PSTAT 131/231: Introduction to Statistical Machine Learning

Lecture 1
Statistical Machine Learning Basics

ISL Chapter 1&2

What is Machine Learning?

Application of learning algorithms and statistical methods to real-world datasets:

What is Machine Learning?

Application of learning algorithms and statistical methods to real-world datasets:

i.e. how can a machine learn from data?

What is Machine Learning?

Application of learning algorithms and statistical methods to real-world datasets:

i.e. how can a machine learn from data?

Nontrivial extraction of implicit, previously unknown, and potentially useful pattern from data

What is Machine Learning?

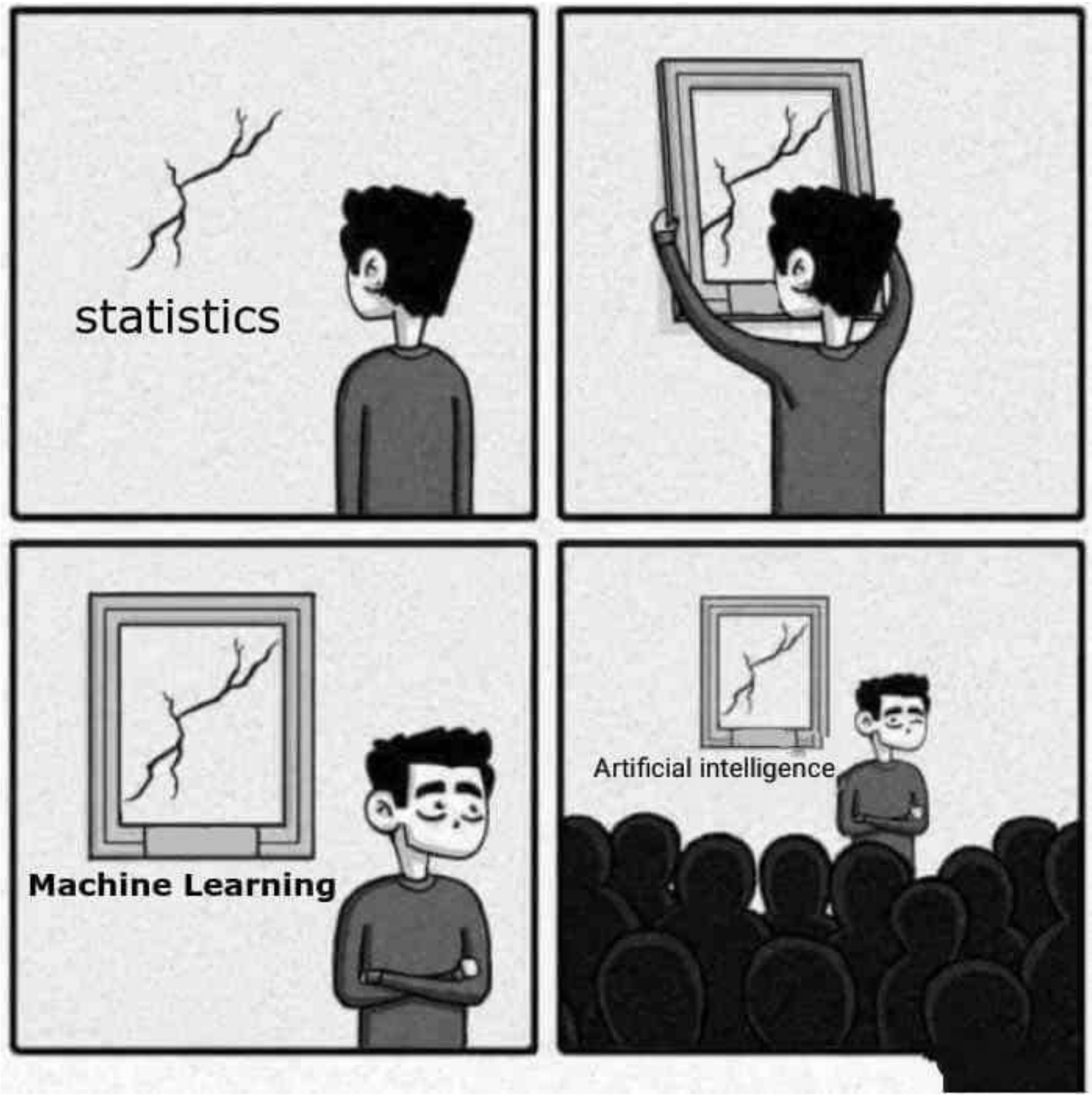
Application of learning algorithms and statistical methods to real-world datasets:

i.e. how can a machine learn from data?

Nontrivial extraction of implicit, previously unknown, and potentially useful pattern from data

- **We will cover many methods that try to achieve this goal**
- **No single method is the best choice for all datasets**
- **Appropriate method depends on goal**

Machine Learning vs Statistical Learning



Machine Learning vs Statistical Learning

Machine Learning vs Statistical Learning

arose as a subfield of Artificial Intelligence in CS

arose as a subfield of Statistics

Machine Learning vs Statistical Learning

Machine Learning	Statistical Learning
arose as a subfield of Artificial Intelligence in CS	arose as a subfield of Statistics
algorithmic point of view	statistical point of view

Machine Learning	vs	Statistical Learning
arose as a subfield of Artificial Intelligence in CS		**arose as a subfield of Statistics**
algorithmic point of view		**statistical point of view**
focus on algorithms and prediction accuracy		**focus on interpretability and inference**

Machine Learning vs Statistical Learning

Machine Learning	Statistical Learning
arose as a subfield of Artificial Intelligence in CS	arose as a subfield of Statistics
algorithmic point of view	statistical point of view
focus on algorithms and prediction accuracy	focus on interpretability and inference

Have much overlap: Both focus on supervised/unsupervised learning problems

Machine Learning vs Statistical Learning

Machine Learning	Statistical Learning
arose as a subfield of Artificial Intelligence in CS	arose as a subfield of Statistics
algorithmic point of view	statistical point of view
focus on algorithms and prediction accuracy	focus on interpretability and inference

Have much overlap: Both focus on supervised/unsupervised learning problems

Machine Learning from a Statistical Perspective

Machine Learning in Industry

Netflix Challenge

- **100,480,507 ratings from 480,189 users to 17,770 movies**
- **<user, movie, date of grade> (about 3 million such triplets)**
- **Predict User rating for films, winner takes 1 million dollars!**

Machine Learning in Industry

Netflix Challenge

- 100,480,507 ratings from 480,189 users to 17,770 movies
- <user, movie, date of grade> (about 3 million such triplets)
- Predict User rating for films, winner takes 1 million dollars!

Machine Learning in Research

Machine Learning in Research

High energy physics experiments

Machine Learning in Research

High energy physics experiments

Microarrays for gene expression (20K+ features)

Machine Learning in Research

High energy physics experiments

Microarrays for gene expression (20K+ features)

Genome-wide association studies (GWAS) (Millions of features)

Machine Learning in Research

High energy physics experiments

Microarrays for gene expression (20K+ features)

Genome-wide association studies (GWAS) (Millions of features)

Unprecedented large scale of data

Machine Learning in Research

High energy physics experiments

Microarrays for gene expression (20K+ features)

Genome-wide association studies (GWAS) (Millions of features)

Unprecedented large scale of data

+

Machine Learning in Research

High energy physics experiments

Microarrays for gene expression (20K+ features)

Genome-wide association studies (GWAS) (Millions of features)

Unprecedented large scale of data

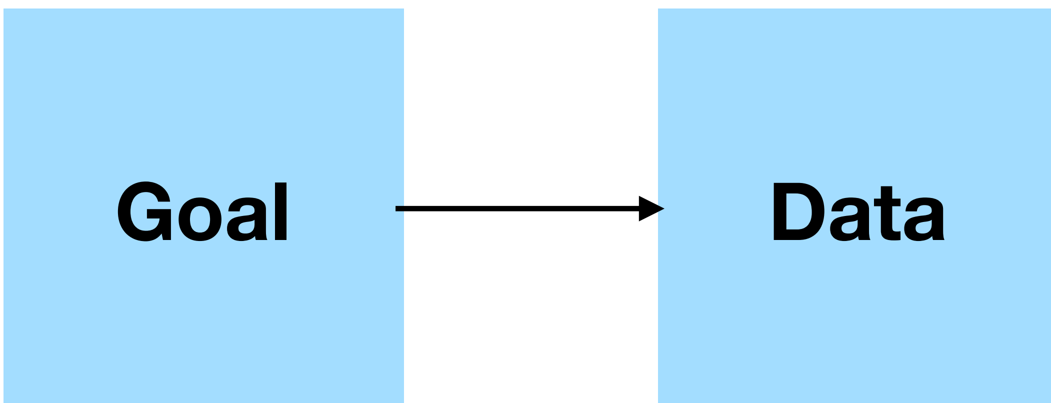
+

Highly complicated underlying structure

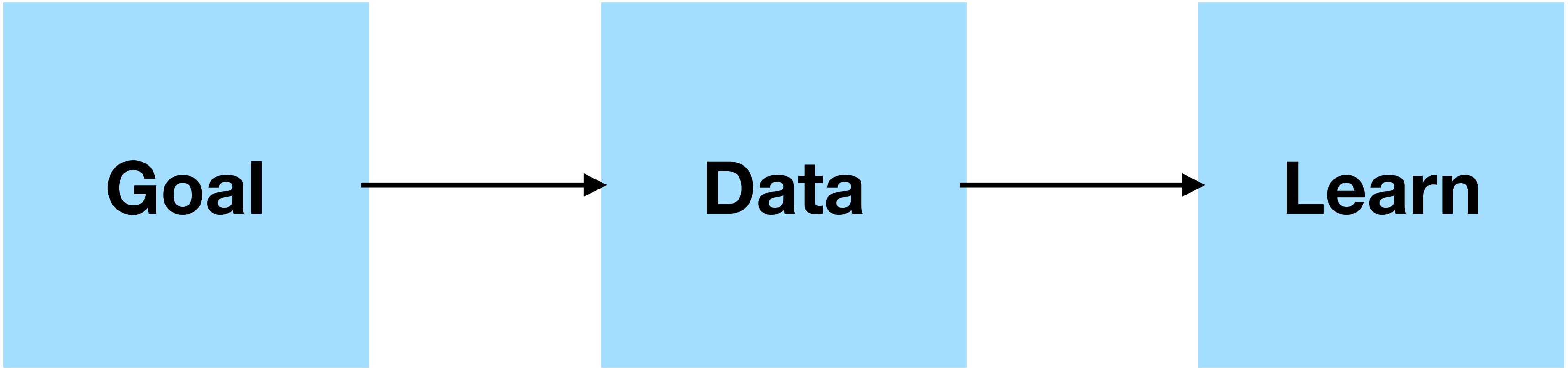
A (very) rough ML workflow

Goal

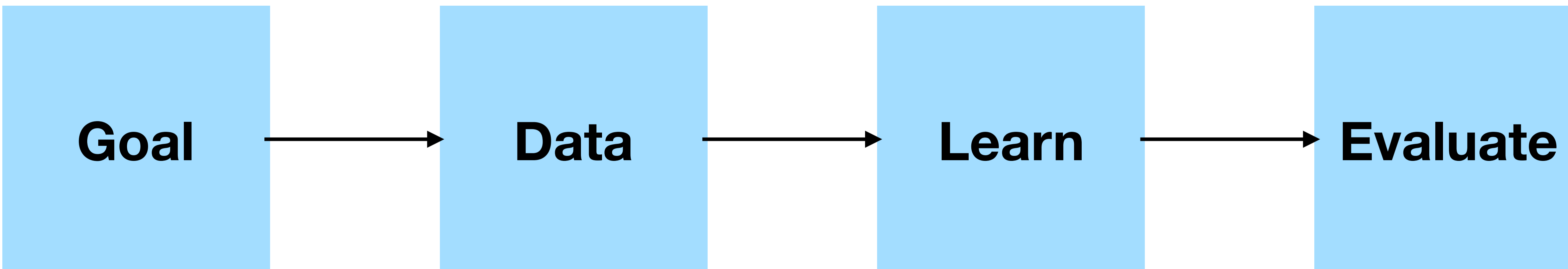
A (very) rough ML workflow



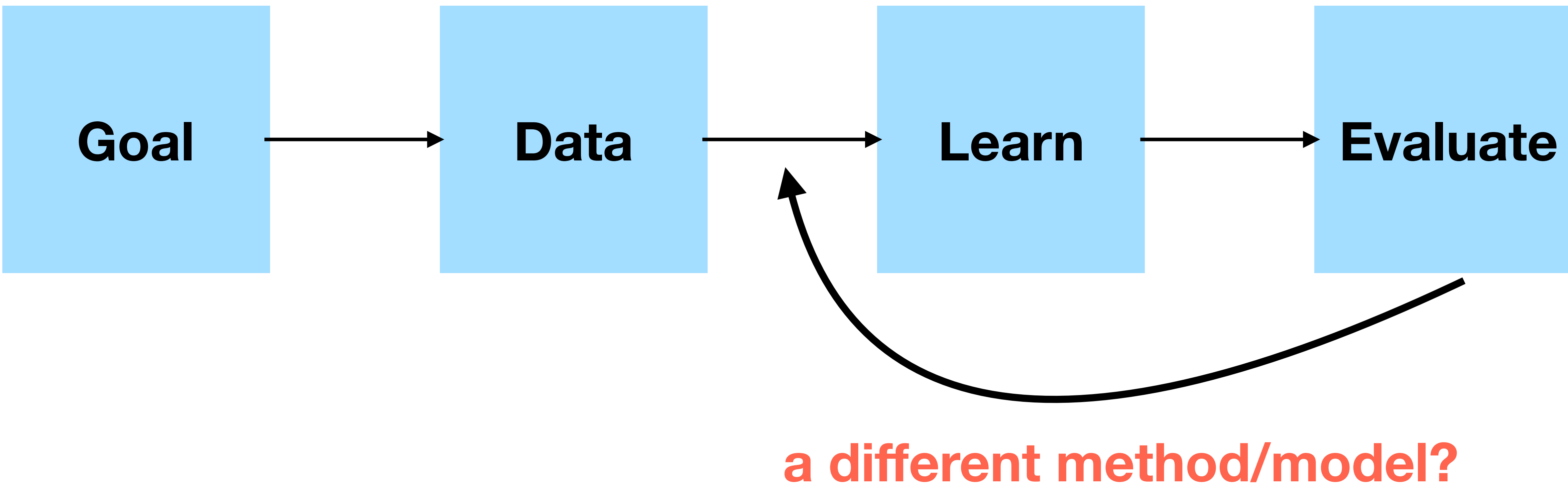
A (very) rough ML workflow



A (very) rough ML workflow



A (very) rough ML workflow



An example…

On population level…

$$Y \qquad\qquad (X_1, \ldots, X_p)$$

An example...

On population level...

$$Y \qquad (X_1, \ldots, X_p)$$

Response

(output, target, dependent variable)

An example…

On population level…

$$Y \qquad\qquad (X_1, \ldots, X_p)$$

Response

(output, target, dependent variable)

Predictors

(inputs, features, covariates, independent variables)

An example...

On population level...

$$Y \qquad\qquad (X_1, \ldots, X_p)$$

Response

(output, target, dependent variable)

Predictors

(inputs, features, covariates, independent variables)

$Y =$ **wage**

An example…

On population level…

$$Y \qquad\qquad (X_1, \ldots, X_p)$$

Response

(output, target, dependent variable)

Predictors

(inputs, features, covariates, independent variables)

$Y =$ **wage**

$X_1 =$ **employee's age**

$X_2 =$ **education level**

$X_3 =$ **calendar year**

An example…

On population level…

$$Y \qquad (X_1, \ldots, X_p)$$

Response

(output, target, dependent variable)

Predictors

(inputs, features, covariates, independent variables)

$Y =$ **wage**

$X_1 =$ **employee's age**

$X_2 =$ **education level**

$X_3 =$ **calendar year**

Goal: understand how an employee's wage depend on his/her age, education level, and the calendar year

Data

On observation level…

wage | **age, education, year**

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_9 \end{pmatrix} \qquad \begin{pmatrix} x_{11}, x_{12}, x_{13} \\ x_{21}, x_{22}, x_{23} \\ x_{31}, x_{32}, x_{33} \\ \vdots \; , \; \vdots \; , \; \vdots \\ x_{91}, x_{92}, x_{93} \end{pmatrix}$$

Goal: understand how an employee's wage depend on his/her age, education level, and the calendar year

Data

On observation level…

$$
\begin{array}{lc}
 & \textbf{wage} \\
\textbf{Mike} & \begin{pmatrix} \boxed{y_1} \\ y_2 \\ y_3 \\ \vdots \\ y_9 \end{pmatrix}
\end{array}
\qquad\qquad
\begin{array}{c}
\textbf{age, education, year} \\
\begin{pmatrix} x_{11}, x_{12}, x_{13} \\ x_{21}, x_{22}, x_{23} \\ x_{31}, x_{32}, x_{33} \\ \vdots \;,\; \vdots \;,\; \vdots \\ x_{91}, x_{92}, x_{93} \end{pmatrix}
\end{array}
$$

Goal: understand how an employee's wage depend on his/her age, education level, and the calendar year

Data

On observation level…

wage

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_9 \end{pmatrix}$$

age, education, year

$$\begin{pmatrix} x_{11}, x_{12}, x_{13} \\ x_{21}, x_{22}, x_{23} \\ x_{31}, x_{32}, x_{33} \\ \vdots\ ,\ \vdots\ ,\ \vdots \\ x_{91}, x_{92}, x_{93} \end{pmatrix}$$

Goal: understand how an employee's wage depend on his/her age, education level, and the calendar year

Data

On observation level…

$$
\begin{array}{ccc}
 & \text{wage} & \\
\text{John} & \begin{pmatrix} y_1 \\ y_2 \\ \boxed{y_3} \\ \vdots \\ y_9 \end{pmatrix} &
\end{array}
\qquad
\begin{array}{c}
\text{age, education, year} \\
\begin{pmatrix} x_{11}, x_{12}, x_{13} \\ x_{21}, x_{22}, x_{23} \\ x_{31}, x_{32}, x_{33} \\ \vdots\ ,\ \vdots\ ,\ \vdots \\ x_{91}, x_{92}, x_{93} \end{pmatrix}
\end{array}
$$

Goal: understand how an employee's wage depend on his/her age, education level, and the calendar year

Data

On observation level…

wage

$$\begin{array}{rc} & \multirow{5}{*}{} \\ \end{array}
\begin{array}{r} \\ \\ \text{John} \\ \\ \\ \end{array}
\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_9 \end{pmatrix}$$

age, education, year

$$\begin{pmatrix} x_{11}, x_{12}, x_{13} \\ x_{21}, x_{22}, x_{23} \\ x_{31}, x_{32}, x_{33} \\ \vdots\ , \ \vdots\ , \ \vdots \\ x_{91}, x_{92}, x_{93} \end{pmatrix}$$

Goal: understand how an employee's wage depend on his/her age, education level, and the calendar year

Data

On observation level…

wage **age, education, year**

$$\begin{array}{r} \\ \\ \text{John} \\ \\ \\ \end{array}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_9 \end{pmatrix} \qquad \begin{pmatrix} x_{11}, x_{12}, x_{13} \\ x_{21}, x_{22}, x_{23} \\ x_{31}, x_{32}, x_{33} \\ \vdots \;,\; \vdots \;,\; \vdots \\ x_{91}, x_{92}, x_{93} \end{pmatrix}$$

Goal: understand how an employee's wage depend on his/her age, education level, and the calendar year

Data

On observation level…

$$
\begin{array}{cc}
 & \text{wage} \\
 & \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_9 \end{pmatrix}
\end{array}
\qquad
\begin{array}{c}
\text{age, education, year} \\
\begin{pmatrix} x_{11}, x_{12}, x_{13} \\ x_{21}, x_{22}, x_{23} \\ x_{31}, x_{32}, x_{33} \\ \vdots, \vdots, \vdots \\ x_{91}, x_{92}, x_{93} \end{pmatrix}
\end{array}
$$

John

Goal: understand how an employee's wage depend on his/her age, education level, and the calendar year

Data

On observation level…

wage

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_9 \end{pmatrix}$$

age, education, year

$$\begin{pmatrix} x_{11}, x_{12}, x_{13} \\ x_{21}, x_{22}, x_{23} \\ x_{31}, x_{32}, x_{33} \\ \vdots\ ,\ \vdots\ ,\ \vdots \\ x_{91}, x_{92}, x_{93} \end{pmatrix}$$

Goal: understand how an employee's wage depend on his/her age, education level, and the calendar year

Data

On observation level…

wage

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_9 \end{pmatrix}$$

age, education, year

$$\begin{pmatrix} x_{11}, x_{12}, x_{13} \\ x_{21}, x_{22}, x_{23} \\ x_{31}, x_{32}, x_{33} \\ \vdots\ ,\ \vdots\ ,\ \vdots \\ x_{91}, x_{92}, x_{93} \end{pmatrix}$$

all 9 employees' education levels

Goal: understand how an employee's wage depend on his/her age, education level, and the calendar year

Learn

$$Y = f(X_1, \ldots, X_p) + \varepsilon$$

Response

Predictors

noise

Key: learn f from observations of $(Y, X_1, \ldots, X_p)$

Evaluate: How well did we learn?

Y **plays the role of supervisor, telling us how well we've learned**

Learner	**Supervisor**

Evaluate: How well did we learn?

Y **plays the role of supervisor, telling us how well we've learned**

Learner	Supervisor
We	Homework&Quiz solution / Professor

Evaluate: How well did we learn?

Y **plays the role of supervisor, telling us how well we've learned**

Learner	Supervisor
We	Homework&Quiz solution / Professor
Machine	Y

Evaluate: How well did we learn?

Y **plays the role of supervisor, telling us how well we've learned**

Learner	Supervisor
We	**Homework&Quiz solution / Professor**
Machine	Y

Supervised learning: need both predictors $(X_1, \ldots, X_p)$ and response Y

Supervised Learning

Supervised learning: need both predictors $(X_1, \ldots, X_p)$ and response Y

What do we care about?

Supervised Learning

Supervised learning: **need both predictors** $(X_1, \ldots, X_p)$ **and response** Y

What do we care about?

Prediction: accurately predict future response given predictors

Supervised Learning

Supervised learning: **need both predictors** $(X_1, \ldots, X_p)$ **and response** Y

What do we care about?

Prediction: accurately predict future response given predictors

Estimation: understand how predictors affect response

Supervised Learning

Supervised learning: need both predictors $(X_1, \ldots, X_p)$ and response Y

What do we care about?

Prediction: accurately predict future response given predictors

Estimation: understand how predictors affect response

Model selection: find the "*best*" model for response given predictors

Supervised Learning

Supervised learning: need both predictors $(X_1, \ldots, X_p)$ and response Y

What do we care about?

Prediction: accurately predict future response given predictors

Estimation: understand how predictors affect response

Model selection: find the "*best*" model for response given predictors

Inference: assess the quality of our predictions and (or) estimation

Regression vs Classification

Regression	Classification

Regression vs Classification

Regression

Classification

Y **is quantitative**

Numerical values
(e.g price, blood pressure)

Regression vs Classification

Regression

Y is quantitative

Numerical values
(e.g price, blood pressure)

Classification

Y is qualitative

Categorical values
(e.g survived/died, digit 0-9)

Regression vs Classification

Regression

Classification

Regression	Classification
Y is quantitative	**Y is qualitative**
Numerical values (e.g price, blood pressure)	**Categorical values (e.g survived/died, digit 0-9)**

The only difference

Regression vs Classification

Regression	Classification
Y is quantitative	**Y is qualitative**
Numerical values (e.g price, blood pressure)	**Categorical values (e.g survived/died, digit 0-9)**

The only difference

Regression and Classification problems have a lot in common

Regression vs Classification

Regression

Classification

Regression	Classification
Y is quantitative	**Y is qualitative**
Numerical values (e.g price, blood pressure)	**Categorical values (e.g survived/died, digit 0-9)**

The only difference

Regression and Classification problems have a lot in common

Both are supervised learning problems (need Y)

Unsupervised Learning

Unsupervised learning: learn without a teacher!

NO Response

$$(X_1, \ldots, X_p)$$

Predictors

Generally a more challenging situation…

Unsupervised Learning

Unsupervised learning: learn without a teacher!

NO Response

$$(X_1, \ldots, X_p)$$

Predictors

Generally a more challenging situation…

goal is more fuzzy: find groups of samples that behave similarly, find predictors that behave similarly…

Unsupervised Learning

Unsupervised learning: learn without a teacher!

NO Response

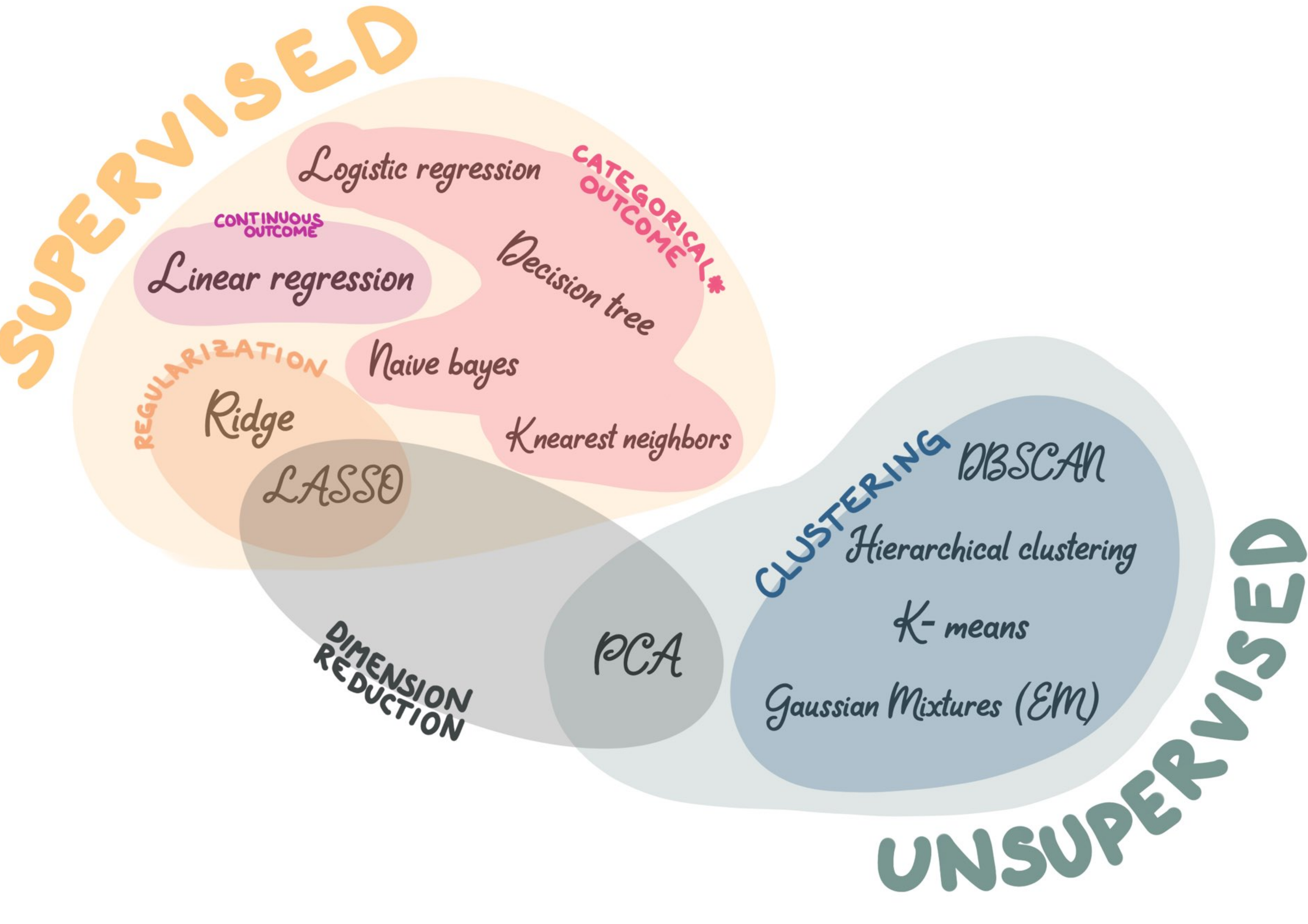
$$(X_1, \ldots, X_p)$$

Predictors

Generally a more challenging situation…

goal is more fuzzy: find groups of samples that behave similarly, find predictors that behave similarly…

harder to evaluate how well we learned…



credit: Chelsea Parlett Pelleriti

Supervised Learning

Unsupervised Learning

Supervised Learning

Unsupervised Learning

Linear regression

Supervised Learning

Unsupervised Learning

Linear regression

Logistic regression

Supervised Learning

Unsupervised Learning

Linear regression

Logistic regression

k-Nearest Neighbors

Supervised Learning

Unsupervised Learning

Linear regression

Logistic regression

k-Nearest Neighbors

decision trees

Supervised Learning | Unsupervised Learning

Linear regression

Logistic regression

k-Nearest Neighbors

decision trees

random forest

Supervised Learning

Unsupervised Learning

Linear regression

Logistic regression

k-Nearest Neighbors

decision trees

random forest

support vector machines

Supervised Learning

Unsupervised Learning

Linear regression

Logistic regression

k-Nearest Neighbors

decision trees

random forest

support vector machines

neural network

Supervised Learning

Unsupervised Learning

Linear regression

Logistic regression

k-Nearest Neighbors

decision trees

random forest

support vector machines

neural network

⋮

Supervised Learning	Unsupervised Learning
Linear regression	**Principle Component Analysis**
Logistic regression	
k-Nearest Neighbors	
decision trees	
random forest	
support vector machines	
neural network	
⋮	

Supervised Learning	Unsupervised Learning
Linear regression	Principle Component Analysis
Logistic regression	k-means Clustering
k-Nearest Neighbors	
decision trees	
random forest	
support vector machines	
neural network	
⋮	

Supervised Learning	Unsupervised Learning
Linear regression	**Principle Component Analysis**
Logistic regression	**k-means Clustering**
k-Nearest Neighbors	**hierarchical Clustering**
decision trees	
random forest	
support vector machines	
neural network	
⋮	

Supervised Learning

Linear regression

Logistic regression

k-Nearest Neighbors

decision trees

random forest

support vector machines

neural network

⋮

Unsupervised Learning

Principle Component Analysis

k-means Clustering

hierarchical Clustering

⋮

Comments

It is important to…

Comments

It is important to…

> **understand the ideas behind these methods, in order to know how and when to use them**

Comments

It is important to…

understand the ideas behind these methods, in order to know how and when to use them

understand the simpler method first, in order to understand the more complicated ones

Comments

It is important to…

understand the ideas behind these methods, in order to know how and when to use them

understand the simpler method first, in order to understand the more complicated ones

accurately evaluate the performance of a method, in order to know how to improve upon it

Next…

Homework 1 will be out next week

Set up your R working environment well in advance of hw due date

Lab sessions start next Wednesday

TAs’ and ULAs’ OH start next week

My OH starts today!