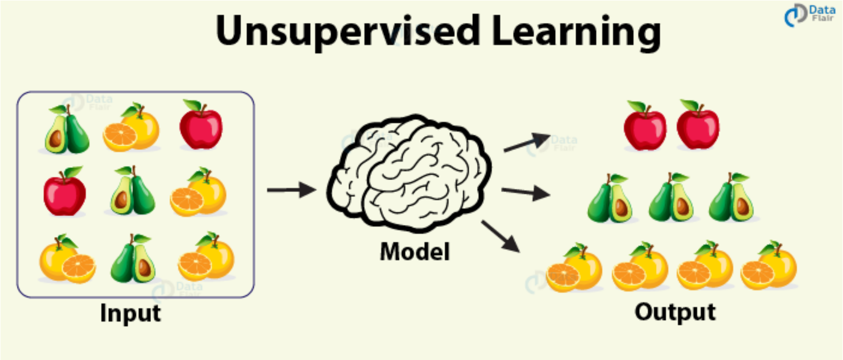
- Risk Assessment
- Score Prediction
- Sales Predictions
- Advertisement

Examples of Regression Problems

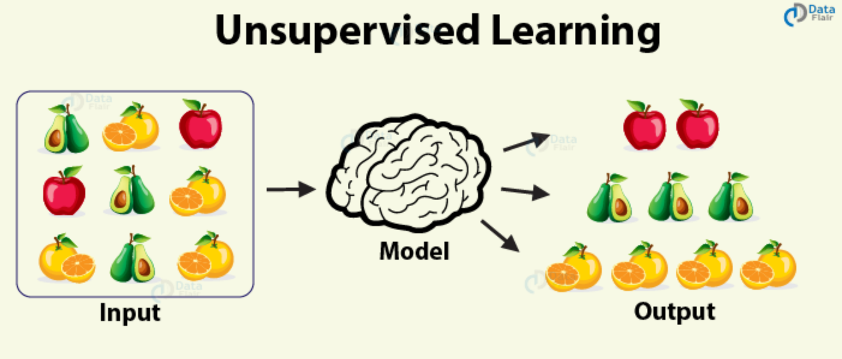
Examples

- Risk Assessment
- Score Prediction
- Sales Predictions
- Advertisement

Unsupervised Learning Introduction

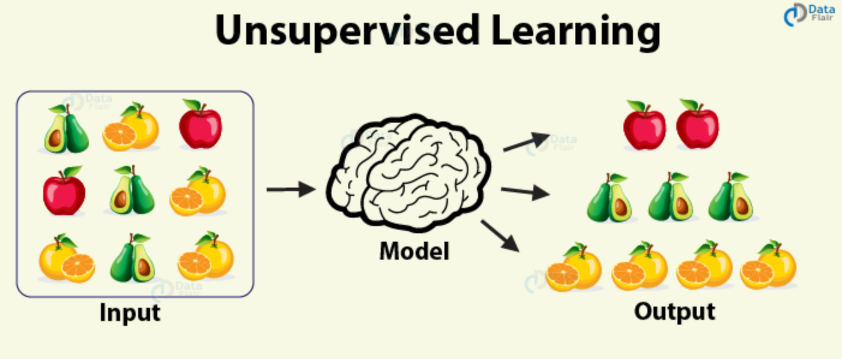


Unsupervised Learning Introduction



Unsupervised Learning

Unsupervised Learning Introduction



Unsupervised Learning

Unsupervised Learning is is a type of learning where the Input data is not labeled and does not have a known result.

Dimensional Reduction

Dimensionality Reduction

Dimensional Reduction

Dimensionality Reduction

Its the transformation of data from a high-dimensional into a low dimensional space so that the later representation Keeps some meaningful properties of the original data.

Dimensional Reduction

Dimensionality Reduction

Its the transformation of data from a high-dimensional into a low dimensional space so that the later representation Keeps some meaningful properties of the original data.

Reasons for Dimensionality Reduction

Dimensional Reduction

Dimensionality Reduction

Its the transformation of data from a high-dimensional into a low dimensional space so that the later representation Keeps some meaningful properties of the original data.

Reasons for Dimensionality Reduction

Dimensional Reduction

Dimensionality Reduction

Its the transformation of data from a high-dimensional into a low dimensional space so that the later representation Keeps some meaningful properties of the original data.

Reasons for Dimensionality Reduction

- Visualization

Dimensional Reduction

Dimensionality Reduction

Its the transformation of data from a high-dimensional into a low dimensional space so that the later representation Keeps some meaningful properties of the original data.

Reasons for Dimensionality Reduction

- Visualization
- Reduce computational load

Dimensional Reduction

Dimensionality Reduction

Its the transformation of data from a high-dimensional into a low dimensional space so that the later representation Keeps some meaningful properties of the original data.

Reasons for Dimensionality Reduction

- Visualization
- Reduce computational load
- Reduce variance in estimation(Regularize the problem.)

Dimensional Reduction

Dimensionality Reduction

Its the transformation of data from a high-dimensional into a low dimensional space so that the later representation Keeps some meaningful properties of the original data.

Reasons for Dimensionality Reduction

- Visualization
- Reduce computational load
- Reduce variance in estimation(Regularize the problem.)

Methods of Dimensionality Reduction.

Dimensional Reduction

Dimensionality Reduction

Its the transformation of data from a high-dimensional into a low dimensional space so that the later representation Keeps some meaningful properties of the original data.

Reasons for Dimensionality Reduction

- Visualization
- Reduce computational load
- Reduce variance in estimation(Regularize the problem.)

Methods of Dimensionality Reduction.

- Principal Component Analysis

Dimensional Reduction

Dimensionality Reduction

Its the transformation of data from a high-dimensional into a low dimensional space so that the later representation Keeps some meaningful properties of the original data.

Reasons for Dimensionality Reduction

- Visualization
- Reduce computational load
- Reduce variance in estimation(Regularize the problem.)

Methods of Dimensionality Reduction.

- Principal Component Analysis
- Canonical Correlation Analysis

Dimensional Reduction

Dimensionality Reduction

Its the transformation of data from a high-dimensional into a low dimensional space so that the later representation Keeps some meaningful properties of the original data.

Reasons for Dimensionality Reduction

- Visualization
- Reduce computational load
- Reduce variance in estimation(Regularize the problem.)

Methods of Dimensionality Reduction.

- Principal Component Analysis
- Canonical Correlation Analysis (Face Recognition)

Dimensional Reduction

Dimensionality Reduction

Its the transformation of data from a high-dimensional into a low dimensional space so that the later representation Keeps some meaningful properties of the original data.

Reasons for Dimensionality Reduction

- Visualization
- Reduce computational load
- Reduce variance in estimation(Regularize the problem.)

Methods of Dimensionality Reduction.

- Principal Component Analysis
- Canonical Correlation Analysis (Face Recognition)

Clustering

Clustering

Clustering

Clustering

Is a machine learning technique that finds groups of data points that are more similar to each other than to other data points in other groups.

Clustering

Clustering

Is a machine learning technique that finds groups of data points that are more similar to each other than to other data points in other groups.

Clustering Types

Clustering

Clustering

Is a machine learning technique that finds groups of data points that are more similar to each other than to other data points in other groups.

Clustering Types

- K-Means

Clustering

Clustering

Is a machine learning technique that finds groups of data points that are more similar to each other than to other data points in other groups.

Clustering Types

- K-Means
- K-Medoids

Clustering

Clustering

Is a machine learning technique that finds groups of data points that are more similar to each other than to other data points in other groups.

Clustering Types

- K-Means
- K-Medoids
- Hierarchical clustering

Clustering

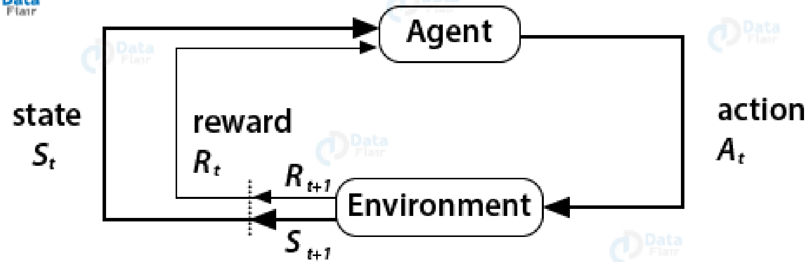
Clustering

Is a machine learning technique that finds groups of data points that are more similar to each other than to other data points in other groups.

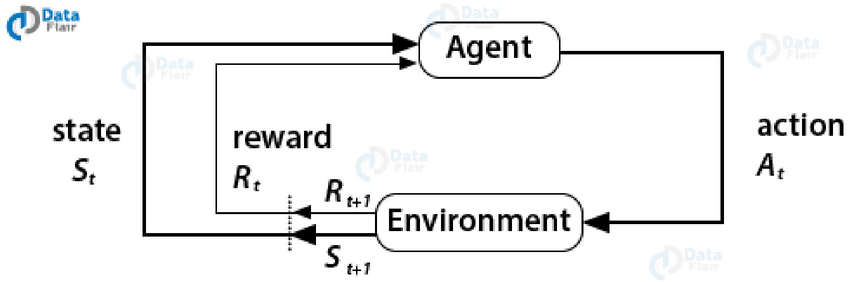
Clustering Types

- K-Means
- K-Medoids
- Hierarchical clustering

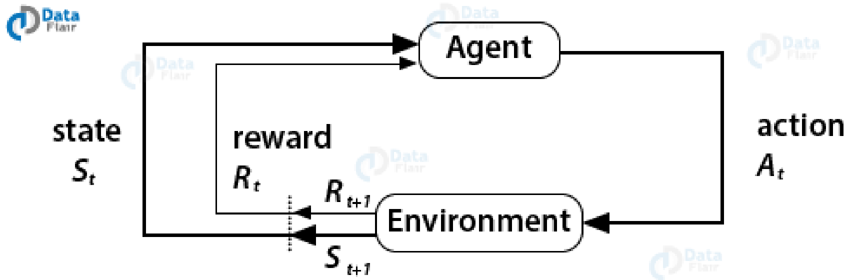
Reinforcement Learning Introduction



Reinforcement Learning Introduction

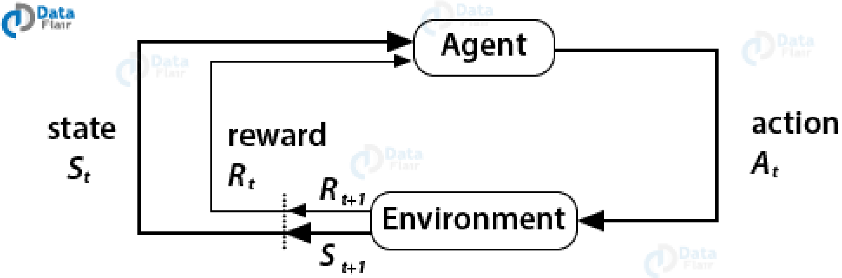


Reinforcement Learning Introduction



Its a machine learning method where by an agent learns to perform certain actions in an environment which lead it to maximum reward.

Reinforcement Learning Introduction



Its a machine learning method where by an agent learns to perform certain actions in an environment which lead it to maximum reward.

Reinforcement Learning Algorithms

Reinforcement Learning Algorithms

- Q-Learning

Reinforcement Learning Algorithms

- Q-Learning
- Temporal Difference (TD)

Reinforcement Learning Algorithms

- Q-Learning
- Temporal Difference (TD)
- Monte-Carlo Tree Search (MCTS)

Reinforcement Learning Algorithms

- Q-Learning
- Temporal Difference (TD)
- Monte-Carlo Tree Search (MCTS)
- Asynchronous Actor-Critic Agents (A3C)

Reinforcement Learning Algorithms

- Q-Learning
- Temporal Difference (TD)
- Monte-Carlo Tree Search (MCTS)
- Asynchronous Actor-Critic Agents (A3C)

Reinforcement Learning Problems

Reinforcement Learning Problems

- Gaming

Reinforcement Learning Problems

- Gaming
- Finance Sector

Reinforcement Learning Problems

- Gaming
- Finance Sector
- Manufacturing

Reinforcement Learning Problems

- Gaming
- Finance Sector
- Manufacturing
- Inventory Management

Reinforcement Learning Problems

- Gaming
- Finance Sector
- Manufacturing
- Inventory Management
- Robot Navigation

Reinforcement Learning Problems

- Gaming
- Finance Sector
- Manufacturing
- Inventory Management
- Robot Navigation

Data Science

What is Data Science?

Data Science

What is Data Science?

Data Science is a step by step process of extracting knowledge from data.

Data Science

What is Data Science?

Data Science is a step by step process of extracting knowledge from data.

Data Science Is OSEMN

Data Science

What is Data Science?

Data Science is a step by step process of extracting knowledge from data.

Data Science Is OSEMN

1. Obtaining data.

Data Science

What is Data Science?

Data Science is a step by step process of extracting knowledge from data.

Data Science Is OSEMN

1. Obtaining data.
2. Scrubbing data.

Data Science

What is Data Science?

Data Science is a step by step process of extracting knowledge from data.

Data Science Is OSEMN

1. Obtaining data.
2. Scrubbing data.
3. Exploring data.

Data Science

What is Data Science?

Data Science is a step by step process of extracting knowledge from data.

Data Science Is OSEMN

1. Obtaining data.
2. Scrubbing data.
3. Exploring data.
4. Modeling data and
5. Interpreting data

Data Science

What is Data Science?

Data Science is a step by step process of extracting knowledge from data.

Data Science Is OSEMN

1. Obtaining data.
2. Scrubbing data.
3. Exploring data.
4. Modeling data and
5. Interpreting data

Data Science Work Flow

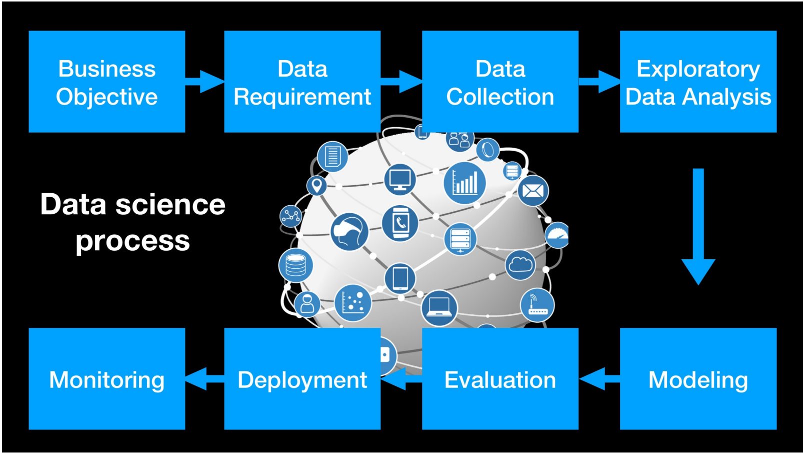


Figure: Data Science Workflow

Performance Metric

The metrics that you choose to evaluate your machine learning algorithms are very important.

Performance Metric

The metrics that you choose to evaluate your machine learning algorithms are very important.

Choice of metric influences how the performance of machine learning algorithms is measured and compared.

Performance Metric

The metrics that you choose to evaluate your machine learning algorithms are very important.

Choice of metric influences how the performance of machine learning algorithms is measured and compared.

Performance Metric Cont..

Performance Metric Cont..

Classification Metric

Performance Metric Cont..

Classification Metric

Classification problems are perhaps the most common type of machine learning problem and as such there are a myriad of metrics that can be used to evaluate predictions for these problems.

Performance Metric Cont..

Classification Metric

Classification problems are perhaps the most common type of machine learning problem and as such there are a myriad of metrics that can be used to evaluate predictions for these problems.

Classification Metric

Performance Metric Cont..

Classification Metric

Classification problems are perhaps the most common type of machine learning problem and as such there are a myriad of metrics that can be used to evaluate predictions for these problems.

Classification Metric

1. Classification Accuracy.

Performance Metric Cont..

Classification Metric

Classification problems are perhaps the most common type of machine learning problem and as such there are a myriad of metrics that can be used to evaluate predictions for these problems.

Classification Metric

1. Classification Accuracy.
2. Area Under ROC Curve(binary).

Performance Metric Cont..

Classification Metric

Classification problems are perhaps the most common type of machine learning problem and as such there are a myriad of metrics that can be used to evaluate predictions for these problems.

Classification Metric

1. Classification Accuracy.
2. Area Under ROC Curve(binary).
3. Confusion Matrix

Performance Metric Cont..

Classification Metric

Classification problems are perhaps the most common type of machine learning problem and as such there are a myriad of metrics that can be used to evaluate predictions for these problems.

Classification Metric

1. Classification Accuracy.
2. Area Under ROC Curve(binary).
3. Confusion Matrix
4. Log loss

Performance Metric Cont..

Classification Metric

Classification problems are perhaps the most common type of machine learning problem and as such there are a myriad of metrics that can be used to evaluate predictions for these problems.

Classification Metric

1. Classification Accuracy.
2. Area Under ROC Curve(binary).
3. Confusion Matrix
4. Log loss
5. Average precision at k.

Performance Metric Cont..

Classification Metric

Classification problems are perhaps the most common type of machine learning problem and as such there are a myriad of metrics that can be used to evaluate predictions for these problems.

Classification Metric

1. Classification Accuracy.
2. Area Under ROC Curve(binary).
3. Confusion Matrix
4. Log loss
5. Average precision at k.
6. Mean average precision at k.

Performance Metric Cont..

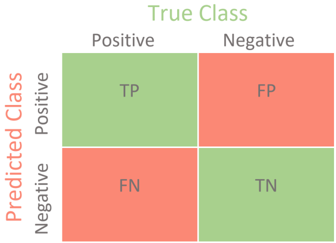
Classification Metric

Classification problems are perhaps the most common type of machine learning problem and as such there are a myriad of metrics that can be used to evaluate predictions for these problems.

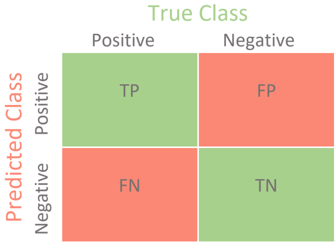
Classification Metric

1. Classification Accuracy.
2. Area Under ROC Curve(binary).
3. Confusion Matrix
4. Log loss
5. Average precision at k.
6. Mean average precision at k.

Confusion Matrix

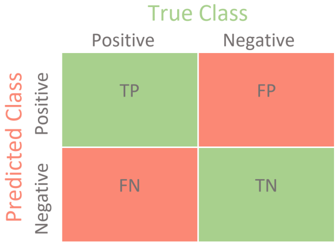


Confusion Matrix



Image

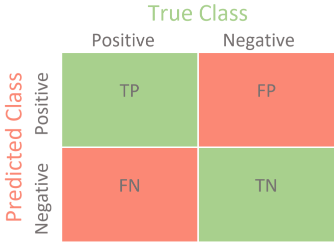
Confusion Matrix



Image

1. True Positives(TP): These are cases in which we predicted Yes and its actually Yes.

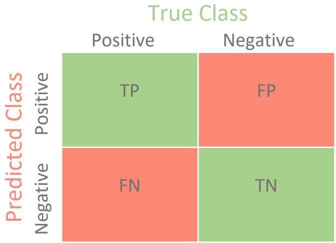
Confusion Matrix



Image

1. True Positives(TP): These are cases in which we predicted Yes and its actually Yes.
2. True Negative(TN): We predicted No and its actually No.

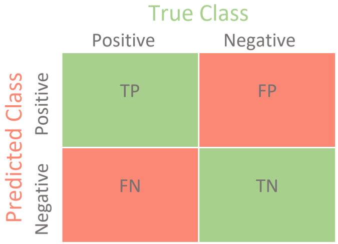
Confusion Matrix



Image

1. True Positives(TP): These are cases in which we predicted Yes and its actually Yes.
2. True Negative(TN): We predicted No and its actually No.
3. False Positive(FP): We predicted Yes, but its actually No.

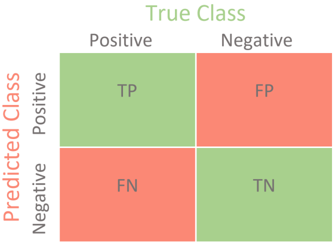
Confusion Matrix



Image

1. True Positives(TP): These are cases in which we predicted Yes and its actually Yes.
2. True Negative(TN): We predicted No and its actually No.
3. False Positive(FP): We predicted Yes, but its actually No.
4. False Negative(FN): We predicted NO , but its actually Yes.

Confusion Matrix



Image

1. True Positives(TP): These are cases in which we predicted Yes and its actually Yes.
2. True Negative(TN): We predicted No and its actually No.
3. False Positive(FP): We predicted Yes, but its actually No.
4. False Negative(FN): We predicted NO , but its actually Yes.

Performance Metric Cont..

Regression Metric

Performance Metric Cont..

Regression Metric

The most common metrics for evaluating predictions on regression machine learning problems:

Performance Metric Cont..

Regression Metric

The most common metrics for evaluating predictions on regression machine learning problems:

1. Mean Absolute Error(MAE).

Performance Metric Cont..

Regression Metric

The most common metrics for evaluating predictions on regression machine learning problems:

1. Mean Absolute Error(MAE).
2. Mean Squared Error(MSE).

Performance Metric Cont..

Regression Metric

The most common metrics for evaluating predictions on regression machine learning problems:

1. Mean Absolute Error(MAE).
2. Mean Squared Error(MSE).
3. Root mean squared error (RMSE)

Performance Metric Cont..

Regression Metric

The most common metrics for evaluating predictions on regression machine learning problems:

1. Mean Absolute Error(MAE).
2. Mean Squared Error(MSE).
3. Root mean squared error (RMSE)
4. Root mean squared logarithmic error (RMSLE)

Performance Metric Cont..

Regression Metric

The most common metrics for evaluating predictions on regression machine learning problems:

1. Mean Absolute Error(MAE).
2. Mean Squared Error(MSE).
3. Root mean squared error (RMSE)
4. Root mean squared logarithmic error (RMSLE)
5. R^2.

Performance Metric Cont..

Regression Metric

The most common metrics for evaluating predictions on regression machine learning problems:

1. Mean Absolute Error(MAE).
2. Mean Squared Error(MSE).
3. Root mean squared error (RMSE)
4. Root mean squared logarithmic error (RMSLE)
5. R^2.
6. Mean percentage error (MPE).

Performance Metric Cont..

Regression Metric

The most common metrics for evaluating predictions on regression machine learning problems:

1. Mean Absolute Error(MAE).
2. Mean Squared Error(MSE).
3. Root mean squared error (RMSE)
4. Root mean squared logarithmic error (RMSLE)
5. R^2.
6. Mean percentage error (MPE).
7. Mean absolute percentage error (MAPE)

Performance Metric Cont..

Regression Metric

The most common metrics for evaluating predictions on regression machine learning problems:

1. Mean Absolute Error(MAE).
2. Mean Squared Error(MSE).
3. Root mean squared error (RMSE)
4. Root mean squared logarithmic error (RMSLE)
5. R^2.
6. Mean percentage error (MPE).
7. Mean absolute percentage error (MAPE)

The cause of poor performance in machine learning is either overfitting or underfitting the data.

The cause of poor performance in machine learning is either overfitting or underfitting the data.

Over fitting

The cause of poor performance in machine learning is either overfitting or underfitting the data.

Over fitting

Over fitting refers to learning the training data too well at the expense of not generalizing well to new data.

The cause of poor performance in machine learning is either overfitting or underfitting the data.

Over fitting

Over fitting refers to learning the training data too well at the expense of not generalizing well to new data.
It is the most common problem in practice and can be addressed by;

The cause of poor performance in machine learning is either overfitting or underfitting the data.

Over fitting

Over fitting refers to learning the training data too well at the expense of not generalizing well to new data.
It is the most common problem in practice and can be addressed by;

- Using re-sampling methods.

The cause of poor performance in machine learning is either overfitting or underfitting the data.

Over fitting

Over fitting refers to learning the training data too well at the expense of not generalizing well to new data.
It is the most common problem in practice and can be addressed by;

- Using re-sampling methods.
- Hold back a validation data set.

The cause of poor performance in machine learning is either overfitting or underfitting the data.

Over fitting

Over fitting refers to learning the training data too well at the expense of not generalizing well to new data.
It is the most common problem in practice and can be addressed by;

- Using re-sampling methods.
- Hold back a validation data set.
- Increase training set.

The cause of poor performance in machine learning is either overfitting or underfitting the data.

Over fitting

Over fitting refers to learning the training data too well at the expense of not generalizing well to new data.
It is the most common problem in practice and can be addressed by;

- Using re-sampling methods.
- Hold back a validation data set.
- Increase training set.
- feature selection.

The cause of poor performance in machine learning is either overfitting or underfitting the data.

Over fitting

Over fitting refers to learning the training data too well at the expense of not generalizing well to new data.
It is the most common problem in practice and can be addressed by;

- Using re-sampling methods.
- Hold back a validation data set.
- Increase training set.
- feature selection.
- L2(ridge) and L1(lasso) regularization.

The cause of poor performance in machine learning is either overfitting or underfitting the data.

Over fitting

Over fitting refers to learning the training data too well at the expense of not generalizing well to new data.
It is the most common problem in practice and can be addressed by;

- Using re-sampling methods.
- Hold back a validation data set.
- Increase training set.
- feature selection.
- L2(ridge) and L1(lasso) regularization.

Under fitting

Under fitting

Under fitting refers to failing to learn the problem from the training data sufficiently.

Under fitting

Under fitting refers to failing to learn the problem from the training data sufficiently.
And can be addressed by;

Under fitting

Under fitting refers to failing to learn the problem from the training data sufficiently.
And can be addressed by;

- Add features.

Under fitting

Under fitting refers to failing to learn the problem from the training data sufficiently.
And can be addressed by;

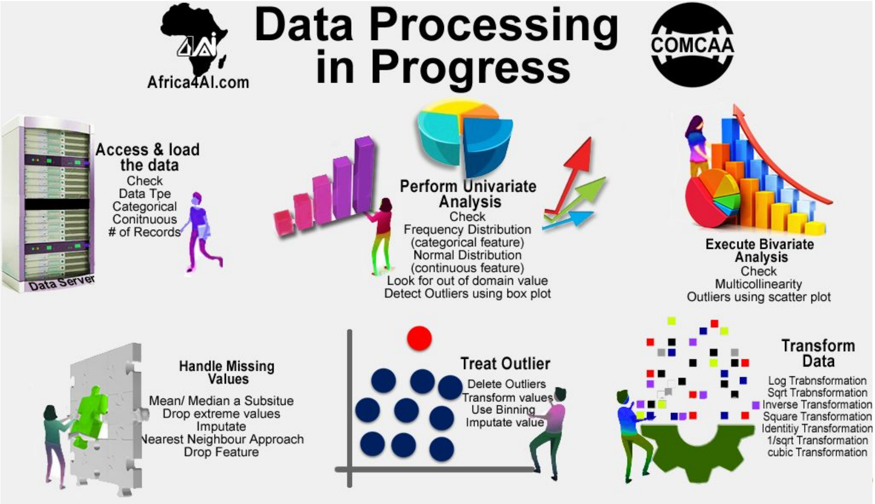
- Add features.
- Increase training time.

Under fitting

Under fitting refers to failing to learn the problem from the training data sufficiently.
And can be addressed by;

- Add features.
- Increase training time.

Data Preprocessing



Univariate plots

- Histogram
- Density Plots

Univariate plots

- Histogram
- Density Plots
- Box and whisker Plots

Univariate plots

- Histogram
- Density Plots
- Box and whisker Plots

Multivariate Plots

- Correlation Matrix Plot.

Univariate plots

- Histogram
- Density Plots
- Box and whisker Plots

Multivariate Plots

- Correlation Matrix Plot.
- Scatter Plot Matrix

Univariate plots

- Histogram
- Density Plots
- Box and whisker Plots

Multivariate Plots

- Correlation Matrix Plot.
- Scatter Plot Matrix

Importance of Features Selection

Importance of Features Selection

Importance of Features Selection

Importance of Features Selection

- **Reduces Overfitting:**

Importance of Features Selection

Importance of Features Selection

- **Reduces Overfitting:** Less redundant data means less opportunity to make decisions based on noise.
- **Improves Accuracy:**

Importance of Features Selection

Importance of Features Selection

- **Reduces Overfitting:** Less redundant data means less opportunity to make decisions based on noise.
- **Improves Accuracy:** Less misleading data means modeling accuracy improves.
- **Reduces Training Time:**

Importance of Features Selection

Importance of Features Selection

- **Reduces Overfitting:** Less redundant data means less opportunity to make decisions based on noise.
- **Improves Accuracy:** Less misleading data means modeling accuracy improves.
- **Reduces Training Time:** Less data means that algorithms train faster.

Implementation

Implementation Steps

Implementation

Implementation Steps

1. Prepare Problem

Implementation

Implementation Steps

1. Prepare Problem
   - Load libraries

Implementation

Implementation Steps

1. Prepare Problem
   - Load libraries
   - Load dataset.

Implementation

Implementation Steps

1. Prepare Problem
   - Load libraries
   - Load dataset.
2. Summarize Data

Implementation

Implementation Steps

1. Prepare Problem
   - Load libraries
   - Load dataset.
2. Summarize Data
   - Descriptive statistics

Implementation

Implementation Steps

1. Prepare Problem
   - Load libraries
   - Load dataset.
2. Summarize Data
   - Descriptive statistics
   - Data visualizations.

Implementation

Implementation Steps

1. Prepare Problem
   - Load libraries
   - Load dataset.
2. Summarize Data
   - Descriptive statistics
   - Data visualizations.
3. Prepare Data:

Implementation

Implementation Steps

1. Prepare Problem
   - Load libraries
   - Load dataset.
2. Summarize Data
   - Descriptive statistics
   - Data visualizations.
3. Prepare Data:
   - Data Cleaning

Implementation

Implementation Steps

1. Prepare Problem
   - Load libraries
   - Load dataset.
2. Summarize Data
   - Descriptive statistics
   - Data visualizations.
3. Prepare Data:
   - Data Cleaning
   - Feature Selection

Implementation

Implementation Steps

1. Prepare Problem
   - Load libraries
   - Load dataset.
2. Summarize Data
   - Descriptive statistics
   - Data visualizations.
3. Prepare Data:
   - Data Cleaning
   - Feature Selection
   - Data Transforms

Implementation

Implementation Steps

1. Prepare Problem
   - Load libraries
   - Load dataset.
2. Summarize Data
   - Descriptive statistics
   - Data visualizations.
3. Prepare Data:
   - Data Cleaning
   - Feature Selection
   - Data Transforms

Implementation Cont....

4. Evaluate Algorithms:

Implementation Cont....

4. Evaluate Algorithms:
   - Split-out validation dataset.

Implementation Cont....

4. Evaluate Algorithms:
   - Split-out validation dataset.
   - Test options and evaluation metric.

Implementation Cont....

4. Evaluate Algorithms:
   - Split-out validation dataset.
   - Test options and evaluation metric.
   - Spot Check Algorithms.

Implementation Cont....

4. Evaluate Algorithms:
   - Split-out validation dataset.
   - Test options and evaluation metric.
   - Spot Check Algorithms.
   - Compare Algorithms.

Implementation Cont....

4. Evaluate Algorithms:
   - Split-out validation dataset.
   - Test options and evaluation metric.
   - Spot Check Algorithms.
   - Compare Algorithms.
5. Improve Accuracy.

Implementation Cont....

4. Evaluate Algorithms:
   - Split-out validation dataset.
   - Test options and evaluation metric.
   - Spot Check Algorithms.
   - Compare Algorithms.
5. Improve Accuracy.
   - Algorithm Tuning.

Implementation Cont....

4. Evaluate Algorithms:
   - Split-out validation dataset.
   - Test options and evaluation metric.
   - Spot Check Algorithms.
   - Compare Algorithms.
5. Improve Accuracy.
   - Algorithm Tuning.
   - Ensembles.

Implementation Cont....

4. Evaluate Algorithms:
   - Split-out validation dataset.
   - Test options and evaluation metric.
   - Spot Check Algorithms.
   - Compare Algorithms.
5. Improve Accuracy.
   - Algorithm Tuning.
   - Ensembles.
6. Finalize Model

Implementation Cont....

4. Evaluate Algorithms:
   - Split-out validation dataset.
   - Test options and evaluation metric.
   - Spot Check Algorithms.
   - Compare Algorithms.
5. Improve Accuracy.
   - Algorithm Tuning.
   - Ensembles.
6. Finalize Model
   - Finalize Model

Implementation Cont....

4. Evaluate Algorithms:
   - Split-out validation dataset.
   - Test options and evaluation metric.
   - Spot Check Algorithms.
   - Compare Algorithms.
5. Improve Accuracy.
   - Algorithm Tuning.
   - Ensembles.
6. Finalize Model
   - Finalize Model
   - Predictions on validation dataset.

Implementation Cont....

4. Evaluate Algorithms:
   - Split-out validation dataset.
   - Test options and evaluation metric.
   - Spot Check Algorithms.
   - Compare Algorithms.
5. Improve Accuracy.
   - Algorithm Tuning.
   - Ensembles.
6. Finalize Model
   - Finalize Model
   - Predictions on validation dataset.
   - Create standalone model on entire training dataset.

Implementation Cont....

4. Evaluate Algorithms:
   - Split-out validation dataset.
   - Test options and evaluation metric.
   - Spot Check Algorithms.
   - Compare Algorithms.
5. Improve Accuracy.
   - Algorithm Tuning.
   - Ensembles.
6. Finalize Model
   - Finalize Model
   - Predictions on validation dataset.
   - Create standalone model on entire training dataset.
   - Save model for later use.

Implementation Cont....

4. Evaluate Algorithms:
   - Split-out validation dataset.
   - Test options and evaluation metric.
   - Spot Check Algorithms.
   - Compare Algorithms.
5. Improve Accuracy.
   - Algorithm Tuning.
   - Ensembles.
6. Finalize Model
   - Finalize Model
   - Predictions on validation dataset.
   - Create standalone model on entire training dataset.
   - Save model for later use.

Solving Problems Using Machine Learning

Solving Problems Using Machine Learning

- Is it "A" or "B"?

Solving Problems Using Machine Learning

Solving Problems Using Machine Learning

- Is it “A" or “B"?
  Classification

Solving Problems Using Machine Learning

Solving Problems Using Machine Learning

- Is it “A" or “B"?
  Classification
- Is this Weird?

Solving Problems Using Machine Learning

Solving Problems Using Machine Learning

- Is it “A" or “B"?
  Classification
- Is this Weird?
  Anomality Detection

Solving Problems Using Machine Learning

Solving Problems Using Machine Learning

- Is it “A" or “B"?
  Classification
- Is this Weird?
  Anomality Detection
- How much or how many?

Solving Problems Using Machine Learning

Solving Problems Using Machine Learning

- Is it “A" or “B"?
  Classification
- Is this Weird?
  Anomality Detection
- How much or how many?
  Regression.

Solving Problems Using Machine Learning

Solving Problems Using Machine Learning

- Is it “A" or “B"?
  Classification
- Is this Weird?
  Anomality Detection
- How much or how many?
  Regression.
- How is this Organized?

Solving Problems Using Machine Learning

Solving Problems Using Machine Learning

- Is it “A" or “B"?
  Classification
- Is this Weird?
  Anomality Detection
- How much or how many?
  Regression.
- How is this Organized?
  Clustering

Digital Native

Sagesse Devoir
Ecole Polytechnique Thiès

Solving Problems Using Machine Learning

Solving Problems Using Machine Learning

- Is it "A" or "B"?
  Classification
- Is this Weird?
  Anomality Detection
- How much or how many?
  Regression.
- How is this Organized?
  Clustering
- What should i do next?

Solving Problems Using Machine Learning

Solving Problems Using Machine Learning

- Is it “A" or “B"?
  Classification
- Is this Weird?
  Anomality Detection
- How much or how many?
  Regression.
- How is this Organized?
  Clustering
- What should i do next?
  Reinforcement Learning

Solving Problems Using Machine Learning

Solving Problems Using Machine Learning

- Is it “A" or “B"?
  Classification
- Is this Weird?
  Anomality Detection
- How much or how many?
  Regression.
- How is this Organized?
  Clustering
- What should i do next?
  Reinforcement Learning

Confusing Terms

- **Data Mining:**

Confusing Terms

- **Data Mining:** Is the discovering of interesting, unexpected, or valuable structures in large data sets.
- **Statistic :**

Confusing Terms

- **Data Mining:** Is the discovering of interesting, unexpected, or valuable structures in large data sets.
- **Statistic :** Is the methodology for collecting the data, treatment and interpretation.
- **Statistic(s):**

Confusing Terms

- **Data Mining:** Is the discovering of interesting, unexpected, or valuable structures in large data sets.
- **Statistic :** Is the methodology for collecting the data, treatment and interpretation.
- **Statistic(s):**Is the collection of data,analyzing and interpretation of data results.

Sources of Data

- Human Generated data .

Sources of Data

- Human Generated data .
- Machine Generated Data.

Sources of Data

- Human Generated data .
- Machine Generated Data.
- Business Generated data.

Sources of Data

- Human Generated data .
- Machine Generated Data.
- Business Generated data.

Types of Data

- Structured

Types of Data

- Structured
- Semi-Structured

Types of Data

- Structured
- Semi-Structured
- Unstructured

Types of Data

- Structured
- Semi-Structured
- Unstructured

Implementation

Programming Language

Implementation

Programming Language

The primary tool to be used for the implementation is;

Implementation

Programming Language

The primary tool to be used for the implementation is; **Python.**(scikit-learn, pandas, matplotlib,seabon.)

Why?

Implementation

Programming Language

The primary tool to be used for the implementation is; **Python.**(scikit-learn, pandas, matplotlib,seabon.)

Why?

- Code readability

Implementation

Programming Language

The primary tool to be used for the implementation is; **Python.**(scikit-learn, pandas, matplotlib,seabon.)

Why?

- Code readability
- Easy to Learn

Implementation

Programming Language

The primary tool to be used for the implementation is; **Python.**(scikit-learn, pandas, matplotlib,seabon.)

Why?

- Code readability
- Easy to Learn
- Well-Supported

Implementation

Programming Language

The primary tool to be used for the implementation is;
Python.(scikit-learn, pandas, matplotlib,seabon.)

Why?

- Code readability
- Easy to Learn
- Well-Supported
- Graphics and Visualization Tools

Implementation

Programming Language

The primary tool to be used for the implementation is; **Python.**(scikit-learn, pandas, matplotlib,seabon.)

Why?

- Code readability
- Easy to Learn
- Well-Supported
- Graphics and Visualization Tools
- Well-Supported

Implementation

Programming Language

The primary tool to be used for the implementation is; **Python.**(scikit-learn, pandas, matplotlib,seabon.)

Why?

- Code readability
- Easy to Learn
- Well-Supported
- Graphics and Visualization Tools
- Well-Supported
- Scalability

Implementation

Programming Language

The primary tool to be used for the implementation is; **Python.**(scikit-learn, pandas, matplotlib,seabon.)

Why?

- Code readability
- Easy to Learn
- Well-Supported
- Graphics and Visualization Tools
- Well-Supported
- Scalability

Train, Test and Validation

Training Set

Train, Test and Validation

Training Set

To fit the parameters i.e weights.

Train,Test and Validation

Training Set

To fit the parameters i.e weights.

Validation Set.

Train, Test and Validation

Training Set

To fit the parameters i.e weights.

Validation Set.

- Part of the training set

Train, Test and Validation

Training Set

To fit the parameters i.e weights.

Validation Set.

- Part of the training set
- For parameter selection

Train, Test and Validation

Training Set

To fit the parameters i.e weights.

Validation Set.

- Part of the training set
- For parameter selection
- To avoid over fitting

Train, Test and Validation

Training Set

To fit the parameters i.e weights.

Validation Set.

- Part of the training set
- For parameter selection
- To avoid over fitting

Testing Set

Train,Test and Validation

Training Set

To fit the parameters i.e weights.

Validation Set.

- Part of the training set
- For parameter selection
- To avoid over fitting

Testing Set

For testing or evaluating the performance of a trained machine learning model,i.e. evaluating the predictive power and generalization.

Train,Test and Validation

Training Set

To fit the parameters i.e weights.

Validation Set.

- Part of the training set
- For parameter selection
- To avoid over fitting

Testing Set

For testing or evaluating the performance of a trained machine learning model,i.e. evaluating the predictive power and generalization.

Recommendations

Books

Recommendations

Books

- The Elements of Statistical Learning: Data Mining, Inference, and Prediction Trevor Hastie, Robert Tibshirani, and Jerome Friedman.

Recommendations

Books

- The Elements of Statistical Learning: Data Mining, Inference, and Prediction Trevor Hastie, Robert Tibshirani, and Jerome Friedman.
- Pattern Recognition of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies (MIT Press).

Recommendations

Books

- The Elements of Statistical Learning: Data Mining, Inference, and Prediction Trevor Hastie, Robert Tibshirani, and Jerome Friedman.
- Pattern Recognition of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies (MIT Press).
- Understanding Machine Learning: From Theory to Algorithms..

Recommendations

Books

- The Elements of Statistical Learning: Data Mining, Inference, and Prediction Trevor Hastie, Robert Tibshirani, and Jerome Friedman.
- Pattern Recognition of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies (MIT Press).
- Understanding Machine Learning: From Theory to Algorithms..

1. https://www.edureka.co/blog/introduction-to-machine-learning/
2. https://analyticsindiamag.com/
3. https://data-flair.training/blogs/machine-learning-tutorial/
4. https://machinelearningmastery.com/machine-learning-with-python/
5. www.africa4ai.com

Conclusion


THANK
YOU
GRACIAS
ARIGATO
SHUKURIA
DANKSCHEEN
TASHAKKUR ATU
SUKSAMA
BIYAN
SHUKRIA
GRAZIE
MEHRBANI
BOLZIN
MERCI
JUSPAXAR