Empirical Learning

Ravi Kothari, Ph.D.
ravi.kothari@ashoka.edu.in

"Let us see what is out there.."

Learning from Examples

Learning from Examples

Learning from Examples

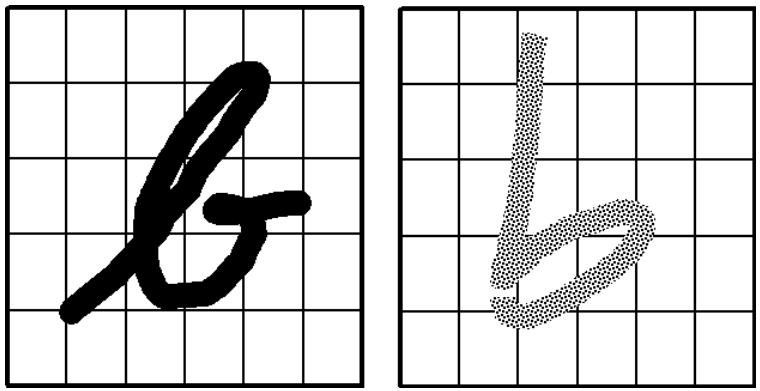
- In many situations, enough is not known (or it is too cumbersome) to construct first-principles based models

Learning from Examples

- In many situations, enough is not known (or it is too cumbersome) to construct first-principles based models
- Try designing an optical character recognition system to recognize characters written by different people, with different instruments, on different types of paper, with a variable amount of soiling etc.

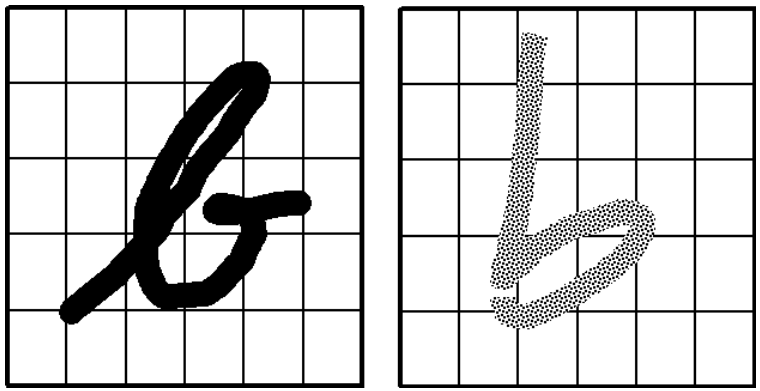
Learning from Examples

- In many situations, enough is not known (or it is too cumbersome) to construct first-principles based models
- Try designing an optical character recognition system to recognize characters written by different people, with different instruments, on different types of paper, with a variable amount of soiling etc.



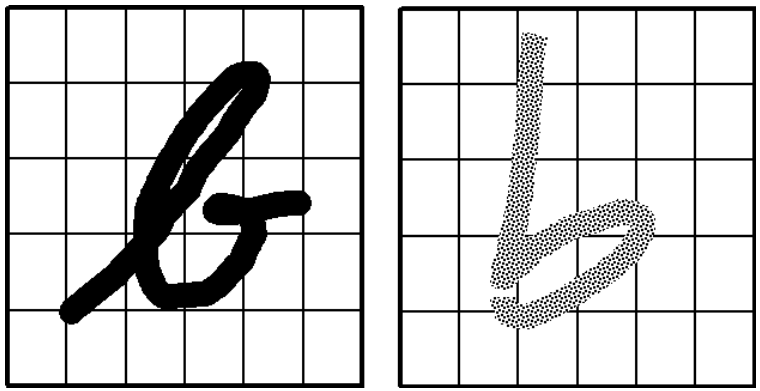
Learning from Examples

- In many situations, enough is not known (or it is too cumbersome) to construct first-principles based models
- Try designing an optical character recognition system to recognize characters written by different people, with different instruments, on different types of paper, with a variable amount of soiling etc.



Learning from Examples

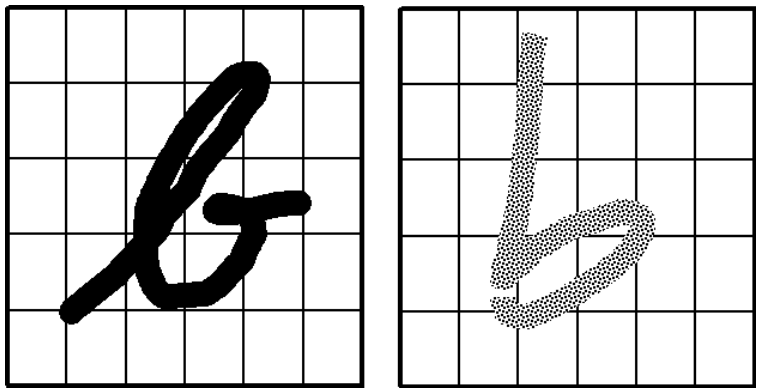
- In many situations, enough is not known (or it is too cumbersome) to construct first-principles based models
- Try designing an optical character recognition system to recognize characters written by different people, with different instruments, on different types of paper, with a variable amount of soiling etc.



- How about a system for *recognizing faces*, for *driverless car*?

Learning from Examples

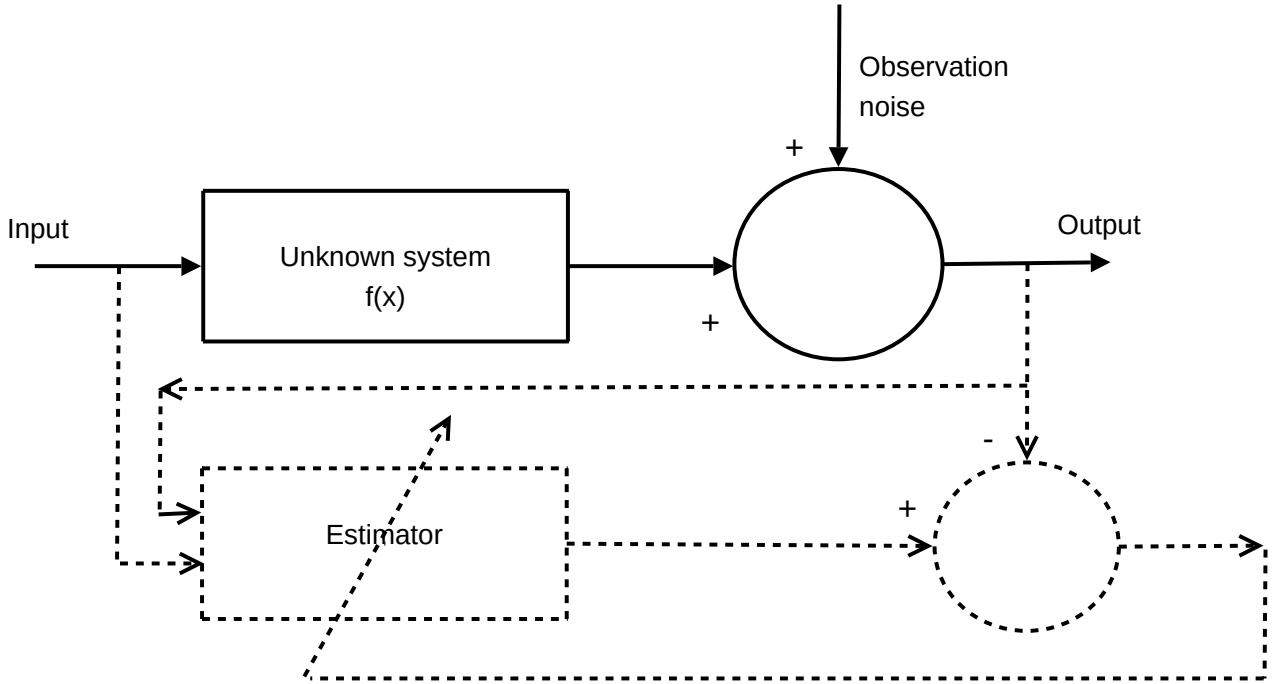
- In many situations, enough is not known (or it is too cumbersome) to construct first-principles based models
- Try designing an optical character recognition system to recognize characters written by different people, with different instruments, on different types of paper, with a variable amount of soiling etc.



- How about a system for *recognizing faces*, for *driverless car*?
- *Empirical model construction* becomes very attractive

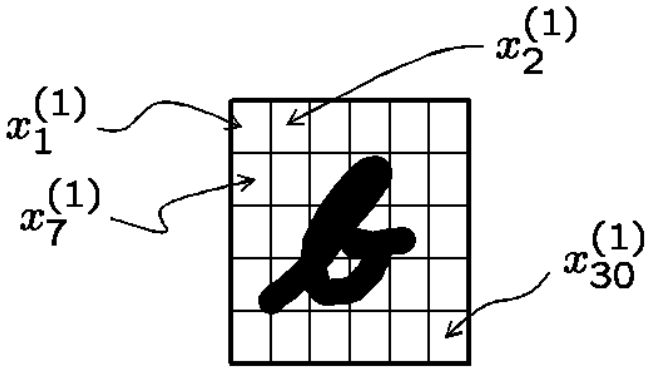
Empirical Learning – A Typical Setting

Empirical Learning – A Typical Setting



Observed Data

Observed Data

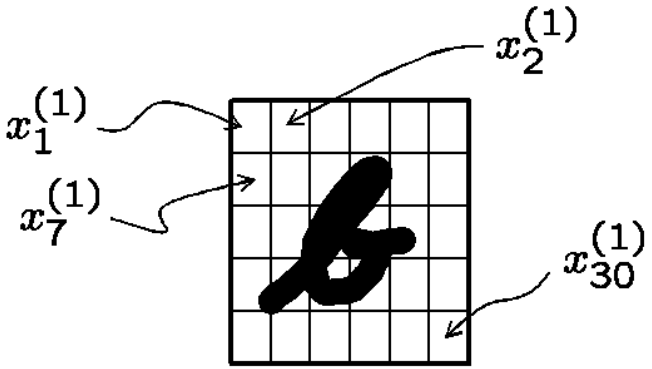


$$\mathbf{x}^{(1)} = [x_1^{(1)}\ x_1^{(1)}\ \ldots\ x_n^{(1)}]$$

0	0	…	1	0	$\mathbf{x}^{(1)}$ A character from John's writing
0	1	…	1	1	$\mathbf{x}^{(2)}$ A character from Mahesh's writing
		⋮			
1	1	…	0	0	$\mathbf{x}^{(N)}$ Another character from John's writing

Often called the input or the feature vector

Observed Data



$$\mathbf{x}^{(1)} = [x_1^{(1)}\ x_1^{(1)}\ \ldots\ x_n^{(1)}]$$

0	0	…	1	0
0	1	…	1	1
		⋮		
1	1	…	0	0

$\mathbf{x}^{(1)}$ A character from John's writing

$\mathbf{x}^{(2)}$ A character from Mahesh's writing

$\mathbf{x}^{(N)}$ Another character from John's writing

Often called the input or the feature vector

- Observed data may be binary, real, categorical, ordered (and noisy!)

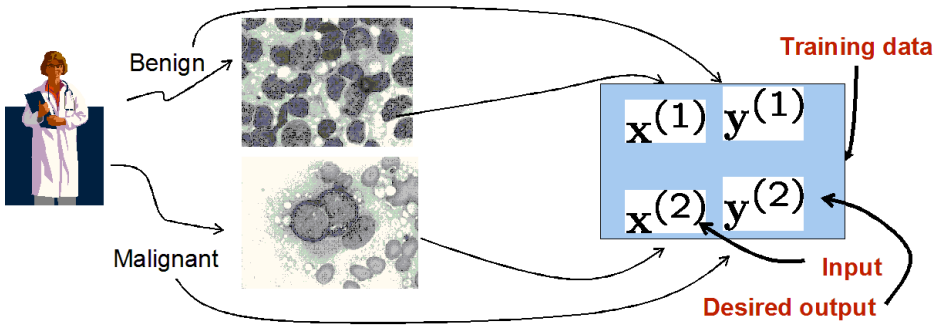
(Un)Supervised Learning

(Un)Supervised Learning

- Supervised learning: A class label is also available

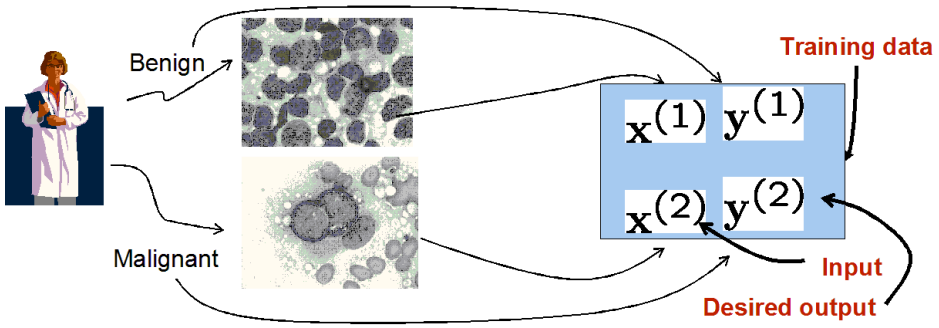
(Un)Supervised Learning

- Supervised learning: A class label is also available



(Un)Supervised Learning

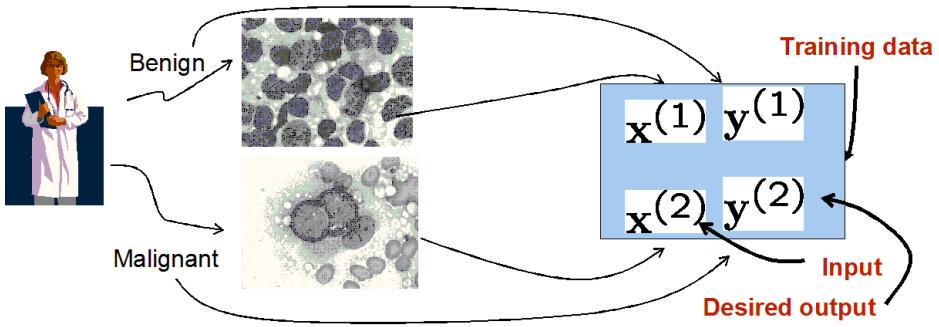
- Supervised learning: A class label is also available



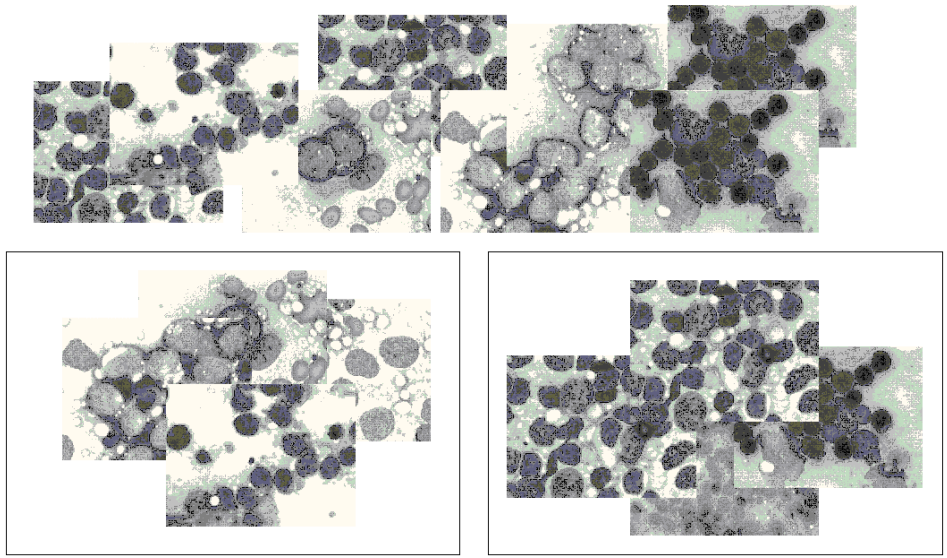
- Unsupervised learning: A class label is not available

(Un)Supervised Learning

- Supervised learning: A class label is also available



- Unsupervised learning: A class label is not available



Some Learning Scenarios

Some Learning Scenarios

- Determining (credit-card or other types of) fraud

Some Learning Scenarios

- Determining (credit-card or other types of) fraud
- Speaker identification, speech recognition

Some Learning Scenarios

- Determining (credit-card or other types of) fraud
- Speaker identification, speech recognition
- Drug design and bio-informatics

Some Learning Scenarios

- Determining (credit-card or other types of) fraud
- Speaker identification, speech recognition
- Drug design and bio-informatics
- Driverless cars

Some Learning Scenarios

- Determining (credit-card or other types of) fraud
- Speaker identification, speech recognition
- Drug design and bio-informatics
- Driverless cars
- Intrusion detection, surveillance

Some Learning Scenarios

- Determining (credit-card or other types of) fraud
- Speaker identification, speech recognition
- Drug design and bio-informatics
- Driverless cars
- Intrusion detection, surveillance
- Automatic Colorization – https://arxiv.org/pdf/1603.08511.pdf

Some Learning Scenarios

- Determining (credit-card or other types of) fraud
- Speaker identification, speech recognition
- Drug design and bio-informatics
- Driverless cars
- Intrusion detection, surveillance
- Automatic Colorization – https://arxiv.org/pdf/1603.08511.pdf
- Automatic Caption Generation – http://cs.stanford.edu/people/karpathy/cvpr2015.pdf

Some Learning Scenarios

- Determining (credit-card or other types of) fraud
- Speaker identification, speech recognition
- Drug design and bio-informatics
- Driverless cars
- Intrusion detection, surveillance
- Automatic Colorization – https://arxiv.org/pdf/1603.08511.pdf
- Automatic Caption Generation – http://cs.stanford.edu/people/karpathy/cvpr2015.pdf
- Machine Translation – https://research.googleblog.com/2015/07/how-google-translate-squeezes-deep.html

Learning – A Formal Definition

Based on N (possibly noisy) observations $\mathcal{X} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$ *of the input and output of a fixed though unknown system* $f(x)$*, construct an estimator* $\hat{f}(x;\theta)$ *so as to minimize,*

$$E\left[\left(L(f(x) - \hat{f}(x;\theta))\right)\right]$$

Aspects of Empirical Models

Aspects of Empirical Models

- How do we know if the estimator is consistent?

Aspects of Empirical Models

- How do we know if the estimator is consistent?
- How do we estimate the rate of convergence?

Aspects of Empirical Models

- How do we know if the estimator is consistent?
- How do we estimate the rate of convergence?
- How do we estimate the prediction error or validate the model?