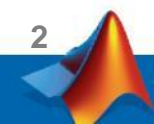# Data Analytics Workshop

Terasoft

Application Engineer
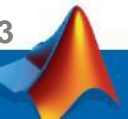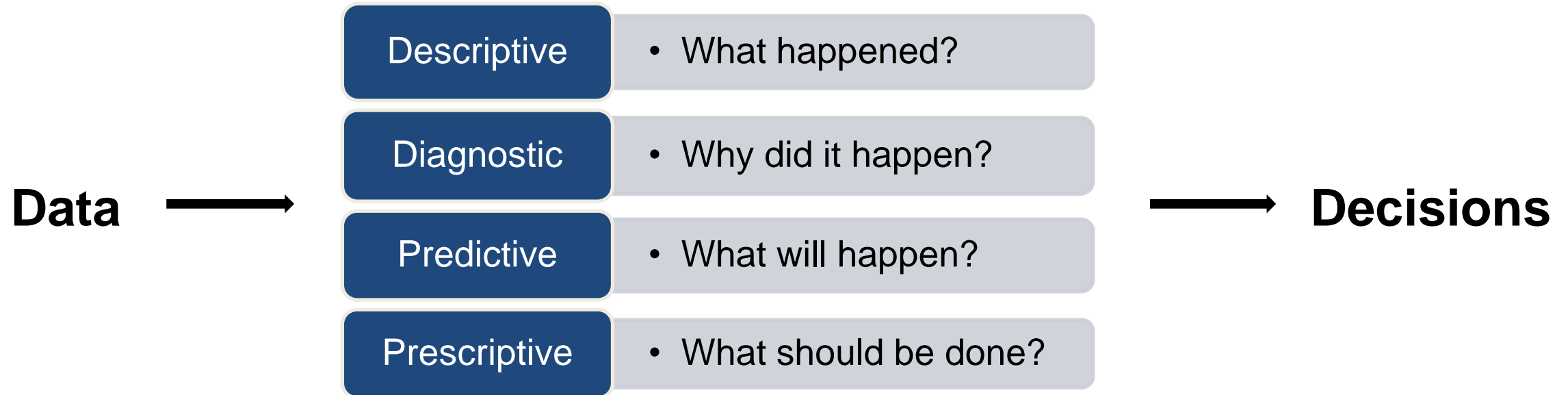
Jeffrey Liu

Outlines

- Introducing the Workflow of Data Analytics
  - From Data to actionable insights

- Data import and preprocessing
  - Import data from external files
  - Understanding the data and preprocessing
    - Merging tables   (exercise 1)
    - Locating and removing missing data   (exercise 2)
    - Grouping statistics   (exercise 3)

- Train a prediction model
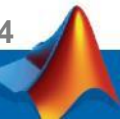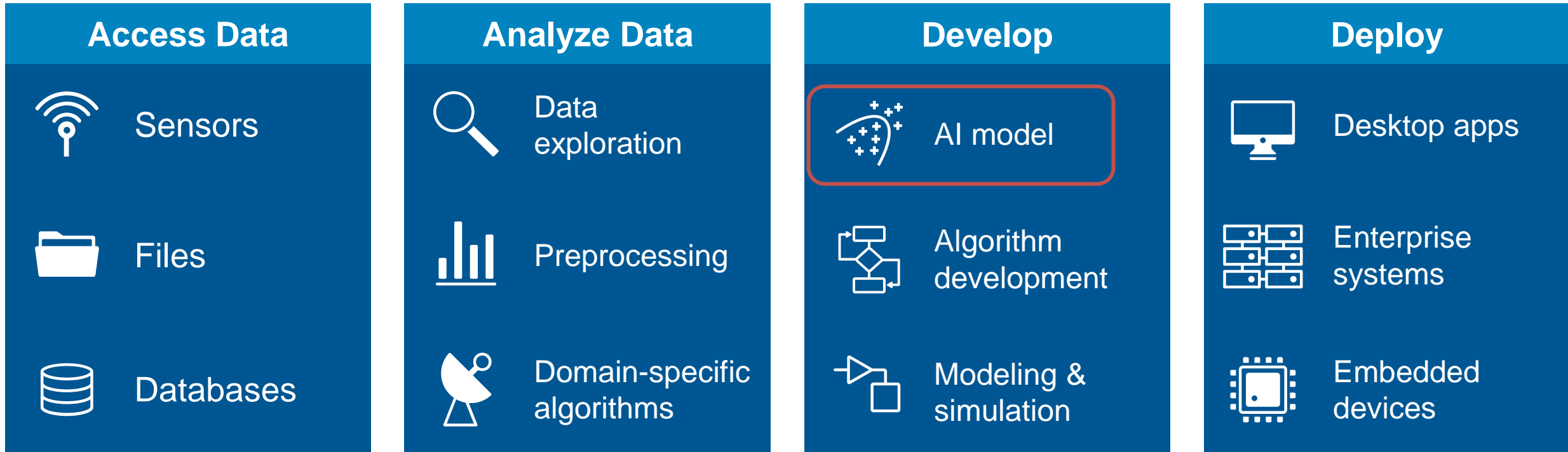 - Classification Learner App   (exercise 4)
  - Hand write code

What is Data Analytics?

Turn large volumes of complex data into actionable information

Data ⟶

Descriptive	• What happened?
Diagnostic	• Why did it happen?
Predictive	• What will happen?
Prescriptive	• What should be done?

⟶ **Decisions**

① MATLAB Analytics work with **business** and **engineering data**

② MATLAB enables **domain experts to do Data Science**

③ MATLAB Analytics can **run anywhere**

Basic Workflow in most of AI / Data Analytics Journey

Access Data	Analyze Data	Develop	Deploy
Sensors	Data exploration	AI model	Desktop apps
Files	Preprocessing	Algorithm development	Enterprise systems
Databases	Domain-specific algorithms	Modeling & simulation	Embedded devices

What is Table?

Extracting Data From Tables

© Terasoft, Inc.

Indexing into Tables


```
>> t2 = t1(6:15,[1 5 8 13])
```

19-by-15

10-by-4

INTRO to the Table Data Type
"Working with Financial Data in MATLAB"


```
>> doc readtable
>> T = table()
>> methods(T)
>> doc table
```

- Tables are data containers that allow for working with mixed datatypes
- **Methods:** join, innerjoin, outerjoin, write, sortrows, width, height, etc.

Merging Data

- T3 = [T1;T2];
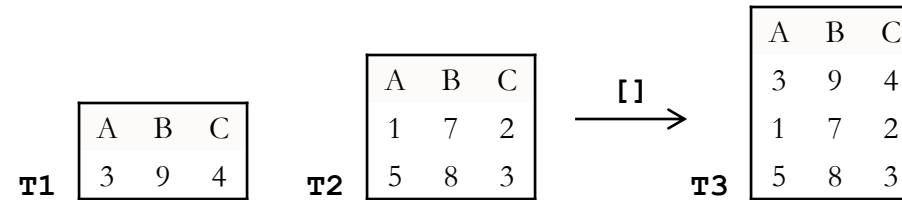
- Join, innerjoin, outerjoin
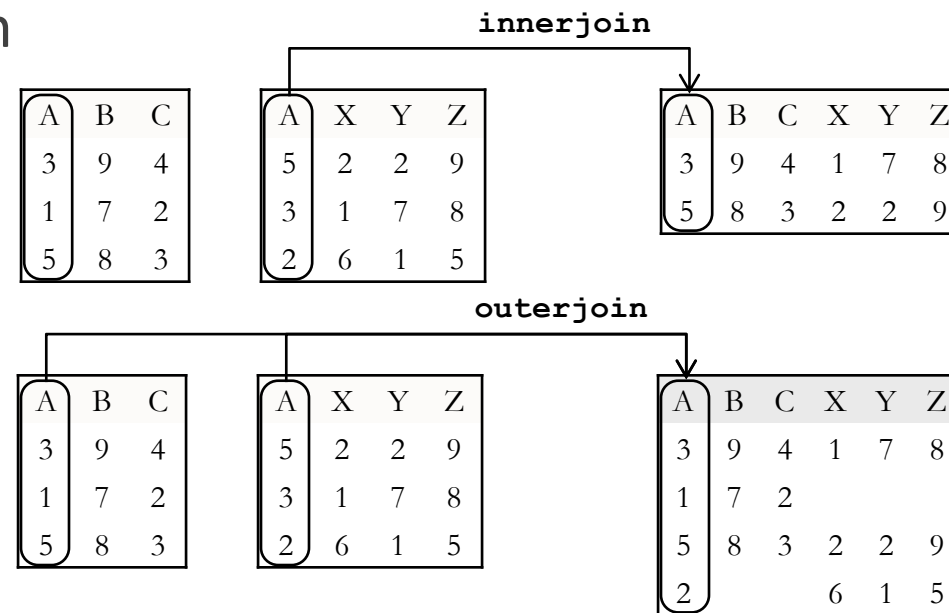
Exercise 1

- 將join_ex.xslx中的兩張表格(TableA, TableB)依照"Date"欄位左右合併。

- Use "innerjoin", "outerjoin" to combine "table" variables
  - Import files with table array

    ```
    >> Variables = readrable(date, data, 'var_name');
    ```
  - Merge two tables
  - innerjoin, outerjoin

    ```
    >> Combined_table = outerjoin(TblA, TblB, 'Keys', 'Date', 'MergeKeys', true);
    ```
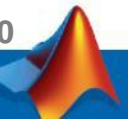
Dealing with Missing Values

Ignore   Delete   Replace
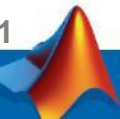
Avoiding NaNs in Calculations

- Several functions are designed to ignore NaNs in calculations

```
nancov              nanmedian           nansum
nanmax              nanmin              nanvar
nanmean             nanstd
```

- if a column contains all NaNs, ignoring them will result in applying the desired function to an empty array.

3	2	4	NaN
1	NaN	2	1
NaN	NaN	6	1

↓ nanmean

2	2	4	1

3	NaN	4	NaN
1	NaN	2	1
NaN	NaN	6	1

↓ nanmean

2	NaN	4	1

Locating Missing(Other) Values

- MATLAB provides numerous "**is***" functions that take an array as input and return a logical output that signifies if the input has a certain characteristic.

isfinite			isnan	
false		Inf		false
true		3		false
false		NaN		true
true		1		false
true		7		false

Numerical comparison

3	NaN	4	NaN
1	NaN	2	1
NaN	NaN	6	1

`>> x == NaN`

Not a Number!

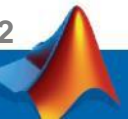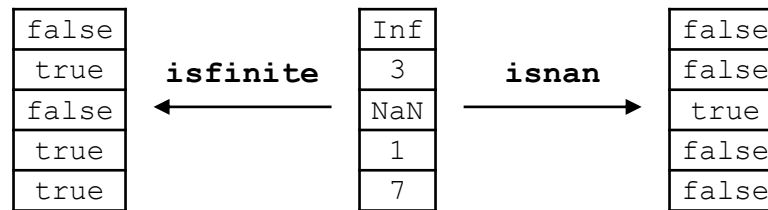| F	F	F	F
F	F	F	F
F	F	F	F

✖

- You can use logical indexing to remove elements from an array.

```
>> idx = isnan(x);
>> x(idx) = [];
```

3
1
7

NaN
NaN

x

3	NaN	4	NaN
1	NaN	2	1
NaN	NaN	6	1

`>> isnan(x)`

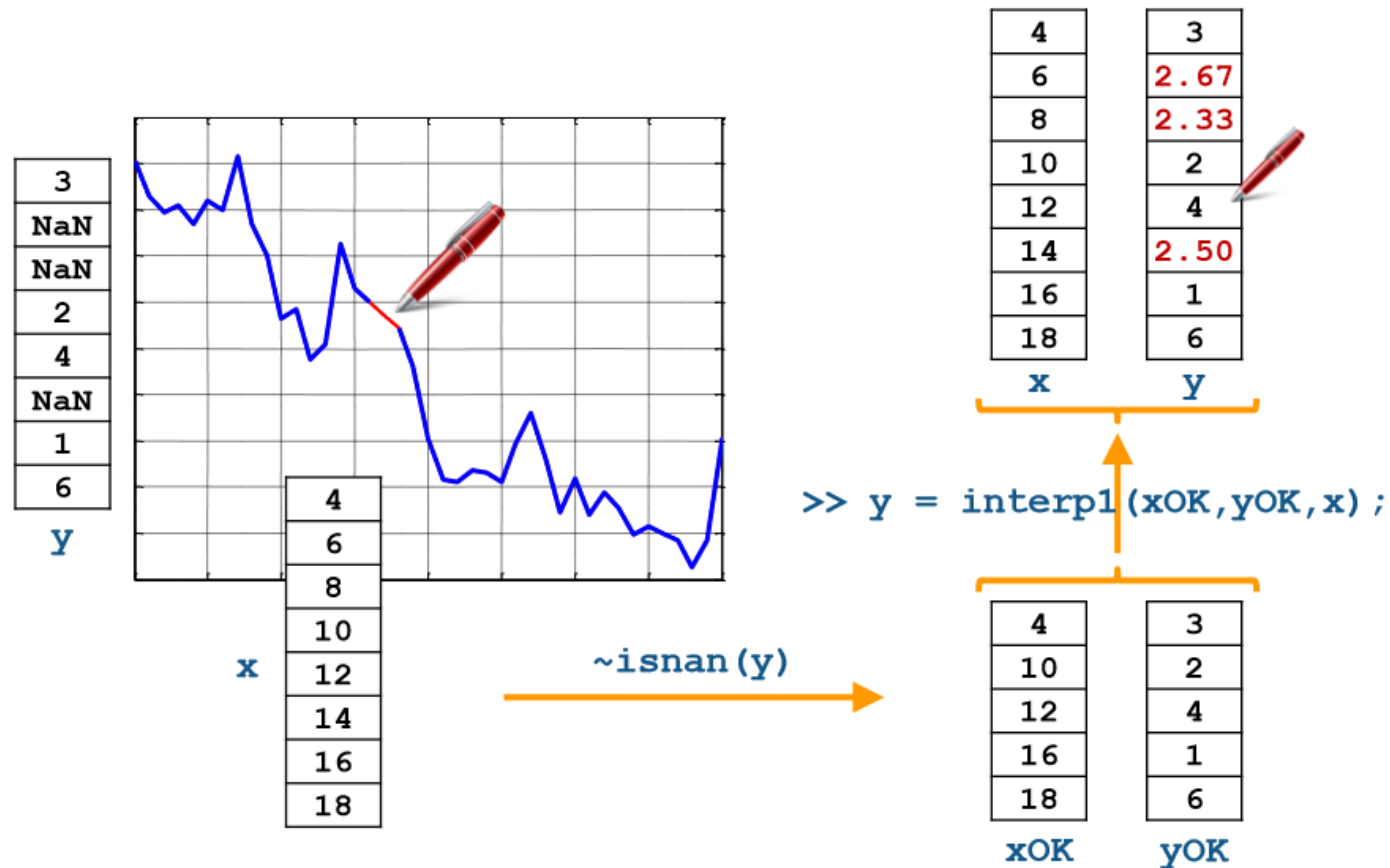| F	T	F	T
F	T	F	F
T	T	F	F

✔

`>> all(isnan(x))`

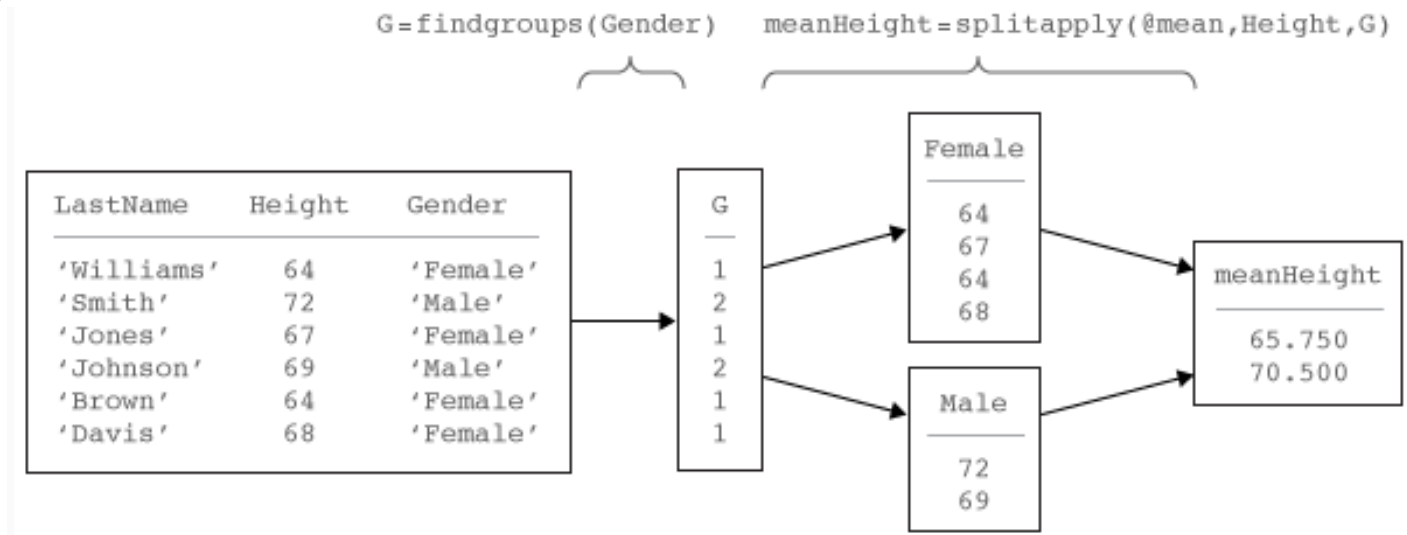| F	T	F	F
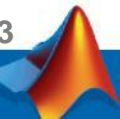
Exercise 2

- 依照下圖宣告變數並得到內插法後結果"y"

Grouping statistics

- Function 'grpstats' : Summary statistics organized by group
  - Syntax : statarray = grpstats(tbl,groupvar,whichstats)

    tbl : data in table or dataset array

    groupvar : Column name for grouping in tbl

    whichstats : Types of summary statistics (numel, std, max, sum…)
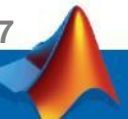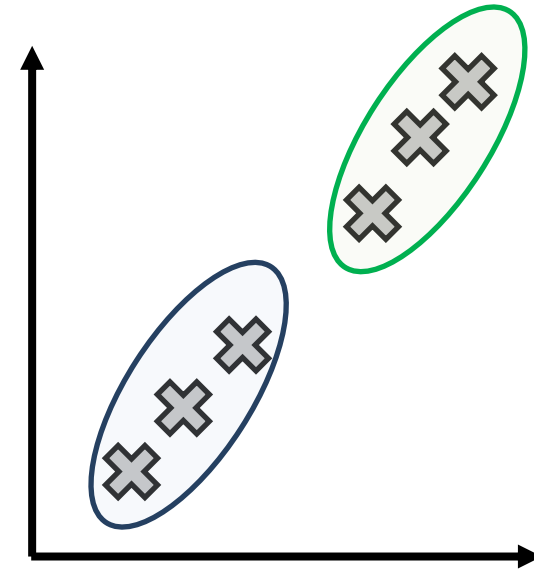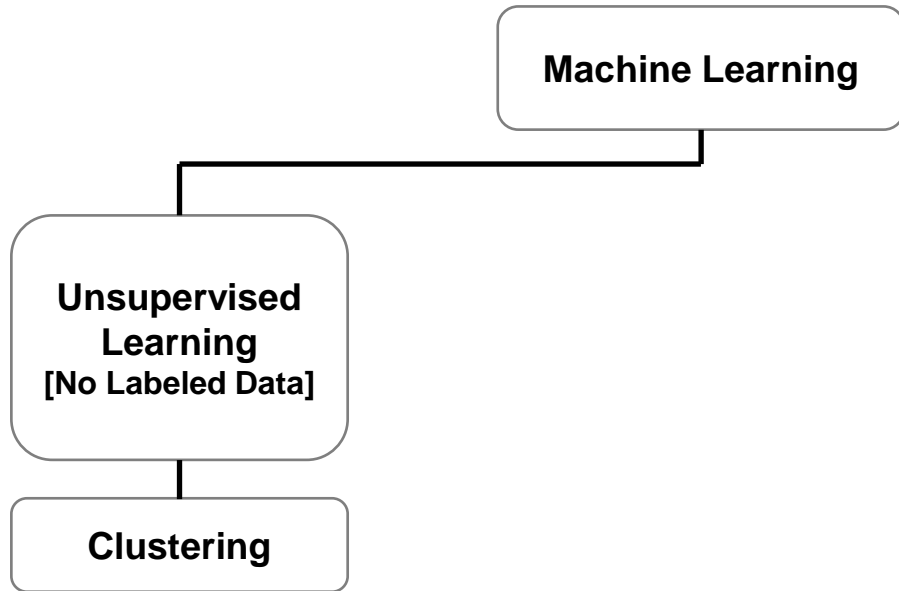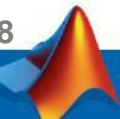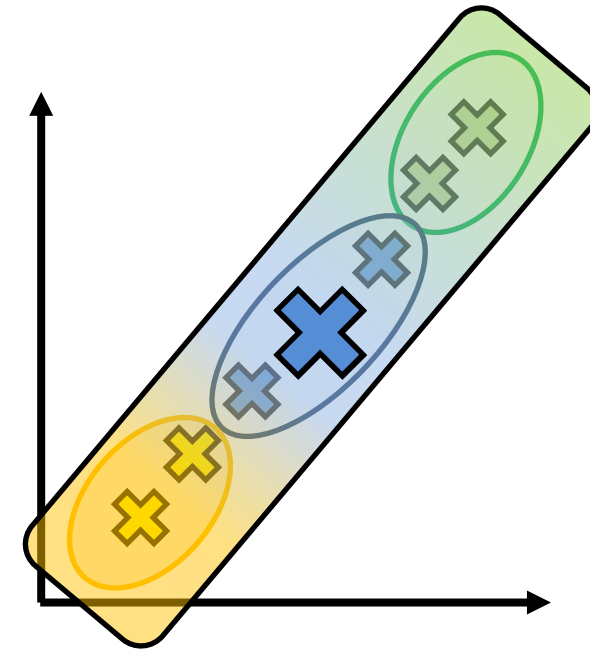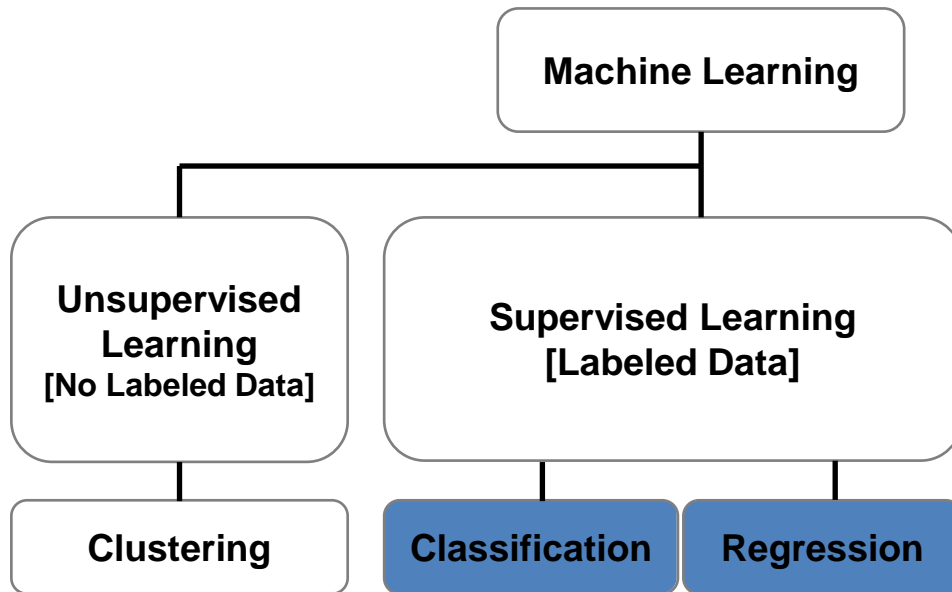
- New Split-apply workflow

Exercise 3

- 在bank-full.csv中依照三種類別(工作、婚姻、購買與否)區分客戶，並計算類別中的平均收入
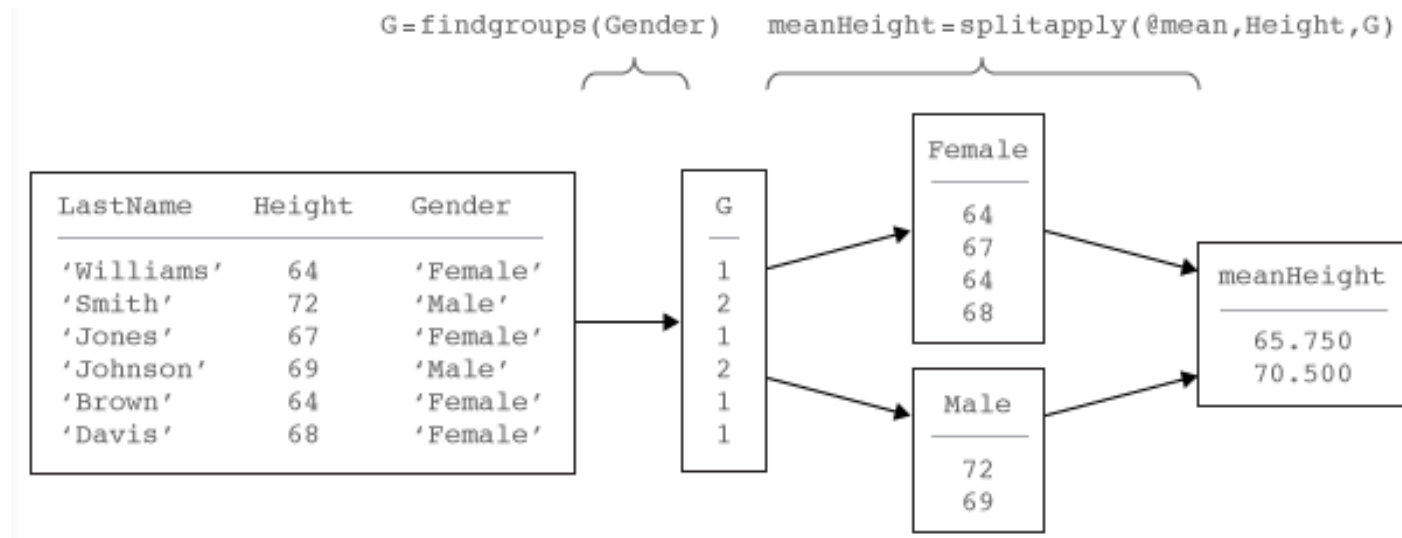
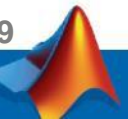# Machine Learning, Deep Learning, and Reinforcement Learning

Machine Learning

Unsupervised Learning
[No Labeled Data]

Clustering

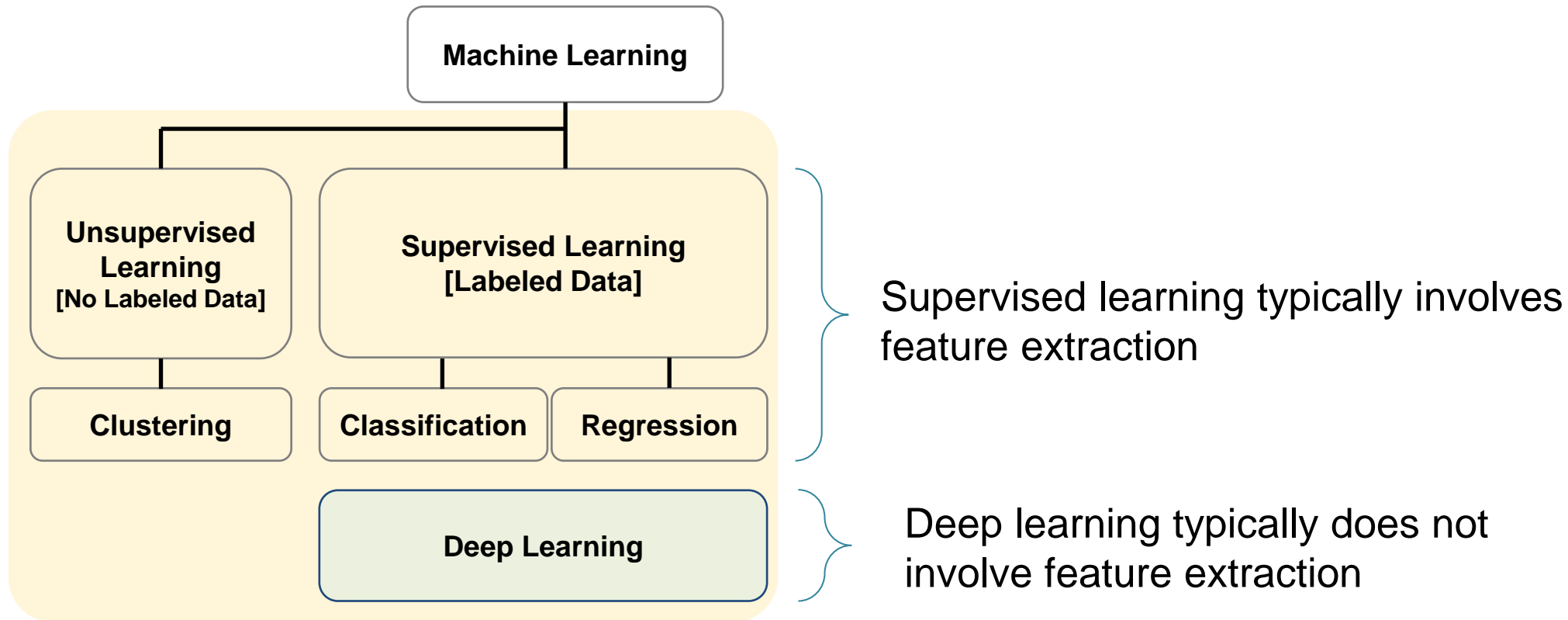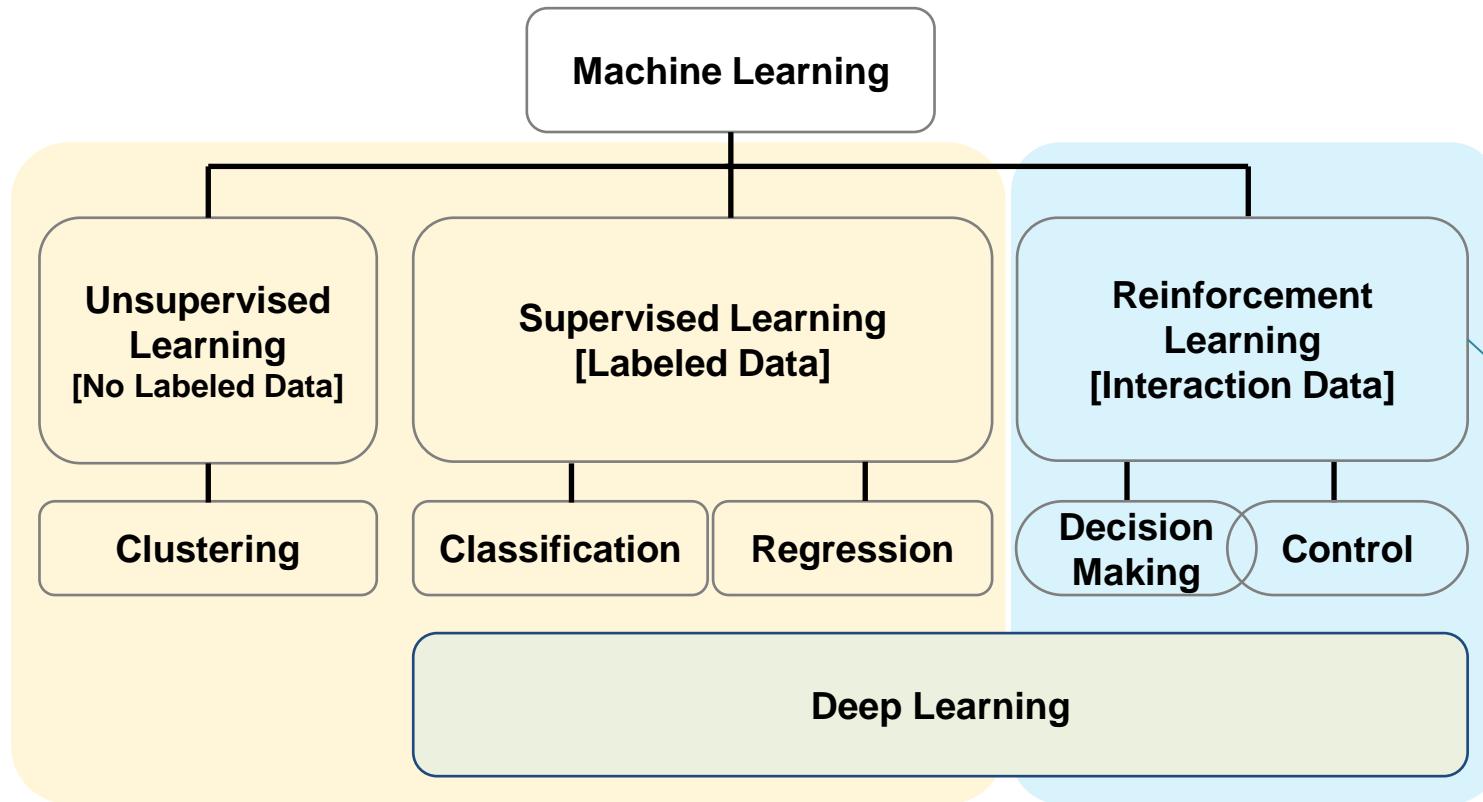# Machine Learning, Deep Learning, and Reinforcement Learning

Machine Learning, Deep Learning, and Reinforcement Learning

Machine Learning

Unsupervised Learning
[No Labeled Data]

Supervised Learning
[Labeled Data]

Clustering

Classification

Regression

Deep Learning

Supervised learning typically involves feature extraction

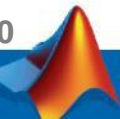Deep learning typically does not involve feature extraction

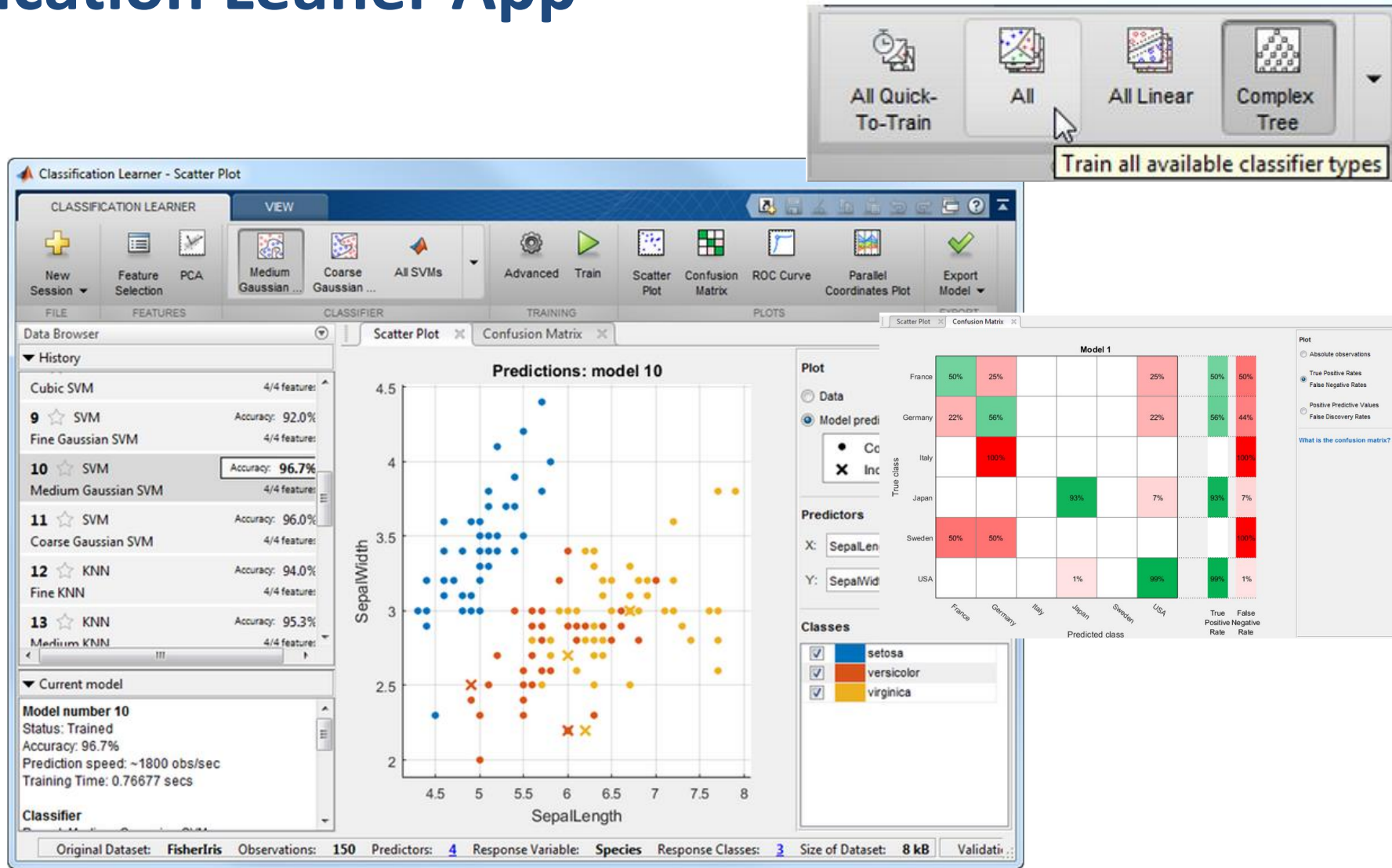# Machine Learning, Deep Learning, and Reinforcement Learning

Reinforcement learning:

- Learning through trial & error [*interaction*]

- Complex problems typically need deep learning [*Deep Reinforcement Learning*]

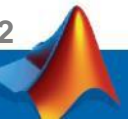- It's about learning a **behavior** or accomplishing a **task**
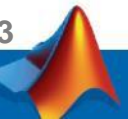
Classification Leaner App

Exercise 4

- 使用bankfull.csv中的資料,將最後一個欄位y(是否購買)當作被預測變數,並利用Classification Learner訓練出一個機器學習模型。

```
>> Data = readtable('bankfull.csv');
>> ClassificationLearner
>> …
>> y_hat = predict(myModel, Data);
```

Other Recourses

- Free online courses
  - MATLAB Onramp
  - Machine Learning Onramp
  - Deep Learning Onramp
  - Image Processing Onramp
  - …
- FUJ Campus Wide License installation

Thanks for your attention