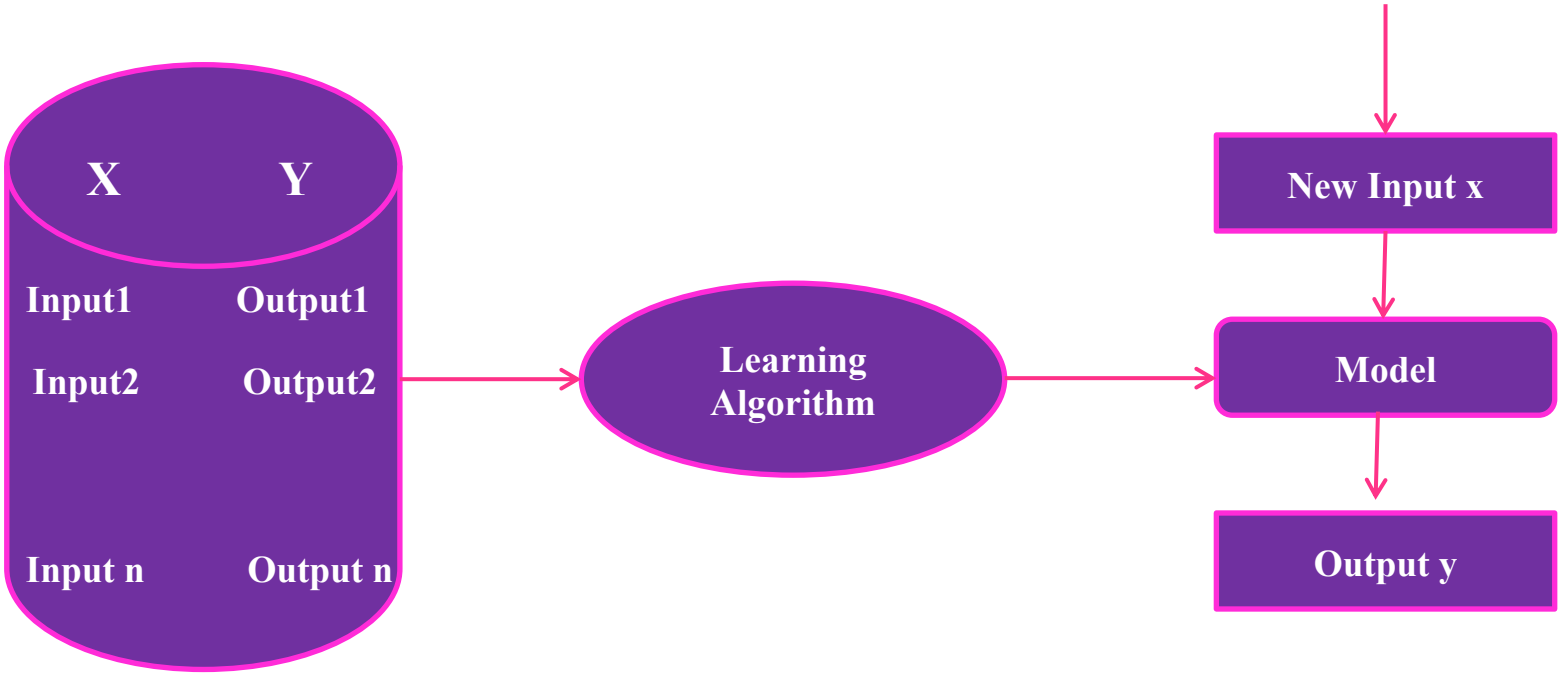
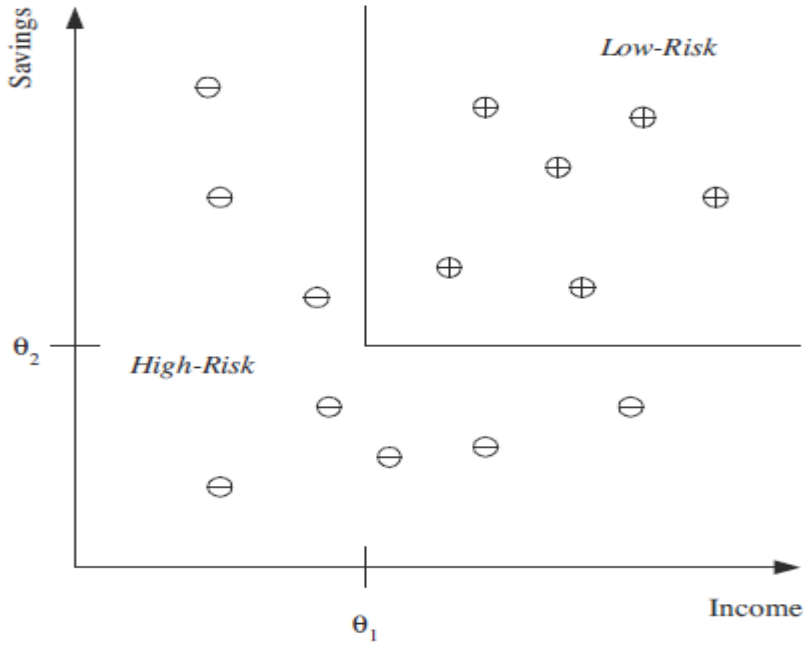
Supervised Learning

Supervised Learning



Supervised Learning

- **Classification**
  - Credit Scoring (Prediction)



Supervised Learning

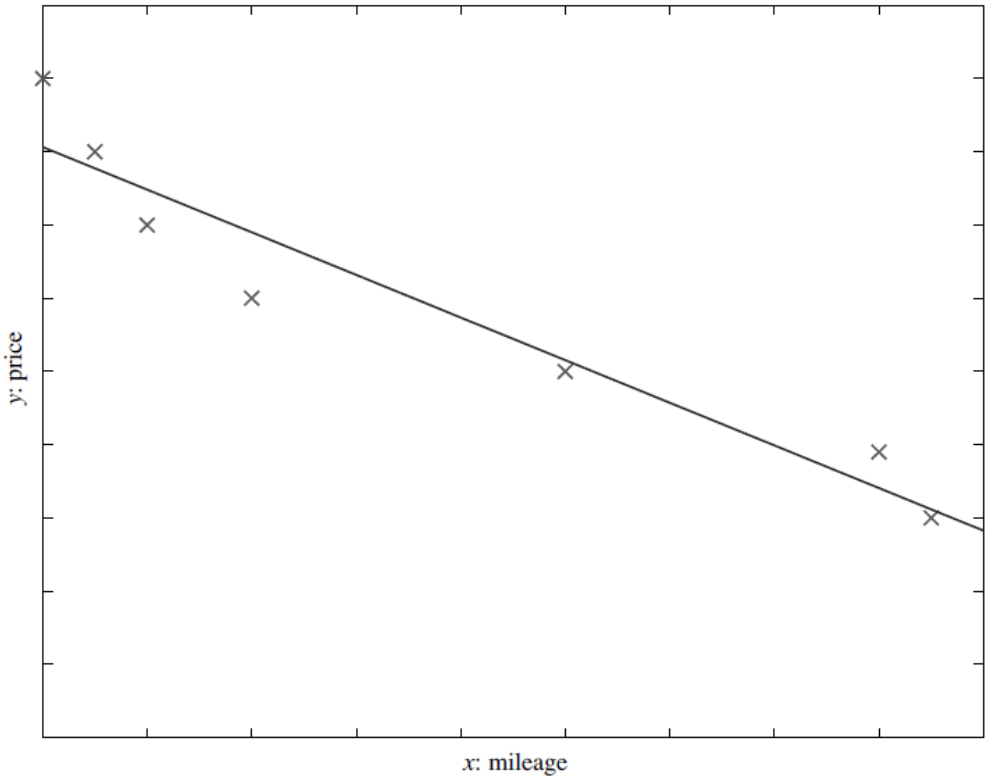
- *optical character recognition*
- *face recognition*
- *medical diagnosis*
- *speech recognition*

Supervised Learning

- Discriminant
  - A function that separates the examples of different classes
- Knowledge extraction
  - Learning a rule from data
  - Rule is a simple model that explains the data
    - Eg: if the discriminant separating low-risk and high-risk customers is identified/learned, then the knowledge of the properties of low-risk customers can be easily understood.
- Compression

Supervised Learning

- **Regression**



Example Data

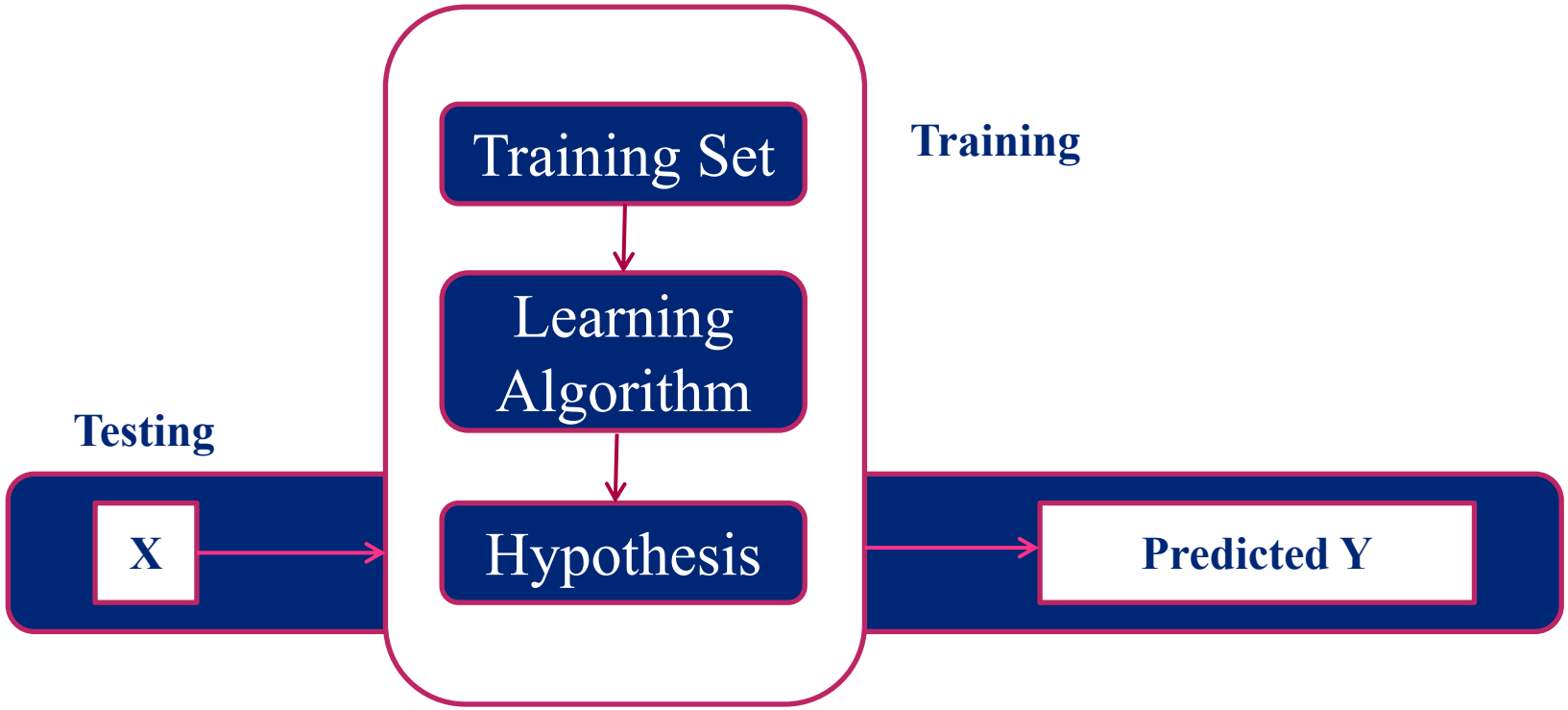
Training Examples:

	Action	Author	Thread	Length	Where
e1	skips	known	new	long	Home
e2	reads	unknown	new	short	Work
e3	skips	unknown	old	long	Work
e4	skips	known	old	long	home
e5	reads	known	new	short	home
e6	skips	known	old	long	work

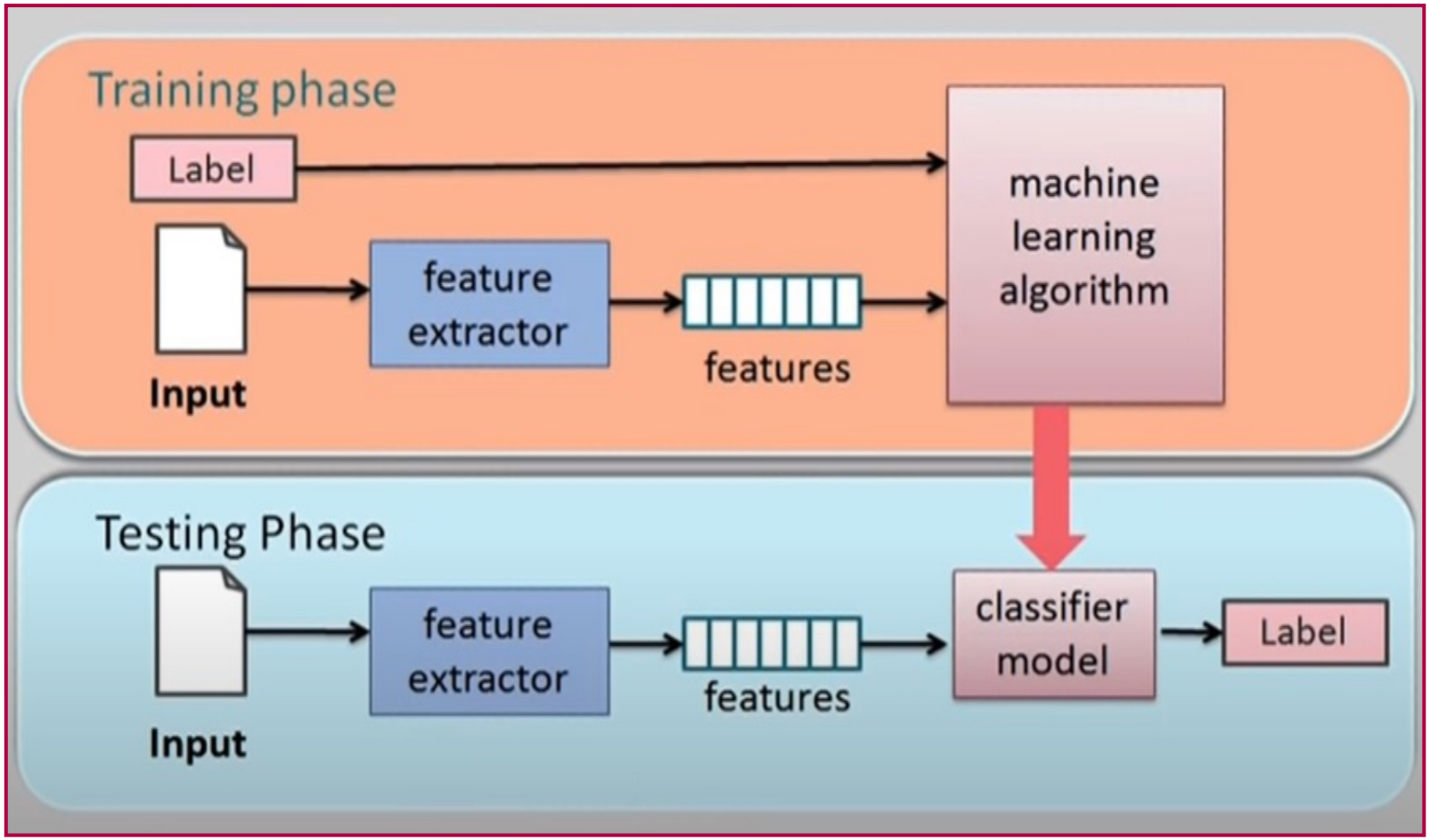
New Examples:

e7	???	known	new	short	work
e8	???	unknown	new	short	work

Schematic Diagram



Detailed Diagram



Classification Learning

- Task T
  - Input
  - Output
- Performance Metric P
- Experience E

Classification Learning

- Task T
  - Input : a set of instances $d_1 \ldots d_n$
  - Output : a set of predictions , $\hat{y}_1, \ldots, \hat{y}_n$
- Performance Metric P
  - Probability(Wrong Prediction) on examples from D
- Experience E
  - A set of labelled examples (x,y)where y is the true label for x
  - Examples should be sampled from some fixed distribution D

Classification Learning

Task	Instance	Labels
medical diagnosis	**patient record:** blood pressure diastolic, blood pressure systolic, age, sex (0 or 1), BMI, cholesterol	{-1,+1} = low, high risk of heart disease
finding company names in text	**a word in context:** capitalized (0,1), word-after-this-equals-Inc, bigram-before-this-equals-acquired-by, …	{first,later,outside} = first word in name, second or later word in name, not in a name
brain-human-interface	**brain state**: neural activity over the last 100ms of 96 neurons	{n,s,e,w,ne,se,nw,sw} = direction you intend to move the cursor

Classification Learning

Task	Instance	Labels	Getting data
medical diagnosis	**patient record:** lab readings	risk of heart disease	wait and look for heart disease
finding company names in text	**a word in context:** capitalized, nearby words, ...	{first,later,outside}	
brain-human-interface	**brain state**: neural activity over the last 100ms of 96 neurons	{n,s,e,w,ne,se,nw,sw}	

Classification Learning

Task	Instance	Labels	Getting data
medical diagnosis	**patient record:** lab readings	risk of heart disease	wait and look for heart disease
finding company names in text	**a word in context:** capitalized, nearby words, ...	{first,later,outside}	text with company names highlighted (say by hand)
brain-human-interface	**brain state**: neural activity over the last 100ms of 96 neurons	{n,s,e,w,ne,se,nw,sw}	

Classification Learning

Task	Instance	Labels	Getting data
medical diagnosis	**patient record:** lab readings	risk of heart disease	wait and look for heart disease
finding company names in text	**a word in context:** capitalized, nearby words, ...	{first,later,outside}	text with company names highlighted (say by hand)
brain-human-interface	**brain state**: neural activity over the last 100ms of 96 neurons	{n,s,e,w,ne,se,nw,sw}	recordings of someone doing known tasks (so direction can be inferred)

Components

Every ML algorithm has three components:

- Representation
- Optimization
- Evaluation

Components

Representation

- **Numerical functions**
  - Linear regression
  - Neural networks
  - Support vector machines
- **Symbolic functions**
  - Decision trees
  - Rules in propositional logic
  - Rules in first-order predicate logic

Components

Representation

- Instance-based functions
  - Nearest-neighbour
  - Case-based
- Probabilistic Graphical Models
  - Naïve Bayes
  - Bayesian networks
  - Hidden-Markov Models (HMMs)
  - Probabilistic Context Free Grammars (PCFGs)
  - Markov networks

Components

Optimization

- **Gradient descent**
  - Perceptron
  - Backpropagation
- **Dynamic Programming**
  - HMM Learning
  - PCFG Learning
- **Divide and Conquer**
  - Decision tree induction
  - Rule learning
- **Evolutionary Computation**
  - Genetic Algorithms (GAs)
  - Genetic Programming (GP)
  - Neuro-evolution

Components

Evaluation

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- K-L divergence

Exercises

- Given 1 million images with their ground truth values. The data consists of images and its corresponding label pairs.

Exercises

- Given 1 million images with their ground truth values. The data consists of images and its corresponding label pairs.
- While looking into the dataset, it is found that the images are from a wide range of categories such as vehicles, plants, animals, human faces and so on.
    - **Which algorithm to use and how?**

Exercises

- Given the task of building an automated taxi.
  - Define the constraints.
  - What are the inputs?
  - What is the output?
  - How can you communicate with the passenger?
  - Do you need to communicate with the other automated taxis, that is, do you need a “language”?