Machine Learning

tanshuqiu

Email: tsq@cqut.edu.cn

How do we learn Machine Learning

Methods

- ✓Previewing our textbooks.
- ✓Listening carefully in our class.
- ✓Doing more practice by yourself.
- ✓Review some important notes.

About our course

Contents

- ✓Total teaching hours: 64h
- ✓The evaluation method: classwork and homework
- ✓Using software: MATLAB

How to get high score

Key points

✓On time to the class

✓On time to submit the classwork and homework

✓Concentrate on the class

✓Do your classwork and homework well

✓Innovativeness

Reduce your score

Key points

✓Late for class or early leave.

✓Absence for no reason.

✓Disturb classroom order.

✓Late for submitting your classwork or homework.

✓**Copy others' classwork or homework**.

✓don't submit your work classwork or homework.

How to get your score

Key points

✓Learning attitude + Daily attendance （20%）

✓Homework except the last one （40%）

✓The last homework (40%)

Course contents

Key points

- ✓Getting Started with MATLAB Machine Learning
- ✓Importing and Organizing Data in MATLAB
- ✓From Data to Knowledge Discovery
- ✓Finding Relationships between Variables - Regression Techniques
- ✓Pattern Recognition through Classification Algorithms
- ✓Identifying Groups of Data Using Clustering Methods

Getting Started with MATLAB Machine Learning

1. Basic information of machine learning
2. Discover the different types of machine learning
3. Choosing the right algorithm
4. How to build machine learning models step by step
5. Statistics and Machine Learning Toolbox

1. Basic information of machine learning

Machine learning is a multidisciplinary field created by intersection and synergy between computer science, statistics, neurobiology, and control theory. Its emergence has played a key role in several fields and has fundamentally changed the vision of software programming.

1. Basic information of machine learning

As it might be expected, machine learning interconnects and coexists with the study of, and research on, human learning. Like humans, whose brain and neurons are the foundation of insight, **Artificial Neural Networks (ANNs)** are the basis of any decision-making activity of the computer.

1. Basic information of machine learning

The extraction of general laws from a set of observed data is called **induction**; it is opposed to **deduction**, in which, starting from general laws, we want to predict the value of a set of variables. The overall process in which, starting from a set of observations, we want to make predictions for new situations is called **inference**.

1. Basic information of machine learning

We can distinguish two types of inductive learning:

Learning by example: Knowledge gained by starting from a set of positive examples that are instances of the concept to be learned and negative examples that are non-instances of the concept.

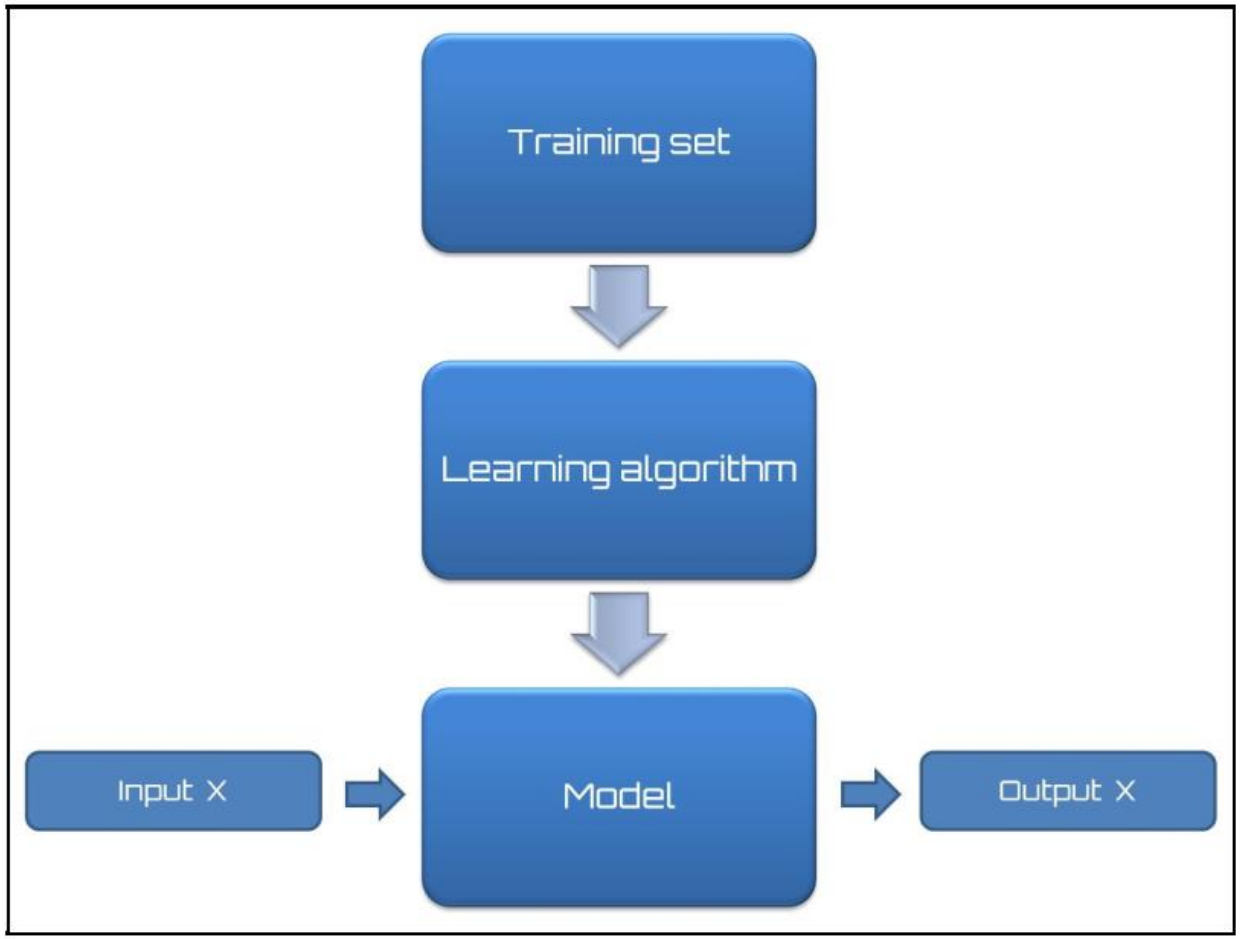
Learning regularity: This is not a concept to learn. The goal is to find regularity (common characteristics) in the instances provided.

2. Discover the different types of machine learning

The power of the machine learning is due to the quality of its algorithm, which have been improved and updated over the years; these are divided into several main types depending on the nature of the signal used for learning or the type of feedback adopted by the system: **Supervised learning, Unsupervised learning, Reinforcement learning.**

2. Discover the different types of machine learning

Supervised learning: The algorithm generates a function that links input values to a desired output through the observation of a set of examples in which each data input has its relative output data, and that is used to construct predictive models.


Training set
Learning algorithm
Input X
Model
Output X

2. Discover the different types of machine learning

All supervised learning algorithms are based on the following thesis: if an algorithm provides an adequate number of examples, it will be able to create a derived function B that will approximate the desired function A.

2. Discover the different types of machine learning

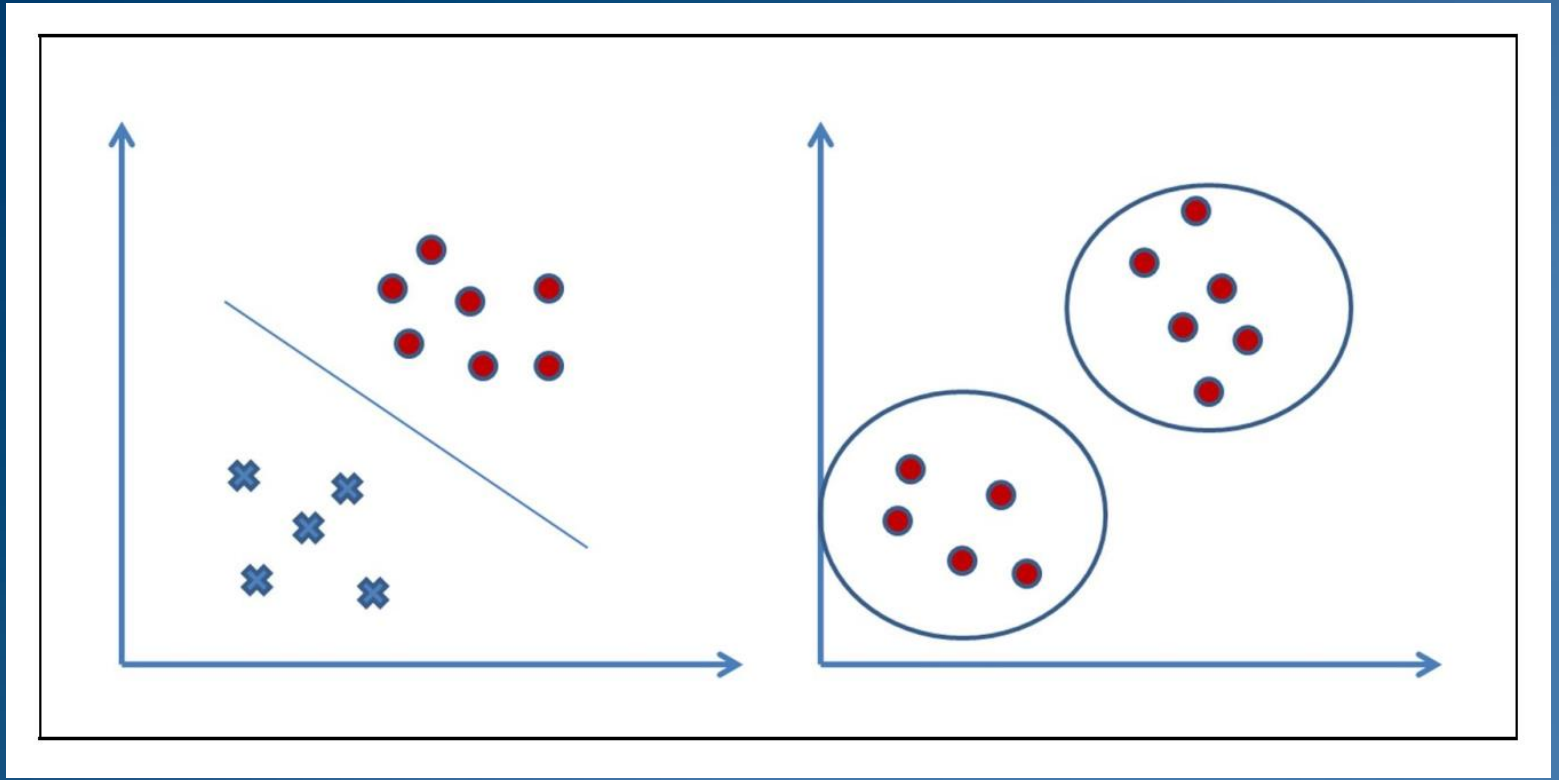
Unsupervised learning: The algorithm tries to derive knowledge from a general input without the help of a set of pre-classified examples that are used to build descriptive models. A typical example of the application of these algorithms are search engines.

2. Discover the different types of machine learning

The aim of unsupervised learning is to automatically extract information from databases. This process occurs **without prior knowledge** of the contents to be analyzed. Unlike supervised learning, there is no information on membership classes of the examples or generally on the output corresponding to a certain input.

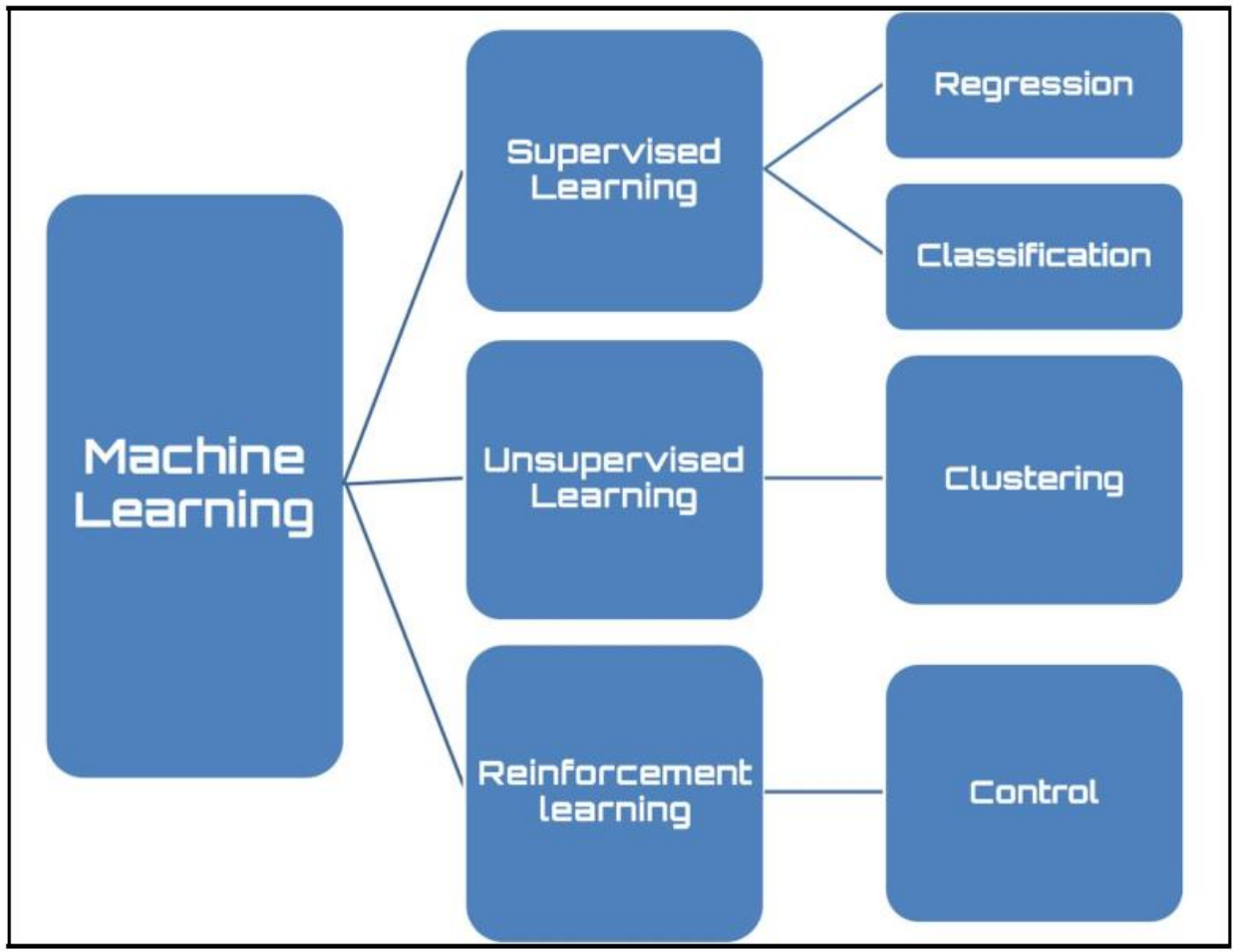
2. Discover the different types of machine learning

The validity of these algorithms depends on the usefulness of the information they can extract from the databases. These algorithms work by comparing data and looking for similarities or differences. Available data concerns only the set of features that describe each example.



2. Discover the different types of machine learning

Reinforcement learning: The algorithm is able to learn depending on the changes that occur in the environment in which it is performed. In fact, since every action has some effect on the environment concerned, the algorithm is driven by the same feedback environment.


Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement learning
Regression
Classification
Clustering
Control

2. Discover the different types of machine learning

Reinforcement learning aims to create algorithms that can learn and adapt to environmental changes. This programming technique is based on the concept of receiving external stimuli depending on the algorithm choices. A correct choice will involve a premium while an incorrect choice will lead to a penalty.

2. Discover the different types of machine learning

The information available is called reinforcement signals. But the system does not give any information on how to update the agent's behavior (that is, weights). You cannot define a cost function or a gradient. The goal of the system is to create the smart agents that have a machinery able to learn from their experience.

3. Choosing the right algorithm

What is the right algorithm for my needs?

It mainly depends on the data available to us: the size, quality, and nature of the data. It depends on what we want to do with the answer. It depends on how the algorithm has been expressed in instructions for the computer. It depends on how much time we have. There is **no best method or one-size-fits-all**. The only way to be sure that the algorithm chosen is the right one is to try it.

3. Choosing the right algorithm

If we start from what we have (data), it is a classification problem, and two options are available:

Classify based on input

Classify based on output

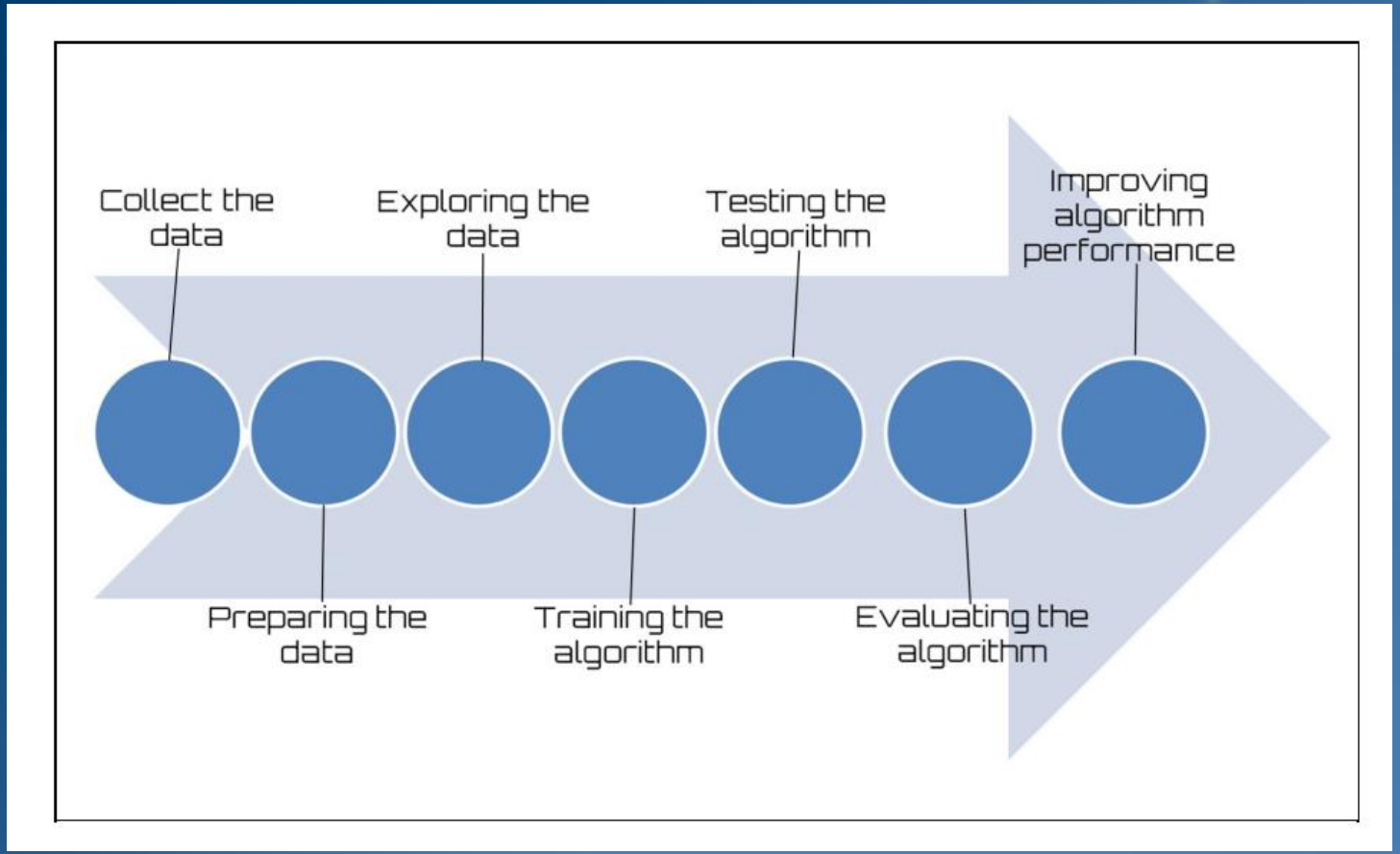
3. Choosing the right algorithm

Classify based on input: We have a supervised learning problem if we can label the input data. If we cannot label the input data but want to find the structure of the system, then it is unsupervised. Finally, if our goal is to optimize an objective function by interacting with the environment, it is a reinforcement learning problem.

3. Choosing the right algorithm

Classify based on output: If our model output is a number, we have to deal with a regression problem. But it is a classification problem if the output of the model is a class. Finally, we have a clustering problem if the output of the model is a set of input groups.

4. How to build machine learning models step by step



4. How to build machine learning models step by step

Preparing the data: We have collected the data; now we have to prepare it for the next step. Once we have this data, we must make sure it is in a format usable by the algorithm we want to use.

4. How to build machine learning models step by step

Exploring the data: At this point, we can look at data to verify that it is actually working and we do not have a bunch of empty values. In this step, through the use of plots, we can recognize patterns or whether there are some data points that are vastly different from the rest of the set. Plotting data in one, two, or three dimensions can also help.

4. How to build machine learning models step by step

Training the algorithm: In this step, the machine learning begins to work with the definition of the model and the next training. The model starts to extract knowledge from large amounts of data that we had available, and that nothing has been explained so far.

4. How to build machine learning models step by step

Testing the algorithm: In this step, we use the information learned in the previous step to see if the model actually works. The evaluation of an algorithm is for seeing how well the model approximates the real system. In the case of supervised learning, we have some known values that we can use to evaluate the algorithm. In unsupervised learning, we may need to use some other metrics to evaluate success.

4. How to build machine learning models step by step

Evaluating the algorithm: We can assess the approximation ability of the model by applying it to real data. The model, preventively trained and tested, is then valued in this phase.

Improving algorithm performance: we have evaluated the performance, and now we are ready to analyze the whole process to identify any possible room for improvement.

5. Statistics and Machine Learning Toolbox

MATLAB is a software platform optimized for solving scientific problems and design. In MATLAB, calculation, visualization, and programming are integrated in an easy-to-use environment, where problems and solutions are expressed in familiar mathematical notation.

MATLAB R2017a - sponsored use

HOME PLOTS APPS

Search Documentation Log In

New Script | New | Open | Find Files | Compare | Import Data | Save Workspace | New Variable | Open Variable | Clear Workspace | Analyze Code | Run and Time | Clear Commands | Layout | Preferences | Set Path | Add-Ons | Help | Community | Request Support | Learn MATLAB

FILE VARIABLE CODE ENVIRONMENT RESOURCES

C: ▸ Users ▸ lavoro ▸ Documents ▸ MATLAB ▸

Current Folder

Name

- mRMR_0.9_compiled
- MIToolbox-2.1.2
- MIRtoolbox1.6.1
- FSLib_v4.0_2016
- Examples
- CQT
- Add-Ons

mRMR_0.9_compiled (Folder)

Workspace

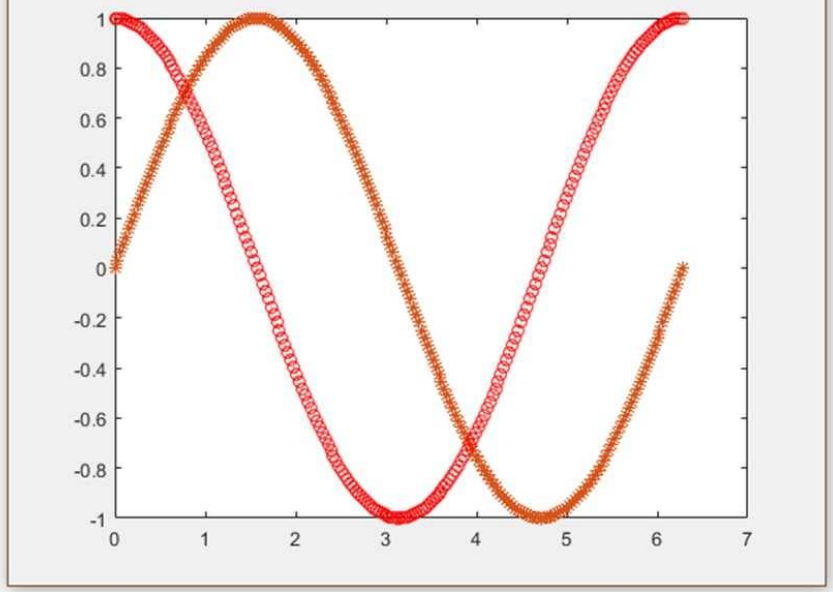
Name	Value
t	1x201 double
y	1x201 double
z	1x201 double

Command Window

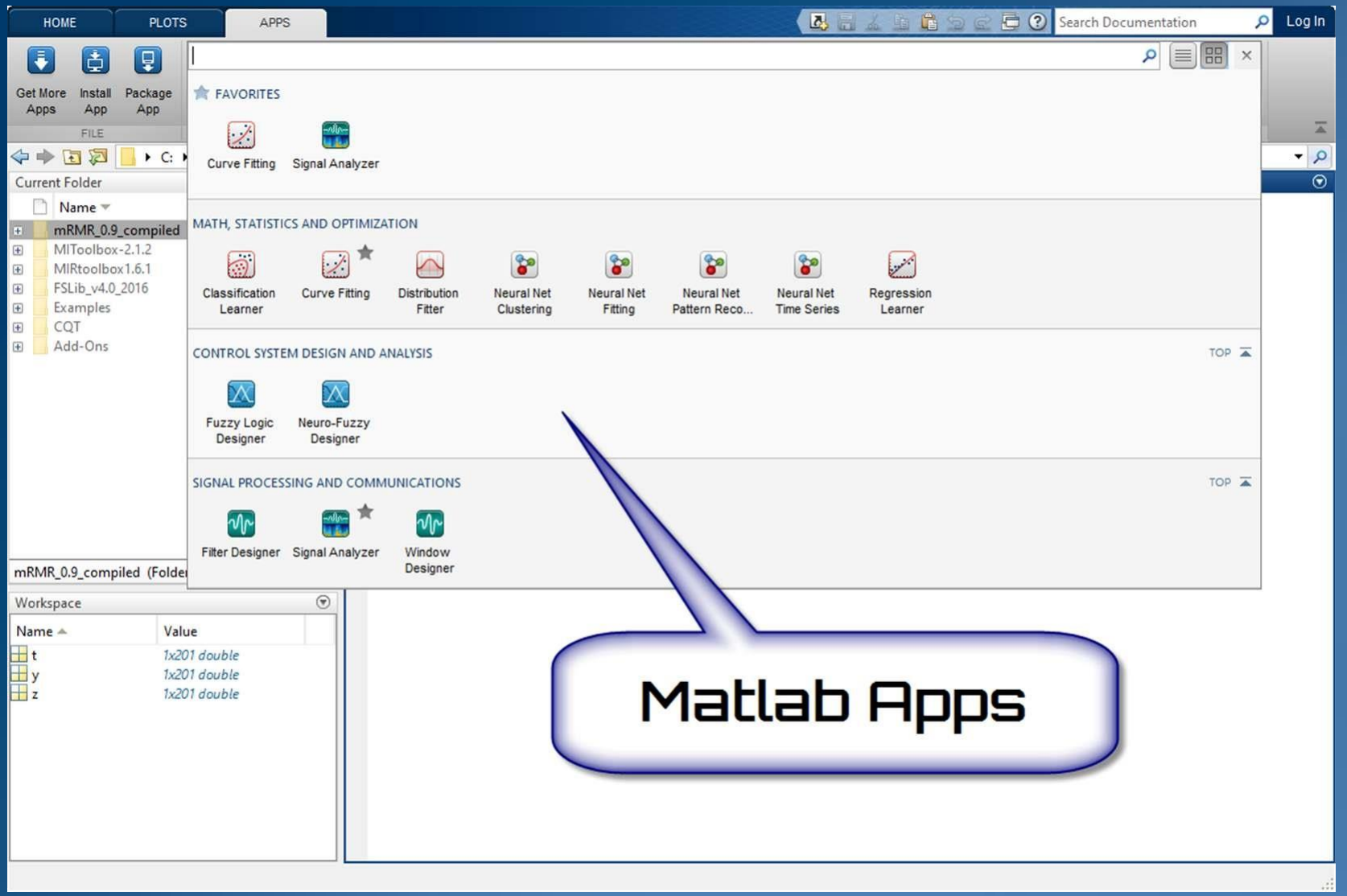
fx >>

Figure 1

File Edit View Insert Tools Desktop Window Help



Click and drag to move Command Window...


HOME
PLOTS
APPS
Search Documentation
Log In
Get More Apps
Install App
Package App
FILE
FAVORITES
Curve Fitting
Signal Analyzer
MATH, STATISTICS AND OPTIMIZATION
Classification Learner
Curve Fitting
Distribution Fitter
Neural Net Clustering
Neural Net Fitting
Neural Net Pattern Reco...
Neural Net Time Series
Regression Learner
CONTROL SYSTEM DESIGN AND ANALYSIS
TOP
Fuzzy Logic Designer
Neuro-Fuzzy Designer
SIGNAL PROCESSING AND COMMUNICATIONS
TOP
Filter Designer
Signal Analyzer
Window Designer
Current Folder
Name
mRMR_0.9_compiled
MIToolbox-2.1.2
MIRtoolbox1.6.1
FSLib_v4.0_2016
Examples
CQT
Add-Ons
Workspace
Name
Value
t
1x201 double
y
1x201 double
z
1x201 double
Matlab Apps

5. Statistics and Machine Learning Toolbox

The Statistics and Machine Learning Toolbox contains all the tools necessary to extract knowledge from large datasets. It provides functions and apps to analyze, describe, and model data.

In this toolbox are developed supervised and unsupervised machine learning algorithms including Support Vector Machines (SVMs), decision trees, kNearest Neighbor (KNN), k-means and so on.

5. Statistics and Machine Learning Toolbox

Supported datatypes

Numeric scalars, vectors, matrices, or arrays having single or double precision entries. These data forms have the single or double datatype.

Cell arrays of character vectors; character, logical, or categorical arrays; or numeric vectors for categorical variables representing grouping data. These data forms have the cellstr, char, logical, categorical, and single or double datatype, respectively

5. Statistics and Machine Learning Toolbox

Supported datatypes

Some functions support tabular arrays for heterogeneous. The table datatype contains variables of any of the datatypes previously listed.

Some functions accept gpuArray input arguments so that they execute on the GPU.

5. Statistics and Machine Learning Toolbox

Unsupported datatypes

Complex numbers.

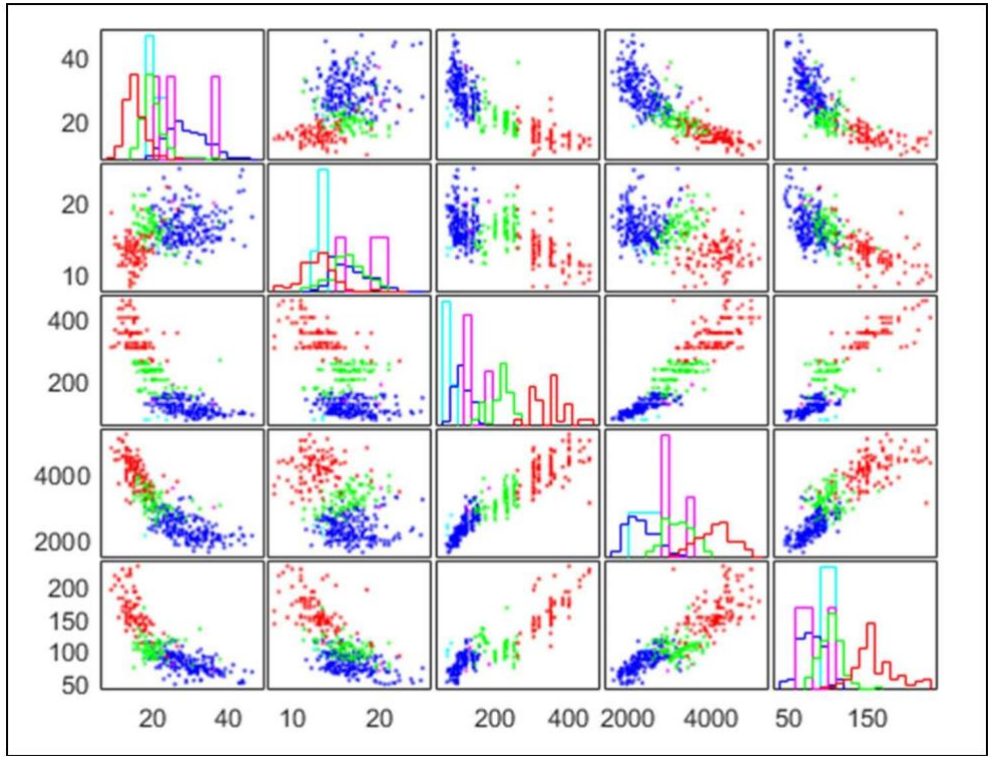
Custom numeric datatypes, for example, a variable that is double precision and an object.

Signed or unsigned numeric integers for non-grouping data, for example, unint8 and int16.

Sparse matrices, for example, matrix ONE such that issparse(ONE) returns 1. To use data that is of datatype sparse, recast the data to a matrix using full.

5. Statistics and Machine Learning Toolbox

Data mining and data visualization



Visualizing multivariate data

5. Statistics and Machine Learning Toolbox

The Statistics and Machine Learning Toolbox includes functions for calculating:

Measures of central tendency, including average, median, and various Means.

Measures of dispersion, including range, variance, standard deviation, and mean or median absolute deviation.

Linear and rank correlation.

Results based on data with missing values.

Percentile and quartile estimates.

Density estimates using a kernel-smoothing function.

5. Statistics and Machine Learning Toolbox

Regression analysis

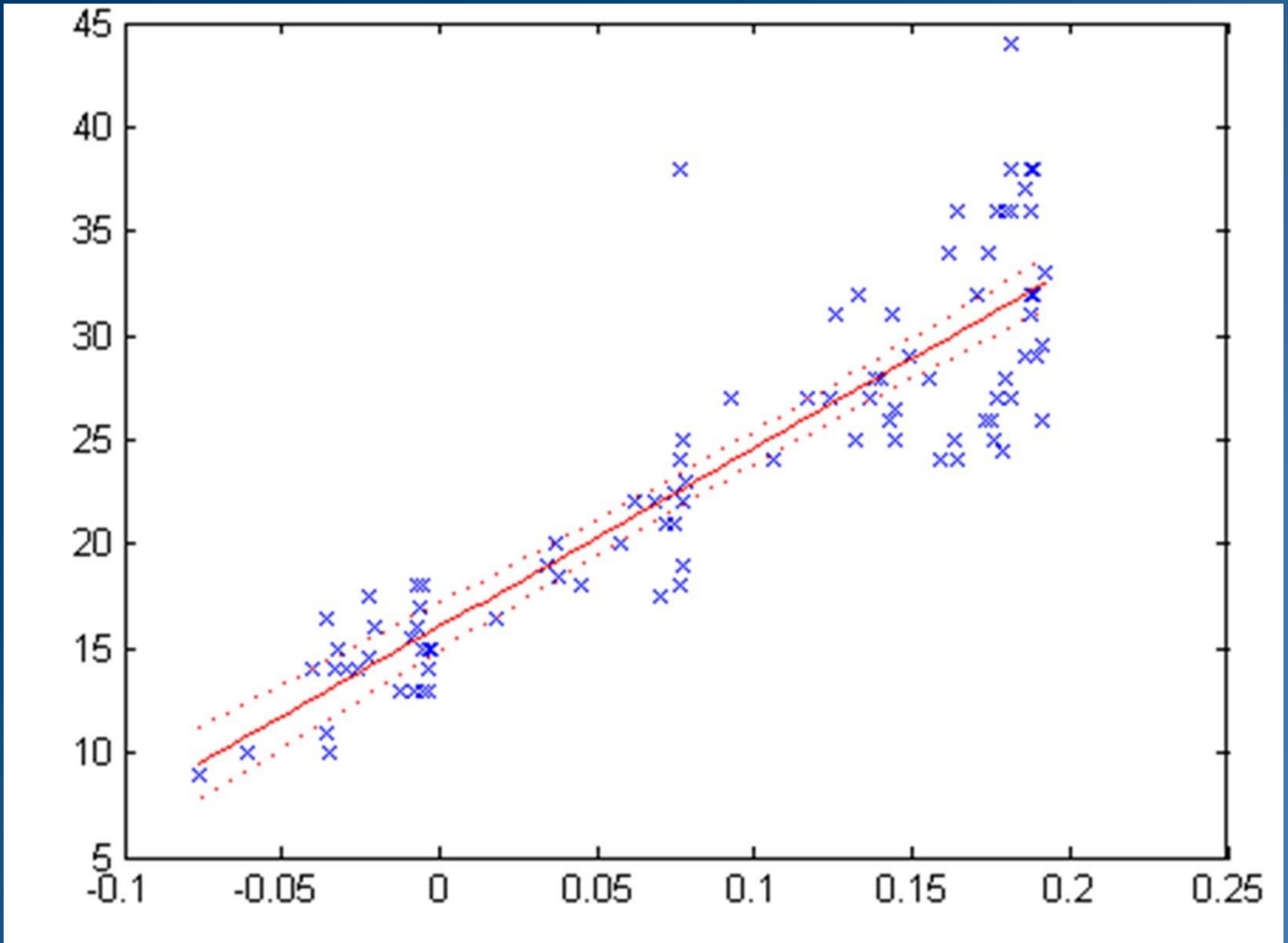
In the Statistics and Machine Learning Toolbox, there are a variety of regression algorithms, including:

Linear regression

Nonlinear regression

Generalized linear models

Mixed-effects models



Scatter plot of linear regression model

5. Statistics and Machine Learning Toolbox

Classification

Classification models are supervised learning methods and are aimed at predicting a categorical target.

The Statistics and Machine Learning Toolbox offers apps and functions that cover a variety of parametric and non-parametric classification algorithms, such as:

Logistic regression, Naive Bayes classification, KNN classification SVM (binary and multiclass classification)

5. Statistics and Machine Learning Toolbox

Classification

We can use it to perform common tasks such as:

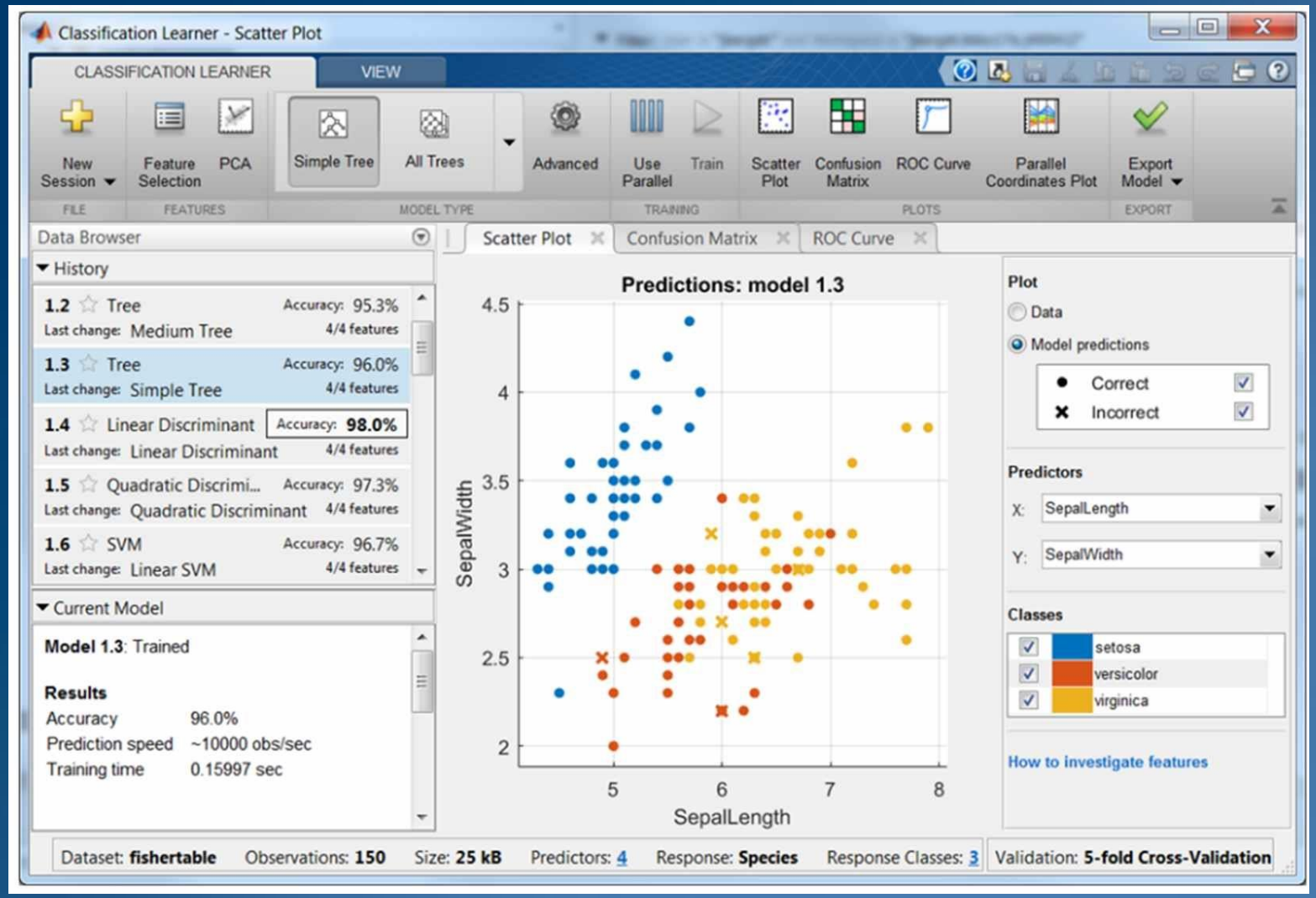
Importing data and specifying cross-validation schemes

Exploring data and selecting features

Training models using several classification algorithms

Comparing and assessing models

Sharing trained models for use in applications such as computer vision and signal processing.



The Classification Learner with a history list containing various classifier types

5. Statistics and Machine Learning Toolbox

Cluster analysis

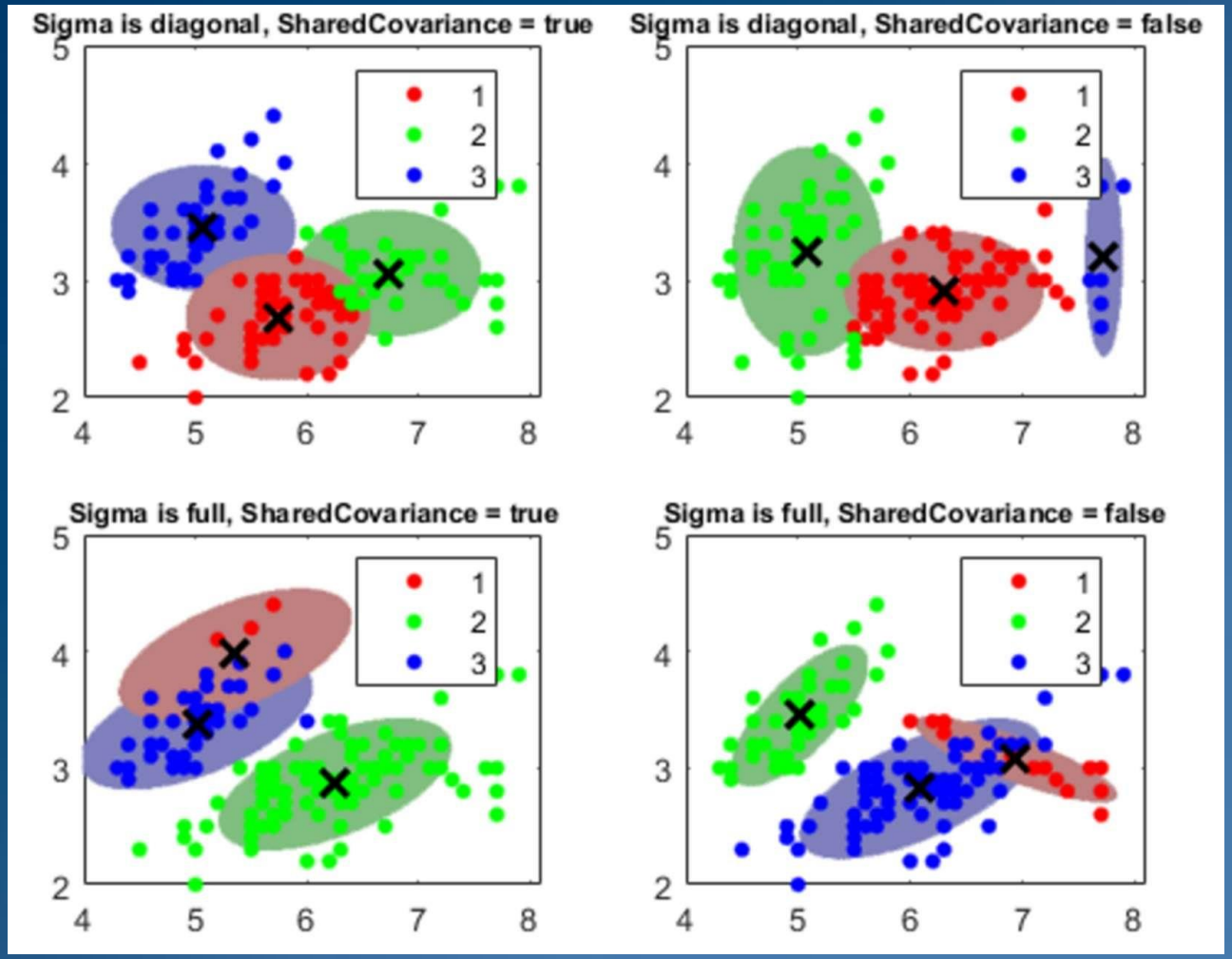
The Statistics and Machine Learning Toolbox provides several algorithms to carry out cluster analysis. Available algorithms include:

k-means

Hierarchical clustering

GMM

HMM



A cluster analysis example