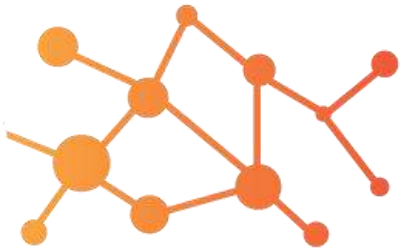


madBlocks

Technology:Innovation:Business

Machine Learning Applications using Python

Madhu Parvathaneni
Director & Certified ML Expert
Madblocks Technologies Pvt Ltd
mad@madblocks.tech

For questions, write us on ai@madblocks.tech

Agenda

madBlocks
Technology:Innovation:Business

Session – 1: Machine Learning Intro

– What is Machine Learning?
– Applications of Machine Learning
– Types of Machine Learning
– Tools and Pre-Requisites

Session – 2: Data Analysis using Python

– MQTT (Data Collection)
– Data Cleaning
– Data Pre-Processing
– Data Analysis Examples

Session – 3: Building Crop Prediction System

– Dataset Review
– Data Cleaning
– Model Training
– API Creation and Demo

madBlocks
Technology:Innovation:Business

Session – 1:
Intro to Machine Learning

For questions, write us on ai@madblocks.tech

Machine Learning

madBlocks
Technology:Innovation:Business

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

- Herbert Simon

Definition by Tom Mitchell (1998):

Machine Learning is the study of algorithms that

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

A well-defined learning task is given by <P, T, E>.

madBlocks
Technology:Innovation:Business

Traditional Programming

Machine Learning

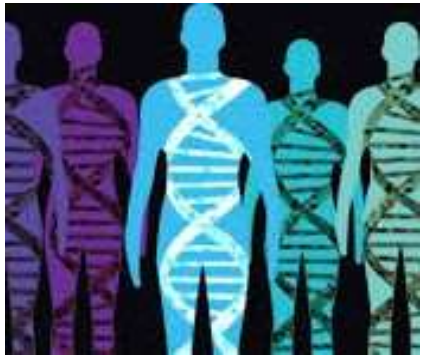
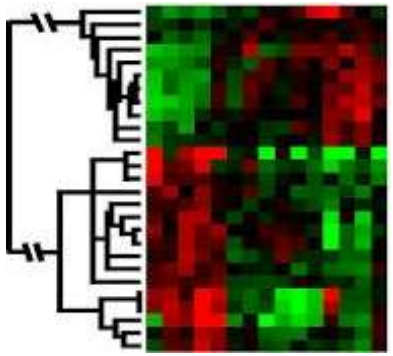
When Do We Use Machine Learning?

madBlocks
Technology:Innovation:Business

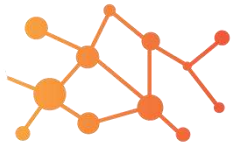
ML is used when:

- 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)



Some more examples of tasks that are best solved by using a learning algorithm



- Recognizing patterns:
  - Facial identities or facial expressions
  - Handwritten or spoken words
  - Medical images
- Generating patterns:
  - Generating images or motion sequences
- Recognizing anomalies:
  - Unusual credit card transactions
  - Unusual patterns of sensor readings in a nuclear power plant
- Prediction:
  - Future stock prices or currency exchange rates

Sample Applications

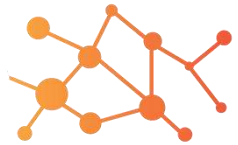
madBlocks
Technology:Innovation:Business

- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration
- Robotics
- Information extraction
- Social networks
- Debugging software
- Healthcare

madBlocks
Technology:Innovation:Business

"Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed." -Arthur Samuel (1959)

Defining the Learning Task



Improve on task T, with respect to performance metric P, based on experience E

T: Playing checkers

P: Percentage of games won against an arbitrary opponent

E: Playing practice games against itself

T: Recognizing hand-written words

P: Percentage of words correctly classified

E: Database of human-labeled images of handwritten words

T: Driving on four-lane highways using vision sensors

P: Average distance traveled before a human-judged error

E: A sequence of images and steering commands recorded while observing a human driver

madBlocks
Technology:Innovation:Business

Types of Learning

Types of Learning

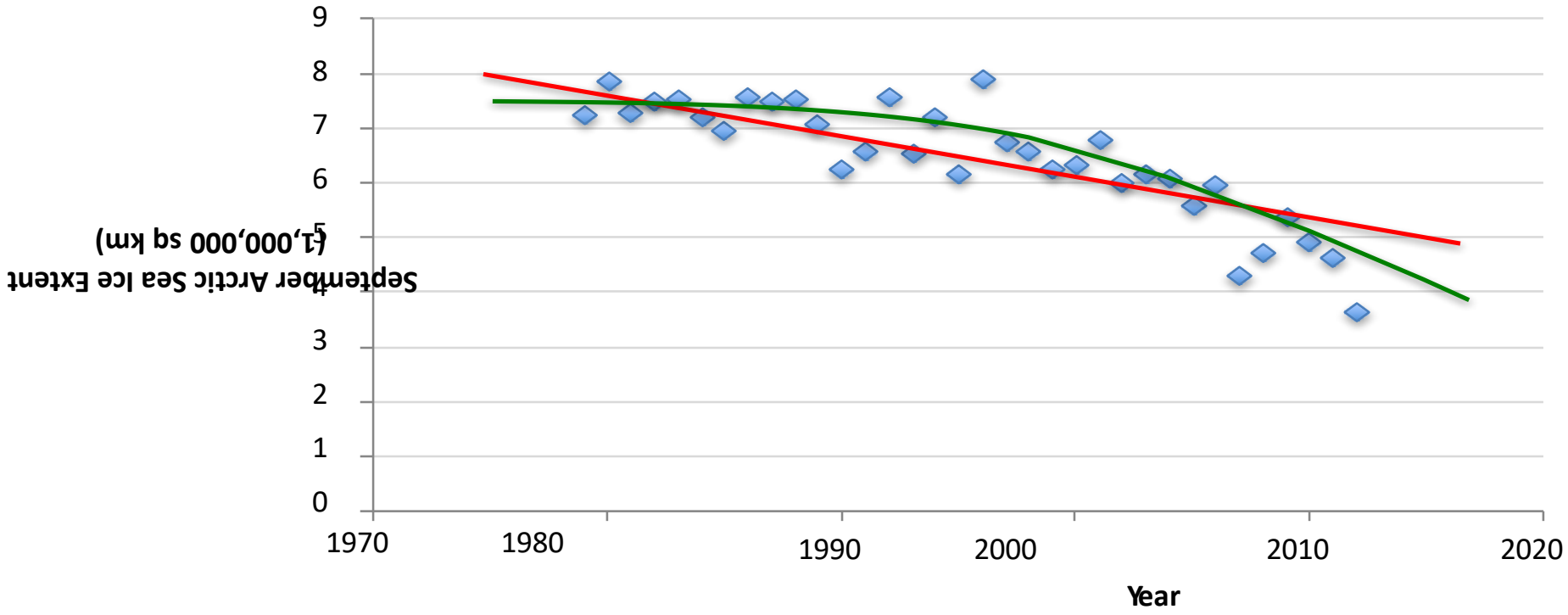
madBlocks
Technology:Innovation:Business

- **Supervised (inductive) learning**
  - Given: training data + desired outputs (labels)
- **Unsupervised learning**
  - Given: training data (without desired outputs)
- **Semi-supervised learning**
  - Given: training data + a few desired outputs
- **Reinforcement learning**
  - Rewards from sequence of actions

Supervised Learning: Regression

madBlocks
Technology:Innovation:Business

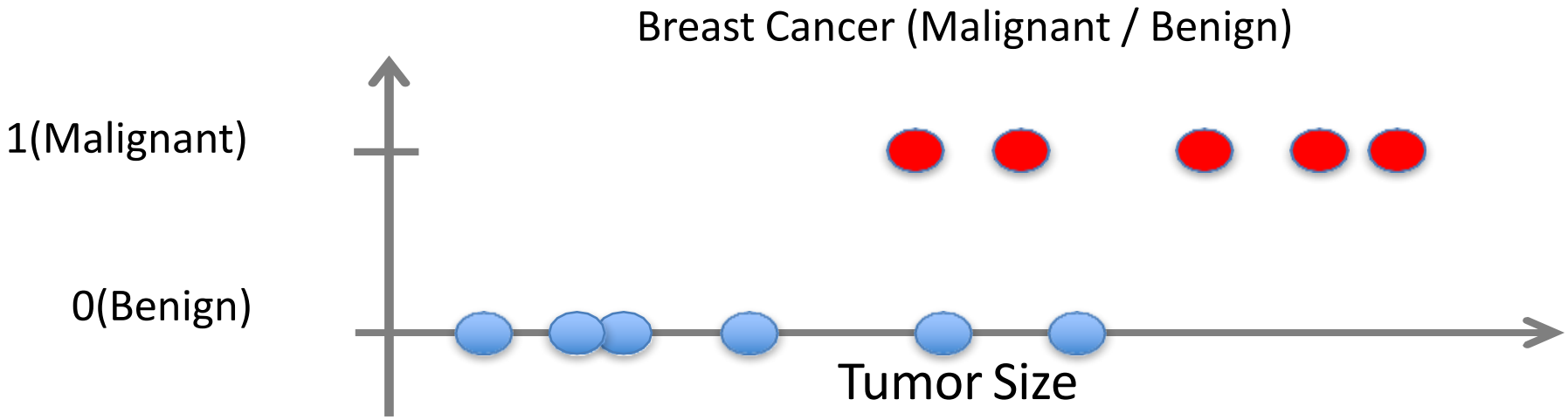
- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is real-valued == regression



Supervised Learning: Classification

madBlocks
Technology:Innovation:Business

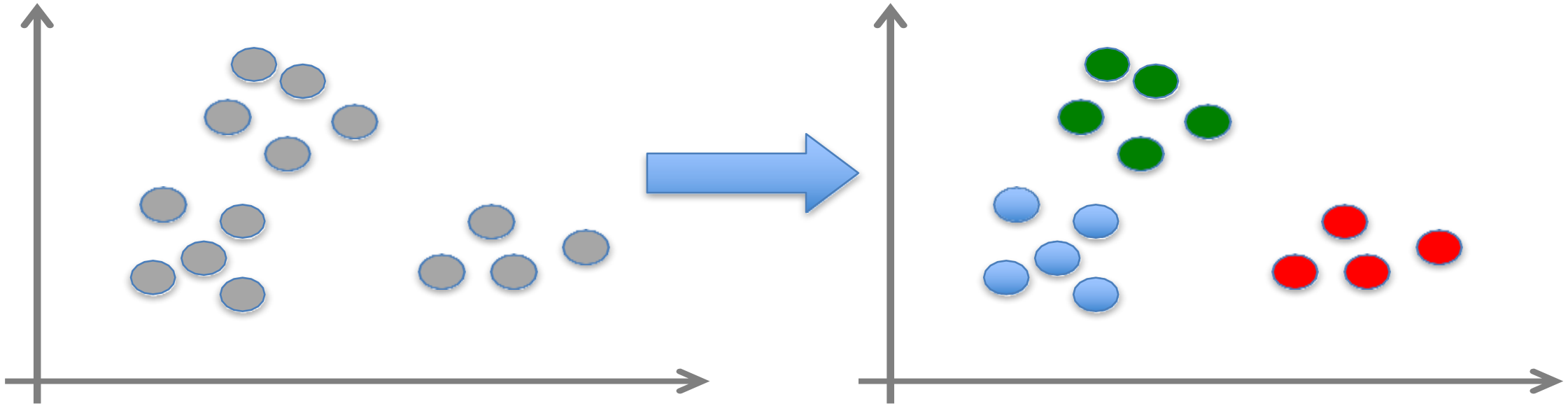
- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is categorical == classification



Unsupervised Learning

madBlocks
Technology:Innovation:Business

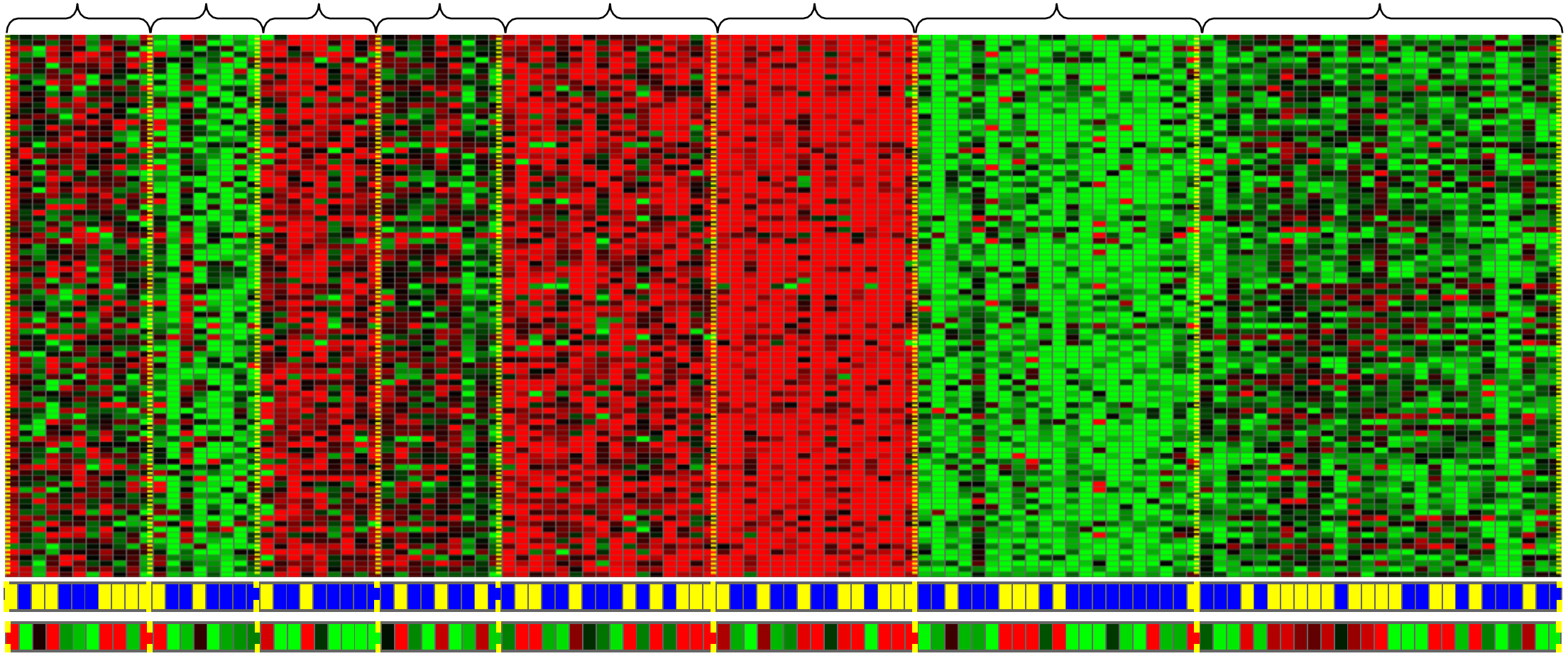
- Given $x_1, x_2, \ldots, x_n$ (without labels)
- Output hidden structure behind the x's
  - E.g., clustering



Unsupervised Learning

madBlocks
Technology:Innovation:Business

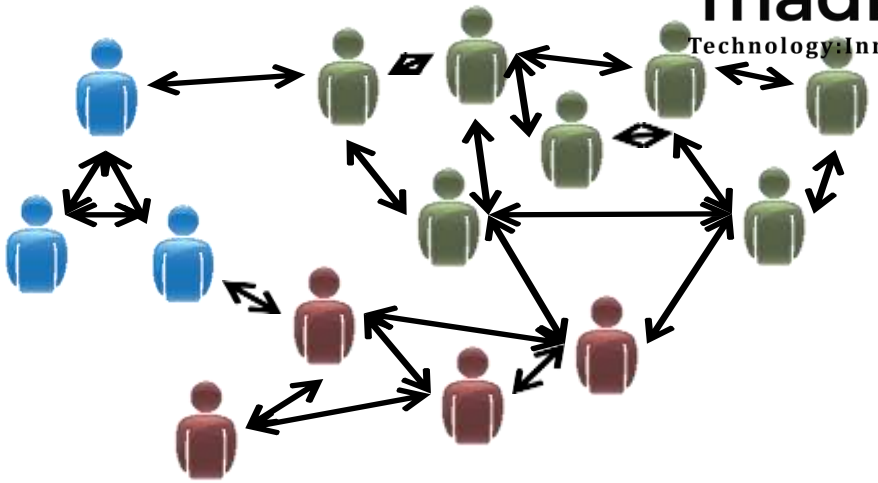
Genomics application: group individuals by genetic similarity



Unsupervised Learning

madBlocks
Technology:Innovation:Business


Organize computing clusters



Social network analysis

Market segmentation

Astronomical data analysis

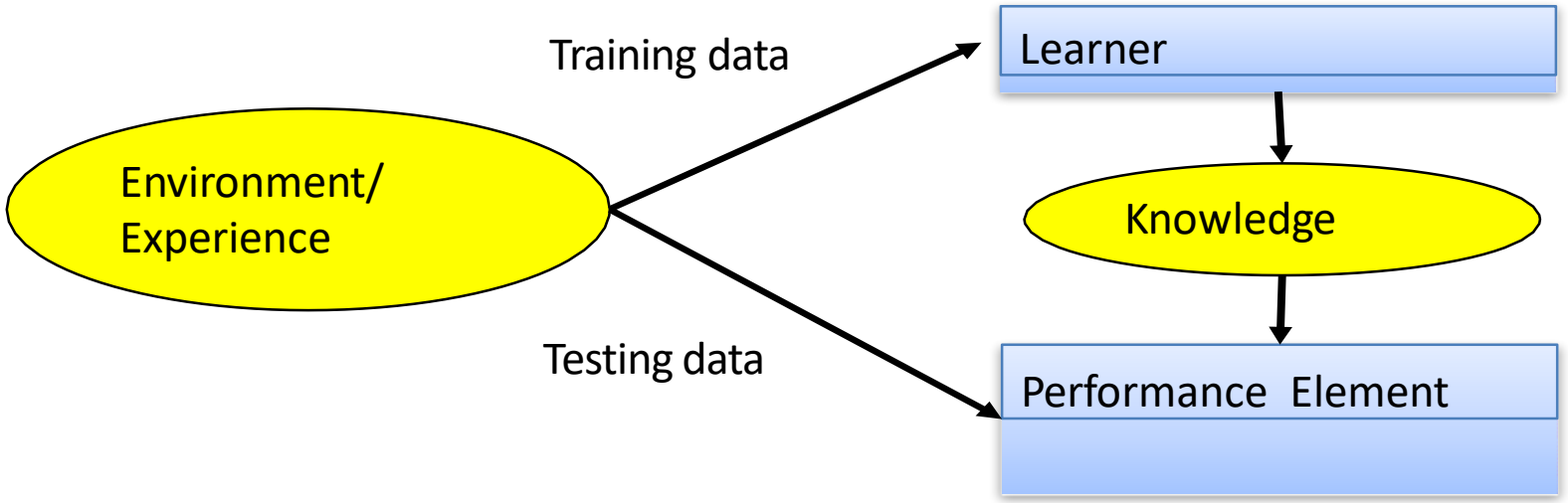
madBlocks
Technology:Innovation:Business

Framing a Learning Problem

Designing a Learning System

madBlocks
Technology:Innovation:Business

- Choose the training experience
- Choose exactly what is to be learned
  - i.e. the ***target function***
- Choose how to represent the target function
- Choose a learning algorithm to infer the target function from the experience



ML in a Nutshell

madBlocks
Technology:Innovation:Business

- Tens of thousands of machine learning algorithms
  - Hundreds new every year

- Every ML algorithm has three components:
  - **Representation**
  - **Optimization**
  - **Evaluation**

Various Function Representations

madBlocks
Technology:Innovation:Business

- Numerical functions
  - Linear regression
  - Neural networks
  - Support vector machines
- Symbolic functions
  - Decision trees
  - Rules in propositional logic
  - Rules in first-order predicate logic
- Instance-based functions
  - Nearest-neighbor
  - Case-based
- Probabilistic Graphical Models
  - Naïve Bayes
  - Bayesian networks
  - Hidden-Markov Models (HMMs)
  - Probabilistic Context Free Grammars (PCFGs)
  - Markov networks

Summary

madBlocks
Technology:Innovation:Business

Pack-Up!

– We gone through the basic understanding of Machine Learning and applications of ML.

– Machine to get learned can be broadly divided into four types (supervised, unsupervised, semisupervised and reinforcement learning).

– Data Pre-Processing is a very important step on the dataset.

madBlocks
Technology:Innovation:Business

Session – 2:
Data Analysis using Python

For questions, write us on blockchain@madblocks.tech

Data Collection

madBlocks
Technology:Innovation:Business

Data Collection

Let’s collect the data from IoT devices:

The data has to be collected from the IoT devices through MQTT Protocol and the received data will be stored in the database server.

The collected data has to be pre-processed for optimizing the model.

MQTT Protocol

madBlocks
Technology:Innovation:Business

Features

Small code footprint,
Ideal if processor or memory resources are limited,
Ideal if bandwidth is low or network is unreliable,
Works on top of TCP/IP

MQTT Protocol

madBlocks
Technology:Innovation:Business

Features

Small code footprint,
Ideal if processor or memory resources are limited,
Ideal if bandwidth is low or network is unreliable,
Works on top of TCP/IP

Pub-Sub Model

Publisher – Sender,
Subscriber - Receiver

MQTT Protocol

madBlocks
Technology:Innovation:Business

Features

Small code footprint,
Ideal if processor or memory resources are limited,
Ideal if bandwidth is low or network is unreliable,
Works on top of TCP/IP

Pub-Sub Model

Publisher – Sender,
Subscriber - Receiver

Broker and Topic

Broker – Server connecting Pub and Sub
Topic – On which topic the messages are shared
Port – 1883 (Default Broker)
Broker – **Your Gateway IP**

For questions, write us on ai@madblocks.tech

MQTT Protocol

madBlocks
Technology:Innovation:Business

Features

Small code footprint,
Ideal if processor or memory resources are limited,
Ideal if bandwidth is low or network is unreliable,
Works on top of TCP/IP

Broker and Topic

Broker – Server connecting Pub and Sub
Topic – On which topic the messages are shared
Port – 1883 (Default Broker)
Broker – **Your Gateway IP**

Pub-Sub Model

Publisher – Sender,
Subscriber - Receiver

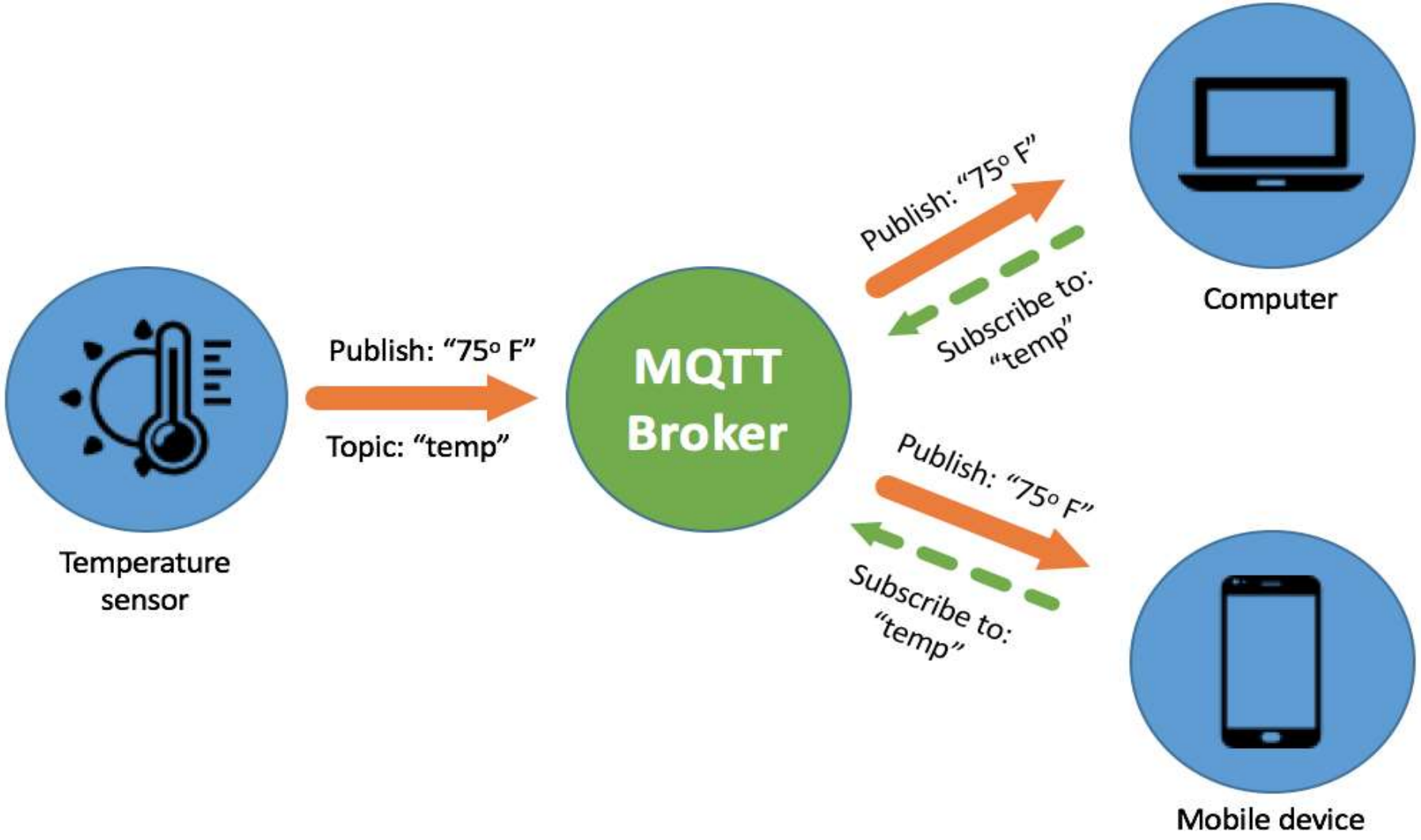
Applications

Smart Home
Smart City
Smart Healthcare
Smart Agriculture
Smart Industries etc….

For questions, write us on ai@madblocks.tech

MQTT Flow

madBlocks
Technology:Innovation:Business



For questions, write us on ai@madblocks.tech

Work Flow

madBlocks
Technology:Innovation:Business

IoT Devices: Sensory Connected Devices

Protocol: MQTT

Publisher: IoT Devices

Subscriber: Google Colab

Step – 1: Create a Publisher Script

Step – 2: Create a Subscriber Script

Step – 3: Store Data in CSV

Step – 4: Pre-Process the data

madBlocks
Technology:Innovation:Business

Session – 3:
Building Crop Prediction System

For questions, write us on ai@madblocks.tech

Problem Statement

madBlocks
Technology:Innovation:Business

The developers team of Anurag University is developing an intelligent crop prediction system for the farming community.

So, let's jump in to support them.

Problem Statement:

Farmers yield various crops as per their knowledge without much study on various parameters which affects the yielding. They follow a void method based on what surrounding neighbors do.

So an engineering group from the university campus wants to develop an assistant to suggest the crop by taking the following parameters into consideration:

1. Temperature
2. Humidity
3. pH
4. rainfall

Explore the Dataset

madBlocks
Technology:Innovation:Business

	temperature	humidity	ph	rainfall	label
0	20.879744	82.002744	6.502985	202.935536	rice
1	21.770462	80.319644	7.038096	226.655537	rice
2	23.004459	82.320763	7.840207	263.964248	rice
3	26.491096	80.158363	6.980401	242.864034	rice
4	20.130175	81.604873	7.628473	262.717340	rice

Each crop has sets of temperature, humidity, ph and rainfall values stored in the dataset.

There were 3100 samples for 31 crops, and now we have to create a model to predict the crop when these values are entered.

For questions, write us on ai@madblocks.tech

Data Pre-Processing

madBlocks
Technology:Innovation:Business

1. Identify the empty fields, if any then try to modify with mean or average of the respective data.

2. Identify the range of data, if the difference is too large then use some standardization methods like Standard Deviation.

3. Identify the categorical data, if found use one-hot encoding and convert them to numerical data.

4. Identify the input variables and target variables from the features of the dataset.

5. Split the data into Training and Testing.

K Nearest Neighbors

madBlocks
Technology:Innovation:Business

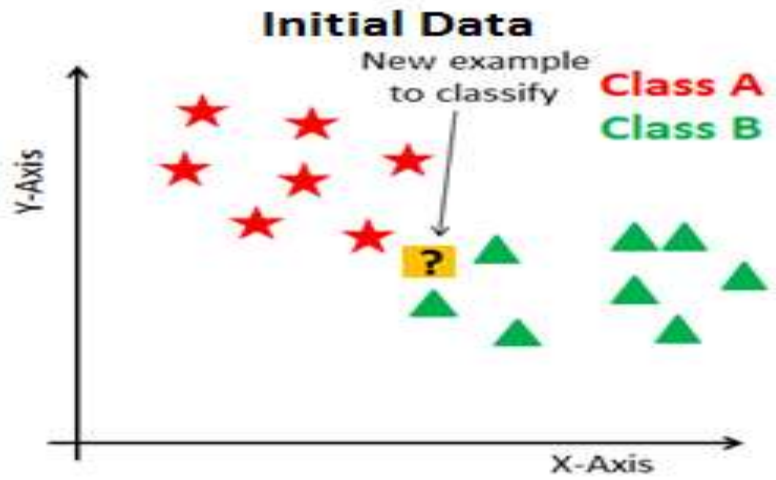
KNN can be used for classification — the output is a class membership (predicts a class — a discrete value).

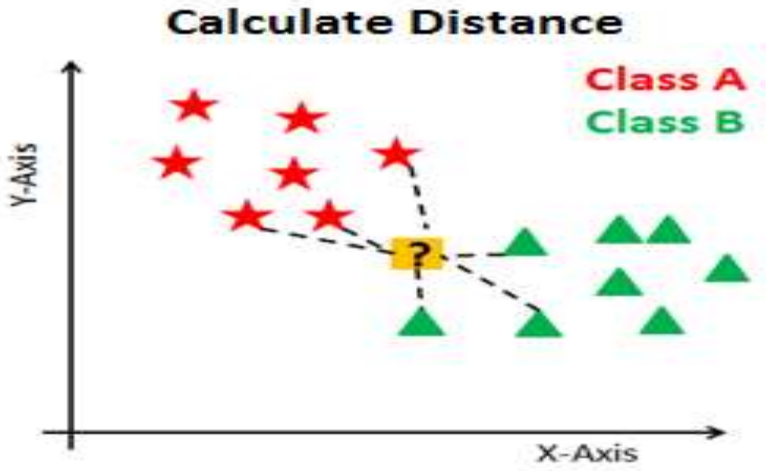
An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors.

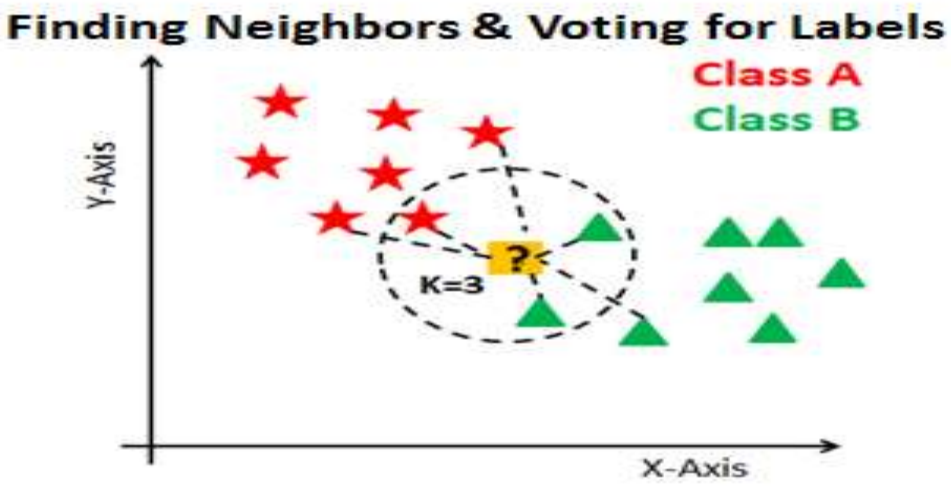
It can also be used for regression — output is the value for the object (predicts continuous values). This value is the average (or median) of the values of its k nearest neighbors.

K Nearest Neighbors

madBlocks
Technology:Innovation:Business







For questions, write us on ai@madblocks.tech

K Nearest Neighbors

madBlocks
Technology:Innovation:Business

Applications of KNN:

Credit ratings —
Should the bank give a loan to an individual?
Would an individual default on his or her loan?
Is that person closer in characteristics to people who defaulted or did not default on their loans?

In political science —
classifying a potential voter to a “will vote” or “will not vote”, or to “vote Democrat” or “vote Republican”.

More advance examples could include handwriting detection (like OCR), image recognition and even video recognition.

K Nearest Neighbors

madBlocks
Technology:Innovation:Business

Pros:

Easy to implement

Simple algorithm — to explain and understand/interpret

High accuracy (relatively) — it is pretty high

Versatile — useful for classification or regression

K Nearest Neighbors

madBlocks
Technology:Innovation:Business

Cons:

Computationally expensive — because the algorithm stores all of the training data

High memory requirement

Stores all (or almost all) of the training data

Prediction stage might be slow (with big N)

Sensitive to irrelevant features and the scale of the data

Work Flow

madBlocks
Technology:Innovation:Business

Process:

1. Create an object to perform KNN Classification
2. Train the Data
3. Test the Data
4. Calculate the accuracy
5. Create API

Summary

madBlocks
Technology:Innovation:Business

Pack-Up!

– We gone through the creation of a model to predict the crop using the dataset.

– We have understood how to train the model using KNN Classification method.

– We have created a simple API to access the model using Flask and ngrok packages.

For questions, write us on ai@madblocks.tech