

Machine Learning – I

Richa Singh

Google classroom code: jwpk5zn

Slides are prepared from several information sources on the web and books

Recap

- Machine Learning is the study of algorithms that
 - improve their performance P
 - at some task T
 - with experience E.A well-defined learning task is given by $\langle P, T, E \rangle$.

Quiz Next Sunday

Recap



Binary classification



Multiclass classification

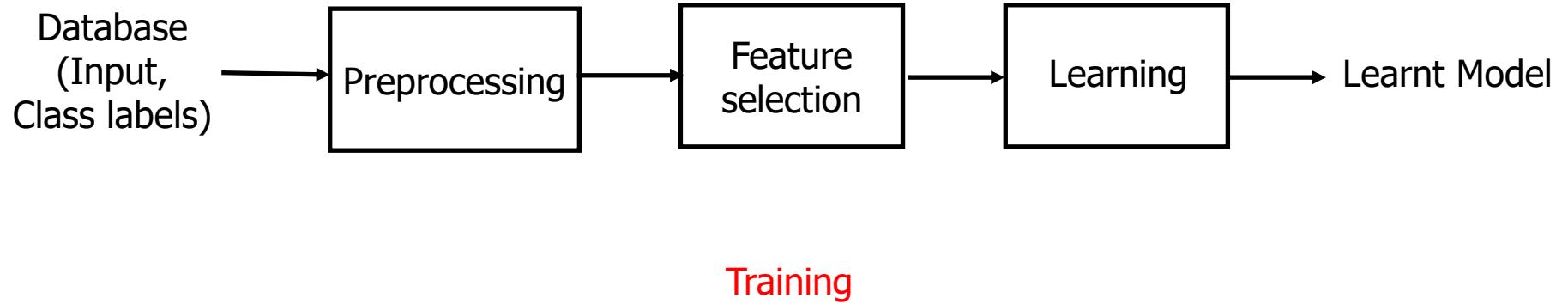


Multilabel classification:
- White shirt
- Black trouser

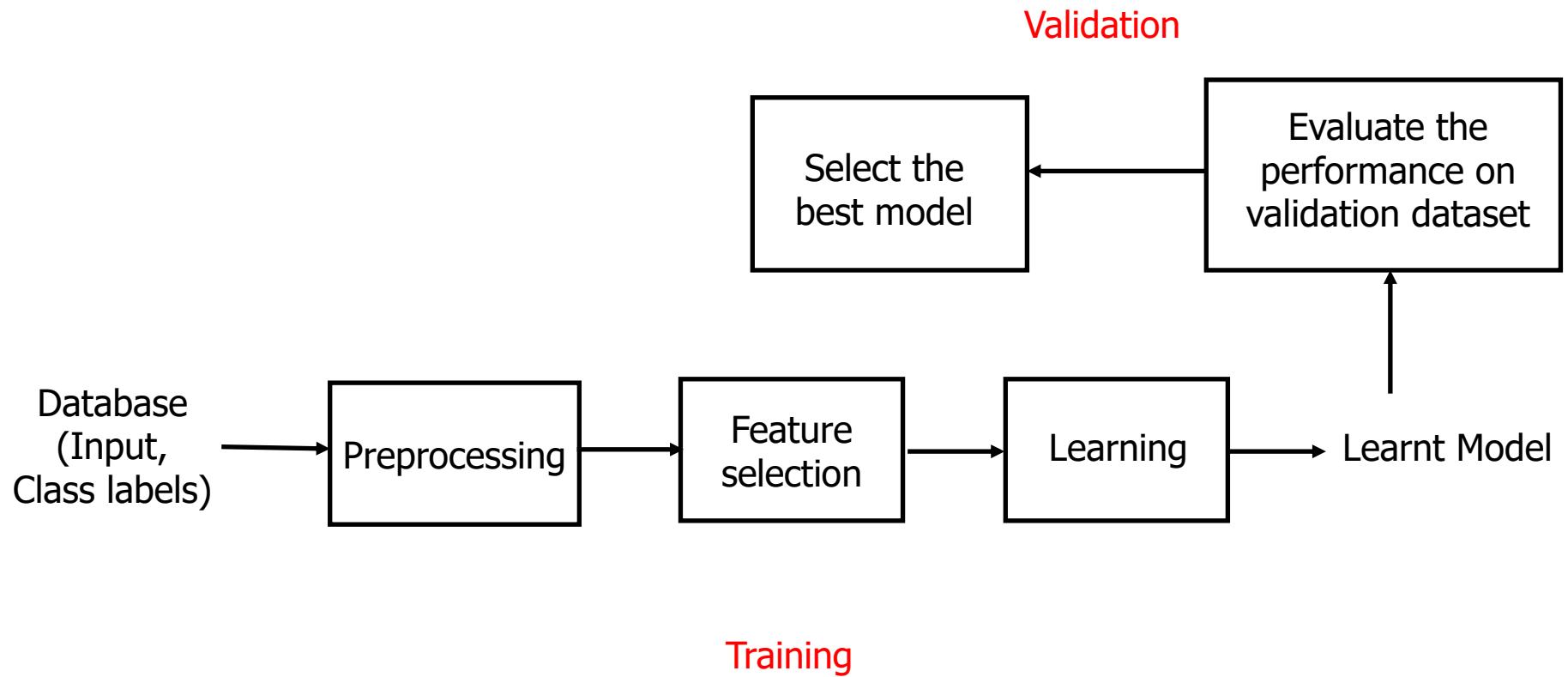
Machine Learning Pipeline

- Three steps:
 - Training
 - Validation
 - Testing

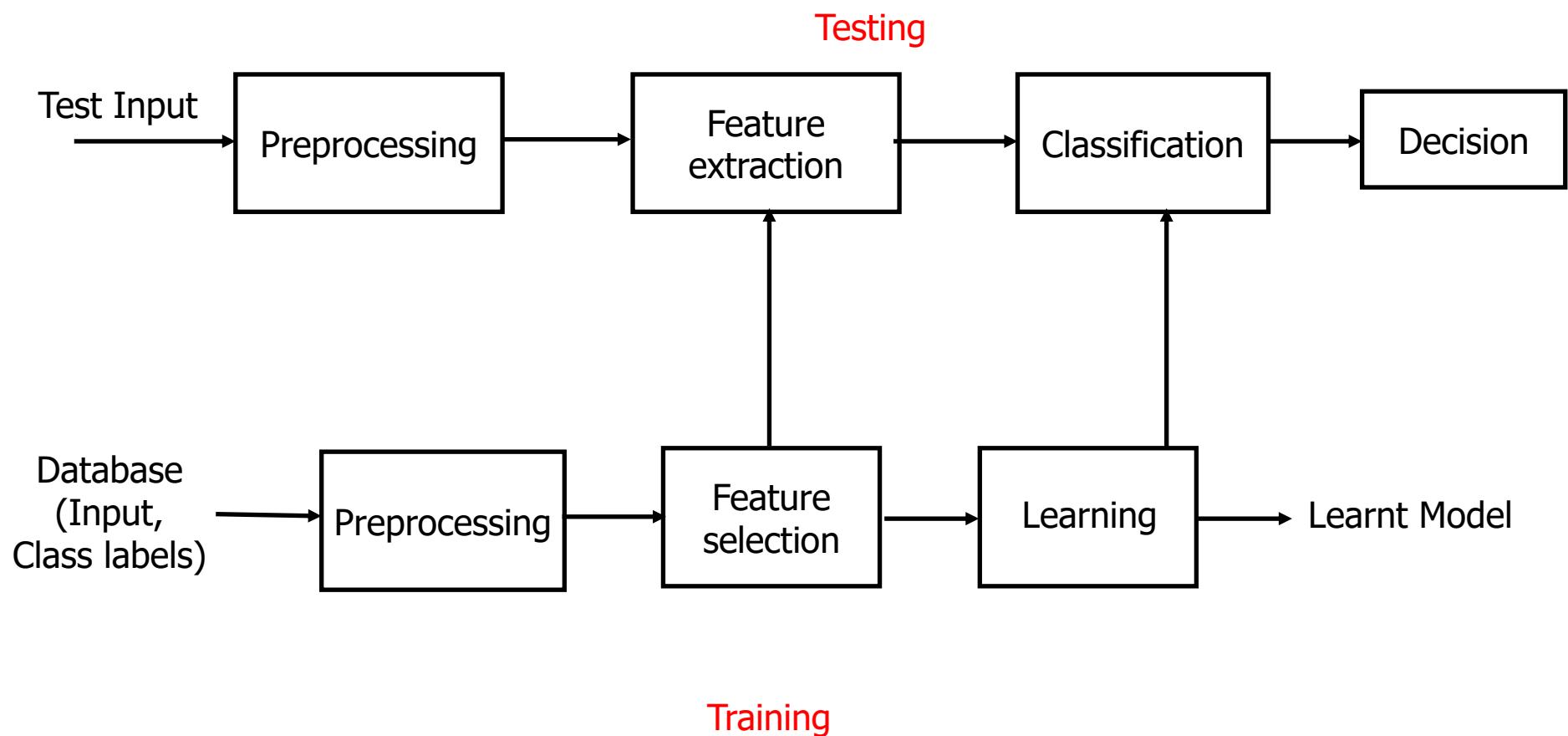
Machine Learning Pipeline



Machine Learning Pipeline



Machine Learning Pipeline



ML in Practice



- Understand domain, prior knowledge, and goals
- Data integration, selection, cleaning, pre-processing, etc.
- Learn models
- Interpret results
- Consolidate and deploy discovered knowledge

Types of ML Paradigms

- Supervised learning
 - Classification
 - Regression
- Unsupervised learning
- Reinforcement learning

Supervised Learning

- Given: training data + desired outputs (labels)
- $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x



Cats



Dogs

Supervised Learning

Classification:

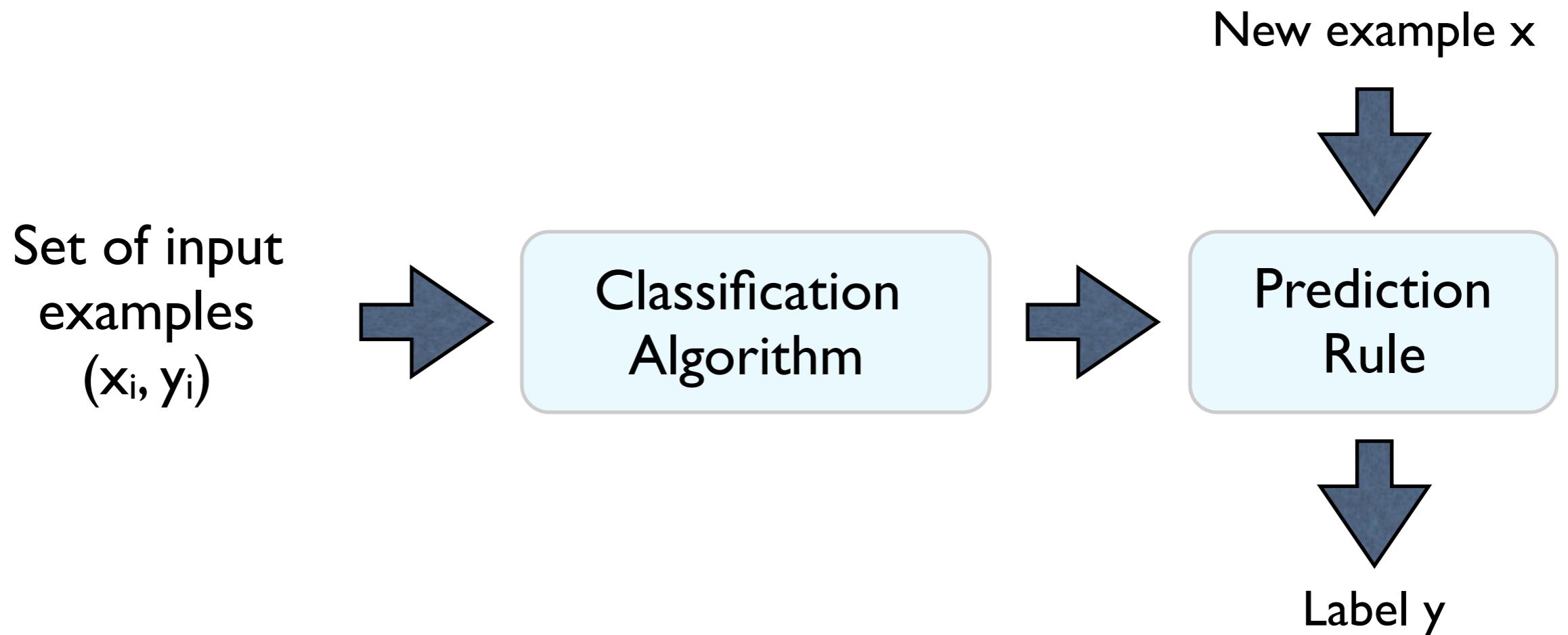
Given labeled data:

$$(x_i, y_i) \quad i=1, \dots, n$$

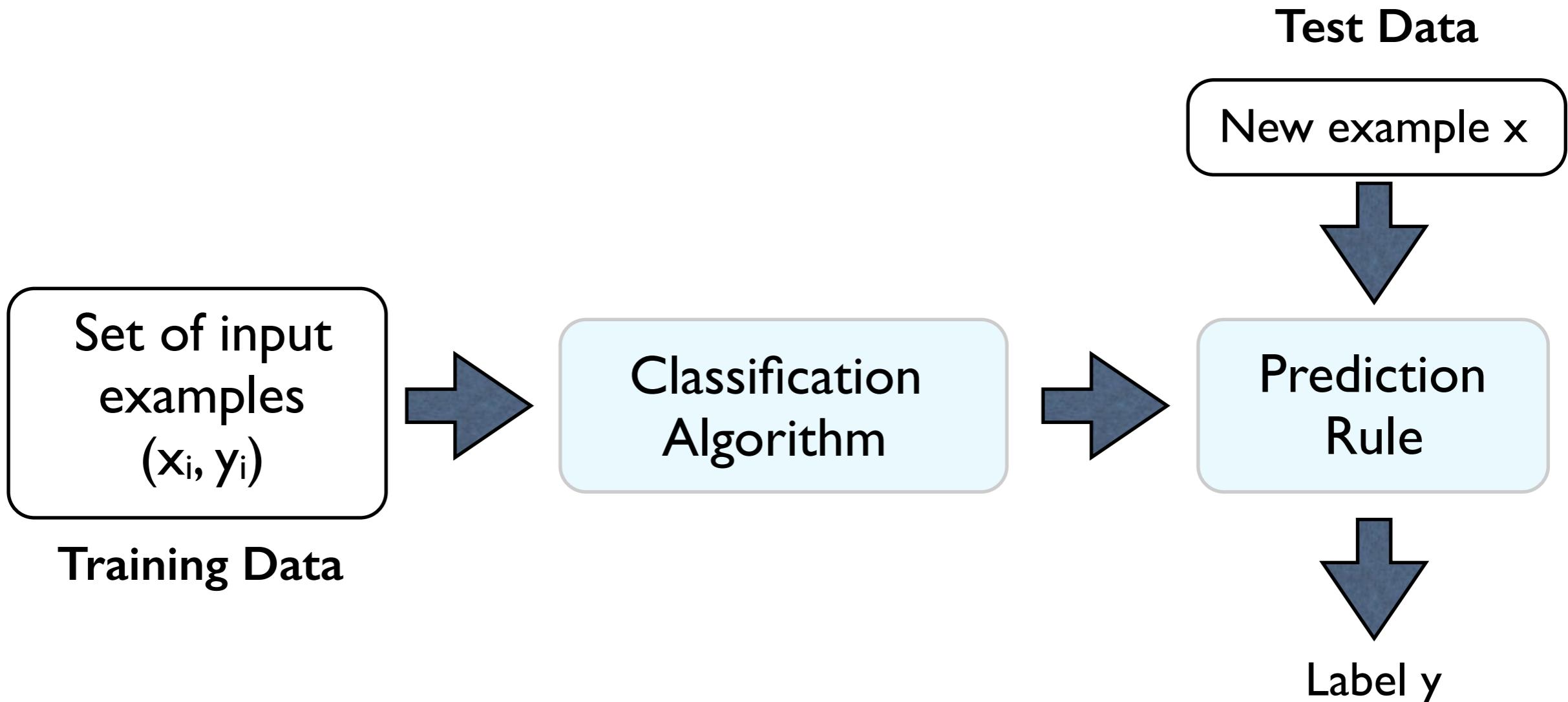
feature vector label

where y is discrete, find a rule to predict y values for unseen x

Typical Classification Algorithm

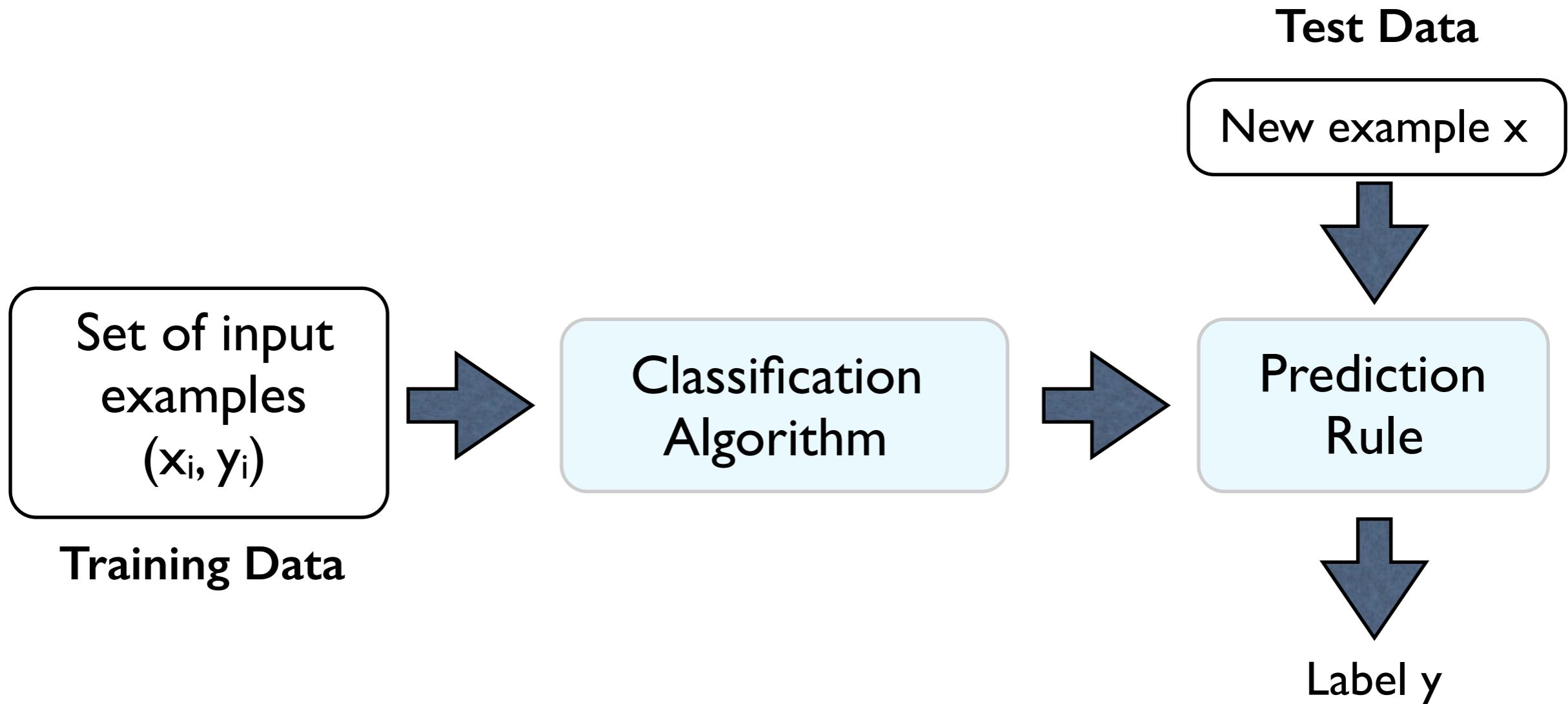


Typical Classification Algorithm



Training and test data must be **separate!**

Typical Classification Algorithm



Performance Measure:

Accuracy (or fraction of correct answers) on **test data**

Supervised Learning

Classification: Given labeled data (x_i, y_i)
where y is **discrete**, predict y values for unseen x

Example I: Predict if a **new** patient has flu or not,
based on **existing** patient data

What is x and y ?

Supervised Learning

Classification: Given labeled data (x_i, y_i)
where y is **discrete**, predict y values for unseen x

Example I: Predict if a patient has flu or not

Fever	Cold	Temperature	Flu?
Yes	No	99F	Yes

↓

I	0	99	+
x		y	

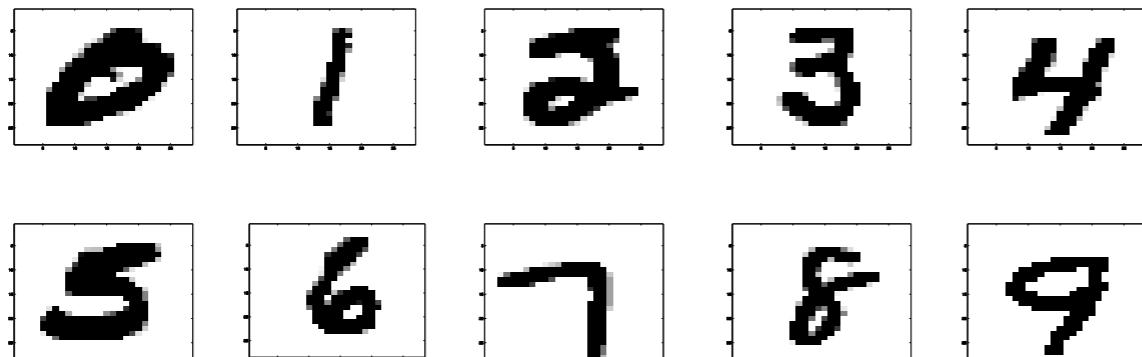
Features: Properties of patient
Label: Flu/No flu

A **binary** (two-label) classification problem

Supervised Learning

Classification: Given labeled data (x_i, y_i)
where y is **discrete**, predict y values for unseen x

Example 2: Which digit in the image ?



Label: 0,1,..,9

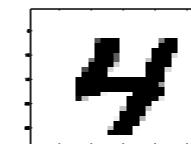
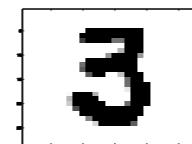
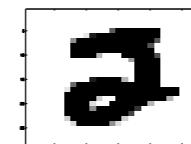
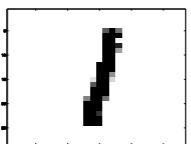
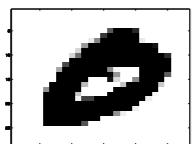
What are the features?

A multiclass classification problem

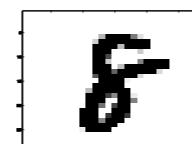
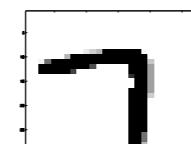
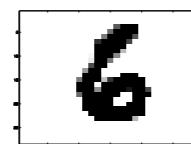
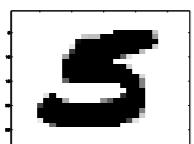
Supervised Learning

Classification: Given labeled data (x_i, y_i)
where y is **discrete**, predict y values for unseen x

Example 2: Which digit in the image ?

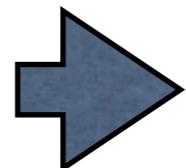


Label: 0,1,..,9



What are the features?

Option: vector of pixel colors



0	0	1	0	1	0	1	...
---	---	---	---	---	---	---	-----

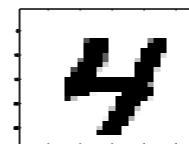
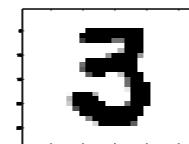
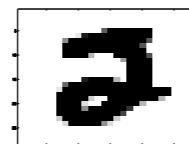
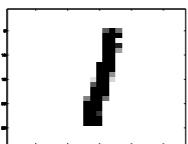
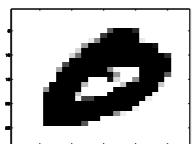
Image

x (0 for white, 1 for black)

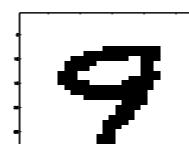
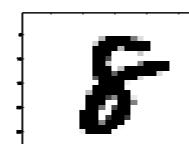
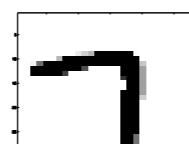
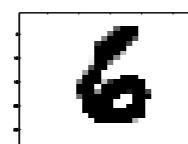
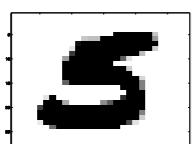
Supervised Learning

Classification: Given labeled data (x_i, y_i)
where y is **discrete**, predict y values for unseen x

Example 2: Which digit in the image ?



Label: 0,1,..,9



What are the features?
Option: vector of pixel colors

There are other options too

Lesson: Choosing features is non-trivial in real applications

Supervised Learning

Classification: Given labeled data (x_i, y_i)
where y is **discrete**, predict y values for unseen x

Example 3: Spam or not?

Email 1

From: Canadian Pharmacy
Subject: Offer ends now!

Email 2

From: Yuncong Chen
Subject: TA meeting

	Pharmacy	offer	meeting	TA	Spam?
Email 1	1	1	0	0	Yes
Email 2	0	0	1	1	No

Label: 0 (not spam), 1 (spam)

Features: Words in the email

Supervised Learning

Regression:

Given data:

$$(x_i, y_i) \quad i=1, \dots, n$$

independent dependent
variable variable

where y is **continuous**, design a rule to predict y values for unseen
 x

Supervised Learning

Regression: Given data (x_i, y_i)
where y is **continuous**, predict y values for unseen x

Example I: Predict house price from properties of house

Bedrooms	Bathrooms	Area	Price
3	2	2000	600K
2	1	1200	400K

x **y**

Independent Variable: Property of house

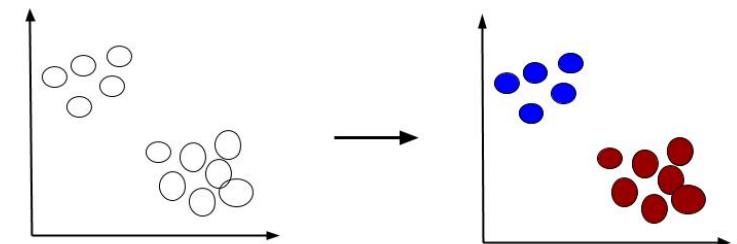
Dependent variable: price

Supervised Learning



Unsupervised Learning

- Given: training data (without labels)
- x_1, x_2, \dots, x_n (without labels)
- Output hidden structure behind the x 's



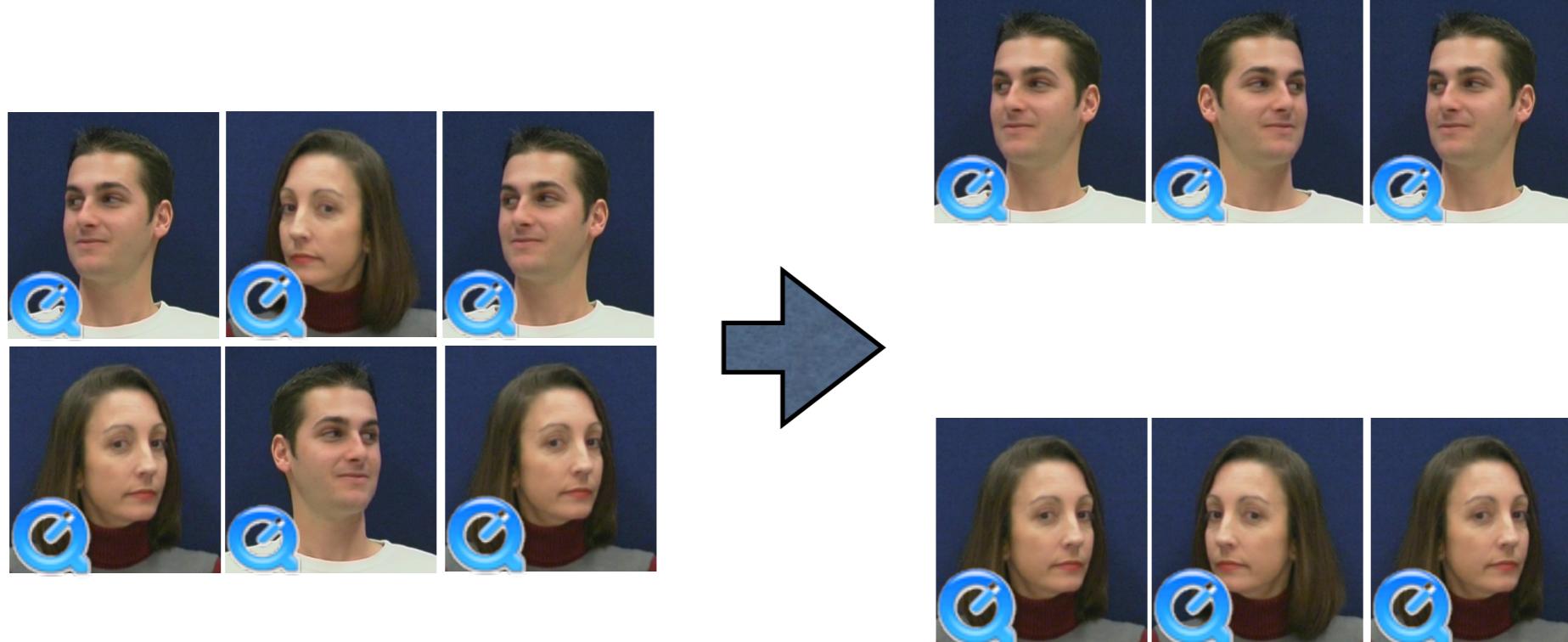
Clustering

Unsupervised Learning

Clustering

Given a set of input objects, group them to clusters by similarity

Example I: Cluster videos by people in them

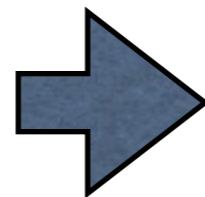
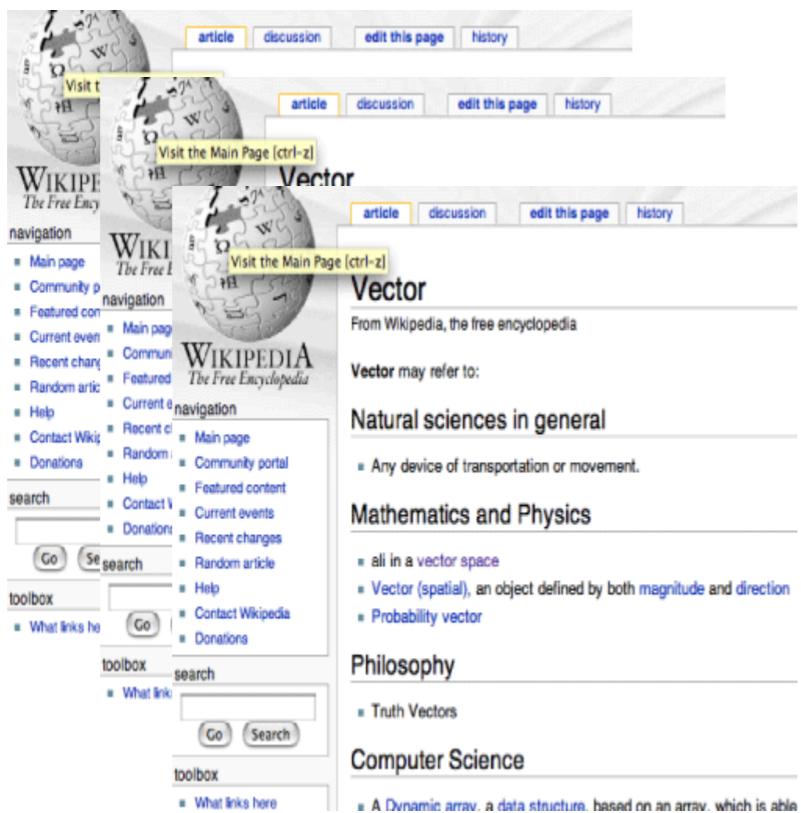


Unsupervised Learning

Clustering

Given a set of input objects, group them to clusters by similarity

Example 2: Cluster documents by topic



Physics

Gravity

Laws of Motion

Electricity

Math

Geometry

Algebra

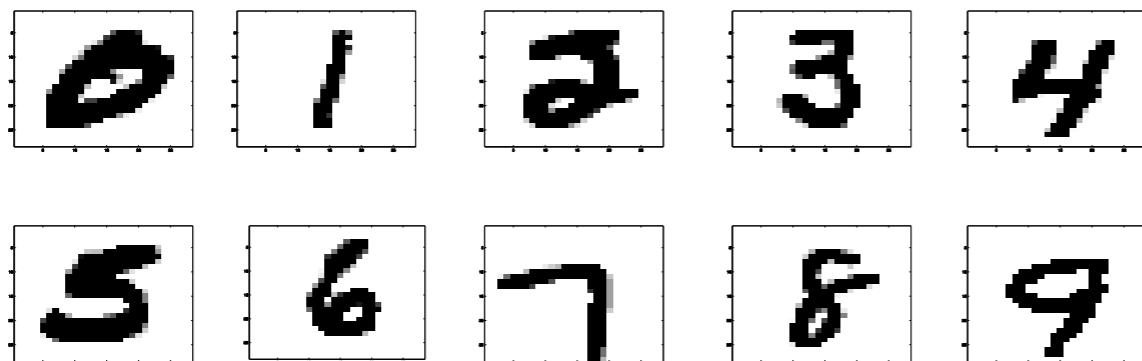
Features: Words in the document

Unsupervised Learning

Dimensionality Reduction

Given high dimensional data, find a good low dimensional representation

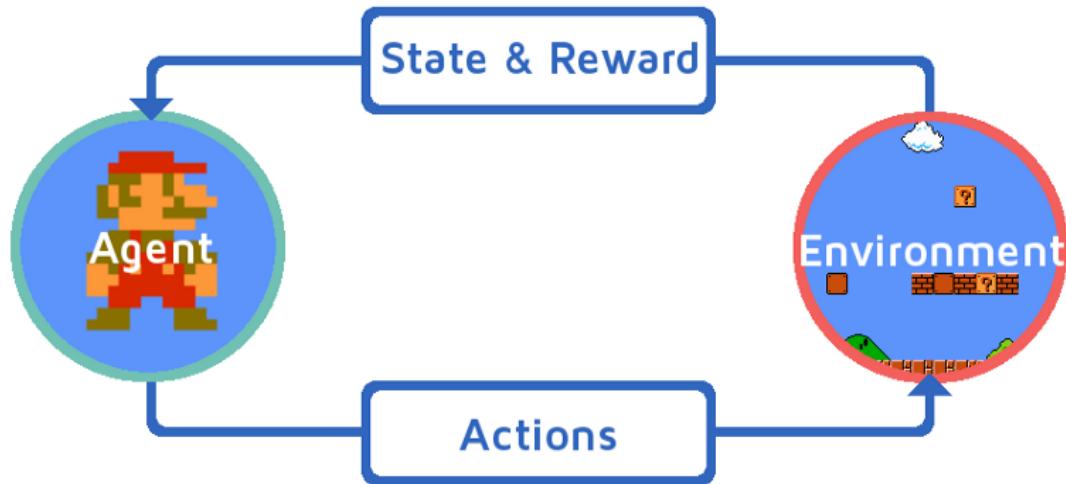
Example I: Images



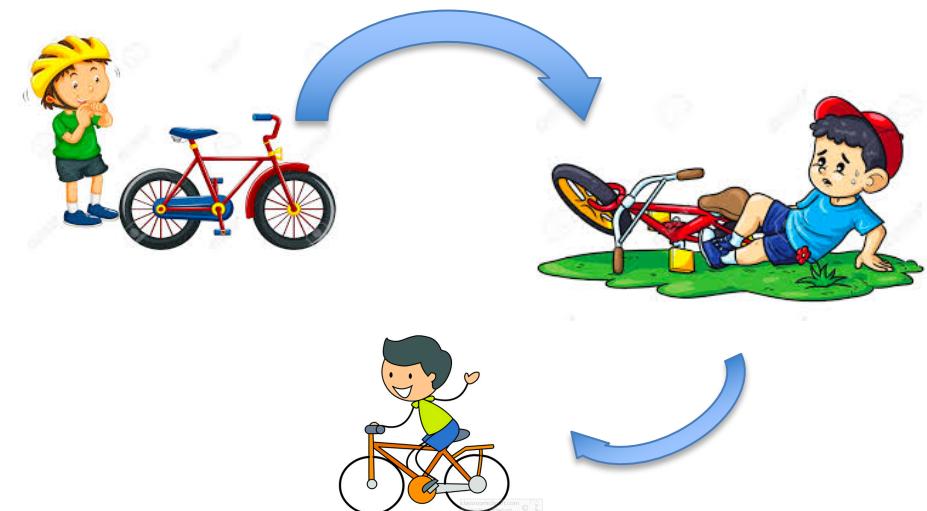
Number of pixels = 768, so 768-dimensional object

Can we find a lower dimensional representation?

Reinforcement Learning



Rewards from sequence of actions



Applications of Learning Paradigms

Supervised

- Person identification
- Object recognition
- Stock prediction

Reinforcement

- Game playing
- Credit assignment

Unsupervised

- Social network analysis
- Dimensionality reduction
- Market segmentation