dads 2009|2010 spring

Theme: **Data Mining for Architecture and Urban Planning**

Lecture I
Data Mining in General

Ceyhun Burak Akgül, PhD in EE

Ahu Sökmenoğlu, M. Arch.

In this lecture

- What is data mining?

- What is it useful for?

- How does it work?

- Does it always work?

- The Context: What is in it for us?

- What will we see/learn/do this semester?

What is data mining?

What is data?

What is data mining?

What is data?

Dictionary.com definition

da·tum [dey-tuhm, dat-uhm, dah-tuhm] ? Show IPA

–noun, plural da·ta [dey-tuh, dat-uh, dah-tuh] ? Show IPA for 1–3, **da·tums** for 4, 5.

1. a single piece of information, as a fact, statistic, or code; an item of data.
2. *Philosophy.*
   a. any fact assumed to be a matter of direct observation.
   b. any proposition assumed or given, from which conclusions may be drawn.
3. Also called **sense datum.** *Epistemology.* the object of knowledge as presented to the mind. Compare IDEATUM.

What is data mining?

What is data?

Dictionary.com definition

da·tum [dey-*tuh*m, dat-*uh*m, dah-*tuh*m] [?] Show IPA

–noun, plural da·ta [dey-*tuh*, dat-*uh*, dah-*tuh*] [?] Show IPA for 1–3, **da·tums** for 4, 5.

1. a single piece of information, as a fact, statistic, or code; an item of data.
2. *Philosophy*.
   a. any fact assumed to be a matter of direct observation.
   b. any proposition assumed or given, from which conclusions may be drawn.
3. Also called **sense datum.** *Epistemology*. the object of knowledge as presented to the mind. Compare IDEATUM.

Also called **sensum.** *Psychology*. the basic unit of an experience resulting from the stimulation of a sense organ; a stimulus or an object of perception or sensation.

What is data mining?

What is data?

- Objective facts about the perceived or sensed
- Measurements
- Physical instantiation/embodiment of *information*

Where do data/information arise from?
  >>

What is data mining?

Sources of information

- Humans
  - Our senses
  - Our likes/dislikes, favorite books, movies, music, …
  - Our identity, political views, web browsing sessions, …
  - Our genome, medical records, …
- Earth and Universe
  - Stars and planets, celestial motions, …
  - Climate, geography, oceans, …
- Organizations/Companies
  - Expenses and revenues, stock rates, …
- …

An evangelical notice about the things to come…

In this course, we will consider

The City itself and our interactions with The City

as our source of information (data)

What is data mining?

"Information is not knowledge."
– Albert Einstein

"We are drowning in information and starving for knowledge."
– Rutherford D. Roger

What might these people have meant?

What is data mining?

"Information is not knowledge."
– Albert Einstein

"We are drowning in information and starving for knowledge."
– Rutherford D. Roger

What might these people have meant?

Information is the *preimage* (or raw form) of **knowledge**

What is data mining?

What is knowledge?

What is data mining?

What is knowledge?

Knowledge in general is the subject matter of epistemology:

- Plato: Knowledge is **justified true belief (JTB)**
  - Is it sufficient?
  - Gettier (1963): There are cases where JTB is not sufficient
- Aristotle considers knowledge from the **causality** perspective: we have knowledge when we know the cause on which a certain fact (data) depends
  - How do we diagnose causality?
- …

What is data mining?

Extracting important patterns and trends from data arising in a specific domain in order to:

- **Predict** the future outcomes of a "system" from past observations (data)

- **Explain** the underlying rules and dynamics that generate the data

- **Understand** "what the data says" about a particular question (hypothesis) of interest

What is data mining?

Extracting important patterns and trends from data arising in a specific domain in order to:

- **Predict** the future outcomes of a "system" from past observations (data)
- **Explain** the underlying rules and dynamics that generate the data
- **Understand** "what the data says" about a particular question (hypothesis) of interest
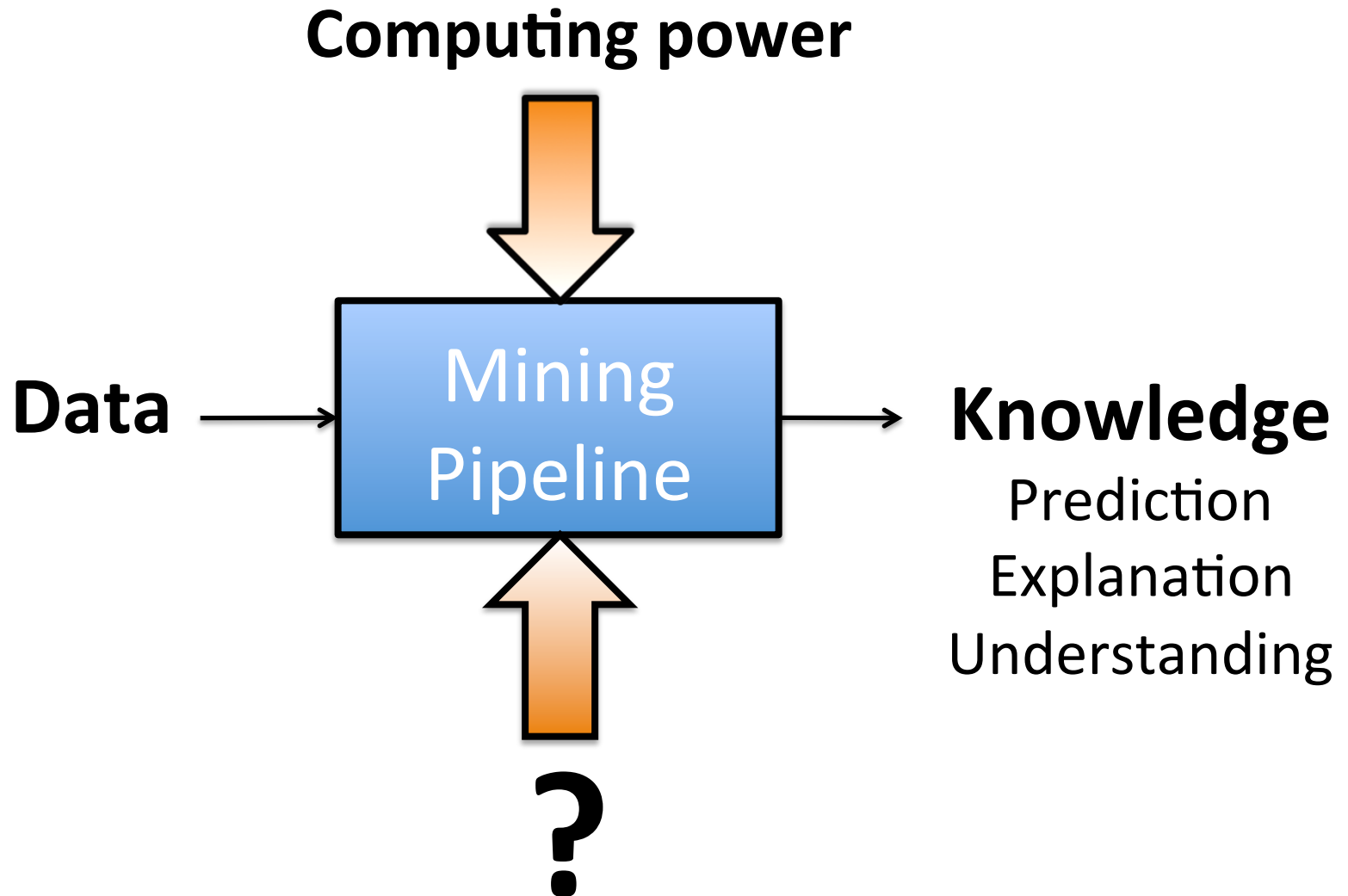
*If we are able to (partially) perform one or more of these tasks, we're said to possess **(partial) operational knowledge** on that particular domain*

What is data mining?

Computing power

Data → **Mining Pipeline** → **Knowledge**
Prediction
Explanation
Understanding

?

What is data mining?

Prediction

- Predict the amount of rain that will fall in Istanbul during March 2010

    Data: monthly records over the past ten years

- Predict whether an elderly patient will develop Alzheimer's within the years to come

    Data: medical records (tests, MR scans, …) and similar about other Alzheimer's patients
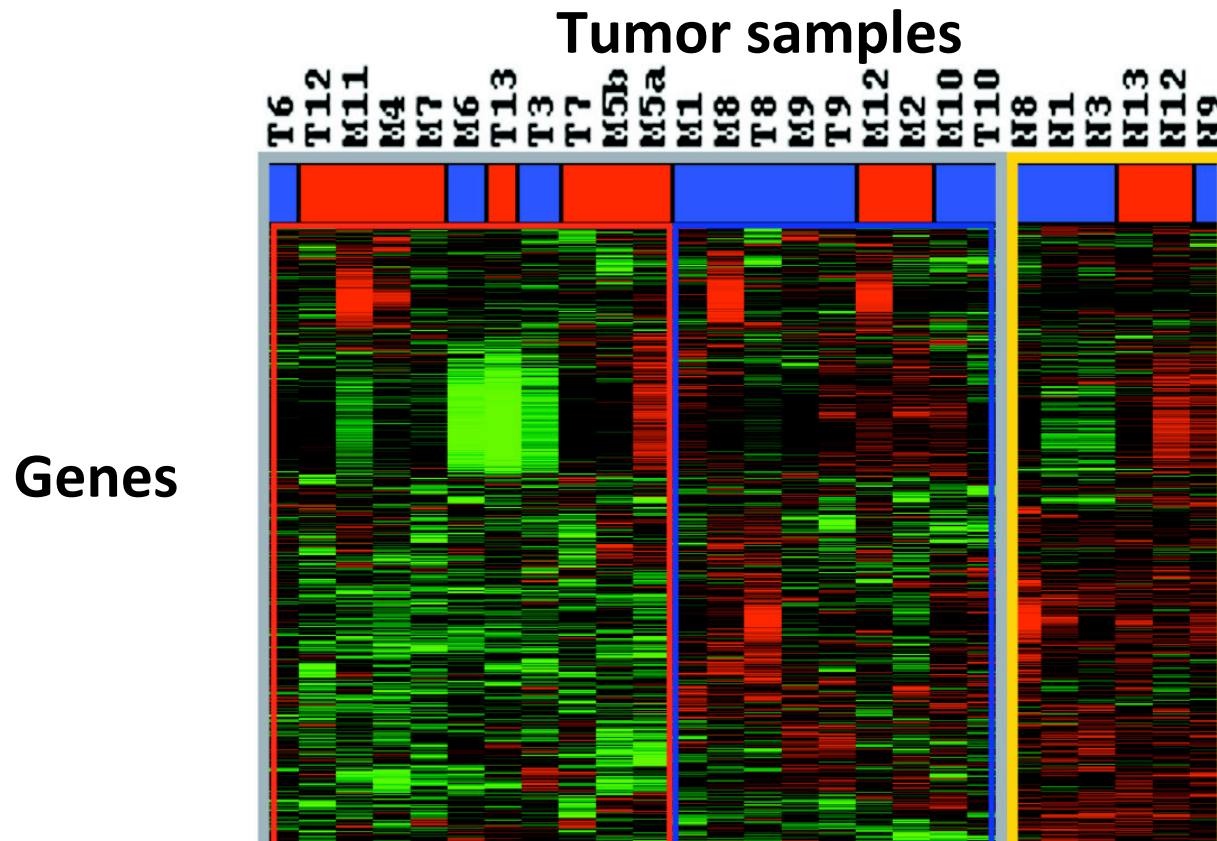
- Predict the price of a stock in six months

    Data: company performance and economic data

- …

What is data mining?

Explanation/Understanding

Tumor samples

Genes

Gene expression profile dataset
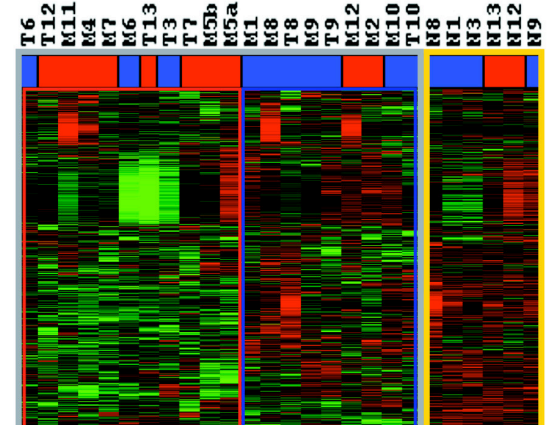
>>

What is data mining?

Explanation/Understanding

Given a gene expression dataset:

- Which samples are most similar to each other, in terms of their expression profiles across genes?

- Which genes are most similar to each other, in terms of their expression profiles across samples?

- Do certain genes show very high (or low) expression for certain cancer samples?

Applications: Classical

- Marketing (e.g., sales analysis)
- Banking (e.g., credit and loan approval)
- Medicine / Biology / Pharmacology
- Manufacturing (e.g., yield analysis)
- Finance (e.g., stock prediction)
- E-Commerce / Web (e.g., hits analysis)

>>

Applications: Less classical

- Multimedia search engines
  - Content search by audiovisual similarity
- Multimedia content management
  - Automatic content categorization and annotation
  - Audiovisual concept detection
- Image-based medical diagnosis
  - Visual biomarker discovery from medical images

How does data mining work?

Computing Power

Data → **Mining Pipeline** → **Knowledge**
Prediction
Explanation
Understanding
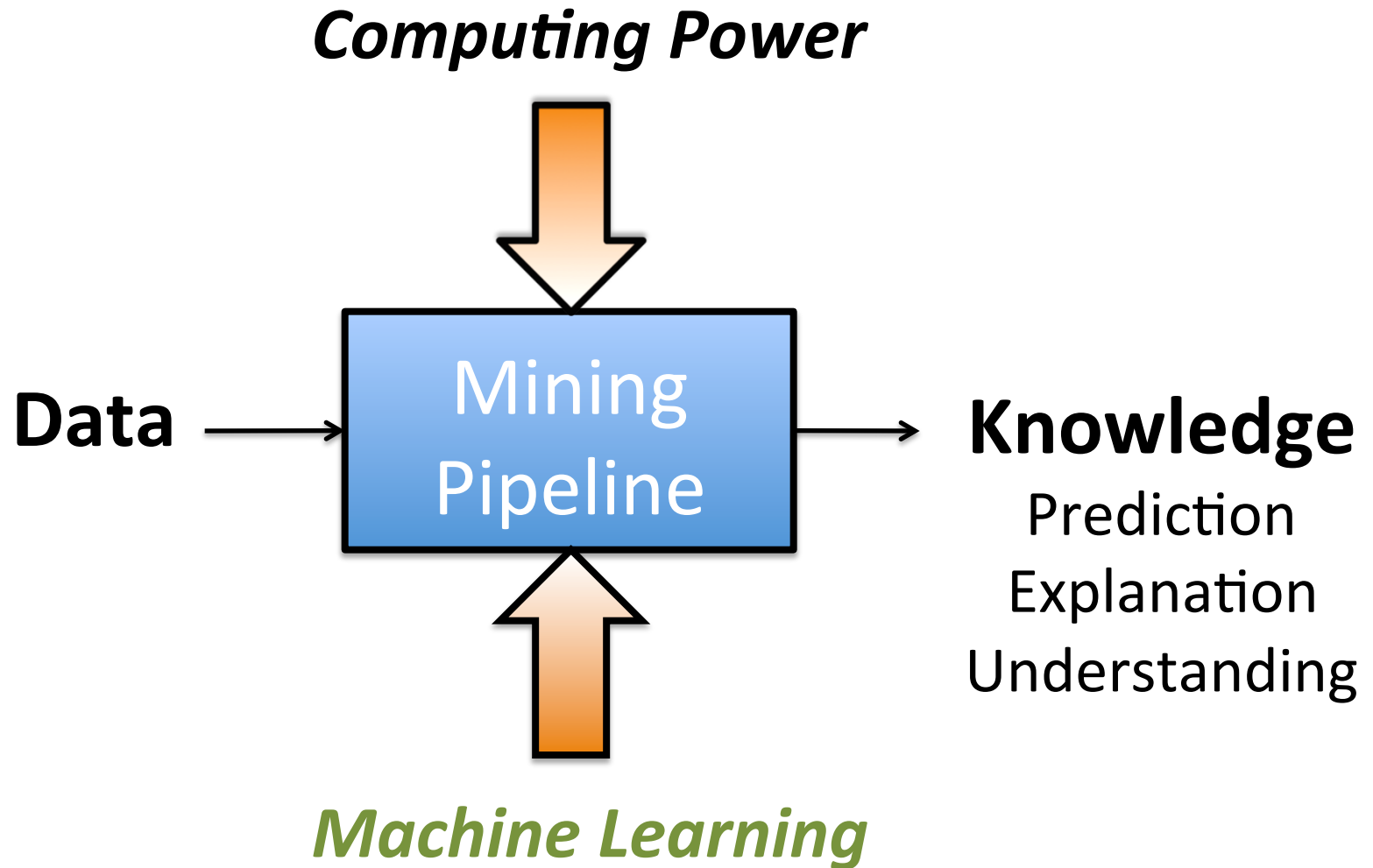
Machine Learning

How does data mining work?
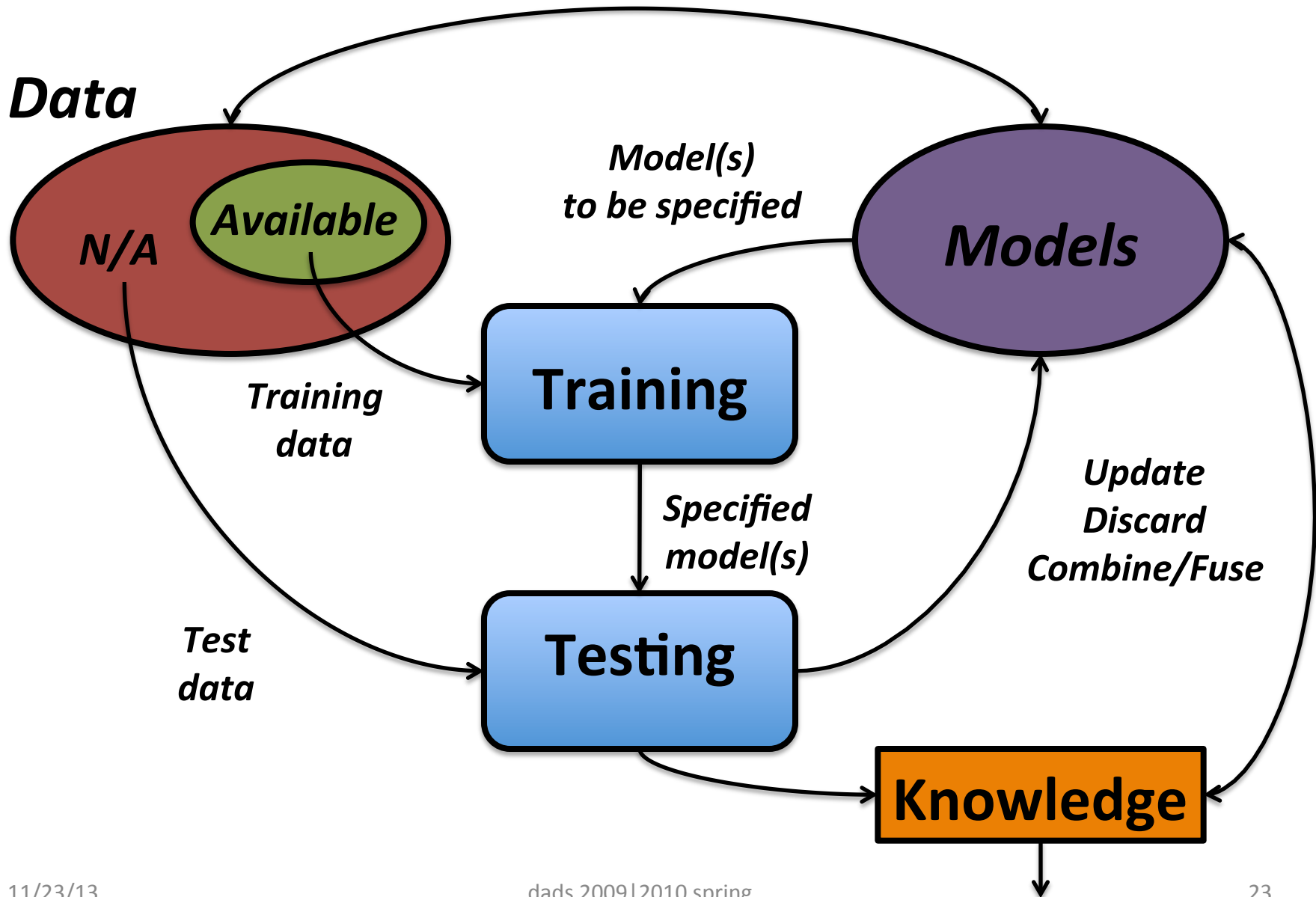
If data are the fuel of data mining,
Machine Learning *is its engine.*

– Not quite like human learning:

Computers have no awareness!

– Though not quite like a calculator neither

Machine learning involves

– Specification of a *statistical* model (generative, discriminative, or both)

– *Training* the model with available data

– *Testing* the model with new (unseen) data

How does data mining work?

Data

N/A

Available

Model(s) to be specified

Models

Training data

Training

Specified model(s)

Test data

Testing

Update Discard Combine/Fuse

Knowledge

Does it always work?

Data mining doesn't work when/if

- You don't have (sufficient) data
- You ask the wrong question
- You just focus on training
- You rely on just one technique
- You mix apples with bananas
- You don't use your intuition
- You use your intuition
- You try to answer every question
- …

The Context

Some example uses of data mining in architecture and urban research:

- Archetypal office building layouts (Hannah, 2007)
- Urban block morphology in terms of shape and density (Laskari, 2007)
- Arabic house typologies (Reffat, 2008)
- Spatio-temporal urban growth patterns and trends for modeling and prediction of urban growth (Liu and Seto, 2008)
- Urban typologies focusing on the aspects of morphology and density for blocks, mobility for streets (Gil *et al.*, 2009).

This semester

week	date	studio
1	**9-Feb**	**-**
2	**16-Feb**	**Introduction: Data Mining in General**
3	23-Feb	Concepts in Data Mining
4	2-Mar	Data Mining Applications in Context Introduction to Semester Project
5	9-Mar	Statistics Primer
6	16-Mar	A Broad Picture of Data Mining Tools Jury Meeting; Semester Project's first concepts & ideas
7	23-Mar	Regression and Classification
8	30-Mar	Clustering, Exploratory Data Analysis, and Visualization Semester Project's review
9	6-Apr	Semester Project's review
10	13-Apr	Semester Project's review
11	20-Apr	Jury Meeting; Presentations
12	27-Apr	Semester Project's review
13	4-May	Semester Project's review
14	11-May	Jury Meeting; Final Presentations

Assignments for next week

Read

- http://en.wikipedia.org/wiki/Data_mining
- Fayyad (2001), The Digital Physics of Data Mining.

Think of and write

A data mining application in your field of expertise: architecture, urban planning, industrial design, …