



LECTURE 2

CEIC6789 NOTES





A PEAK
INTO THE
COLLECTED
DATASET

DATA SCIENCE DIAGRAM: WHAT IS INVOLVED?

DATA

Data collection

- ✧ raw data or primary data
- ✧ Cannot be analyzed

Preprocessing

- ✧ class labeling: numerical, categorical, text, video, audio etc.
- ✧ data cleaning: numbers mixed with text or vice versa
- ✧ missing values

ANALYSIS

Analyze the data

- ✧ What happened?
- ✧ When did it happen?
- ✧ How did it happen?
- ✧ ... and many more!

Extract info and present:

- ✧ metrics
- ✧ KPIs (key performance indicators)
- ✧ Reports
- ✧ Dashboard

ANALYTICS

Branches

- ✧ Traditional methods
- ✧ Machine learning

Techniques

- ✧ Regression (linear, non-linear)
- ✧ Clustering
- ✧ PCA (Principal Component Analysis)
- ✧ Time series
- ✧ SVMs (support vector machines)
- ✧ Neural networks
- ✧ Deep learning
- ✧ Bayesian networks



CHEESE

DATA SCIENCE DIAGRAM: WHAT IS INVOLVED?

DATA

Data collection

- ✧ Raw data or primary data
- ✧ Cannot be analyzed

Preprocessing

- ✧ class labeling: numerical, categorical, text, video, audio etc.
- ✧ data cleaning: numbers mixed with text or vice versa
- ✧ missing values

ANALYSIS

Analyze the data

- ✧ What happened?
- ✧ When did it happen?
- ✧ How did it happen?
- ✧ ... and many more!

Extract info and present:

- ✧ metrics
- ✧ KPIs (key performance indicators)
- ✧ Reports
- ✧ Dashboard

ANALYTICS

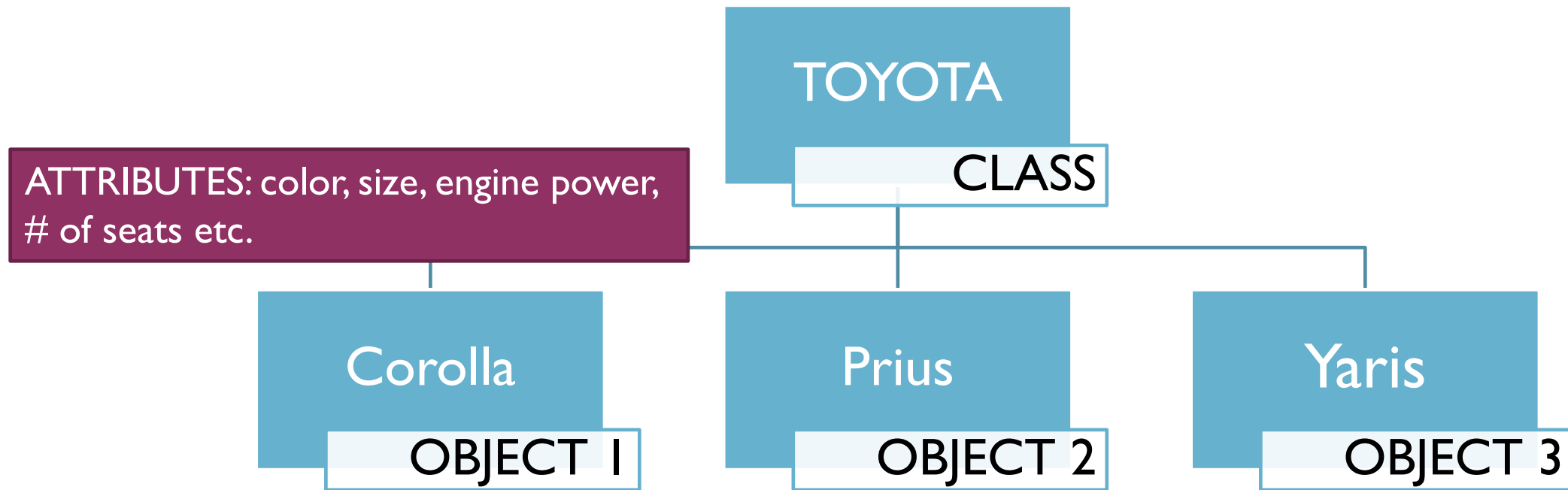
Branches

- ✧ Traditional methods
- ✧ Machine learning

Techniques

- ✧ Regression (linear, non-linear)
- ✧ Clustering
- ✧ PCA (Principal Component Analysis)
- ✧ Time series
- ✧ SVMs (support vector machines)
- ✧ Neural networks
- ✧ Deep learning
- ✧ Bayesian networks

CLASSES, OBJECTS, ATTRIBUTES AND METHODS

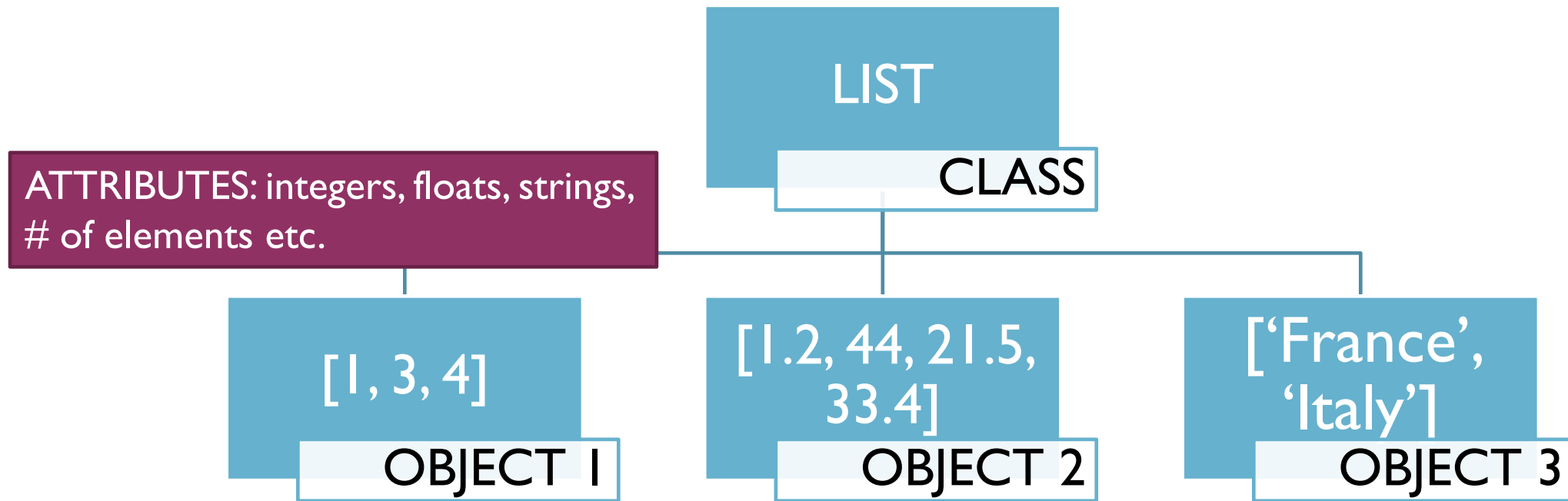


ATTRIBUTES: color, size, engine power,
of seats etc.

METHODS

Corolla.accelerate()
Corolla.stop()
Corolla.turn_right() etc.

PRACTICAL EXAMPLE



METHODS
`[1,3,4].append()`
`[1,3,4].remove()`
`[1,3,4].reverse()` etc.

DIFFERENCE BETWEEN METHODS AND FUNCTIONS

METHOD

Part of a certain class

Can have many parameters,
but the object is one of its
parameters

`object.method()`

FUNCTION

Exists on its own

Can have many parameters

`function()`

MODULES AND PACKAGES

PACKAGE OR LIBRARY

MODULE 1

CLASS 1
CLASS 2
CLASS 3

FUNCTION 1
FUNCTION 2

MODULE 2

CLASS 1
CLASS 2

FUNCTION 1

MODULE 3

CLASS 1
CLASS 2
CLASS 3
CLASS 4

FUNCTION 1

LIBRARIES WE WILL USE

NUMPY

PANDAS

MATPLOTLIB

SKLEARN