MathWorks®
Accelerating the pace of engineering and science

扫描右侧二维码，关注

MATLAB官方微信公众号

MATLAB

© 2016 The MathWorks, Inc. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.

© 2017 The MathWorks, Inc.

MATLAB 和数据分析

R2017a

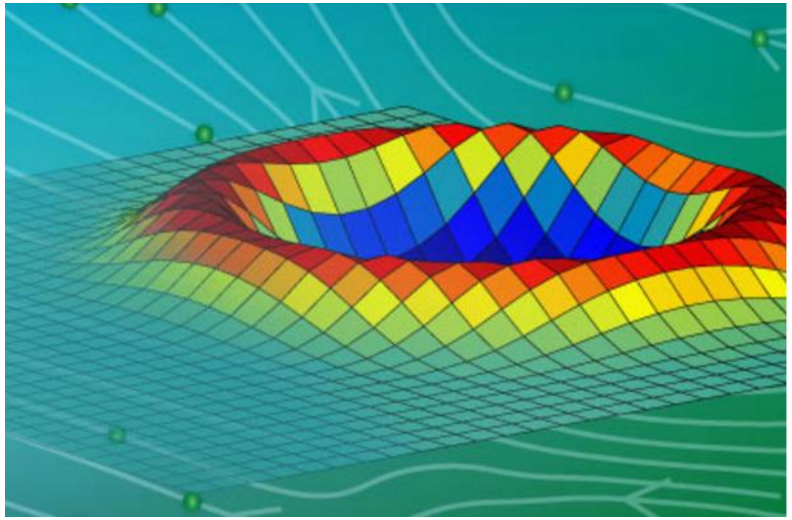
© 2017 The MathWorks, Inc.

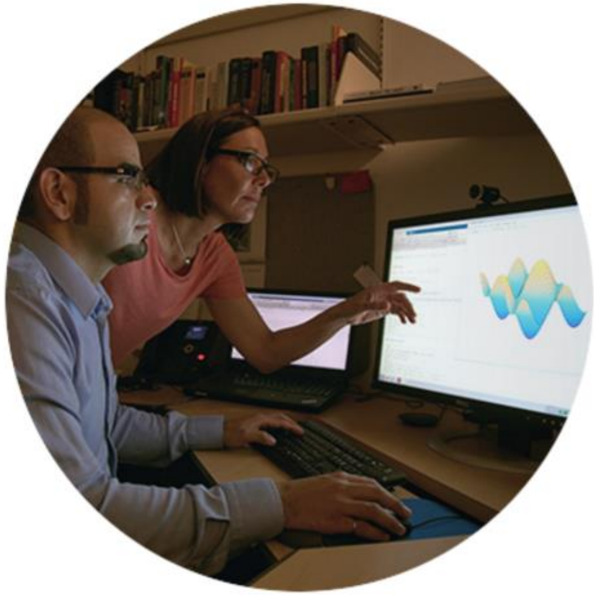
主要内容

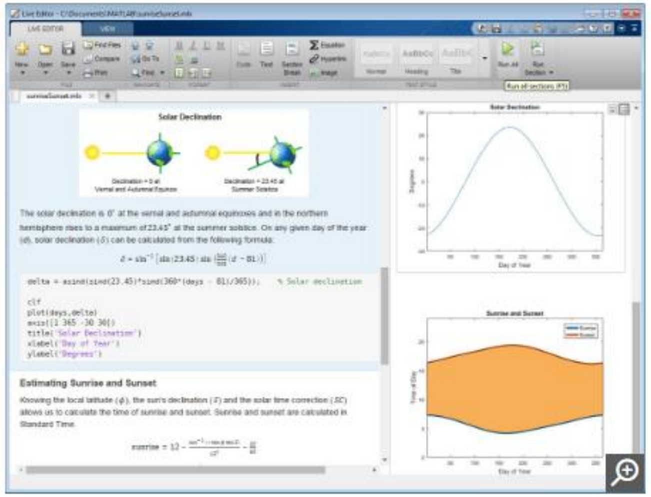
- MATLAB 矩阵运算与数据可视化
- MATLAB 数据处理：流程与优势
- 数据分析应用热点 - 机器学习、深度学习与并行加速

MATLAB 是什么？

MATLAB® 是一种用于算法开发、数据可视化、数据分析以及数值计算的科学计算语言和编程环境。全球数以百万计的工程师和科学家使用 MATLAB® 来分析和设计可改变世界的系统和产品。






LIVE EDITOR
VIEW
New
Open
Save
Find Files
Compare
Print
Go To
Find
Text
Section Break
Equation
Hyperlink
Image
Normal
Heading
Title
Run All
Run Section
Solar Declination
Declination = 0 at Vernal and Autumnal Equinox
Declination = 23.45 at Summer Solstice
Solar Declination
Degrees
Day of Year
Sunrise and Sunset
Time of Day
Day of Year

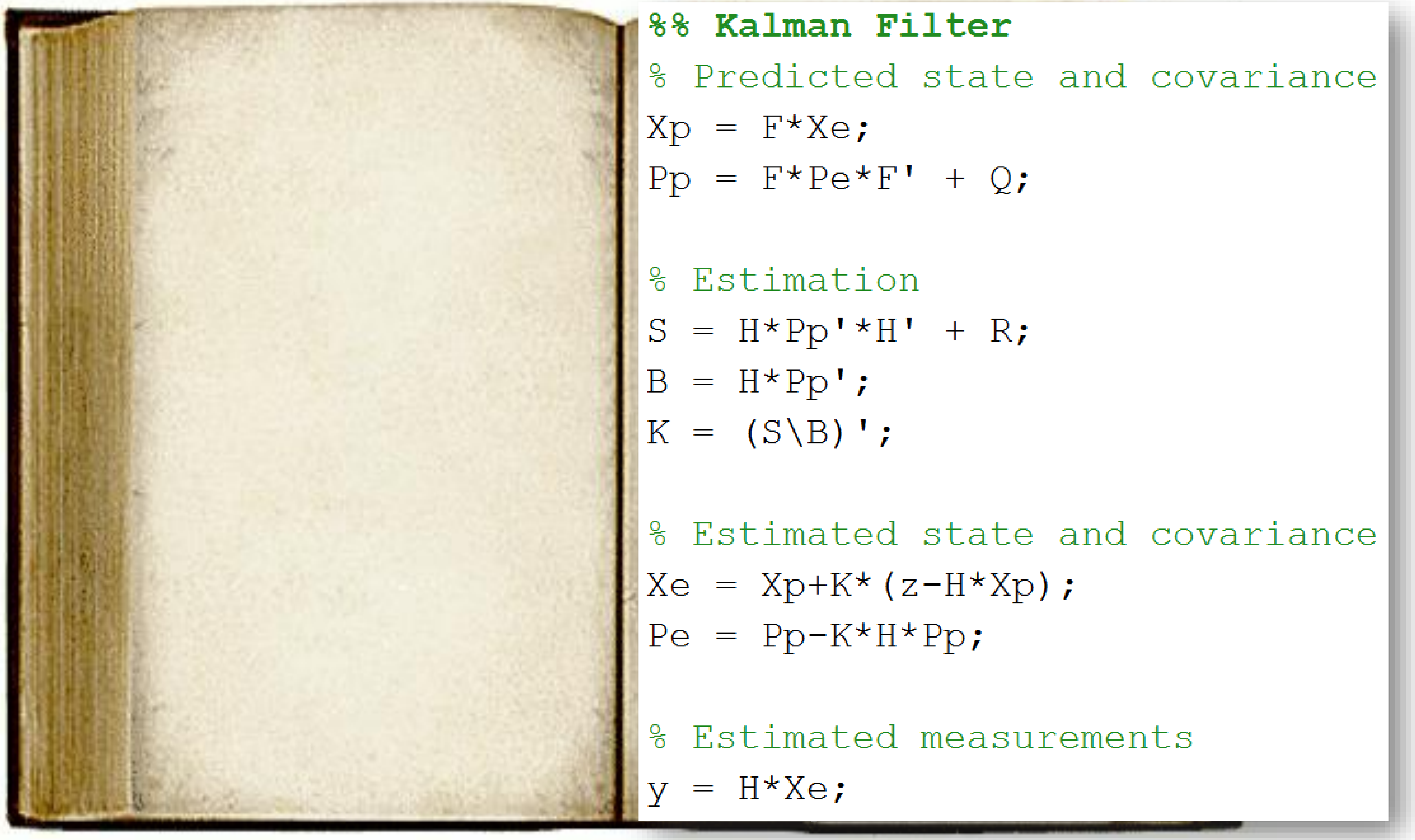
The solar declination is 0° at the vernal and autumnal equinoxes and in the northern hemisphere rises to a maximum of 23.45° at the summer solstice. On any given day of the year (d), solar declination (δ) can be calculated from the following formula:

```
delta = asind(sind(23.45)*sind(360*(days - 81)/365));    % Solar declination

clf
plot(days,delta)
axis([1 365 -30 30])
title('Solar Declination')
xlabel('Day of Year')
ylabel('Degrees')
```

Estimating Sunrise and Sunset

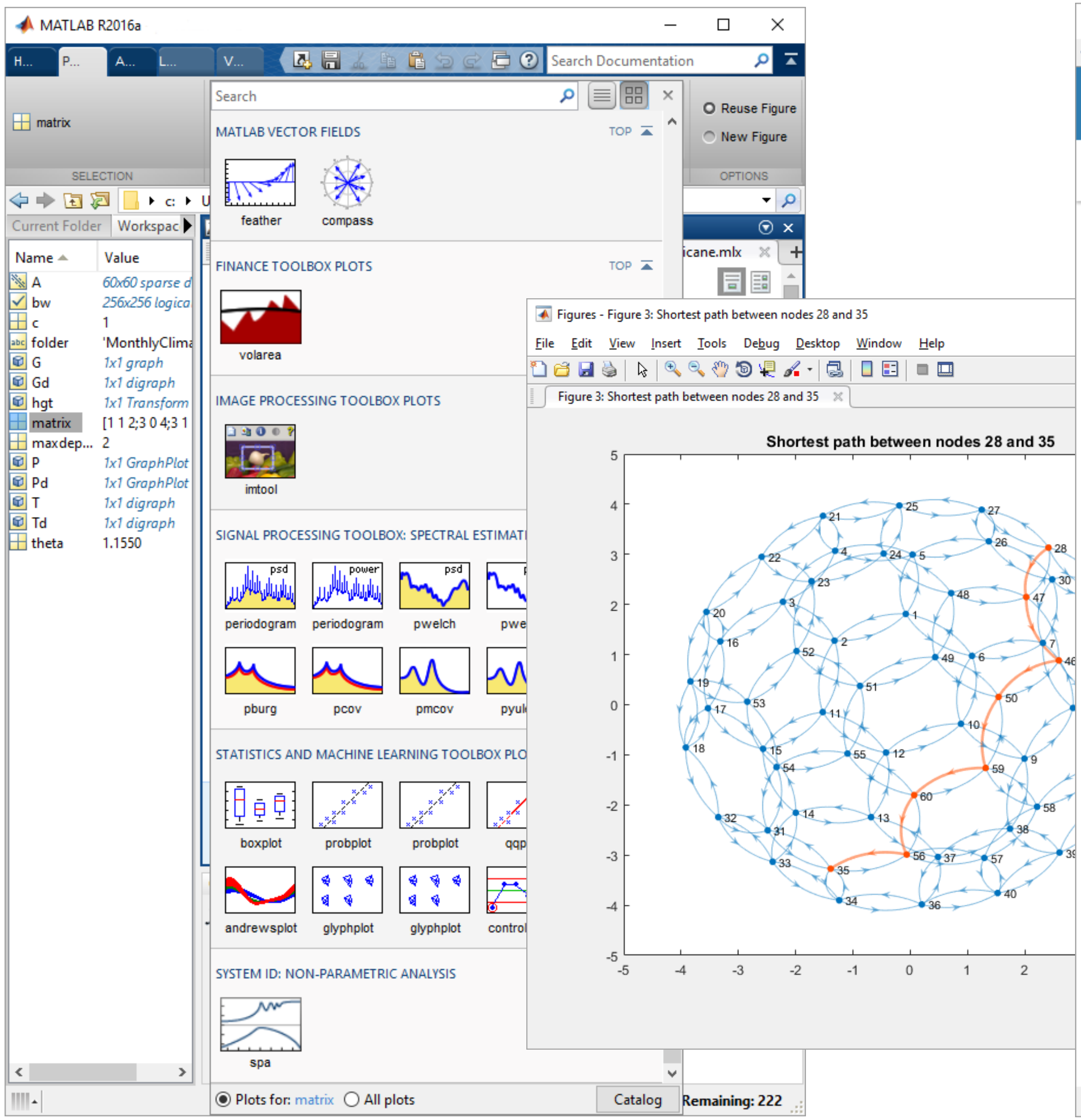
Knowing the local latitude (φ), the sun's declination (δ) and the solar time correction (SC) allows us to calculate the time of sunrise and sunset. Sunrise and sunset are calculated in Standard Time.

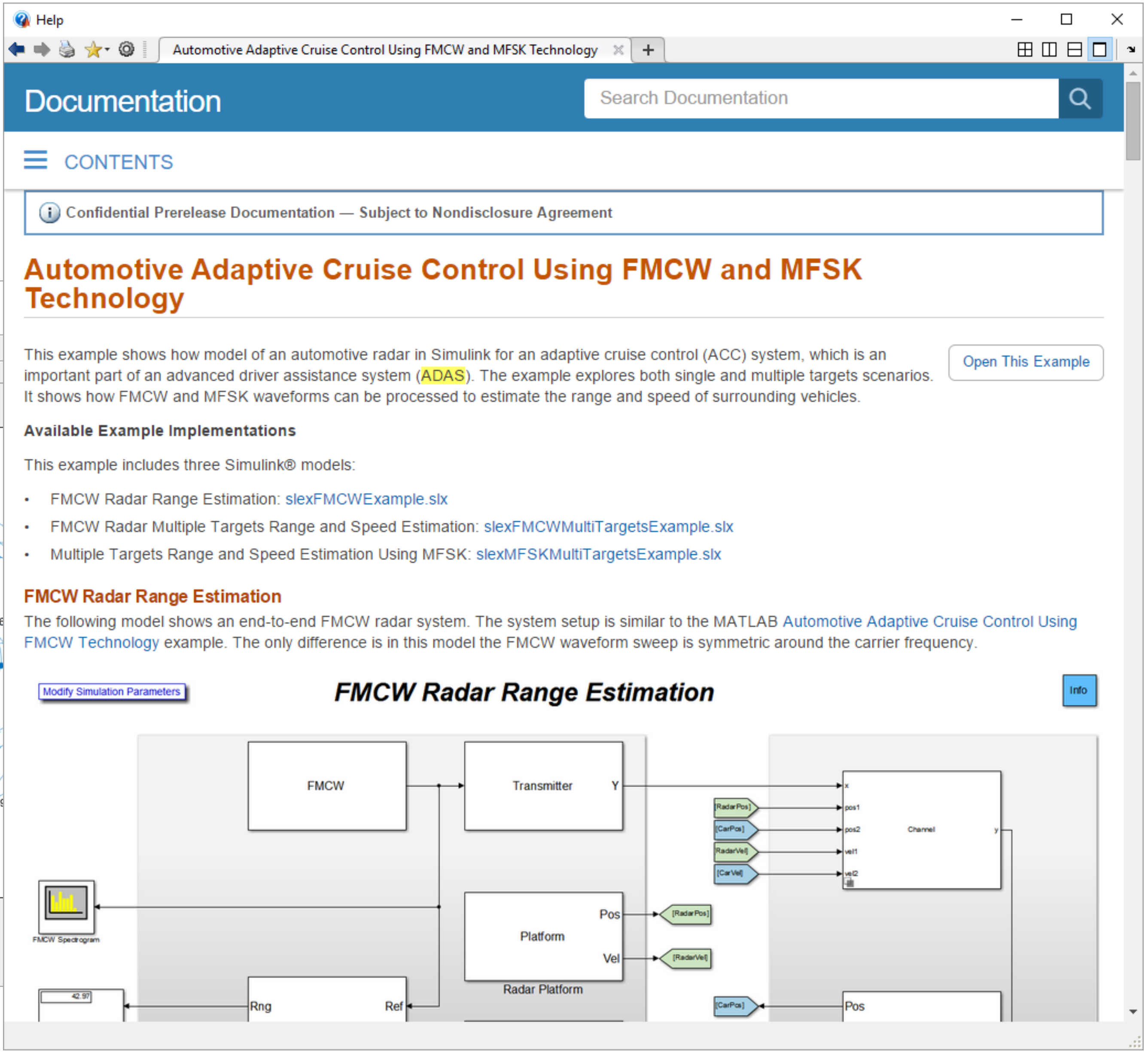
1. MATLAB 是数学语言



```
%% Kalman Filter
% Predicted state and covariance
Xp = F*Xe;
Pp = F*Pe*F' + Q;

% Estimation
S = H*Pp'*H' + R;
B = H*Pp';
K = (S\B)';

% Estimated state and covariance
Xe = Xp+K*(z-H*Xp);
Pe = Pp-K*H*Pp;

% Estimated measurements
y = H*Xe;
```

2. MATLAB 丰富的工具箱


MATLAB R2016a
Search Documentation
matrix
SELECTION
Current Folder
Workspace
Name
Value
A
60x60 sparse d
bw
256x256 logica
c
1
folder
'MonthlyClima
G
1x1 graph
Gd
1x1 digraph
hgt
1x1 Transform
matrix
[1 1 2;3 0 4;3 1
maxdep...
2
P
1x1 GraphPlot
Pd
1x1 GraphPlot
T
1x1 digraph
Td
1x1 digraph
theta
1.1550
Search
MATLAB VECTOR FIELDS
TOP
feather
compass
FINANCE TOOLBOX PLOTS
volarea
IMAGE PROCESSING TOOLBOX PLOTS
imtool
SIGNAL PROCESSING TOOLBOX: SPECTRAL ESTIMAT
periodogram
periodogram
pwelch
pburg
pcov
pmcov
STATISTICS AND MACHINE LEARNING TOOLBOX PLO
boxplot
probplot
probplot
andrewsplot
glyphplot
glyphplot
SYSTEM ID: NON-PARAMETRIC ANALYSIS
spa
○ Reuse Figure
○ New Figure
OPTIONS
● Plots for: matrix ○ All plots
Catalog
Remaining: 222
icane.mlx
Figures - Figure 3: Shortest path between nodes 28 and 35
File Edit View Insert Tools Debug Desktop Window Help
Figure 3: Shortest path between nodes 28 and 35
Shortest path between nodes 28 and 35


Help
Automotive Adaptive Cruise Control Using FMCW and MFSK Technology
Documentation
Search Documentation
CONTENTS
Open This Example

Confidential Prerelease Documentation — Subject to Nondisclosure Agreement

Automotive Adaptive Cruise Control Using FMCW and MFSK Technology

This example shows how model of an automotive radar in Simulink for an adaptive cruise control (ACC) system, which is an important part of an advanced driver assistance system (ADAS). The example explores both single and multiple targets scenarios. It shows how FMCW and MFSK waveforms can be processed to estimate the range and speed of surrounding vehicles.

Available Example Implementations

This example includes three Simulink® models:

- FMCW Radar Range Estimation: slexFMCWExample.slx
- FMCW Radar Multiple Targets Range and Speed Estimation: slexFMCWMultiTargetsExample.slx
- Multiple Targets Range and Speed Estimation Using MFSK: slexMFSKMultiTargetsExample.slx

FMCW Radar Range Estimation

The following model shows an end-to-end FMCW radar system. The system setup is similar to the MATLAB Automotive Adaptive Cruise Control Using FMCW Technology example. The only difference is in this model the FMCW waveform sweep is symmetric around the carrier frequency.

Modify Simulation Parameters
FMCW Radar Range Estimation
Info
FMCW
Transmitter
Channel
Platform
Pos
Vel
Radar Platform
FMCW Spectrogram
Rng
Ref
Pos

3. MATLAB 为工程学而设计

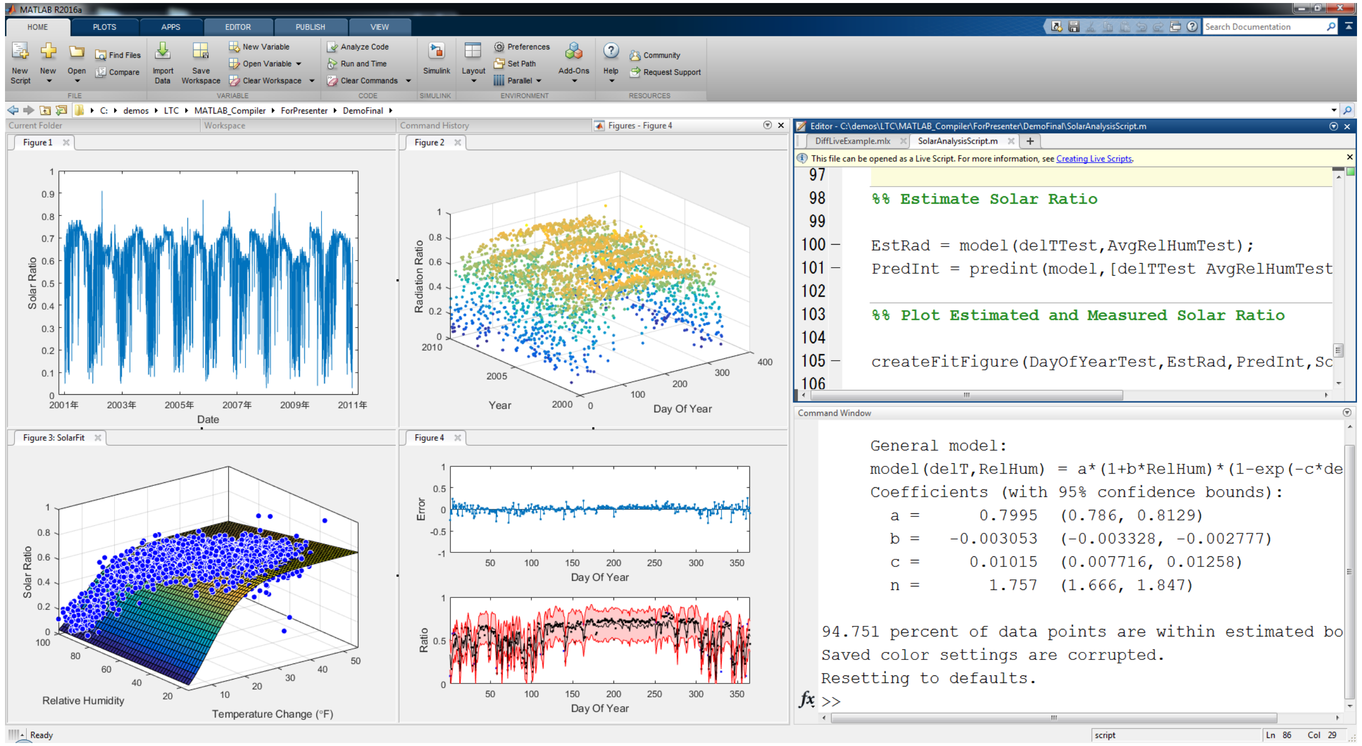
Access

Files

Software

Code & Applications

Hardware


MATLAB R2016a
HOME
PLOTS
APPS
EDITOR
PUBLISH
VIEW
Search Documentation
New Script
New
Open
Find Files
Compare
Import Data
Save Workspace
New Variable
Open Variable
Clear Workspace
Analyze Code
Run and Time
Clear Commands
Simulink
Layout
Preferences
Set Path
Parallel
Add-Ons
Help
Community
Request Support
FILE
VARIABLE
CODE
SIMULINK
ENVIRONMENT
RESOURCES
C: ▸ demos ▸ LTC ▸ MATLAB_Compiler ▸ ForPresenter ▸ DemoFinal ▸
Current Folder
Workspace
Command History
Figures - Figure 4
Figure 1
Solar Ratio
Date
2001年
2003年
2005年
2007年
2009年
2011年
Figure 2
Radiation Ratio
Year
Day Of Year
Figure 3: SolarFit
Solar Ratio
Relative Humidity
Temperature Change (°F)
Figure 4
Error
Day Of Year
Ratio
Day Of Year
Editor - C:\demos\LTC\MATLAB_Compiler\ForPresenter\DemoFinal\SolarAnalysisScript.m
DiffLiveExample.mlx
SolarAnalysisScript.m
Command Window
Ready
script
Ln 86
Col 29

This file can be opened as a Live Script. For more information, see Creating Live Scripts.

```
%% Estimate Solar Ratio

EstRad = model(delTTest,AvgRelHumTest);
PredInt = predint(model,[delTTest AvgRelHumTest

%% Plot Estimated and Measured Solar Ratio

createFitFigure(DayOfYearTest,EstRad,PredInt,Sc
```

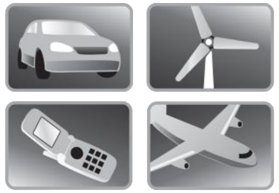
```
    General model:
    model(delT,RelHum) = a*(1+b*RelHum)*(1-exp(-c*de
    Coefficients (with 95% confidence bounds):
      a =      0.7995  (0.786, 0.8129)
      b =   -0.003053  (-0.003328, -0.002777)
      c =     0.01015  (0.007716, 0.01258)
      n =       1.757  (1.666, 1.847)

94.751 percent of data points are within estimated bo
Saved color settings are corrupted.
Resetting to defaults.
>>
```

Share

Reporting and Documentation

PDF .doc .html

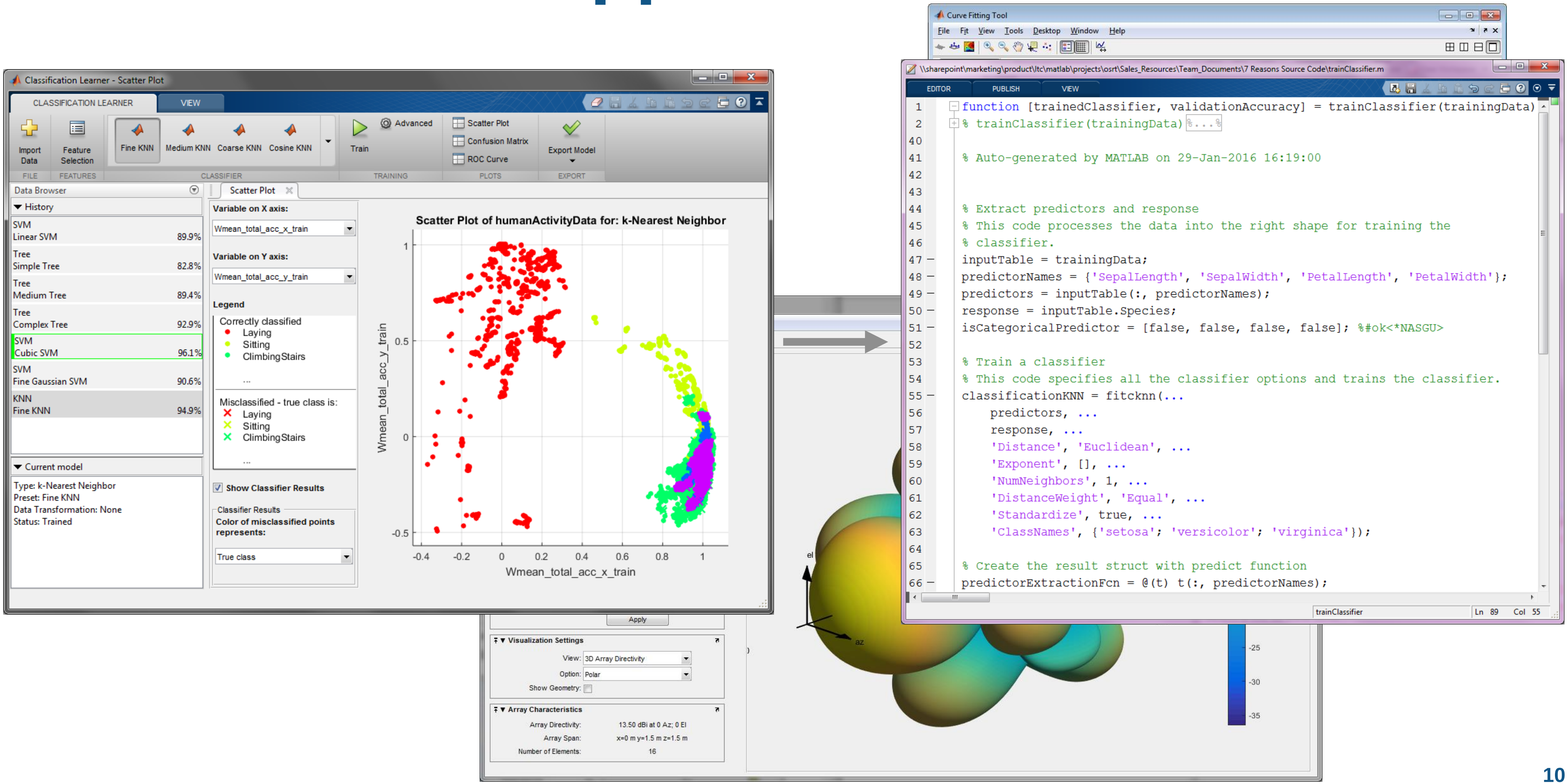
Outputs for Design



Deployment

MATLAB Excel .NET .exe C/C++ Java .dll

4. MATLAB 的 Apps


Classification Learner - Scatter Plot
CLASSIFICATION LEARNER
VIEW
Import Data
Feature Selection
Fine KNN
Medium KNN
Coarse KNN
Cosine KNN
Train
Advanced
Scatter Plot
Confusion Matrix
ROC Curve
Export Model
FILE
FEATURES
CLASSIFIER
TRAINING
PLOTS
EXPORT
Data Browser
Scatter Plot
Scatter Plot of humanActivityData for: k-Nearest Neighbor
Wmean_total_acc_y_train
Wmean_total_acc_x_train
1
0.5
0
-0.5
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
Curve Fitting Tool
File
Fit
View
Tools
Desktop
Window
Help
\\sharepoint\marketing\product\tc\matlab\projects\osrt\Sales_Resources\Team_Documents\7 Reasons Source Code\trainClassifier.m
EDITOR
PUBLISH
VIEW
trainClassifier
Ln 89
Col 55
Apply
el
az
-25
-30
-35

History

SVM Linear SVM	89.9%
Tree Simple Tree	82.8%
Tree Medium Tree	89.4%
Tree Complex Tree	92.9%
SVM Cubic SVM	96.1%
SVM Fine Gaussian SVM	90.6%
KNN Fine KNN	94.9%

Current model

Type: k-Nearest Neighbor
Preset: Fine KNN
Data Transformation: None
Status: Trained

Variable on X axis: Wmean_total_acc_x_train

Variable on Y axis: Wmean_total_acc_y_train

Legend

Correctly classified
- ● Laying
- ● Sitting
- ● ClimbingStairs
- ...

Misclassified - true class is:
- ✕ Laying
- ✕ Sitting
- ✕ ClimbingStairs
- ...

☑ Show Classifier Results

Classifier Results

Color of misclassified points represents: True class

```
function [trainedClassifier, validationAccuracy] = trainClassifier(trainingData)
% trainClassifier(trainingData) %...%

% Auto-generated by MATLAB on 29-Jan-2016 16:19:00


% Extract predictors and response
% This code processes the data into the right shape for training the
% classifier.
inputTable = trainingData;
predictorNames = {'SepalLength', 'SepalWidth', 'PetalLength', 'PetalWidth'};
predictors = inputTable(:, predictorNames);
response = inputTable.Species;
isCategoricalPredictor = [false, false, false, false]; %#ok<*NASGU>

% Train a classifier
% This code specifies all the classifier options and trains the classifier.
classificationKNN = fitcknn(...
    predictors, ...
    response, ...
    'Distance', 'Euclidean', ...
    'Exponent', [], ...
    'NumNeighbors', 1, ...
    'DistanceWeight', 'Equal', ...
    'Standardize', true, ...
    'ClassNames', {'setosa'; 'versicolor'; 'virginica'});

% Create the result struct with predict function
predictorExtractionFcn = @(t) t(:, predictorNames);
```

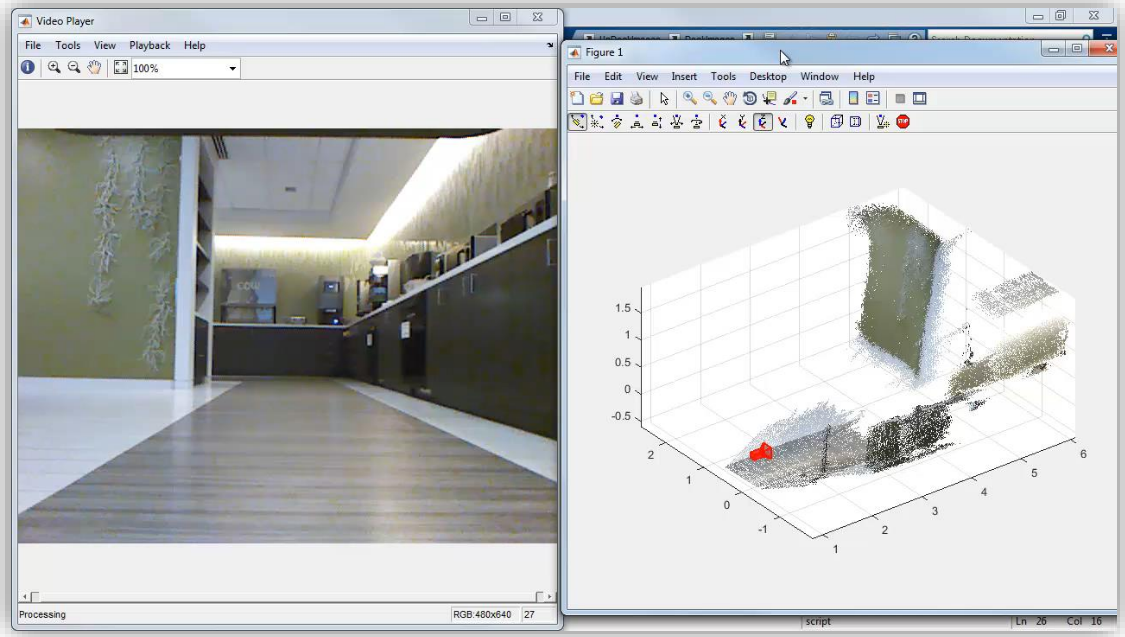
Visualization Settings

View: 3D Array Directivity
Option: Polar
Show Geometry: ☐

Array Characteristics

Array Directivity: 13.50 dBi at 0 Az; 0 El
Array Span: x=0 m y=1.5 m z=1.5 m
Number of Elements: 16

5. MATLAB 快速而高效


Video Player
File Tools View Playback Help
100%
Processing
RGB:480x640
Figure 1
File Edit View Insert Tools Desktop Window Help
1.5
1
0.5
0
-0.5
2
1
0
-1
1
2
3
4
5
6

快速3D 数据并行计算

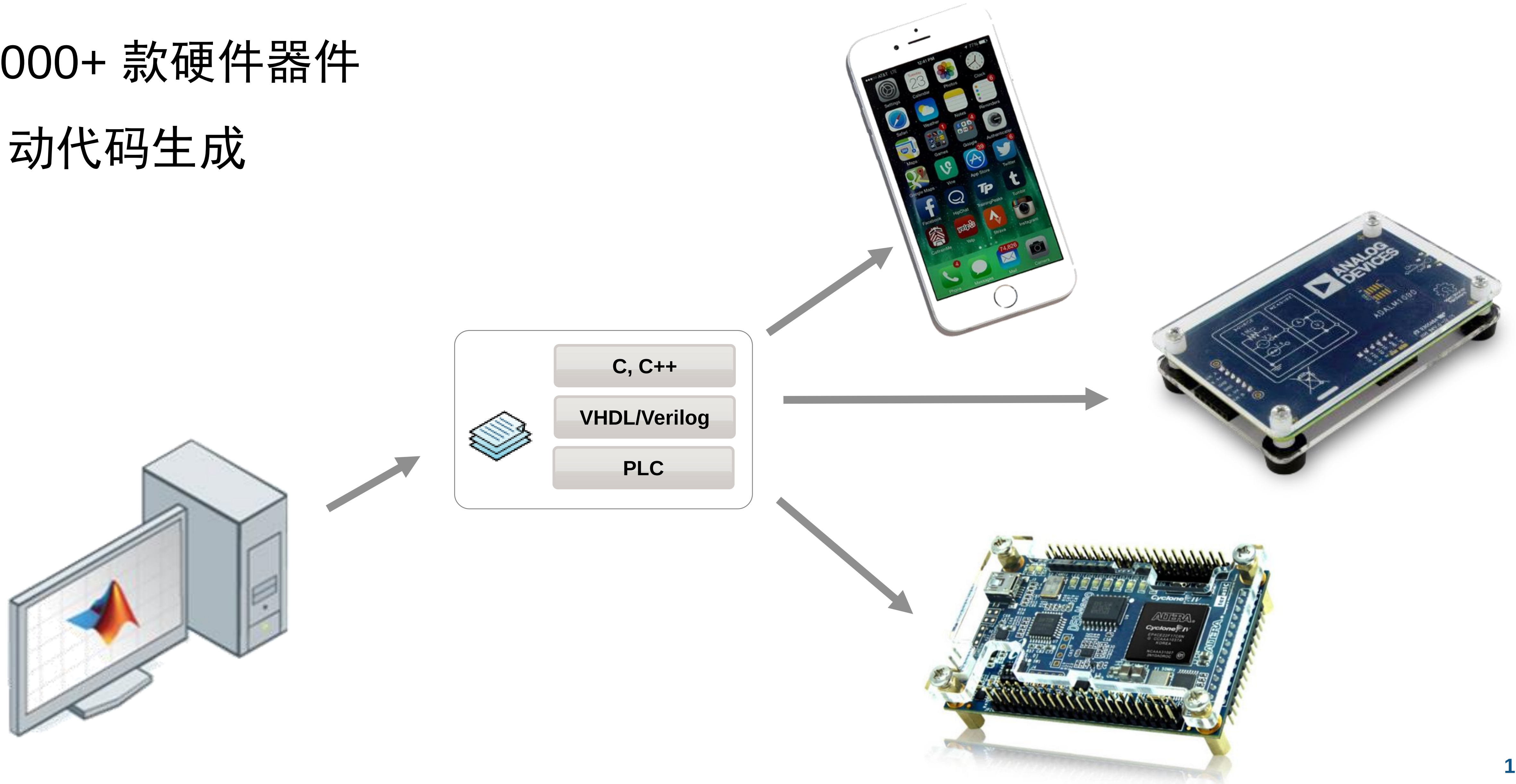
```
parfor m = 1:numel(freq)
    AR(m) = axialRatio(sp, freq(m), 0, 90);
    RL(m) = returnLoss(sp, freq(m));
end
```

```
x = cos(pi*(0:N)/N);
x = gpuArray(x);
y = x';
dt = 6/N^2;
[xx,yy] = meshgrid(x,y);
vv = exp(-40*((xx-.4).^2 + yy.^2));
```

6. MATLAB 高度集成

1,000+ 款硬件器件

自动代码生成



C, C++

VHDL/Verilog

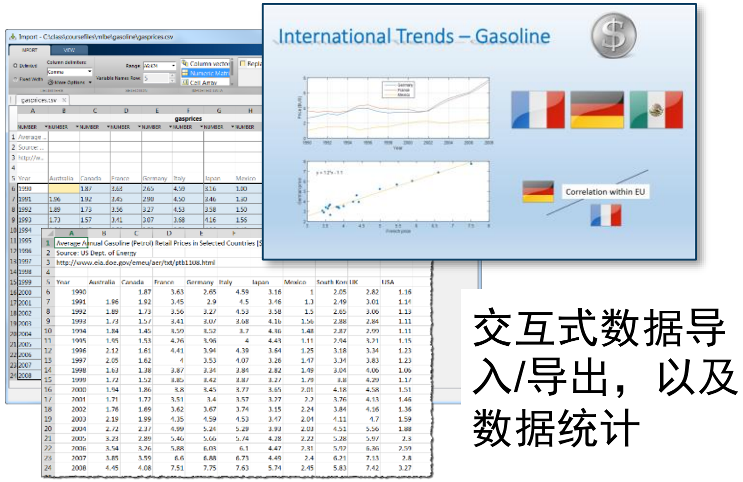
PLC

7. MATLAB 可信性

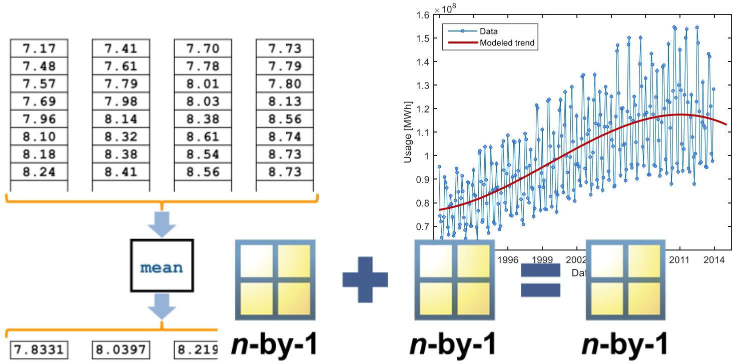
每天100- 200 万次测试

MATLAB
矩阵运算与数据可视化

MATLAB 可以做什么?

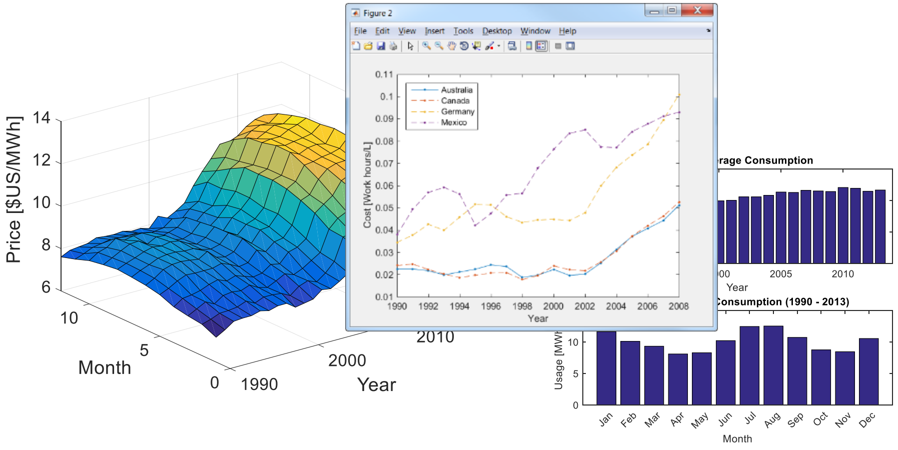

International Trends – Gasoline
Correlation within EU
Average Annual Gasoline (Petrol) Retail Prices in Selected Countries
Source: US Dept. of Energy

交互式数据导入/导出，以及数据统计

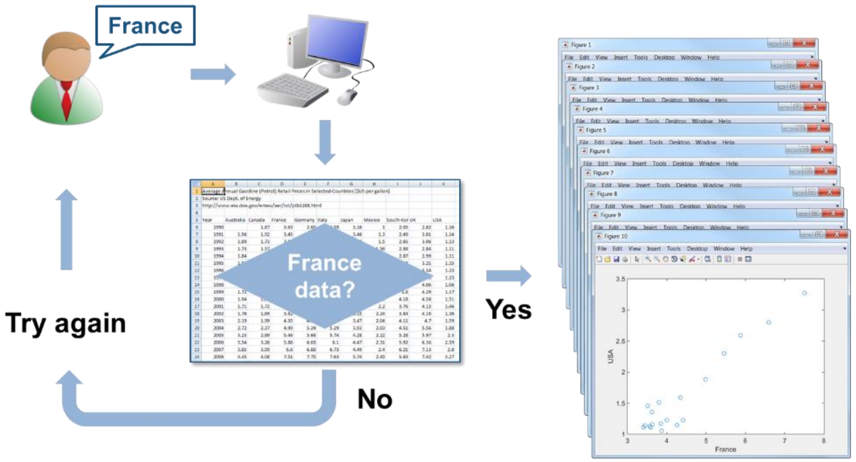

7.17
7.48
7.57
7.69
7.96
8.10
8.18
8.24
7.41
7.61
7.79
7.98
8.14
8.32
8.38
8.41
7.70
7.78
8.01
8.03
8.38
8.61
8.54
8.56
7.73
7.79
7.80
8.13
8.56
8.74
8.73
8.73
mean
7.8331
8.0397
8.219
n-by-1
n-by-1
n-by-1
Data
Modeled trend
Usage [MWh]
1996
1999
2002
2011
2014

数据集合上的计算与分析

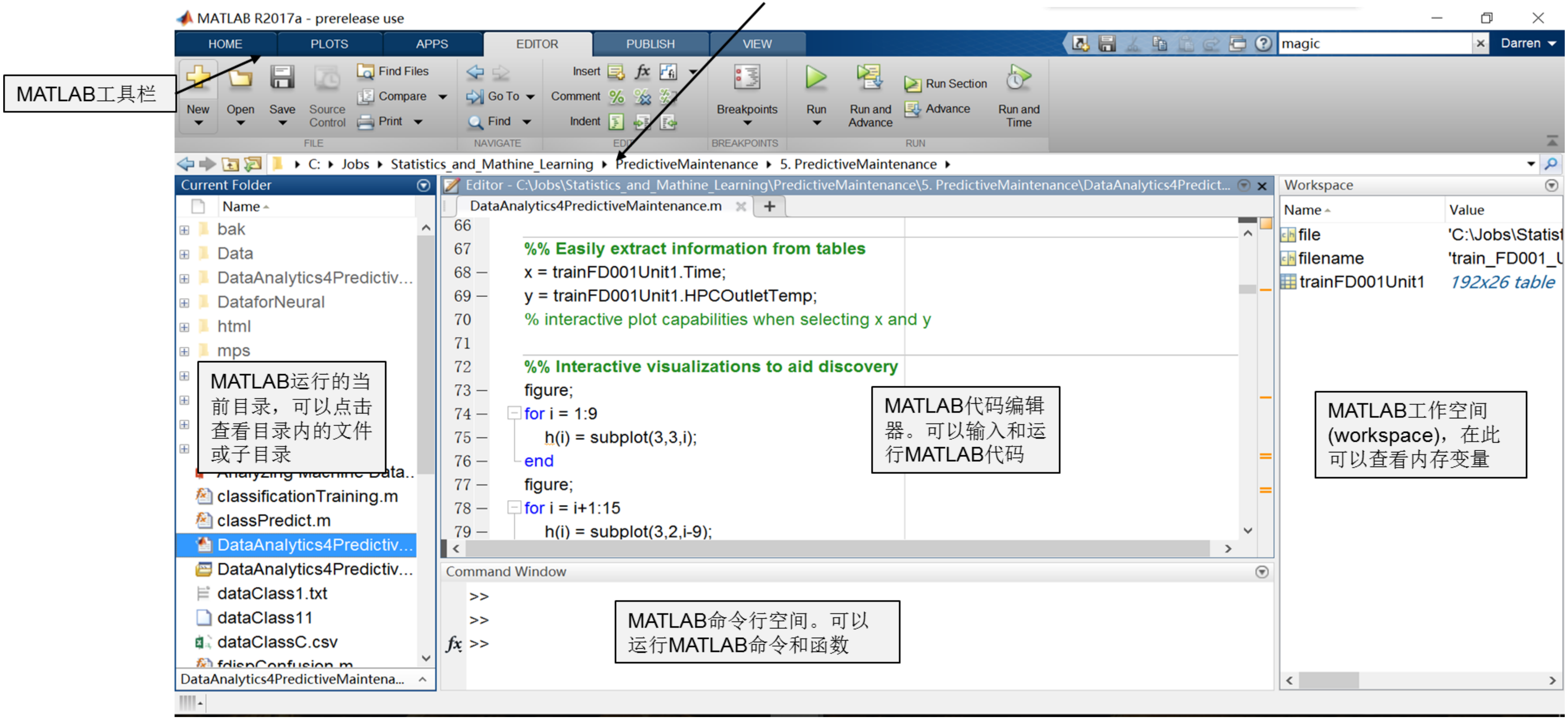
数据可视化



通过编写MATLAB代码自动完成复杂任务


France
France data?
Yes
No
Try again

MATLAB 客户端界面


MATLAB R2017a - prerelease use
HOME
PLOTS
APPS
EDITOR
PUBLISH
VIEW
magic
Darren
New
Open
Save
Source Control
Find Files
Compare
Print
FILE
Go To
Find
NAVIGATE
Insert
Comment
Indent
EDIT
Breakpoints
BREAKPOINTS
Run
Run and Advance
Run Section
Advance
Run and Time
RUN
C: ▸ Jobs ▸ Statistics_and_Mathine_Learning ▸ PredictiveMaintenance ▸ 5. PredictiveMaintenance ▸
Current Folder
Name
bak
Data
DataAnalytics4Predictiv...
DataforNeural
html
mps
Analyzing Machine Data...
classificationTraining.m
classPredict.m
DataAnalytics4Predictiv...
DataAnalytics4Predictiv...
dataClass1.txt
dataClass11
dataClassC.csv
fdispConfusion.m
DataAnalytics4PredictiveMaintena...
Editor - C:\Jobs\Statistics_and_Mathine_Learning\PredictiveMaintenance\5. PredictiveMaintenance\DataAnalytics4Predict...
DataAnalytics4PredictiveMaintenance.m
Workspace
Command Window

显示MATLAB运行的当前路径

MATLAB工具栏

MATLAB运行的当前目录，可以点击查看目录内的文件或子目录

MATLAB代码编辑器。可以输入和运行MATLAB代码

MATLAB工作空间(workspace)，在此可以查看内存变量

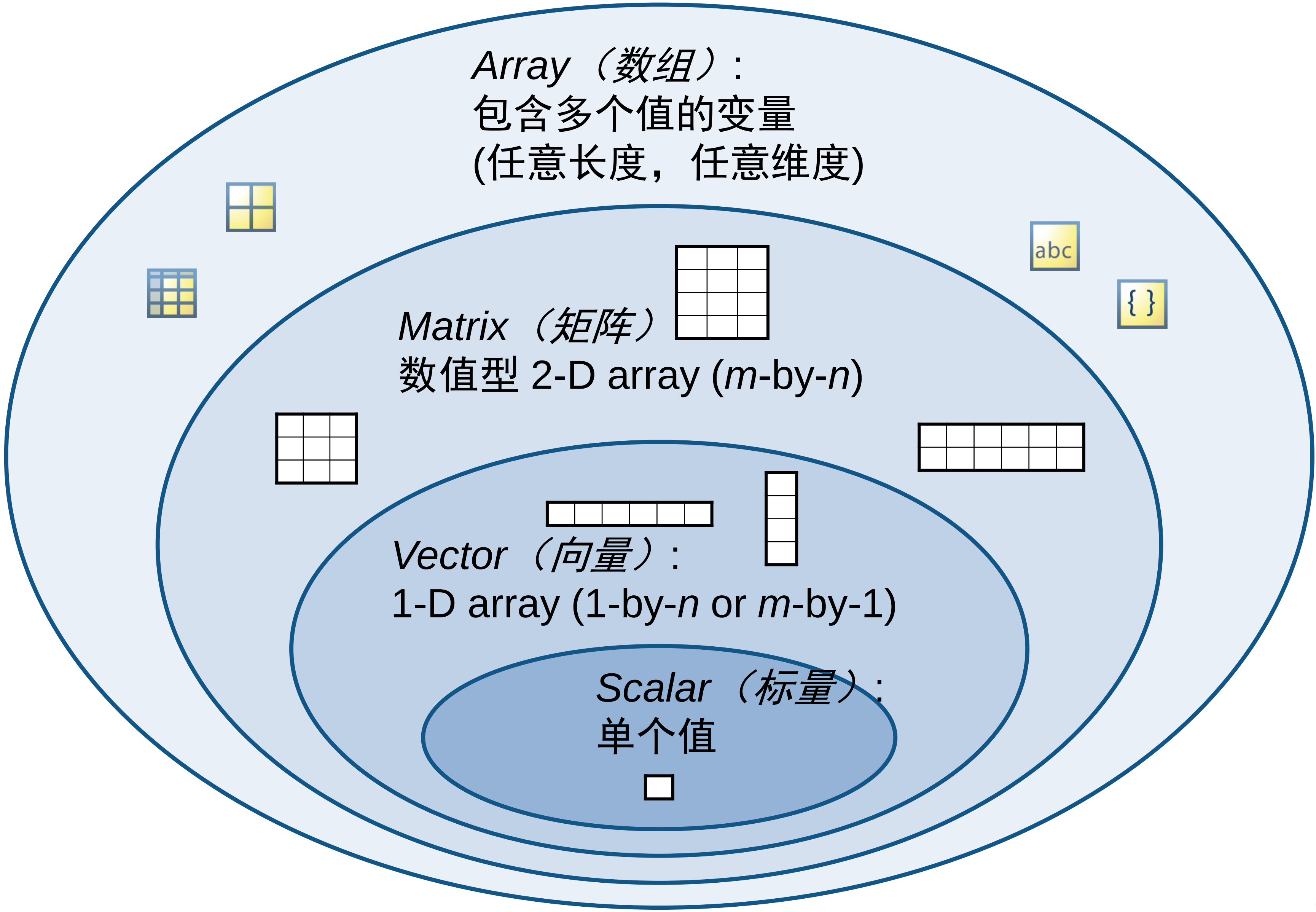
MATLAB命令空间。可以运行MATLAB命令和函数

```
%% Easily extract information from tables
x = trainFD001Unit1.Time;
y = trainFD001Unit1.HPCOutletTemp;
% interactive plot capabilities when selecting x and y

%% Interactive visualizations to aid discovery
figure;
for i = 1:9
    h(i) = subplot(3,3,i);
end
figure;
for i = i+1:15
    h(i) = subplot(3,2,i-9);
```

Name	Value
file	'C:\Jobs\Statist
filename	'train_FD001_U
trainFD001Unit1	*192x26 table*

```
>>
>>
fx >>
```

MATLAB 中的数据结构



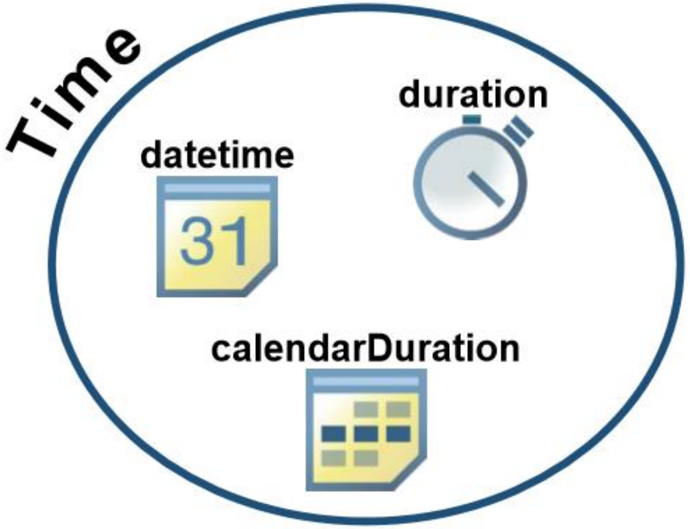
Array（数组）：
包含多个值的变量
(任意长度，任意维度)

Matrix（矩阵）
数值型 2-D array (*m*-by-*n*)

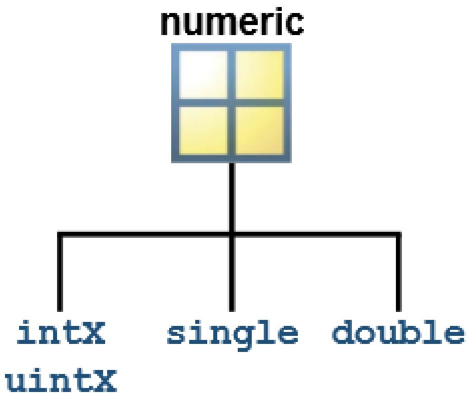
Vector（向量）：
1-D array (1-by-*n* or *m*-by-1)

Scalar（标量）：
单个值

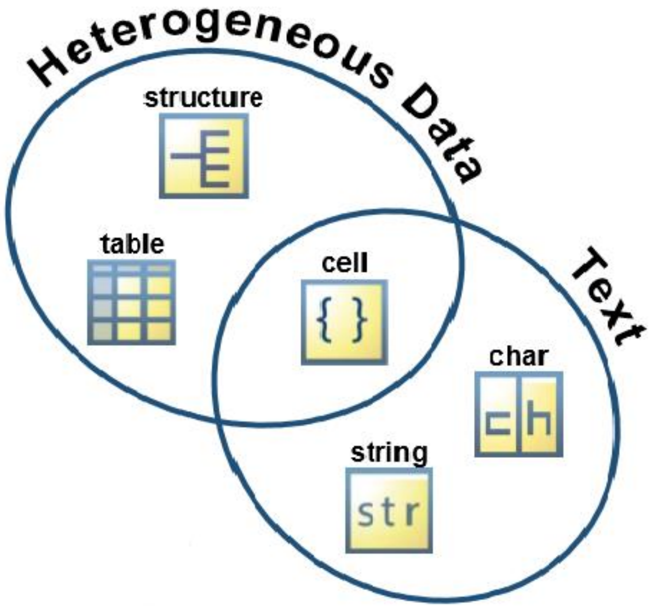
MATLAB 中的数据类型


Time
duration
datetime
31
calendarDuration

时间型


numeric
intX
uintX
single
double

数值型


Heterogeneous Data
structure
table
cell
{ }
Text
char
ch
string
str

异构型和字符型

categorical

分类型

function handle

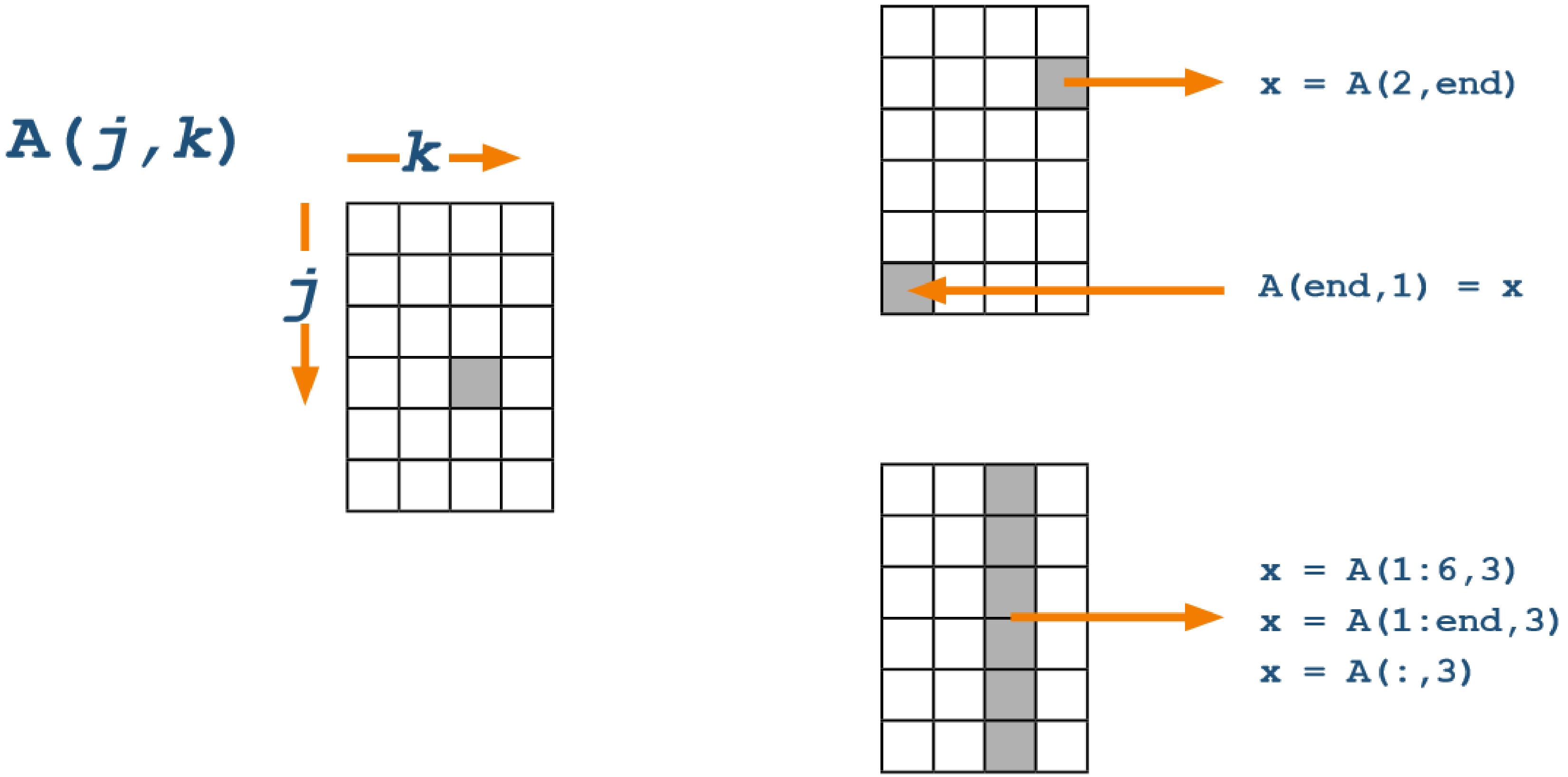
函数句柄型

logical

逻辑型

MATLAB 矩阵计算

- 矩阵化的数据访问



`x = A(2,end)`

`A(end,1) = x`

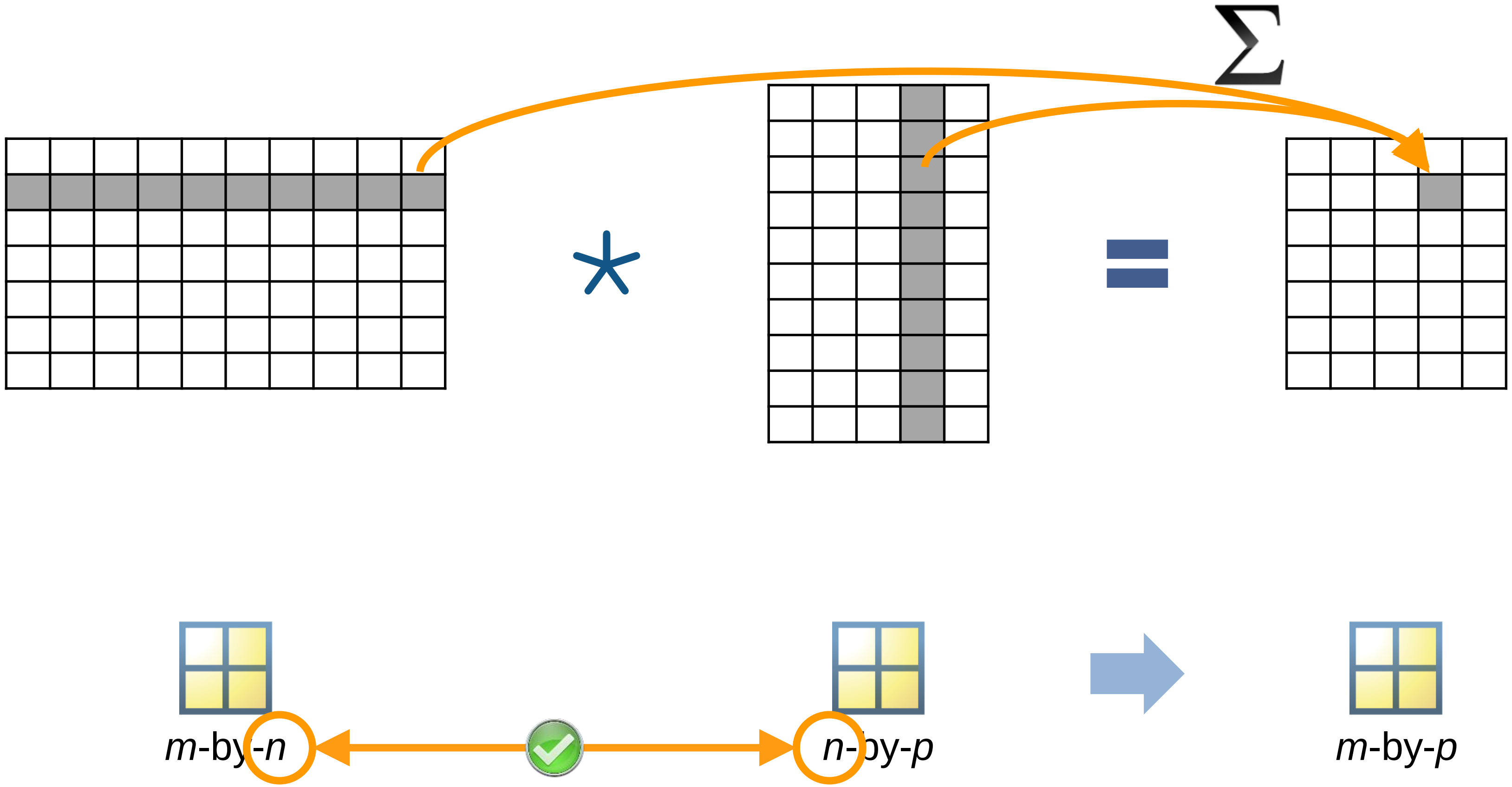
`x = A(1:6,3)`

`x = A(1:end,3)`

`x = A(:,3)`

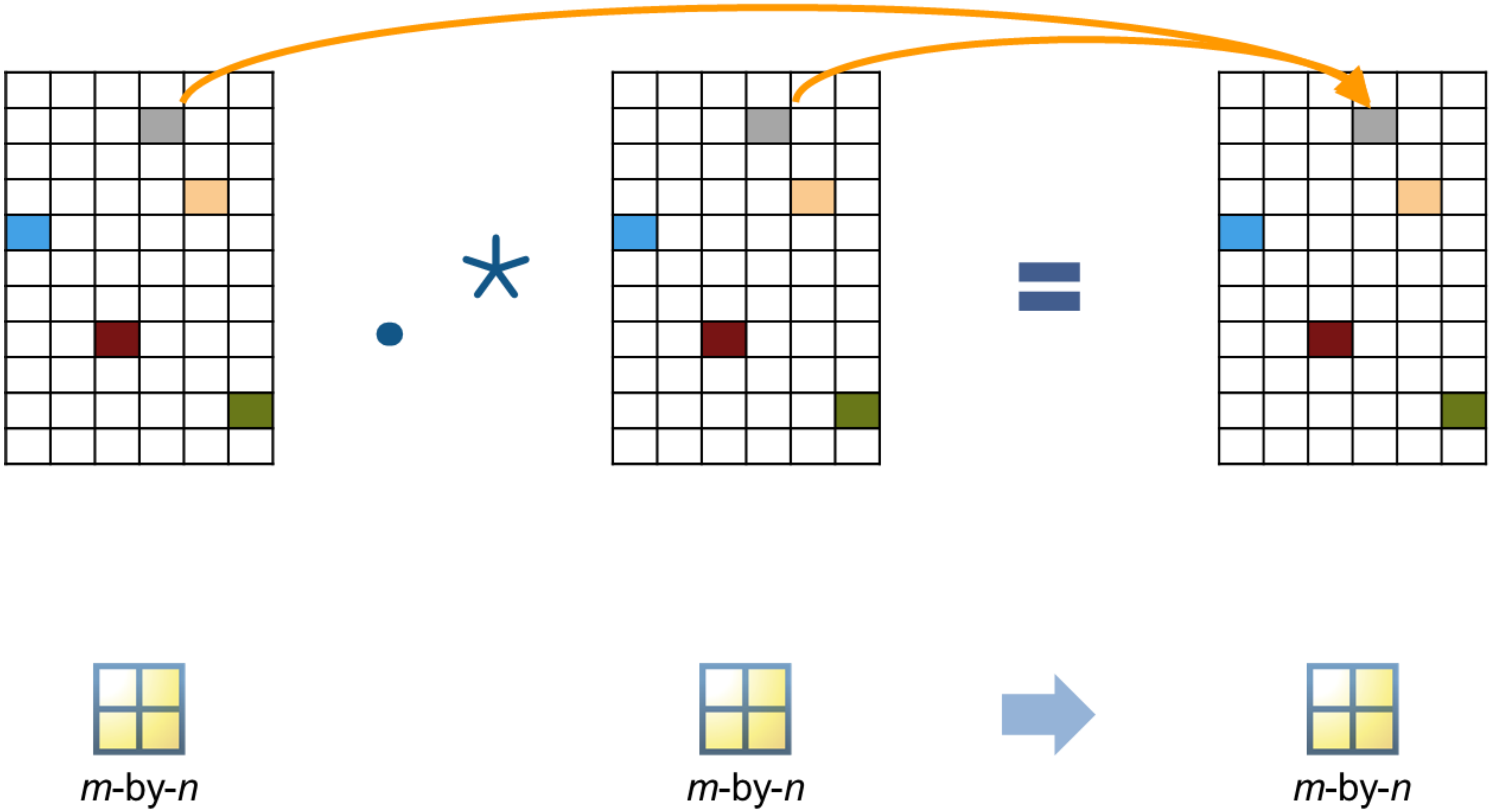
MATLAB 矩阵计算

- **矩阵乘法**


Σ
*
=
m-by-*n*
n-by-*p*
m-by-*p*

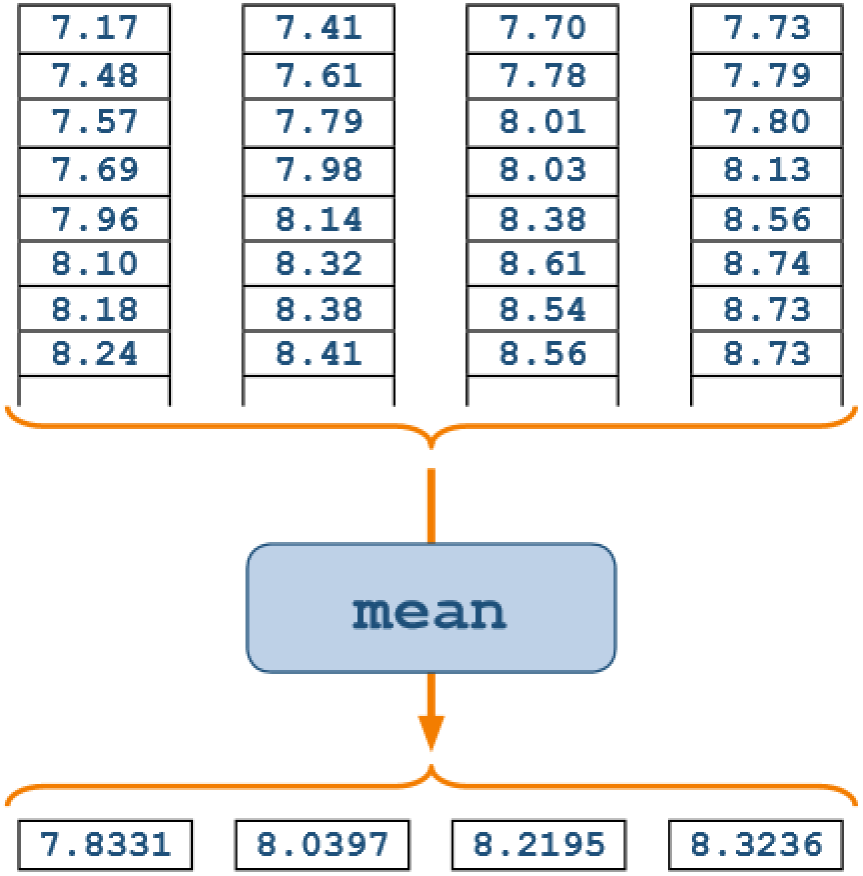
MATLAB 矩阵计算

- **矩阵点乘法**


.*
=
m-by-n
m-by-n
m-by-n

MATLAB 矩阵计算

- **向量化操作的内置函数**



7.17	7.41	7.70	7.73
7.48	7.61	7.78	7.79
7.57	7.79	8.01	7.80
7.69	7.98	8.03	8.13
7.96	8.14	8.38	8.56
8.10	8.32	8.61	8.74
8.18	8.38	8.54	8.73
8.24	8.41	8.56	8.73

7.8331	8.0397	8.2195	8.3236

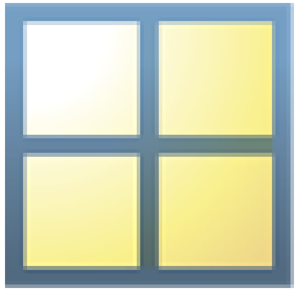
sin
sind
sinh
asin
exp
log
log2
log10
sqrt
nthroot
abs
angle
floor
ceil
round
mod

max
min
mean
median
std
sum
prod
diff
gradient
cumsum
cumprod
corrcoef
cov

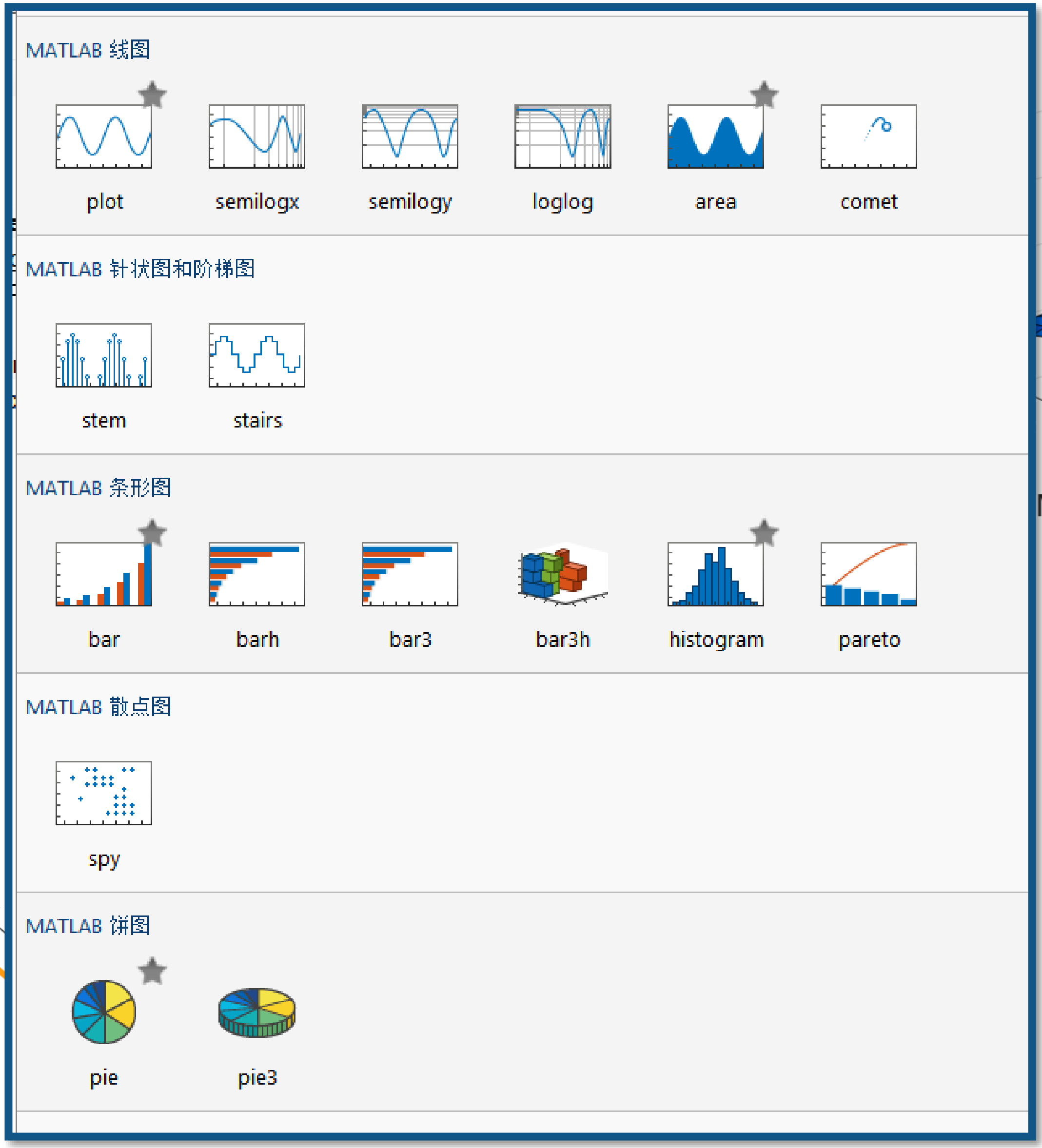
```
i = 0;
for t = 0:.01:10
    i = i + 1;
    y(i) = sin(t);
end
```

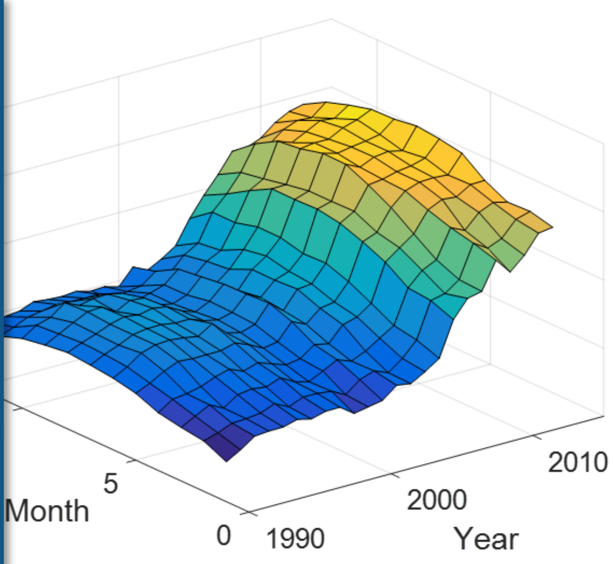
```
t = 0:.01:10;
y = sin(t);
```

矩阵可视化



m-by-*n*


MATLAB 线图
plot
semilogx
semilogy
loglog
area
comet
MATLAB 针状图和阶梯图
stem
stairs
MATLAB 条形图
bar
barh
bar3
bar3h
histogram
pareto
MATLAB 散点图
spy
MATLAB 饼图
pie
pie3


Month
5
0
1990
2000
2010
Year

MATLAB 数据分析

铁路

汽车

航空

零售

金融

互联网

物流

健康管理

医疗设备

清洁能源

石油天然气

工业自动化

非公路汽车

Data
Analytics

状态监测分析

预测分析

车队分析

过程分析

健康监测分析

资产分析

零售分析

运营分析

风险分析

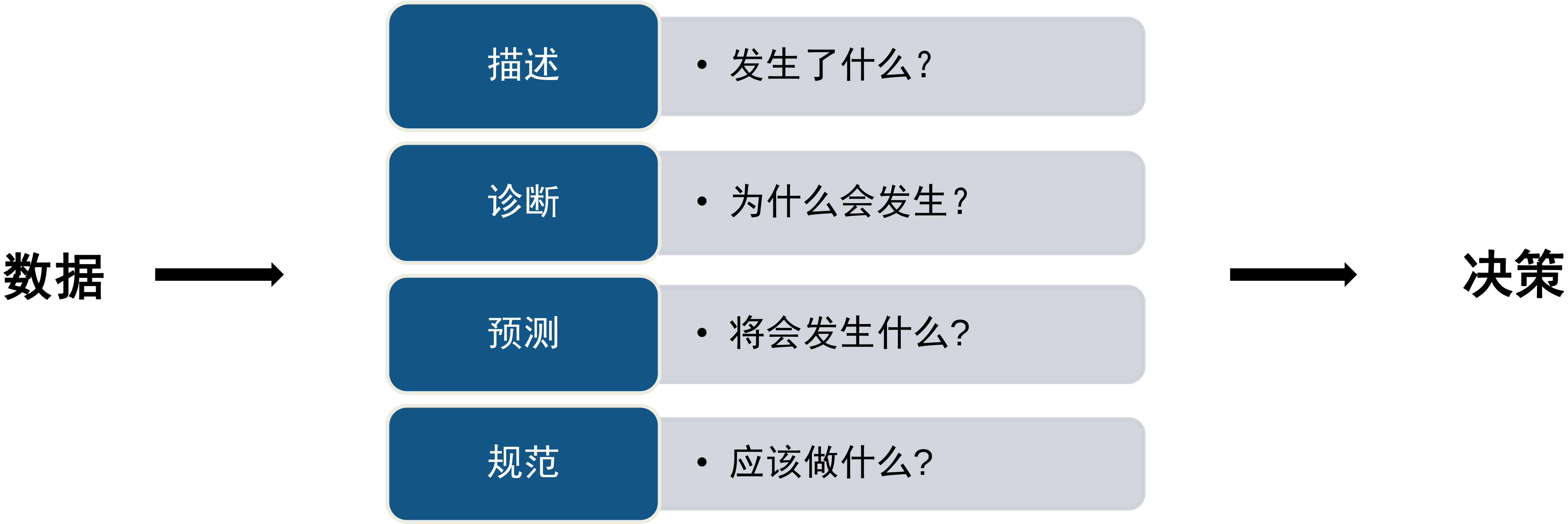
供应链分析

制造过程分析

健康管理分析

为什么需要数据分析?

将大量复杂数据转化为可操作的信息



数据 →

描述
- 发生了什么?

诊断
- 为什么会发生?

预测
- 将会发生什么?

规范
- 应该做什么?

→ 决策

MATLAB 令数据分析变得简单

数据获取与探索

数据预处理

开发预测模型

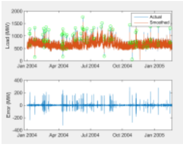
分析结果分享与实现

文件

数据库

传感器

处理散乱的数据



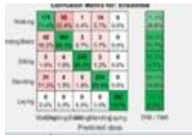
数据压缩与转换

特征提取

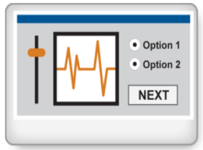
数学建模

参数优化

模型验证



桌面 Apps


Option 1
Option 2
NEXT

企业级系统集成

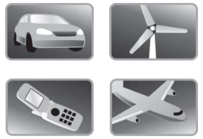
MATLAB Excel

.NET

.exe C/C++

Java .dll

嵌入式系统与硬件



数据分析与探索

数据获取与探索

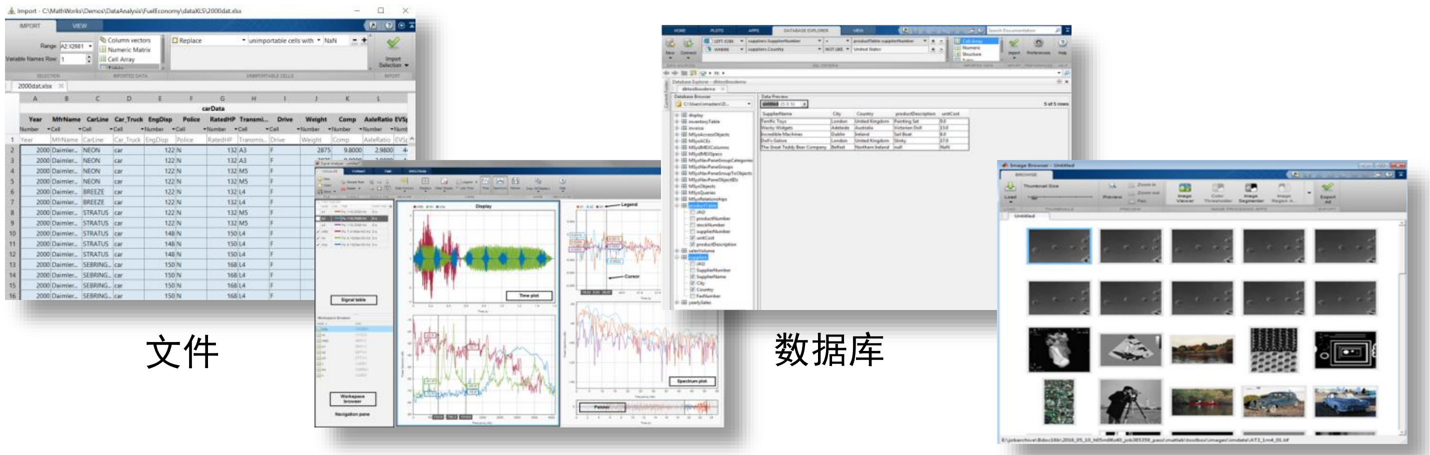
文件

数据库

传感器

1

MATLAB 支持商用数据和工程数据



文件

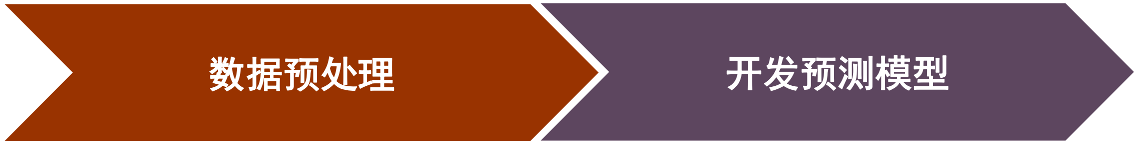
信号

数据库

图像

- 选择并点击多个工具可直接获取各类数据源
- 高性能大数据环境
- 利用内置的数据预处理算法，获取传感器、图像、音频、视频等实时数据

数据处理与模型开发

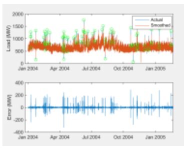


数据预处理 → **开发预测模型**

MATLAB 协助各领域专家从容应对数据科学

2

处理散乱的数据

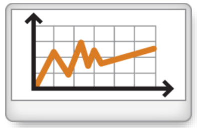


数据压缩与转换

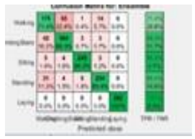
特征提取

数学建模

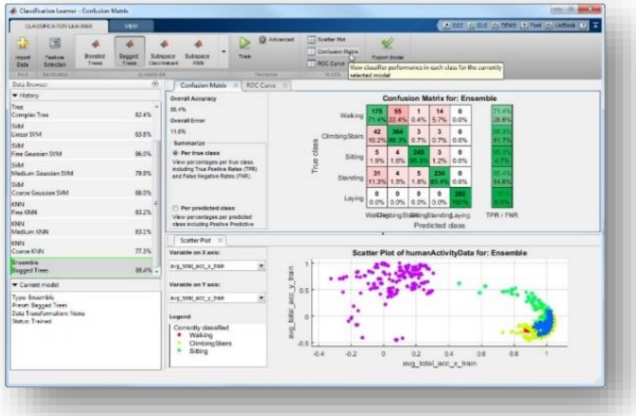
参数优化



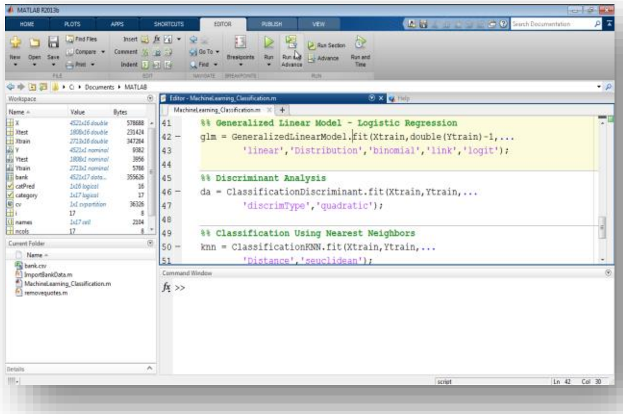
模型验证



Apps


Classification Learner - Confusion Matrix
Confusion Matrix for: Ensemble
Scatter Plot of humanActivityData for: Ensemble
Walking
ClimbingStairs
Sitting
Standing
Laying
True class
Predicted class

代码


Editor - MachineLearning_Classification.m
Command Window

```
%% Generalized Linear Model - Logistic Regression
glm = GeneralizedLinearModel.fit(Xtrain,double(Ytrain)-1,...
        'linear','Distribution','binomial','link','logit');

%% Discriminant Analysis
da = ClassificationDiscriminant.fit(Xtrain,Ytrain,...
        'discrimType','quadratic');

%% Classification Using Nearest Neighbors
knn = ClassificationKNN.fit(Xtrain,Ytrain,...
        'Distance','seuclidean');
```

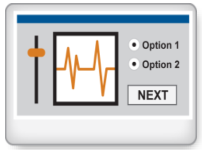
- Apps 简单易用
- 广泛领域的工具，以方便特定领域的分析
- 案例和视频协助快速入门

- MATLAB 代码自动生成
- 大数据高速运算

应用发布与部署

分析结果分享与部署

桌面 Apps

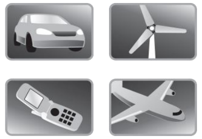


Option 1
Option 2
NEXT

企业级系统集成

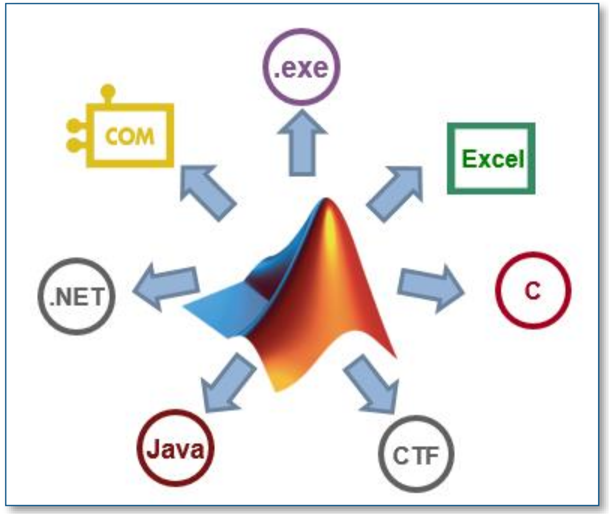
MATLAB Excel
.NET C/C++
.exe
Java .dll

嵌入式系统与硬件



3

MATLAB 分析应用可以在任一平台运行



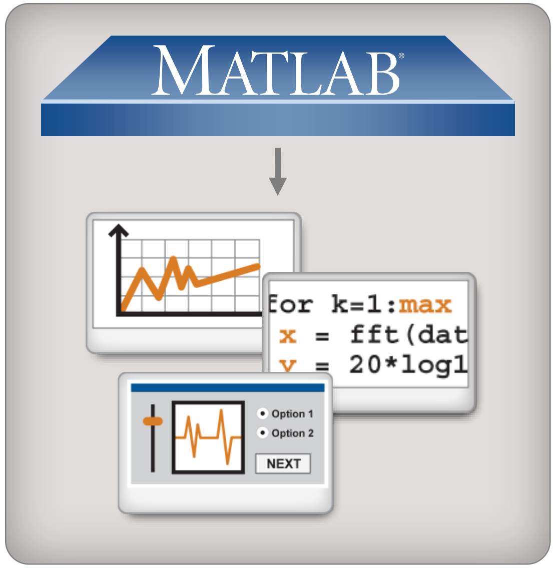
.exe
COM
Excel
.NET
C
Java
CTF

- 选择并点击实现部署
- 桌面和服务器应用统一的实现过程

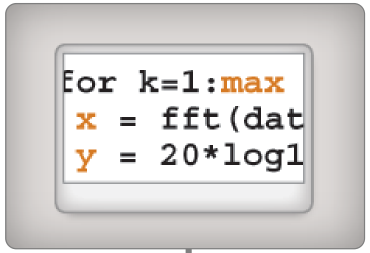
分析结果分享与部署

3

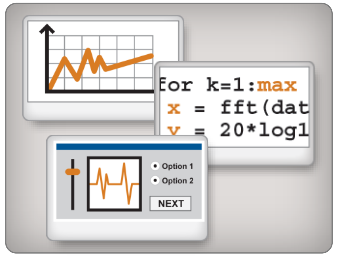
MATLAB 分析应用可以在任一平台运行


MATLAB
Option 1
Option 2
NEXT

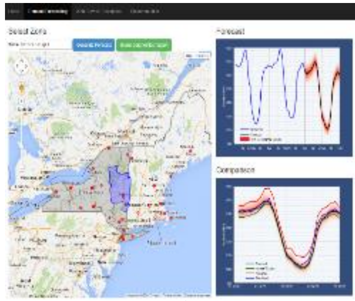
```
for k=1:max
x = fft(dat
y = 20*log1
```



```
for k=1:max
x = fft(dat
y = 20*log1
```


Option 1
Option 2
NEXT

```
for k=1:max
x = fft(dat
y = 20*log1
```

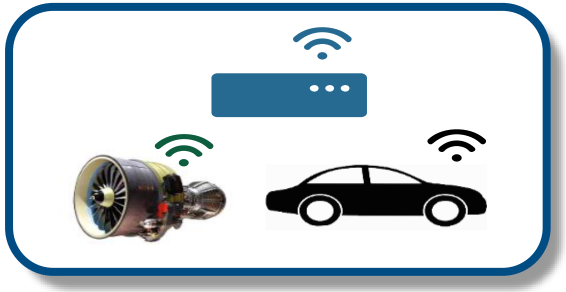


嵌入式系统与硬件

C, C++

HDL

PLC



系统平台

Standalone Application

Excel Add-in

Hadoop/ Spark

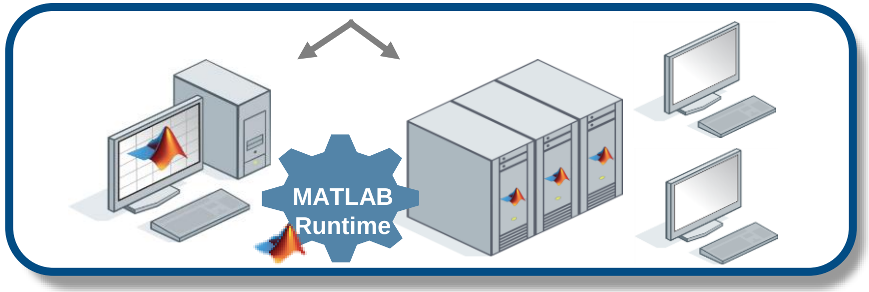
C/C++

Java

Python

.NET

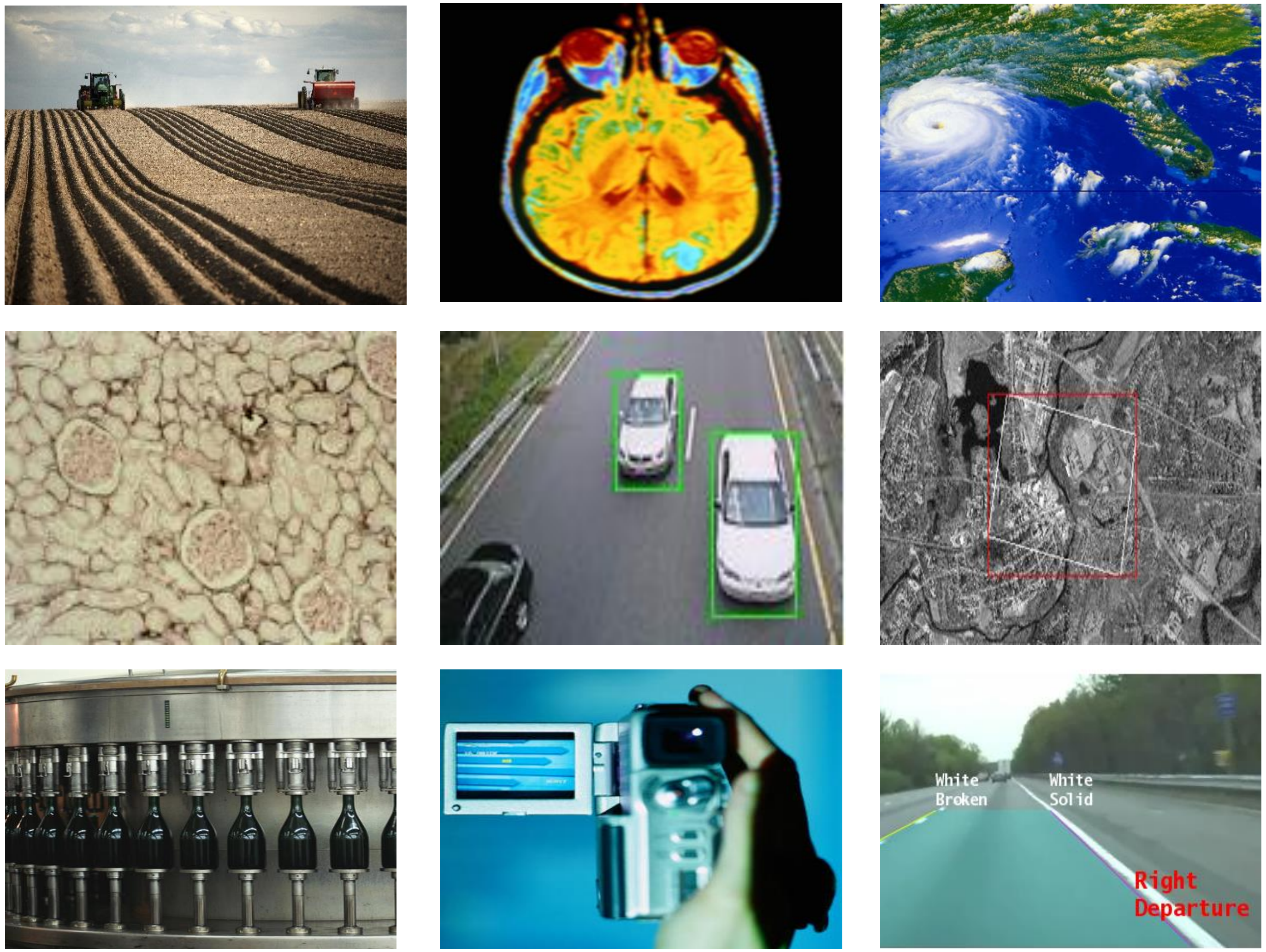
MATLAB Production Server


MATLAB Runtime

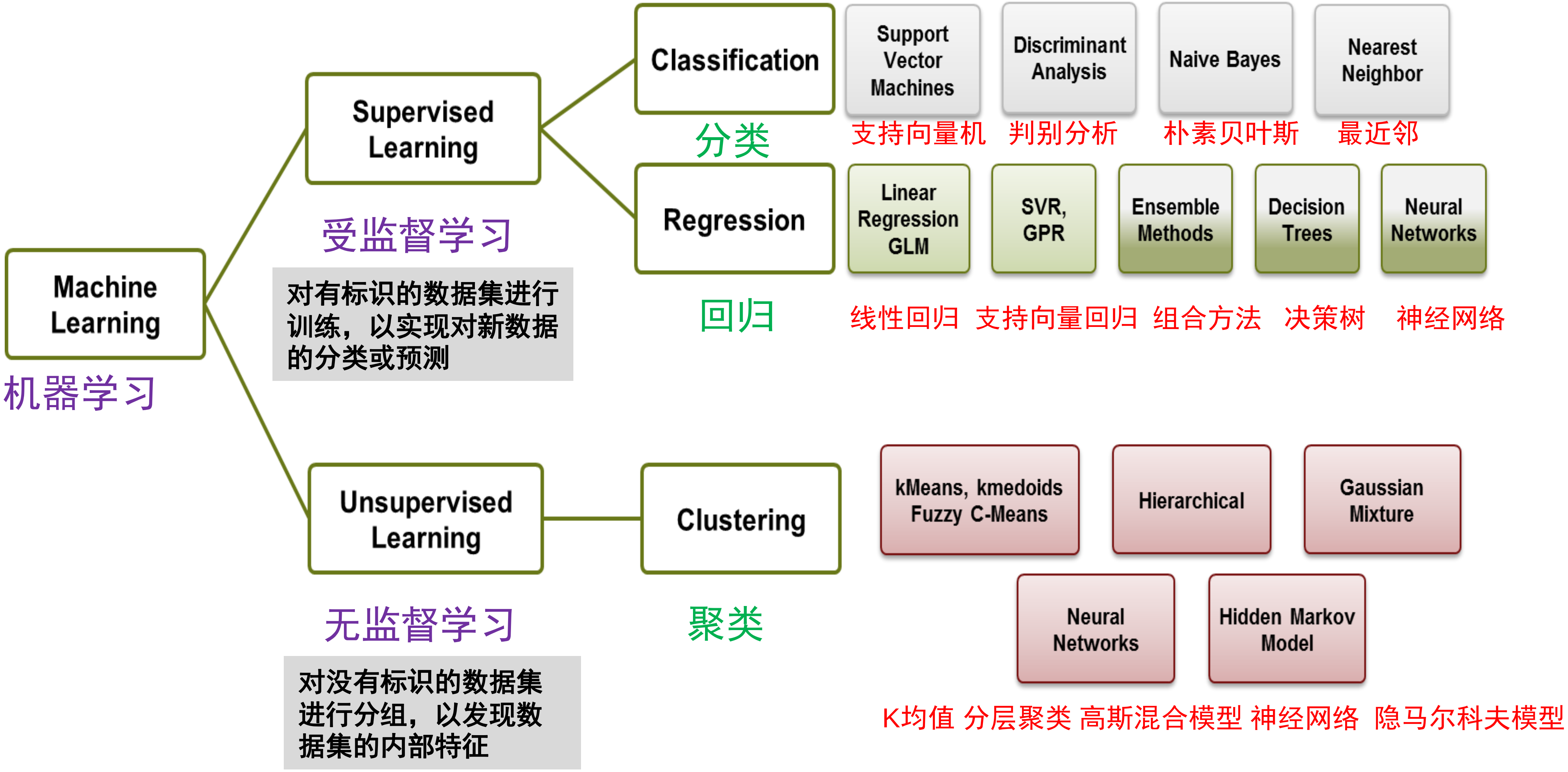
MATLAB
机器学习、大数据与并行加速

机器学习无处不在

- 图像识别
- 语音识别
- 股票预测
- 医疗诊断
- 数据分析
- 机器人
- 更多……


White
Broken
White
Solid
Right
Departure

机器学习概述



Machine Learning

机器学习

Supervised Learning

受监督学习

对有标识的数据集进行训练，以实现对新数据的分类或预测

Classification

分类

Support Vector Machines	Discriminant Analysis	Naive Bayes	Nearest Neighbor
支持向量机	判别分析	朴素贝叶斯	最近邻

Regression

回归

Linear Regression GLM	SVR, GPR	Ensemble Methods	Decision Trees	Neural Networks
线性回归	支持向量回归	组合方法	决策树	神经网络

Unsupervised Learning

无监督学习

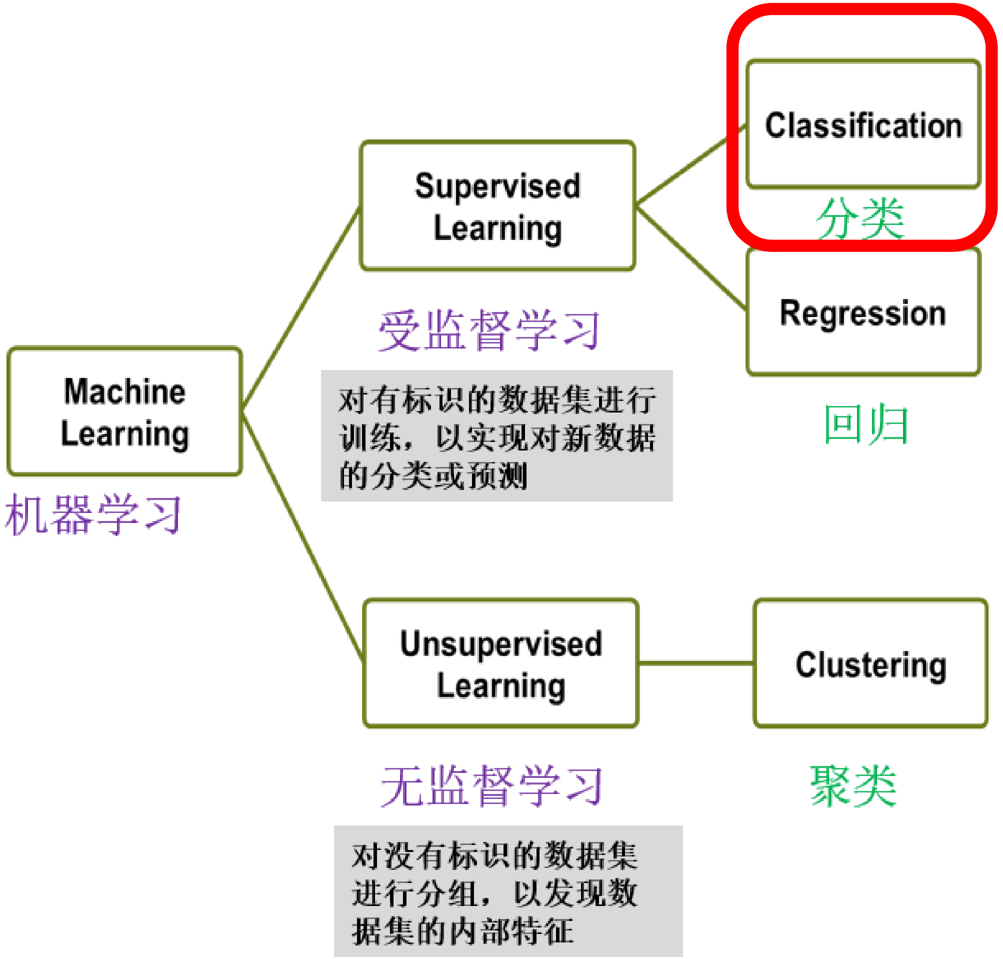
对没有标识的数据集进行分组，以发现数据集的内部特征

Clustering

聚类

kMeans, kmedoids Fuzzy C-Means	Hierarchical	Gaussian Mixture	Neural Networks	Hidden Markov Model
K均值	分层聚类	高斯混合模型	神经网络	隐马尔科夫模型

机器学习：分类



Machine Learning

机器学习

Supervised Learning

受监督学习

对有标识的数据集进行训练，以实现对新数据的分类或预测

Classification

分类

Regression

回归

Unsupervised Learning

无监督学习

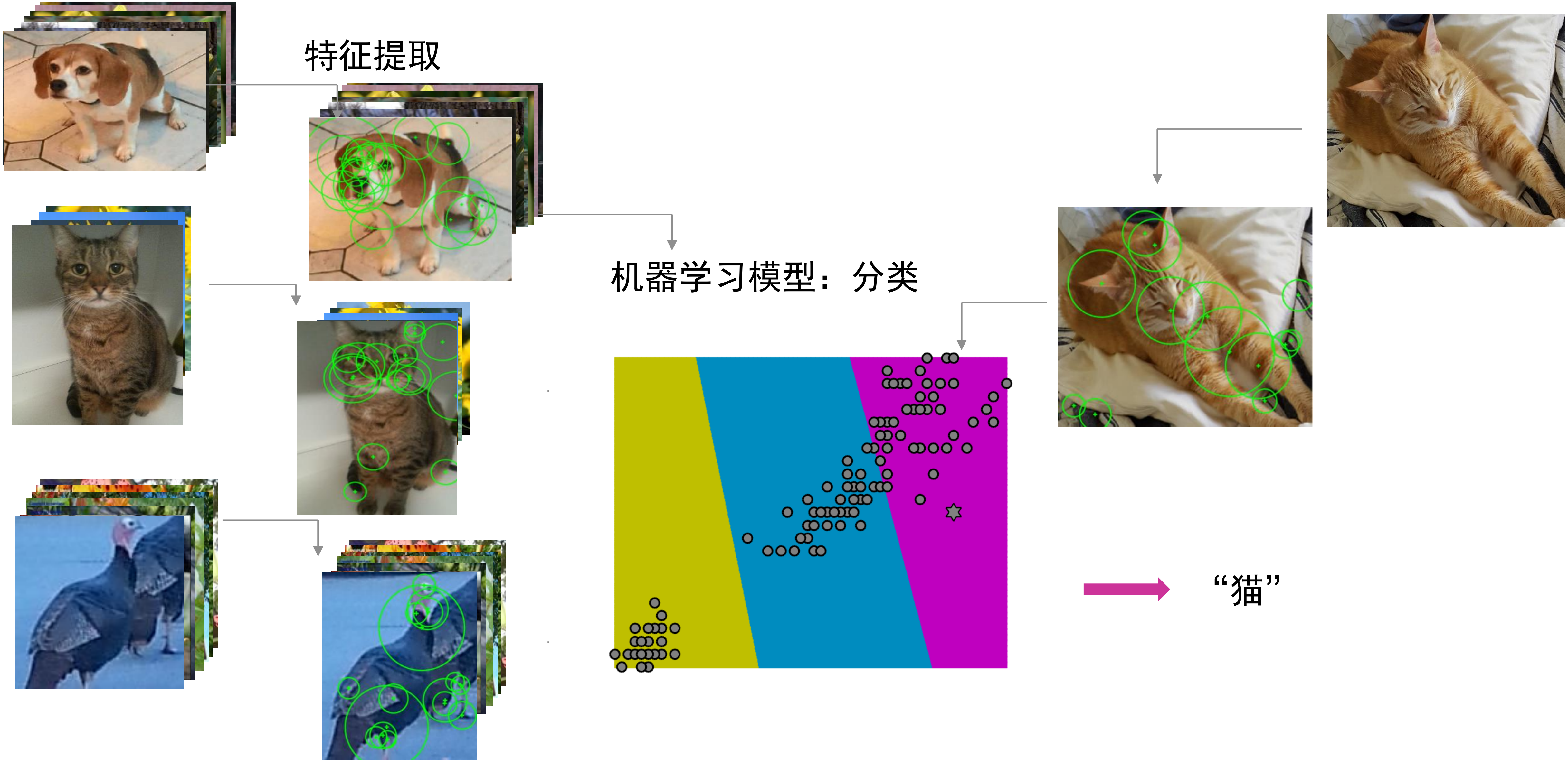
对没有标识的数据集进行分组，以发现数据集的内部特征

Clustering

聚类

目标:
利用有标签的数据训练一个分类模型。

机器学习流程



训练数据

特征提取

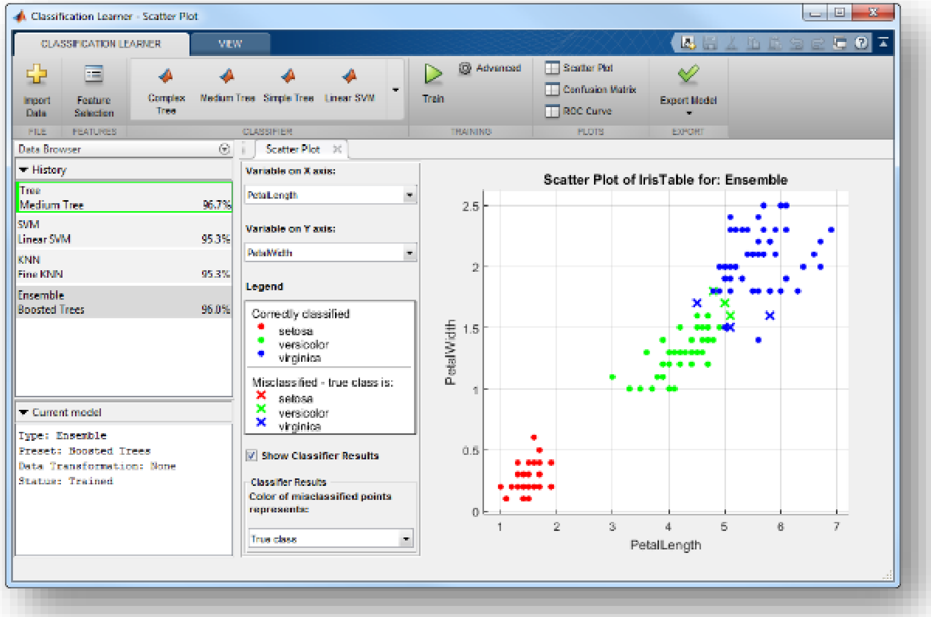
机器学习模型：分类

测试数据

“猫”

统计与机器学习工具箱

- 高效、易用
  - **Classification Learner app**
  - **Regression Learner app**
  - 并行计算选项
- 额外算法
  - Multiclass SVM, kmediods clustering
- 支持大数据 Tall Array
- 支持代码生成


Classification Learner - Scatter Plot
CLASSIFICATION LEARNER
VIEW
Import Data
Feature Selection
Complex Tree
Medium Tree
Simple Tree
Linear SVM
Train
Advanced
Scatter Plot
Confusion Matrix
ROC Curve
Export Model
FILE
FEATURES
CLASSIFIER
TRAINING
PLOTS
EXPORT
Data Browser
History
Scatter Plot
Variable on X axis:
PetalLength
Variable on Y axis:
PetalWidth
Legend
Correctly classified
selosa
versicolor
virginica
Misclassified - true class is:
selosa
versicolor
virginica
☑ Show Classifier Results
Classifier Results
Color of misclassified points represents:
True class
Scatter Plot of IrisTable for: Ensemble
PetalWidth
PetalLength
0
0.5
1
1.5
2
2.5
1
2
3
4
5
6
7

Tree
Medium Tree 96.7%
SVM
Linear SVM 95.3%
KNN
Fine KNN 95.3%
Ensemble
Boosted Trees 96.0%

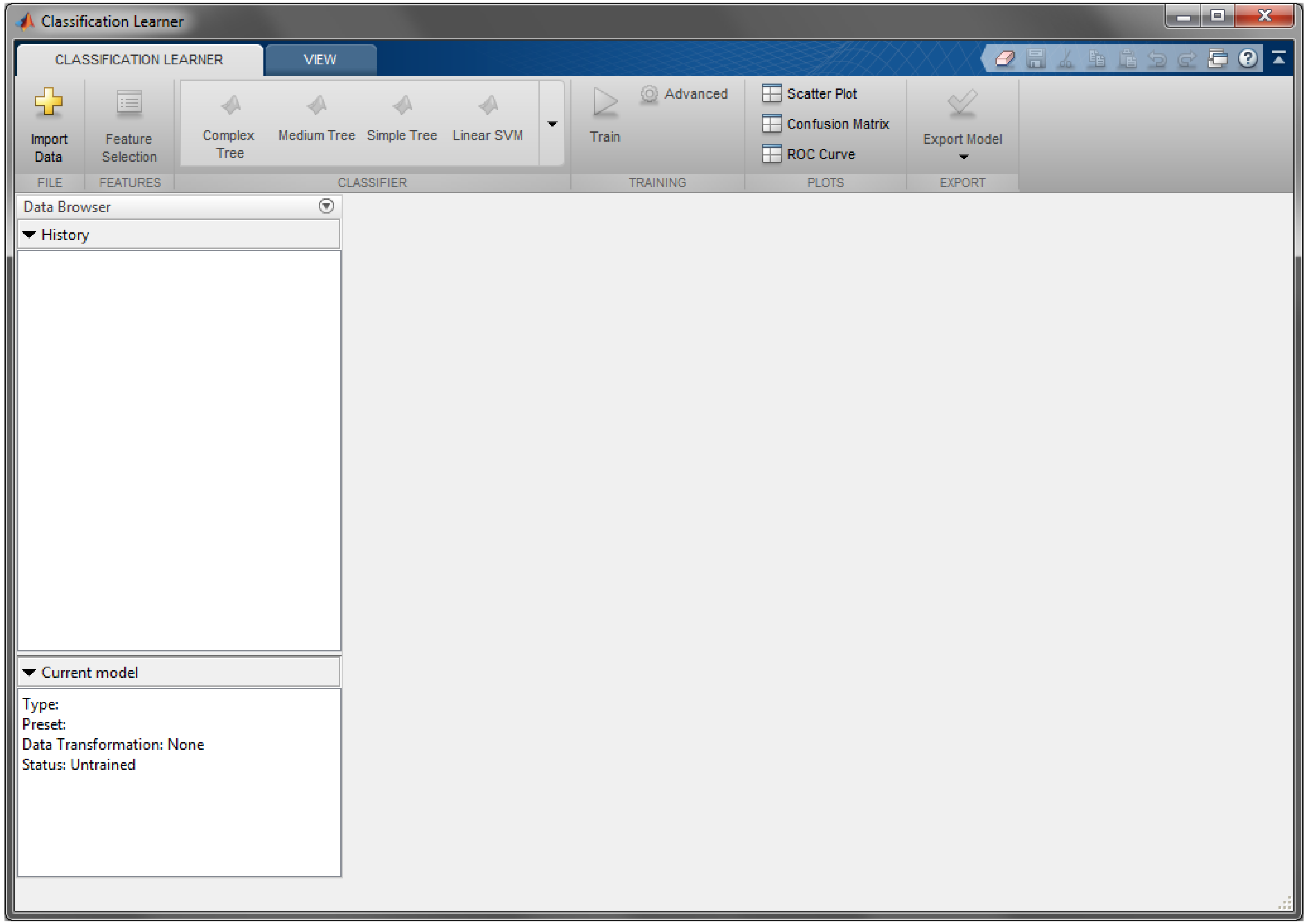
Current model

```
Type: Ensemble
Preset: Boosted Trees
Data Transformation: None
Status: Trained
```

> ***“如果没有这个 图形App，我不会想要尝试机器学习”***

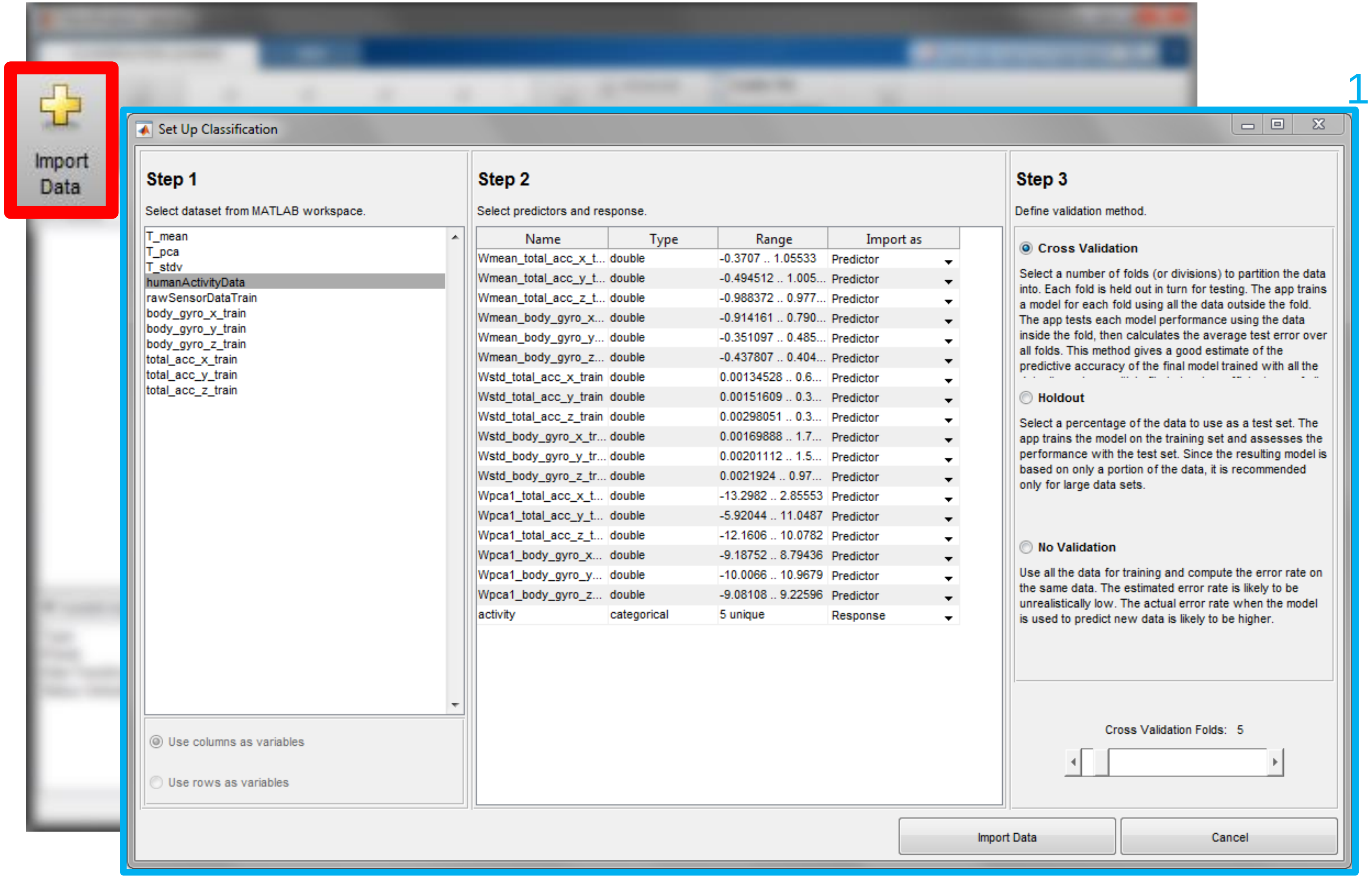
Demo: 鸢尾花分类

使用 Classification Learner App 训练一个模型


Classification Learner
CLASSIFICATION LEARNER
VIEW
Import Data
Feature Selection
Complex Tree
Medium Tree
Simple Tree
Linear SVM
Train
Advanced
Scatter Plot
Confusion Matrix
ROC Curve
Export Model
FILE
FEATURES
CLASSIFIER
TRAINING
PLOTS
EXPORT
Data Browser
History
Current model

Type:
Preset:
Data Transformation: None
Status: Untrained

使用 Classification Learner App 训练一个模型



Step 1

Select dataset from MATLAB workspace.

- T_mean
- T_pca
- T_stdv
- humanActivityData
- rawSensorDataTrain
- body_gyro_x_train
- body_gyro_y_train
- body_gyro_z_train
- total_acc_x_train
- total_acc_y_train
- total_acc_z_train

● Use columns as variables

○ Use rows as variables

Step 2

Select predictors and response.

Name	Type	Range	Import as
Wmean_total_acc_x_t...	double	-0.3707 .. 1.05533	Predictor
Wmean_total_acc_y_t...	double	-0.494512 .. 1.005...	Predictor
Wmean_total_acc_z_t...	double	-0.988372 .. 0.977...	Predictor
Wmean_body_gyro_x...	double	-0.914161 .. 0.790...	Predictor
Wmean_body_gyro_y...	double	-0.351097 .. 0.485...	Predictor
Wmean_body_gyro_z...	double	-0.437807 .. 0.404...	Predictor
Wstd_total_acc_x_train	double	0.00134528 .. 0.6...	Predictor
Wstd_total_acc_y_train	double	0.00151609 .. 0.3...	Predictor
Wstd_total_acc_z_train	double	0.00298051 .. 0.3...	Predictor
Wstd_body_gyro_x_tr...	double	0.00169888 .. 1.7...	Predictor
Wstd_body_gyro_y_tr...	double	0.00201112 .. 1.5...	Predictor
Wstd_body_gyro_z_tr...	double	0.0021924 .. 0.97...	Predictor
Wpca1_total_acc_x_t...	double	-13.2982 .. 2.85553	Predictor
Wpca1_total_acc_y_t...	double	-5.92044 .. 11.0487	Predictor
Wpca1_total_acc_z_t...	double	-12.1606 .. 10.0782	Predictor
Wpca1_body_gyro_x...	double	-9.18752 .. 8.79436	Predictor
Wpca1_body_gyro_y...	double	-10.0066 .. 10.9679	Predictor
Wpca1_body_gyro_z...	double	-9.08108 .. 9.22596	Predictor
activity	categorical	5 unique	Response

Step 3

Define validation method.

● **Cross Validation**

Select a number of folds (or divisions) to partition the data into. Each fold is held out in turn for testing. The app trains a model for each fold using all the data outside the fold. The app tests each model performance using the data inside the fold, then calculates the average test error over all folds. This method gives a good estimate of the predictive accuracy of the final model trained with all the

○ **Holdout**

Select a percentage of the data to use as a test set. The app trains the model on the training set and assesses the performance with the test set. Since the resulting model is based on only a portion of the data, it is recommended only for large data sets.

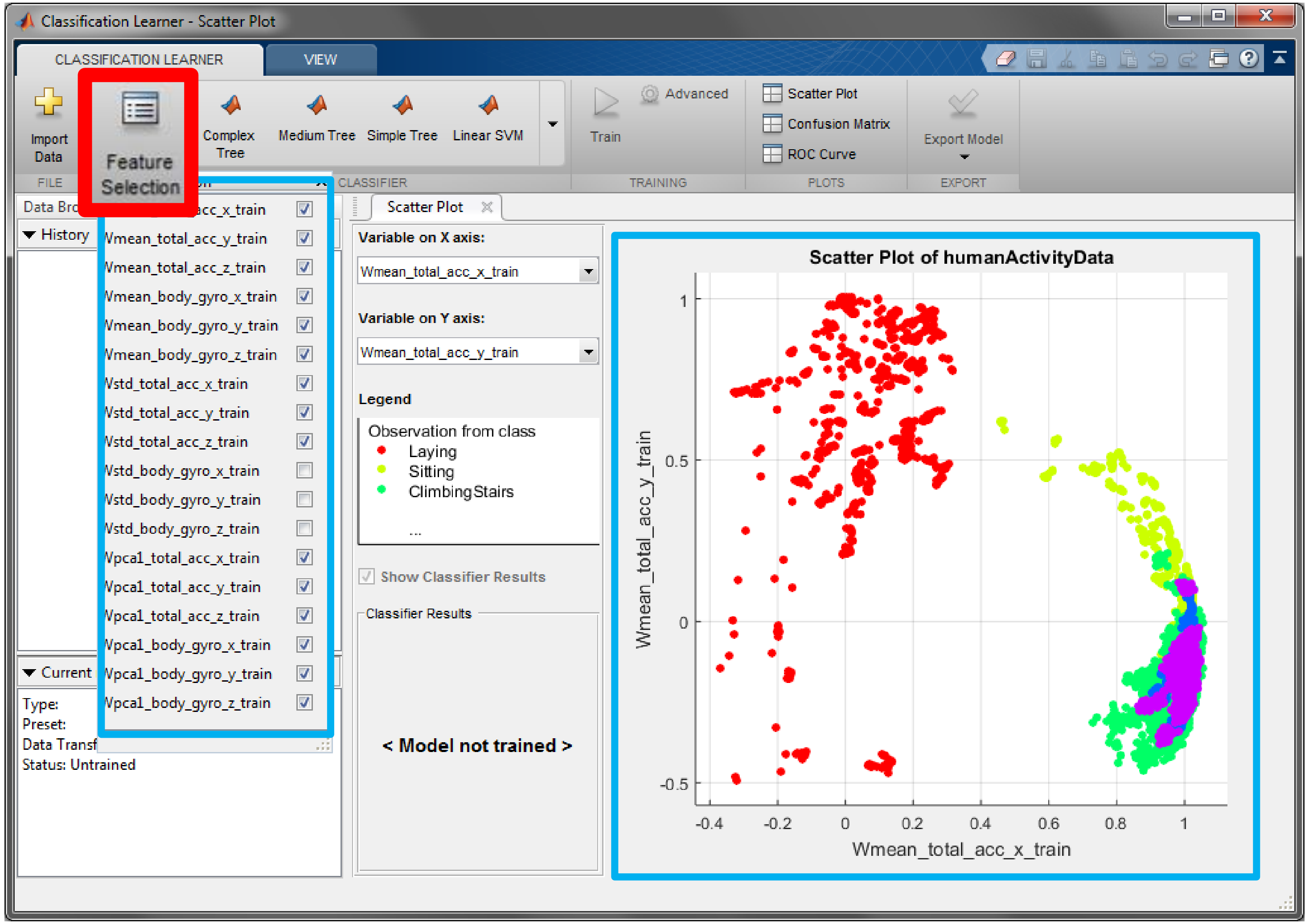
○ **No Validation**

Use all the data for training and compute the error rate on the same data. The estimated error rate is likely to be unrealistically low. The actual error rate when the model is used to predict new data is likely to be higher.

Cross Validation Folds: 5

1. 导入数据并设置 Cross-validation

使用 Classification Learner App 训练一个模型


Classification Learner - Scatter Plot
CLASSIFICATION LEARNER
VIEW
Import Data
Feature Selection
Complex Tree
Medium Tree
Simple Tree
Linear SVM
Train
Advanced
Scatter Plot
Confusion Matrix
ROC Curve
Export Model
FILE
CLASSIFIER
TRAINING
PLOTS
EXPORT
Data Bro
History
Current
Scatter Plot
Scatter Plot of humanActivityData
Wmean_total_acc_y_train
Wmean_total_acc_x_train
1
0.5
0
-0.5
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

acc_x_train ☑
Wmean_total_acc_y_train ☑
Vmean_total_acc_z_train ☑
Vmean_body_gyro_x_train ☑
Vmean_body_gyro_y_train ☑
Vmean_body_gyro_z_train ☑
Vstd_total_acc_x_train ☑
Vstd_total_acc_y_train ☑
Vstd_total_acc_z_train ☑
Vstd_body_gyro_x_train ☐
Vstd_body_gyro_y_train ☐
Vstd_body_gyro_z_train ☐
Vpca1_total_acc_x_train ☑
Vpca1_total_acc_y_train ☑
Vpca1_total_acc_z_train ☑
Vpca1_body_gyro_x_train ☑
Vpca1_body_gyro_y_train ☑
Vpca1_body_gyro_z_train ☑

Type:
Preset:
Data Transf
Status: Untrained

Variable on X axis: Wmean_total_acc_x_train

Variable on Y axis: Wmean_total_acc_y_train

Legend

Observation from class
● Laying
● Sitting
● ClimbingStairs
...

☑ Show Classifier Results

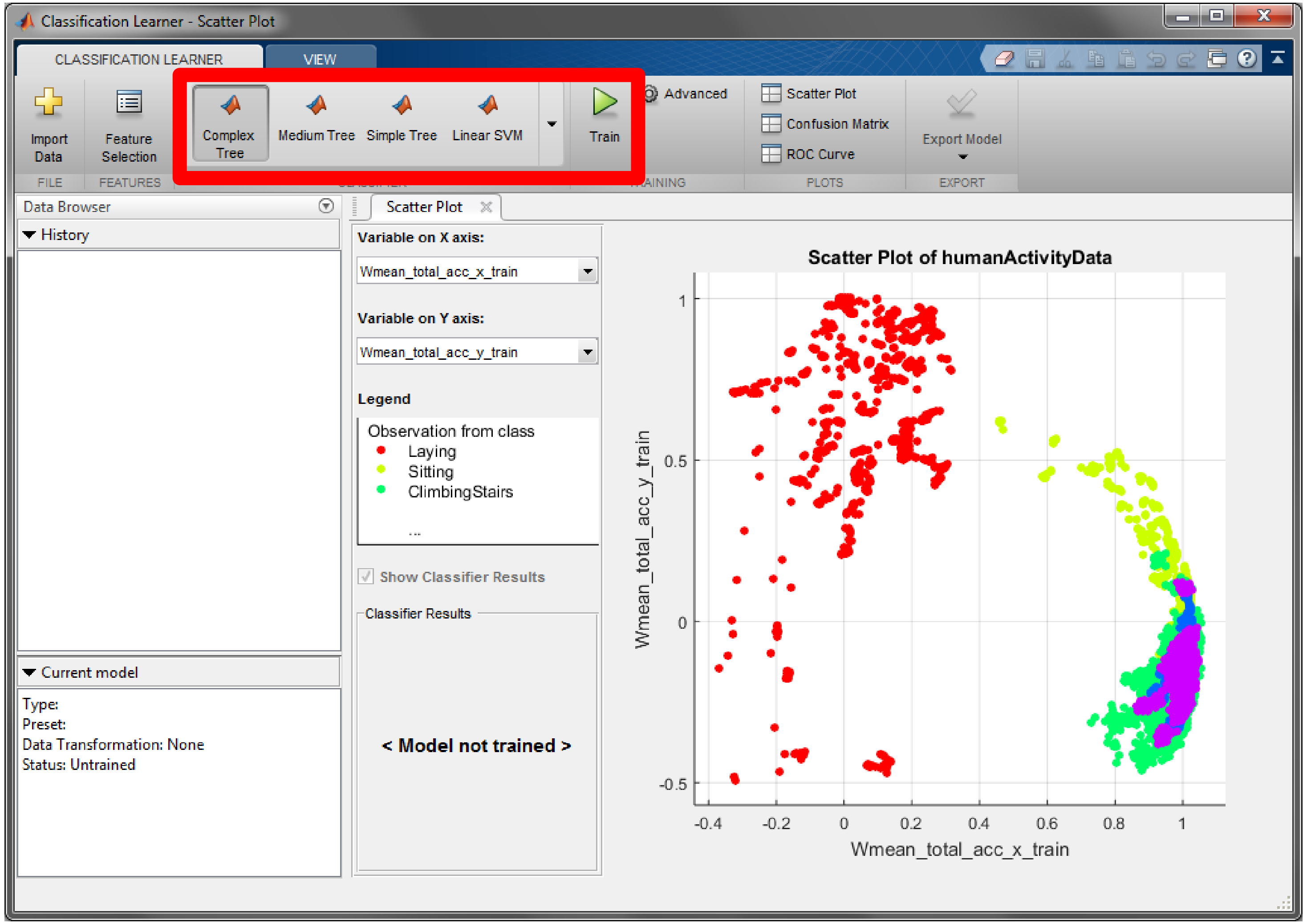
Classifier Results

< Model not trained >

1. 导入数据并设置 Cross-validation

2. 数据探索与特征选择

使用 Classification Learner App 训练一个模型


Classification Learner - Scatter Plot
CLASSIFICATION LEARNER
VIEW
Import Data
Feature Selection
Complex Tree
Medium Tree
Simple Tree
Linear SVM
Train
Advanced
Scatter Plot
Confusion Matrix
ROC Curve
Export Model
FILE
FEATURES
TRAINING
PLOTS
EXPORT
Data Browser
History
Current model
Scatter Plot
Legend
Observation from class
Laying
Sitting
ClimbingStairs
...
Show Classifier Results
Classifier Results
Scatter Plot of humanActivityData
Wmean_total_acc_y_train
1
0.5
0
-0.5
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
Wmean_total_acc_x_train

Variable on X axis: Wmean_total_acc_x_train

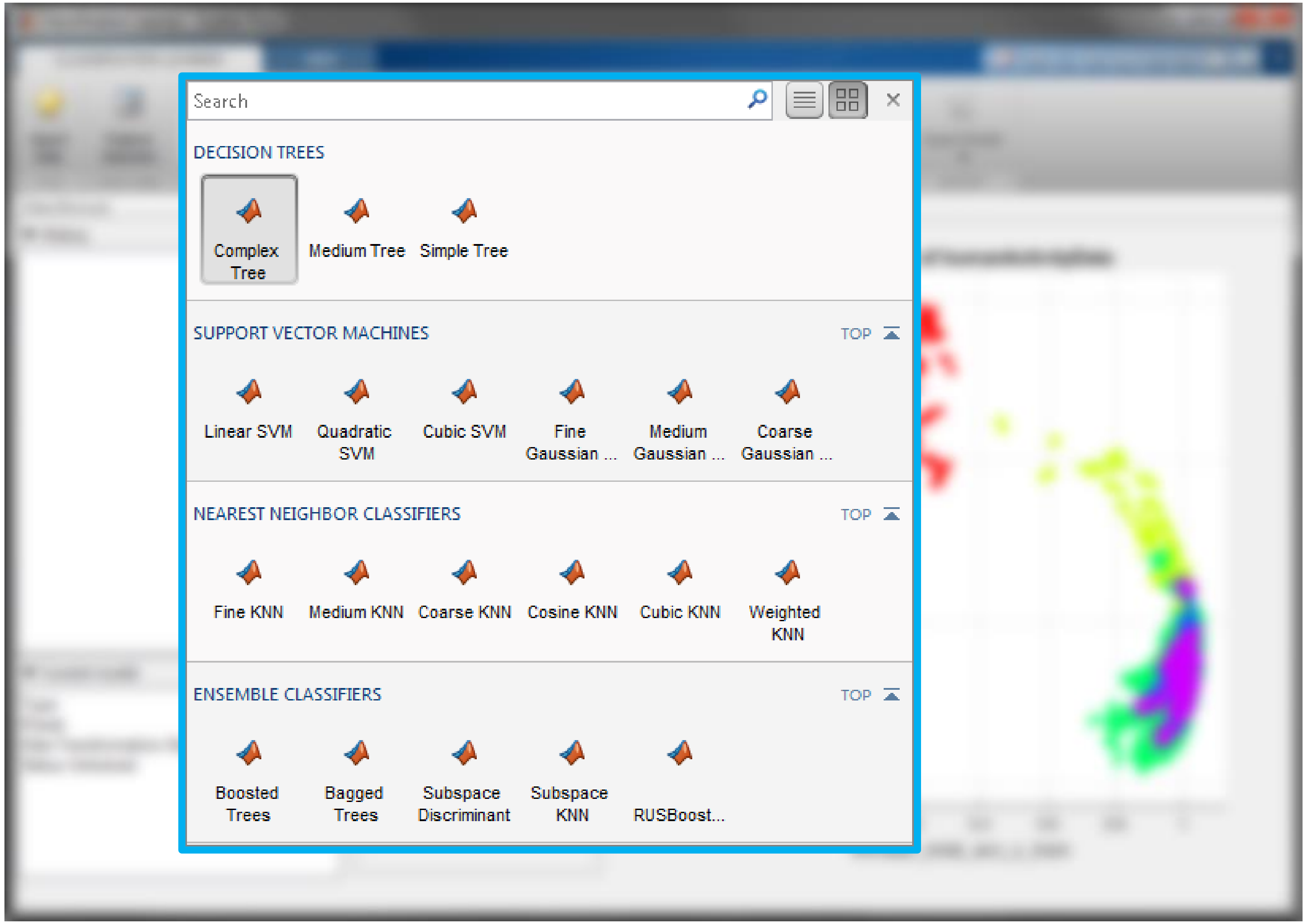
Variable on Y axis: Wmean_total_acc_y_train

Type:
Preset:
Data Transformation: None
Status: Untrained

< Model not trained >

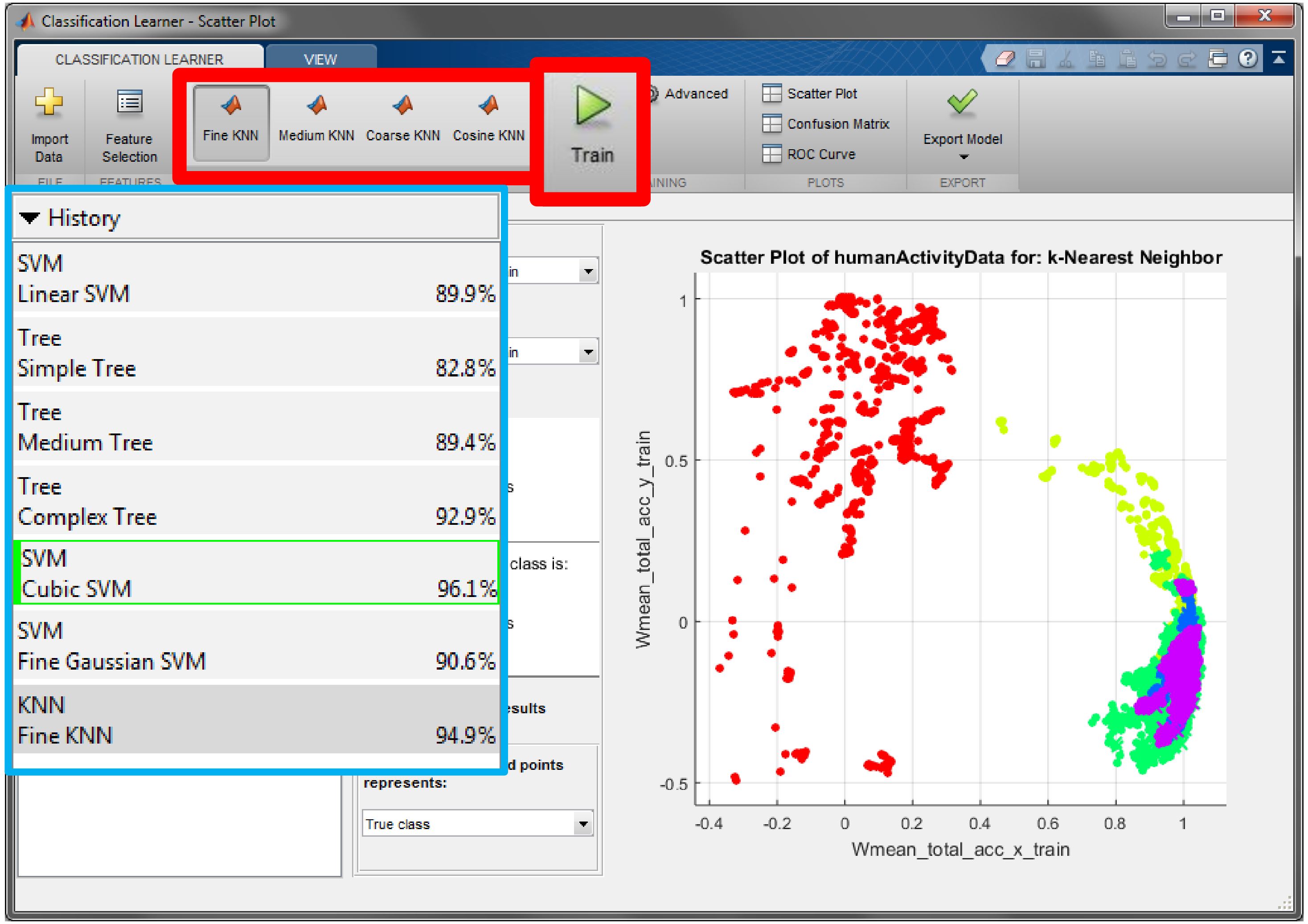
1. 导入数据并设置 Cross-validation
2. 数据探索与特征选择
3. 训练多个模型

使用 Classification Learner App 训练一个模型


Search
DECISION TREES
Complex Tree
Medium Tree
Simple Tree
SUPPORT VECTOR MACHINES
TOP
Linear SVM
Quadratic SVM
Cubic SVM
Fine Gaussian ...
Medium Gaussian ...
Coarse Gaussian ...
NEAREST NEIGHBOR CLASSIFIERS
TOP
Fine KNN
Medium KNN
Coarse KNN
Cosine KNN
Cubic KNN
Weighted KNN
ENSEMBLE CLASSIFIERS
TOP
Boosted Trees
Bagged Trees
Subspace Discriminant
Subspace KNN
RUSBoost...

1. 导入数据并设置 Cross-validation
2. 数据探索与特征选择
3. 训练多个模型

使用 Classification Learner App 训练一个模型

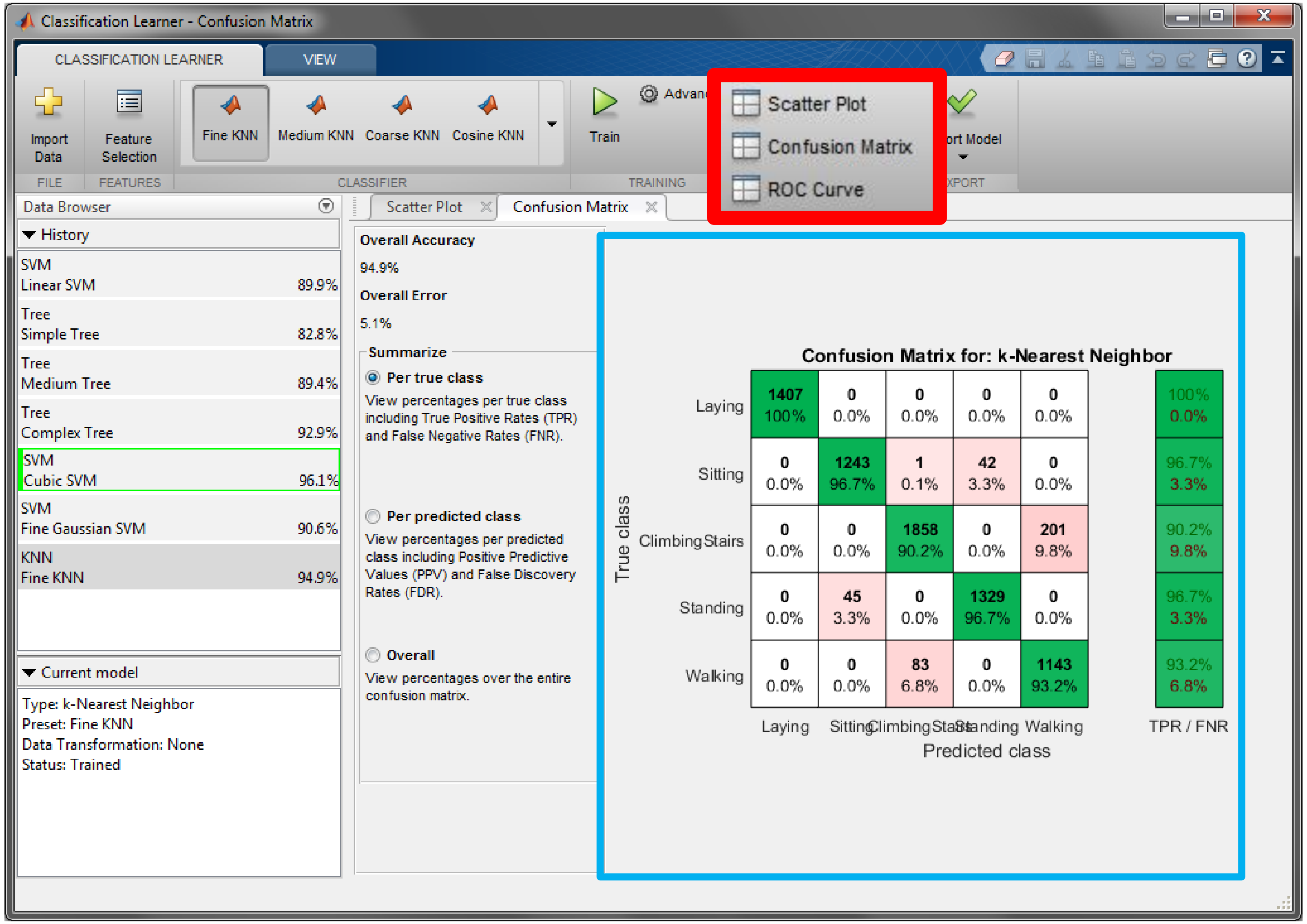


Classification Learner - Scatter Plot
CLASSIFICATION LEARNER
VIEW
Import Data
Feature Selection
Fine KNN
Medium KNN
Coarse KNN
Cosine KNN
Train
Advanced
Scatter Plot
Confusion Matrix
ROC Curve
Export Model
FILE
FEATURES
PLOTS
EXPORT
History
Scatter Plot of humanActivityData for: k-Nearest Neighbor
Wmean_total_acc_y_train
Wmean_total_acc_x_train
1
0.5
0
-0.5
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
True class

SVM Linear SVM	89.9%
Tree Simple Tree	82.8%
Tree Medium Tree	89.4%
Tree Complex Tree	92.9%
SVM Cubic SVM	96.1%
SVM Fine Gaussian SVM	90.6%
KNN Fine KNN	94.9%

1. 导入数据并设置 Cross-validation
2. 数据探索与特征选择
3. 训练多个模型

使用 Classification Learner App 训练一个模型


Classification Learner - Confusion Matrix
CLASSIFICATION LEARNER
VIEW
Import Data
Feature Selection
Fine KNN
Medium KNN
Coarse KNN
Cosine KNN
Train
Advan
Scatter Plot
Confusion Matrix
ROC Curve
ort Model
FILE
FEATURES
CLASSIFIER
TRAINING
XPORT
Data Browser
Scatter Plot
Confusion Matrix

History

SVM Linear SVM	89.9%
Tree Simple Tree	82.8%
Tree Medium Tree	89.4%
Tree Complex Tree	92.9%
SVM Cubic SVM	96.1%
SVM Fine Gaussian SVM	90.6%
KNN Fine KNN	94.9%

Current model

Type: k-Nearest Neighbor
Preset: Fine KNN
Data Transformation: None
Status: Trained

Overall Accuracy

94.9%

Overall Error

5.1%

Summarize

● **Per true class**

View percentages per true class including True Positive Rates (TPR) and False Negative Rates (FNR).

○ **Per predicted class**

View percentages per predicted class including Positive Predictive Values (PPV) and False Discovery Rates (FDR).

○ **Overall**

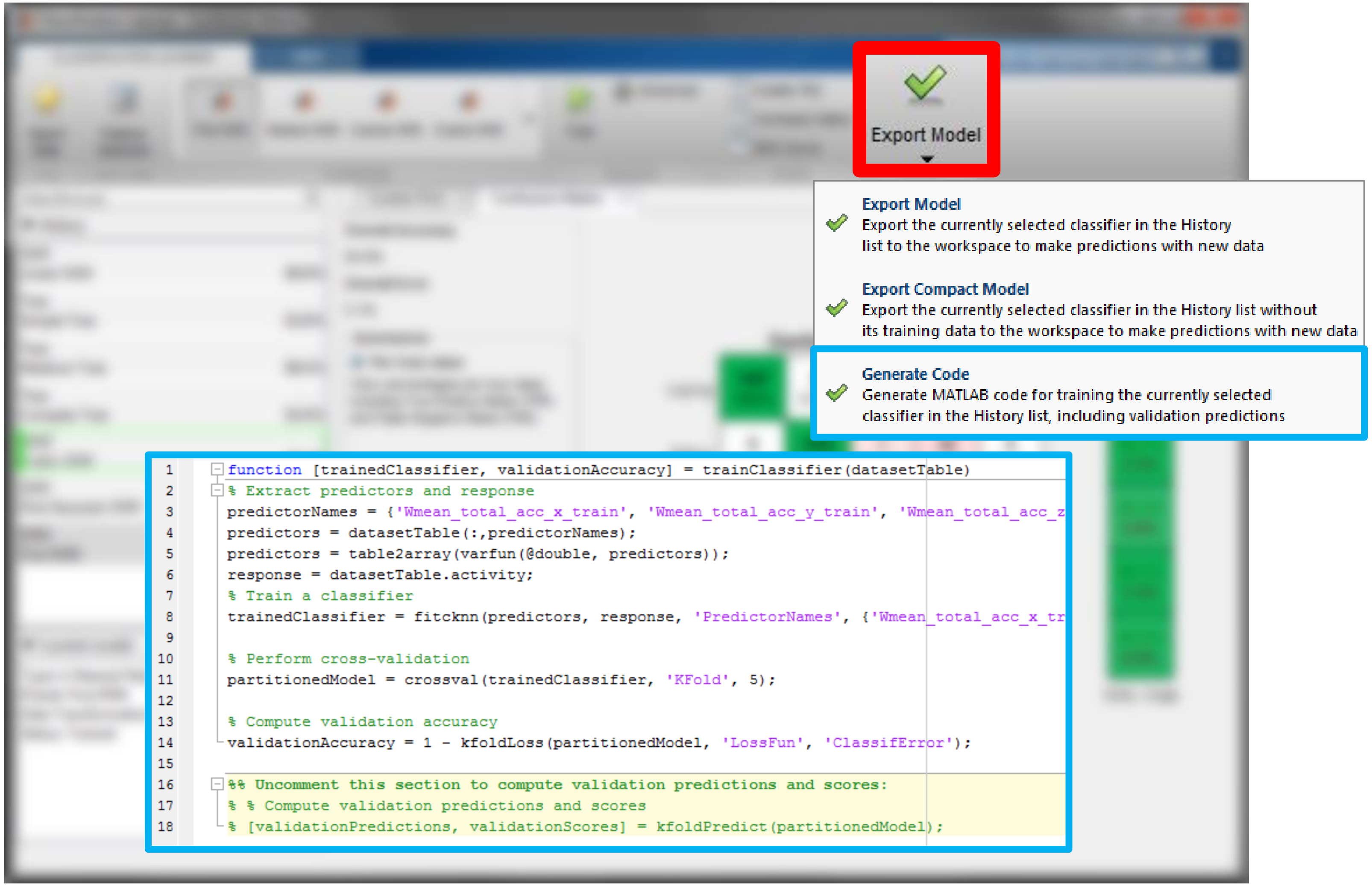
View percentages over the entire confusion matrix.

Confusion Matrix for: k-Nearest Neighbor

True class \ Predicted class	Laying	Sitting	ClimbingStairs	Standing	Walking	TPR / FNR
Laying	1407 100%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
Sitting	0 0.0%	1243 96.7%	1 0.1%	42 3.3%	0 0.0%	96.7% 3.3%
ClimbingStairs	0 0.0%	0 0.0%	1858 90.2%	0 0.0%	201 9.8%	90.2% 9.8%
Standing	0 0.0%	45 3.3%	0 0.0%	1329 96.7%	0 0.0%	96.7% 3.3%
Walking	0 0.0%	0 0.0%	83 6.8%	0 0.0%	1143 93.2%	93.2% 6.8%

1. 导入数据并设置 Cross-validation
2. 数据探索与特征选择
3. 训练多个模型
4. 结果比较与评断

使用 Classification Learner App 训练一个模型



Export Model
Export the currently selected classifier in the History list to the workspace to make predictions with new data

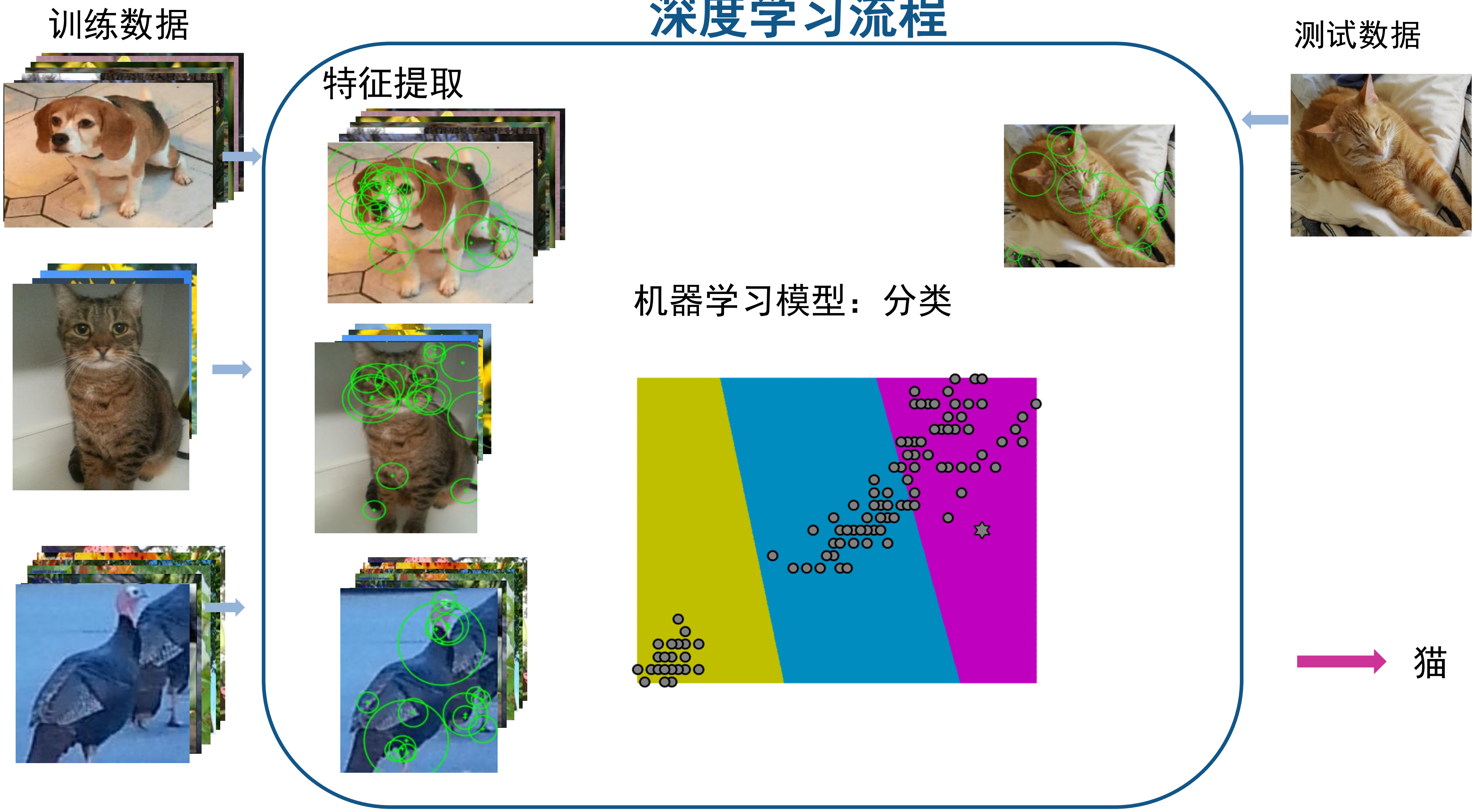
Export Compact Model
Export the currently selected classifier in the History list without its training data to the workspace to make predictions with new data

Generate Code
Generate MATLAB code for training the currently selected classifier in the History list, including validation predictions

```
function [trainedClassifier, validationAccuracy] = trainClassifier(datasetTable)
% Extract predictors and response
predictorNames = {'Wmean_total_acc_x_train', 'Wmean_total_acc_y_train', 'Wmean_total_acc_z
predictors = datasetTable(:,predictorNames);
predictors = table2array(varfun(@double, predictors));
response = datasetTable.activity;
% Train a classifier
trainedClassifier = fitcknn(predictors, response, 'PredictorNames', {'Wmean_total_acc_x_tr

% Perform cross-validation
partitionedModel = crossval(trainedClassifier, 'KFold', 5);

% Compute validation accuracy
validationAccuracy = 1 - kfoldLoss(partitionedModel, 'LossFun', 'ClassifError');

%% Uncomment this section to compute validation predictions and scores:
% % Compute validation predictions and scores
% [validationPredictions, validationScores] = kfoldPredict(partitionedModel);
```

1. 导入数据并设置 Cross-validation
2. 数据探索与特征选择
3. 训练多个模型
4. 结果比较与评断
5. 模型分享或自动代码转换

深度学习流程



训练数据

特征提取

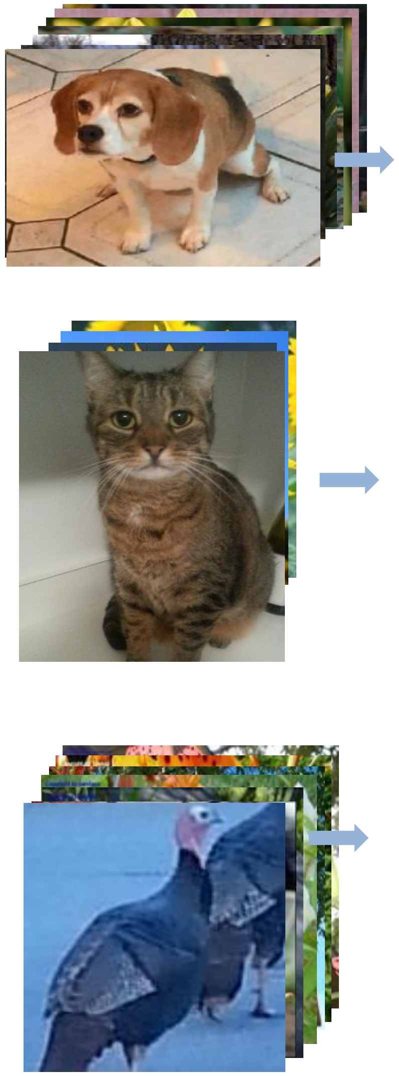
机器学习模型：分类

测试数据

猫

深度学习流程

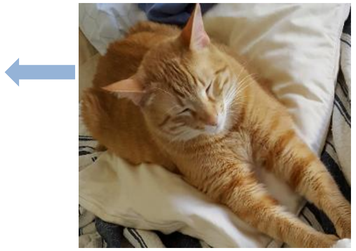
训练数据



卷积神经网络 (CNN)

特征学习与分类

测试数据



猫

10行 MATLAB 代码实现深度学习

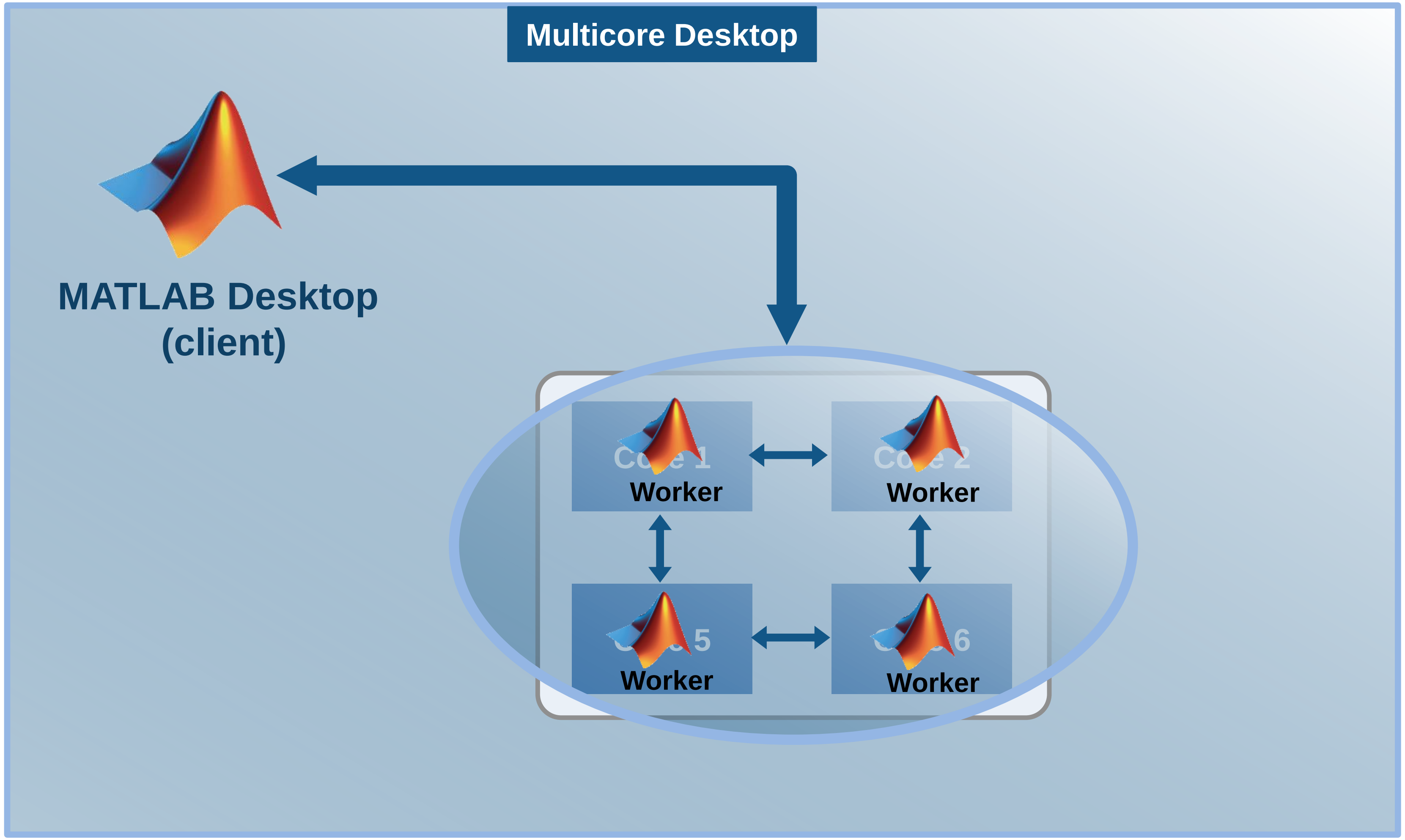
```
camera = webcam;  % Connect to the camera
nnet = alexnet;   % Load the neural net

while true
    picture = camera.snapshot;              % Take a picture
    picture = imresize(picture,[227,227]);  % Resize the picture

    label = classify(nnet, picture);     % Classify the picture

    image(picture);      % Show the picture
    title(char(label));  % Show the label
    drawnow;
end
```

MATLAB 并行计算的实际应用

- 为什么要使用 MATLAB 并行计算?
  - 利用更多硬件的计算能力
  - 加快工作流程，最小化，无需更改原始代码的代码
  - 专注于您的工程和研究，而不是计算

并行计算

多核电脑



Multicore Desktop

MATLAB Desktop
(client)

Worker

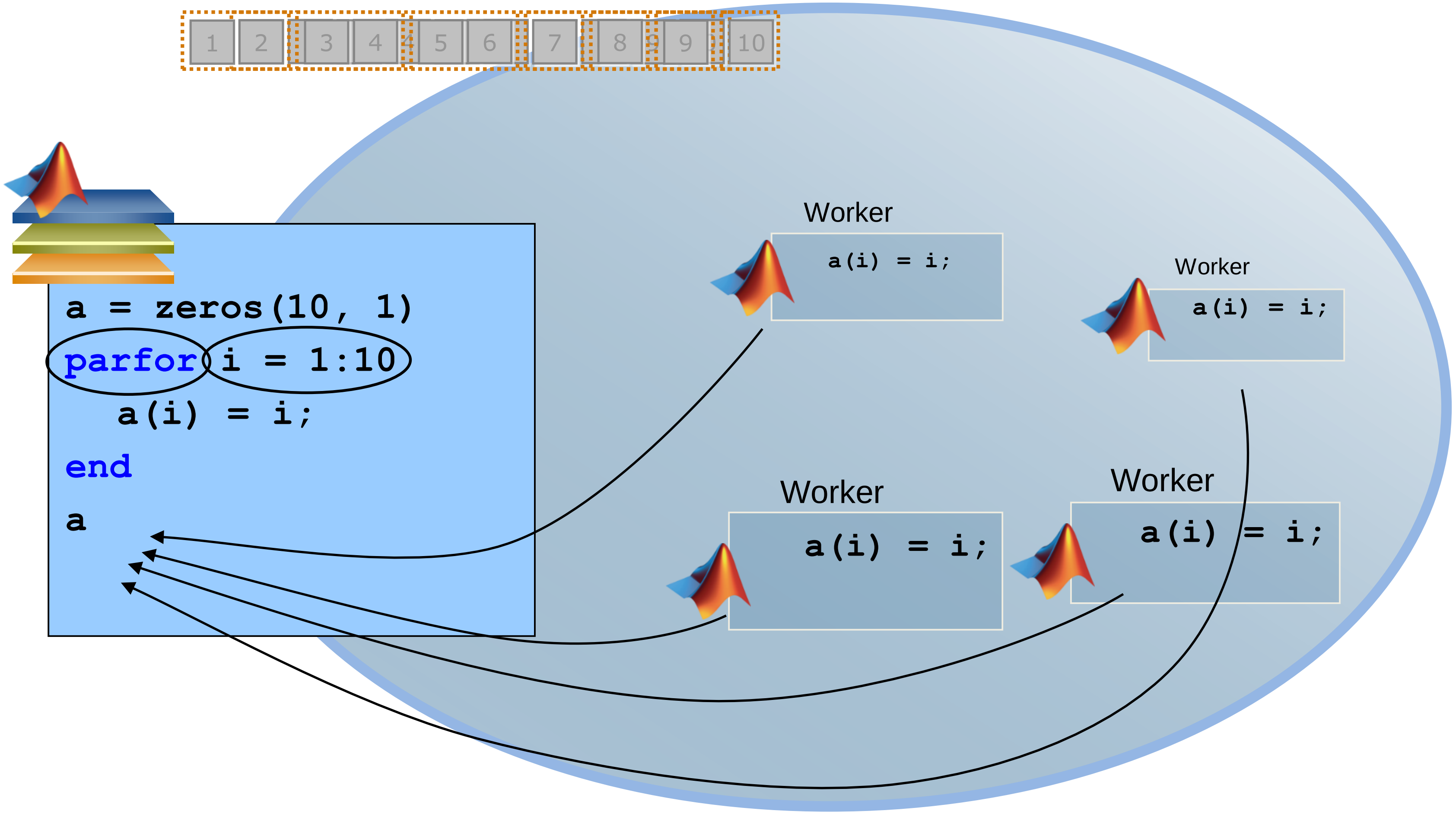
Worker

Worker

Worker

MATLAB multicore

Parfor 循环

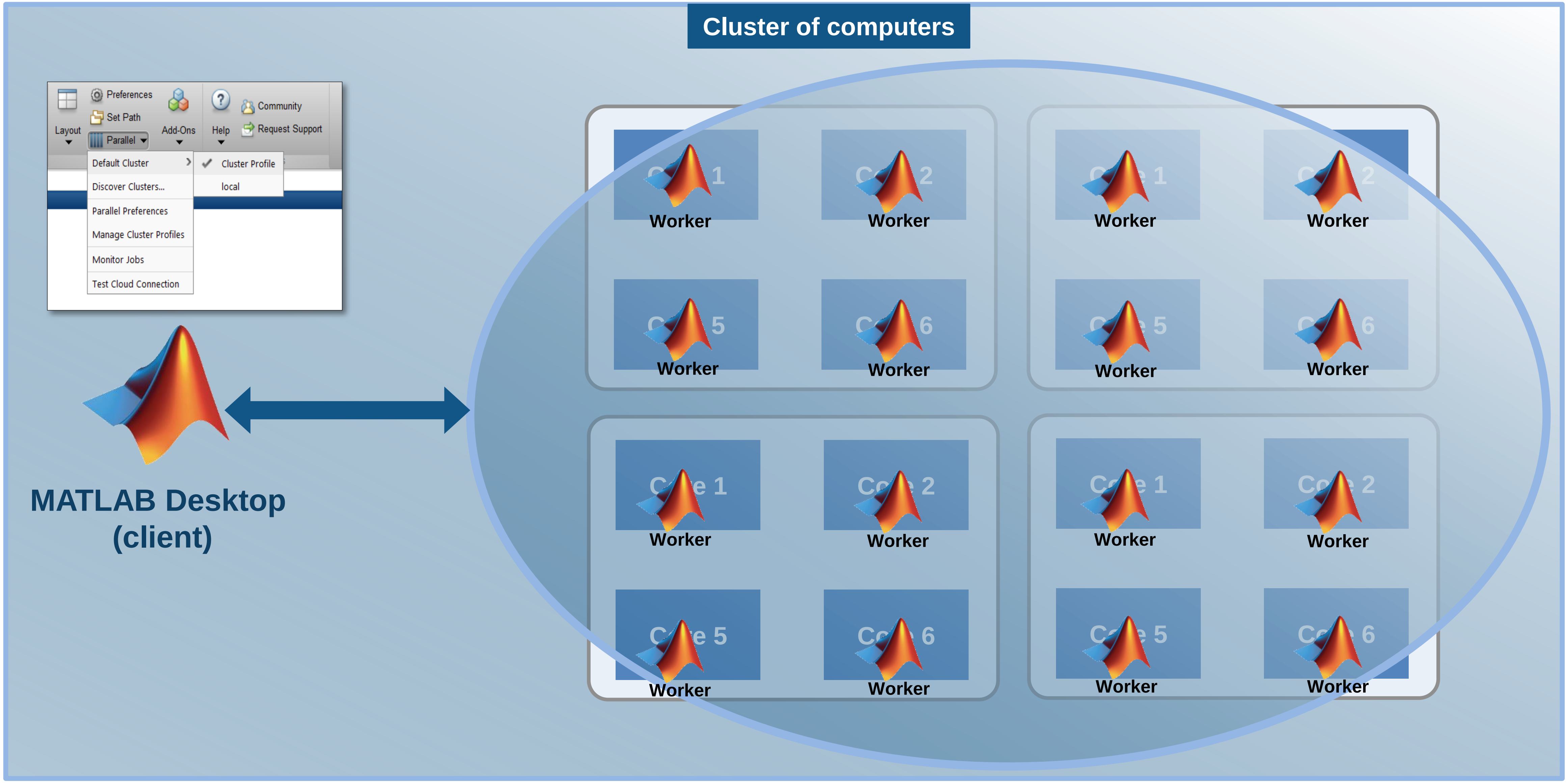


```
a = zeros(10, 1)
parfor i = 1:10
   a(i) = i;
end
a
```

1
2
3
4
5
6
7
8
9
10
Worker
a(i) = i;
Worker
a(i) = i;
Worker
a(i) = i;
Worker
a(i) = i;

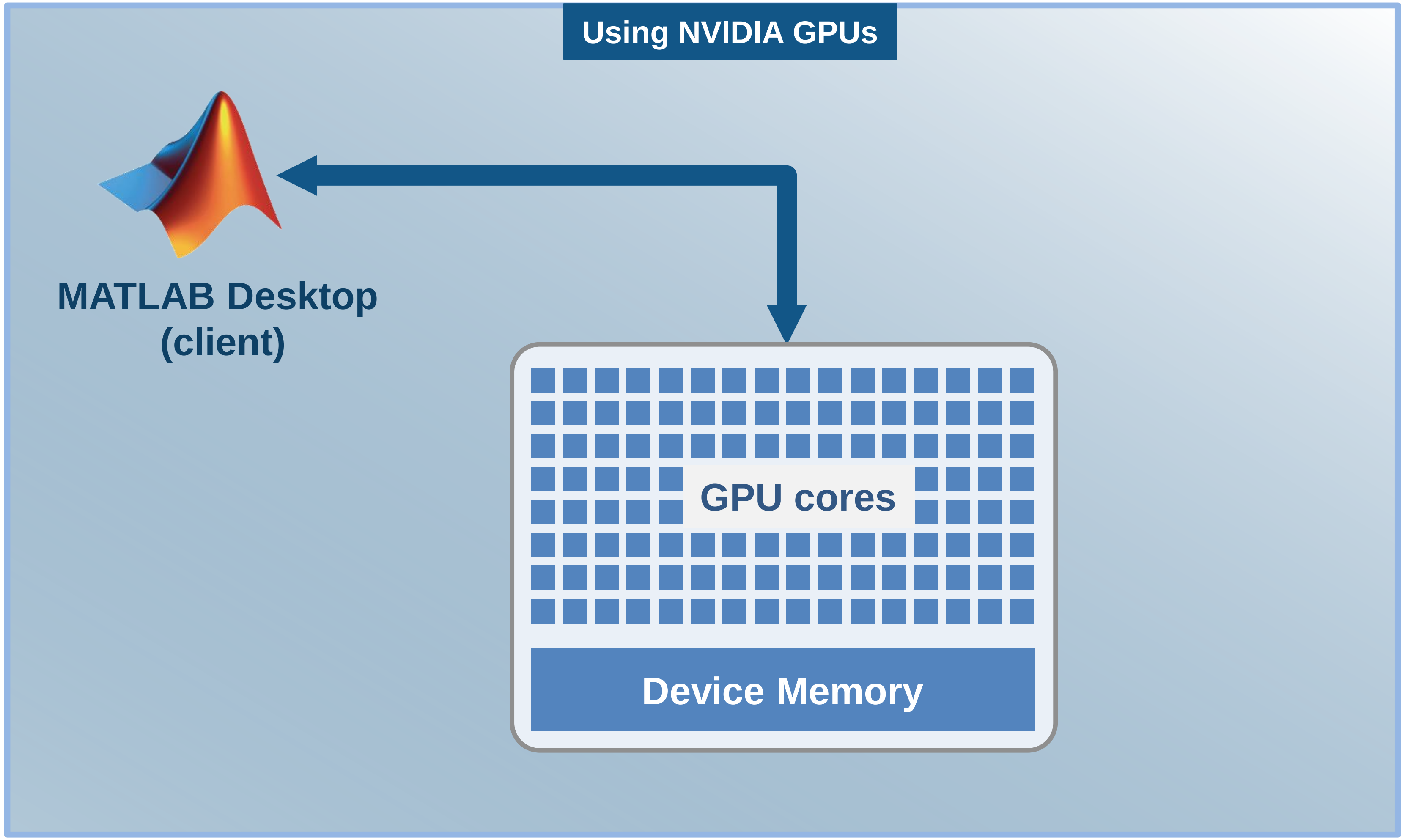
并行计算

集群


Cluster of computers
Layout
Preferences
Set Path
Parallel
Add-Ons
Help
Community
Request Support
Default Cluster
Discover Clusters...
Parallel Preferences
Manage Cluster Profiles
Monitor Jobs
Test Cloud Connection
Cluster Profile
local
MATLAB Desktop
(client)
Core 1
Worker
Core 2
Worker
Core 5
Worker
Core 6
Worker
Core 1
Worker
Core 2
Worker
Core 5
Worker
Core 6
Worker
Core 1
Worker
Core 2
Worker
Core 5
Worker
Core 6
Worker
Core 1
Worker
Core 2
Worker
Core 5
Worker
Core 6
Worker

并行计算

NVIDIA GPUs


Using NVIDIA GPUs
MATLAB Desktop
(client)
GPU cores
Device Memory

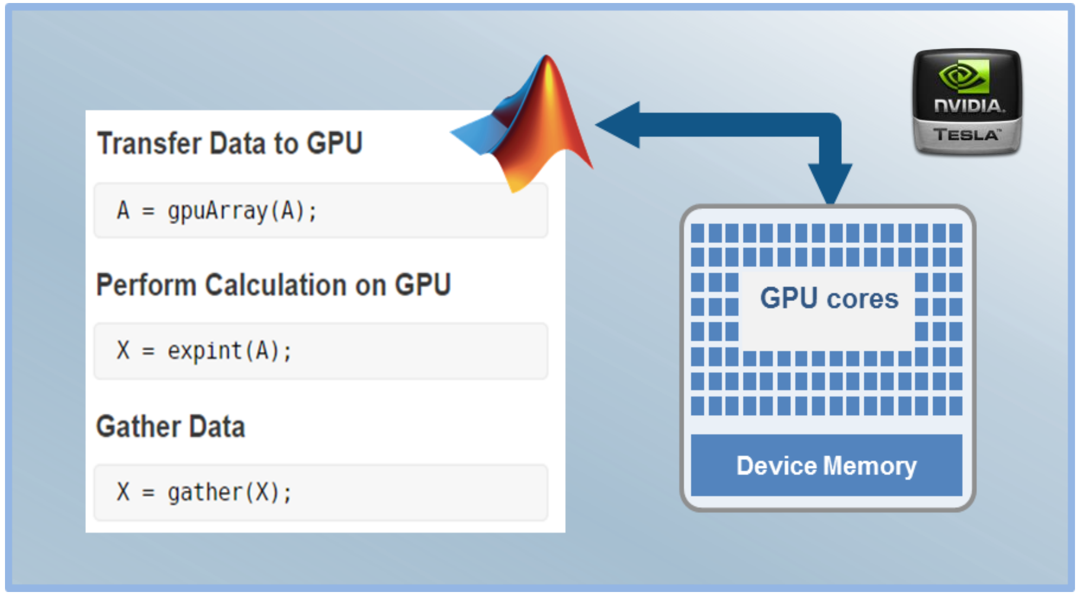
使用 NVIDIA GPU 加速 MATLAB 代码

- 理想问题
  - 大规模并行和/或矢量化操作
  - 计算密集型
  - 元素级运算

- 300+ 具有 GPU 功能的 MATLAB 函数

- 工具箱支持
  - Neural Networks
  - Image Processing
  - Signal Processing

..... Learn More



Transfer Data to GPU

```
A = gpuArray(A);
```

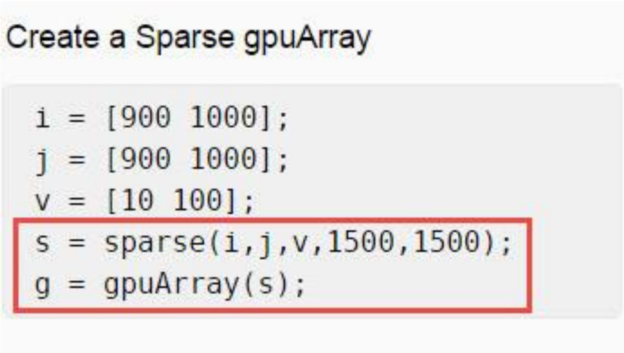
Perform Calculation on GPU

```
X = expint(A);
```

Gather Data

```
X = gather(X);
```

NVIDIA
TESLA
GPU cores
Device Memory



Create a Sparse gpuArray

```
i = [900 1000];
j = [900 1000];
v = [10 100];
s = sparse(i,j,v,1500,1500);
g = gpuArray(s);
```

支持并行计算工具箱 (MATLAB 家族)

通过设置标志或首选项来启用并行计算支持

Image Processing

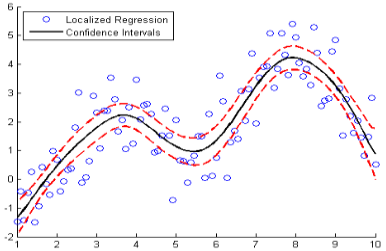
Batch Image Processor, Block Processing, GPU-enabled functions

Original Image of Peppers

Recolored Image of Peppers

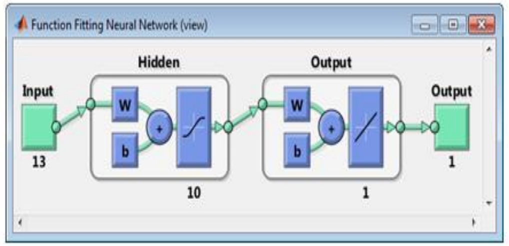
Statistics and Machine Learning

Resampling Methods, k-Means clustering, GPU-enabled functions



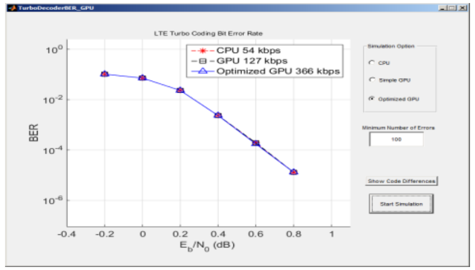
Neural Networks

Deep Learning, Neural Network training and simulation


Function Fitting Neural Network (view)
Hidden
Output
Input
Output
W
b
W
b
13
10
1
1

Signal Processing and Communications

GPU-enabled FFT filtering, cross correlation, BER

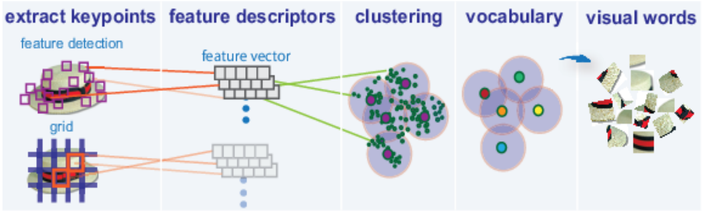

TurboDecoderBER_GPU
LTE Turbo Coding Bit Error Rate
CPU 54 kbps
GPU 127 kbps
Optimized GPU 366 kbps
BER
10^0
10^{-2}
10^{-4}
10^{-6}
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
E_b/N_0 (dB)
Show Code Differences
Start Simulation

Simulation Option
○ CPU
○ Simple GPU
● Optimized GPU

Minimum Number of Errors: 100

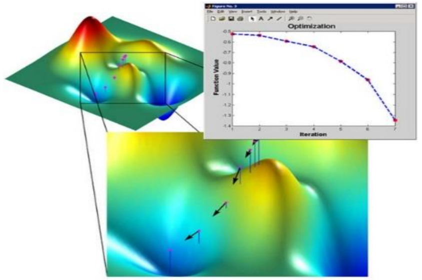
Computer Vision

Parallel-enabled functions in bag-of-words workflow


extract keypoints
feature detection
grid
feature descriptors
feature vector
clustering
vocabulary
visual words

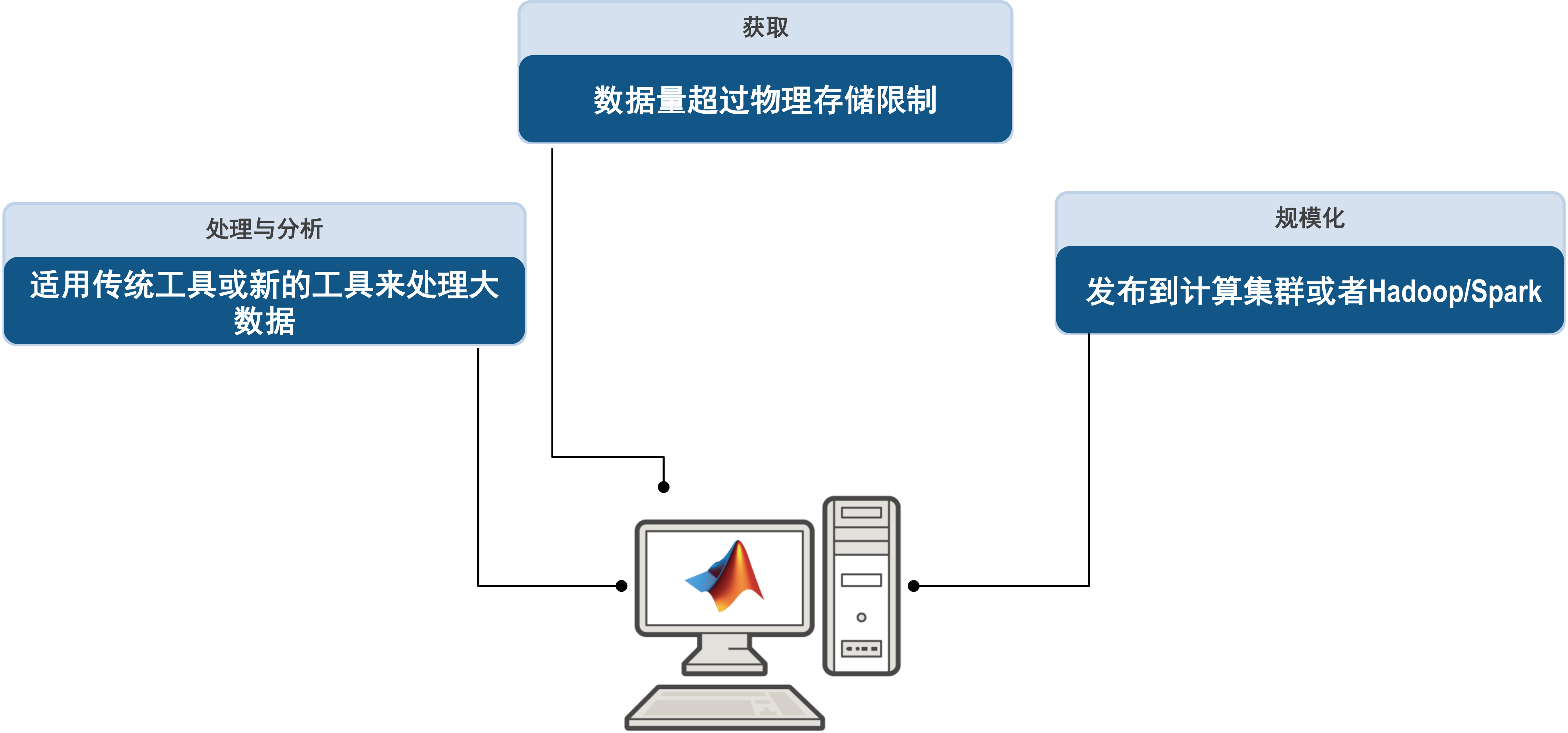
Optimization

Parallel estimation of gradients


Optimization
Function Value
Iteration

[Other parallel-enabled Toolboxes]()

MATLAB 的大数据处理能力



获取

数据量超过物理存储限制

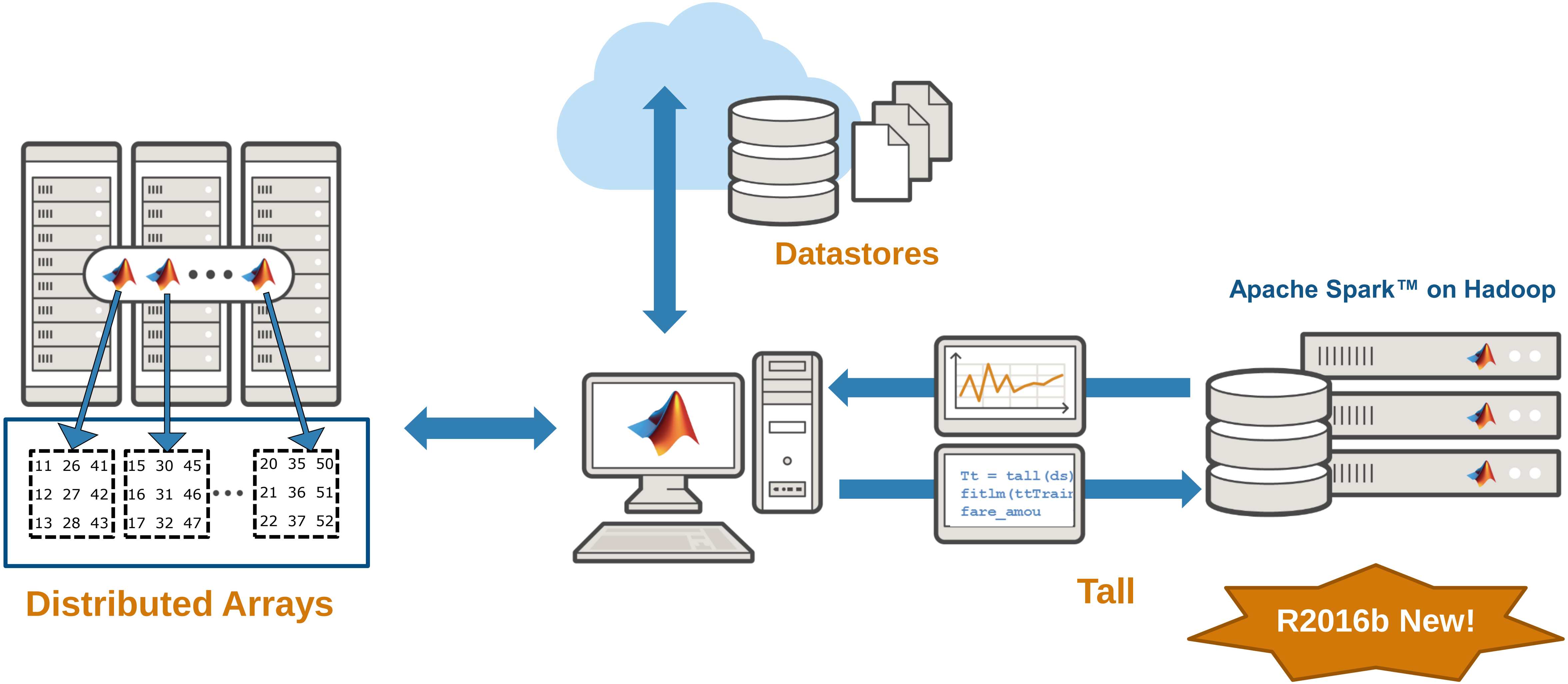
处理与分析

适用传统工具或新的工具来处理大数据

规模化

发布到计算集群或者Hadoop/Spark

MATLAB 的大数据处理能力


11 26 41
12 27 42
13 28 43
15 30 45
16 31 46
17 32 47
20 35 50
21 36 51
22 37 52

Datastores

Apache Spark™ on Hadoop

```
Tt = tall(ds)
fitlm(ttTrai
fare_amou
```

Distributed Arrays

Tall

R2016b New!

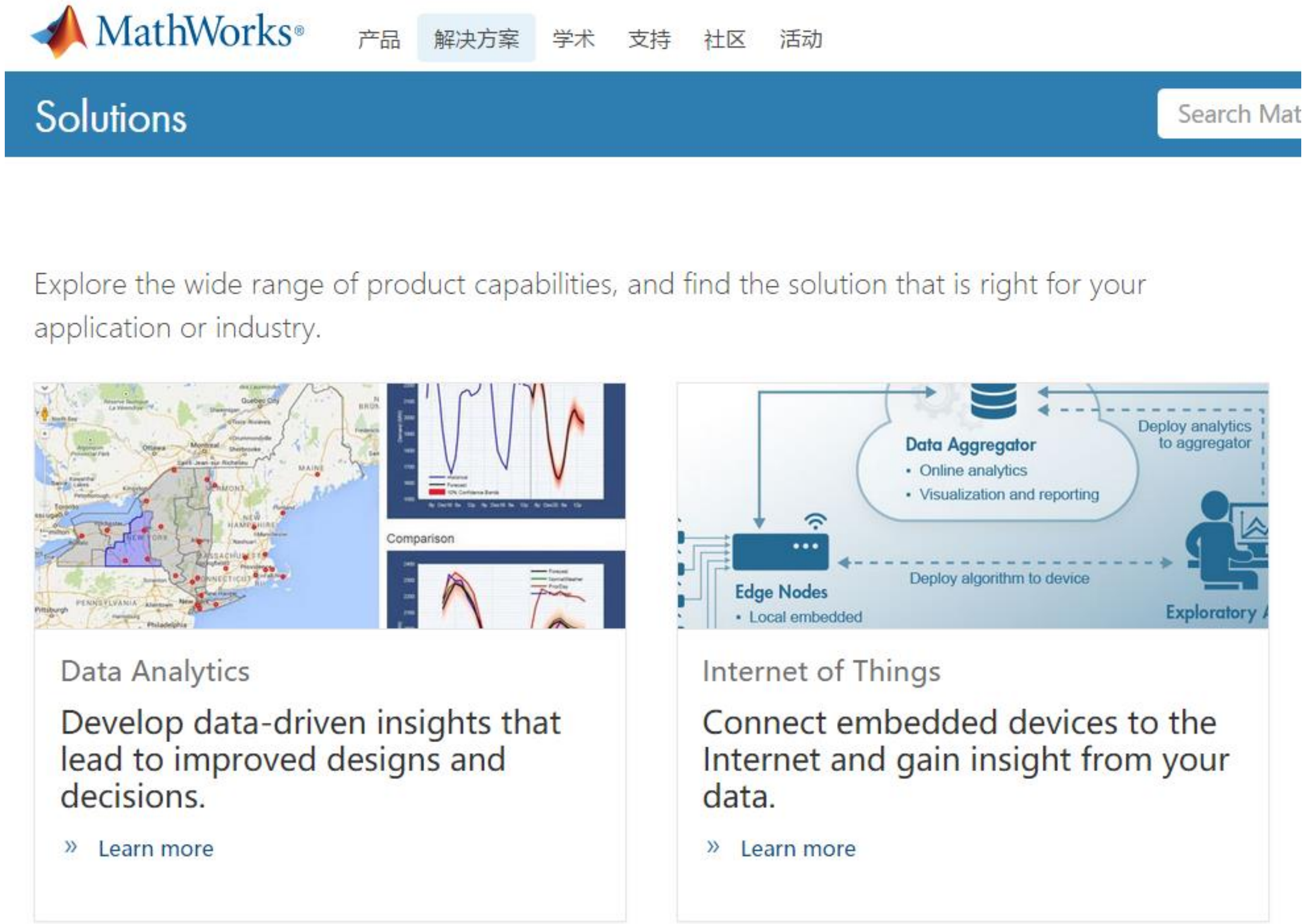
为什么是 MATLAB?

MATLAB

是工程师和科学家最易用、并且最高效的开发平台

更多资源

- **www.mathworks.com/solutions**
  - 白皮书
  - 视频
  - 示例


MathWorks
产品
解决方案
学术
支持
社区
活动
Solutions
Search Mat
Comparison
Data Aggregator
Online analytics
Visualization and reporting
Deploy analytics to aggregator
Deploy algorithm to device
Edge Nodes
Local embedded
Exploratory

Explore the wide range of product capabilities, and find the solution that is right for your application or industry.

Data Analytics

Develop data-driven insights that lead to improved designs and decisions.

» Learn more

Internet of Things

Connect embedded devices to the Internet and gain insight from your data.

» Learn more

扫描右侧二维码，关注

MATLAB官方微信公众号

MATLAB

© 2016 The MathWorks, Inc. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.

谢谢