Brief Intro to Machine Learning

What is Machine Learning?

- From [Simon 83]
  - Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more effectively the next time
- From [Nilsson 96]
  - A machine learns whenever it changes its structure, program, or data in such a manner that its expected future performance improves
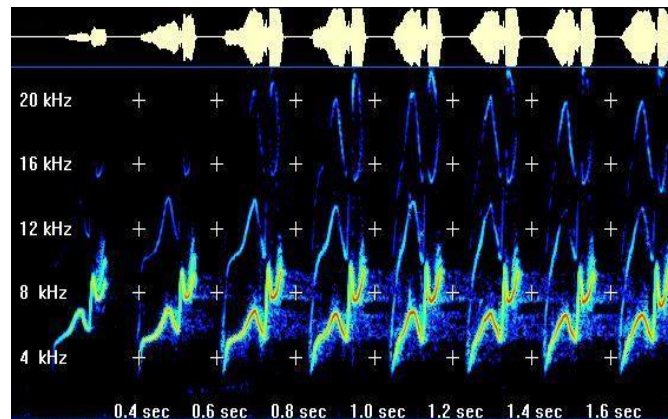
Examples

- Character recognition

- raw data: image
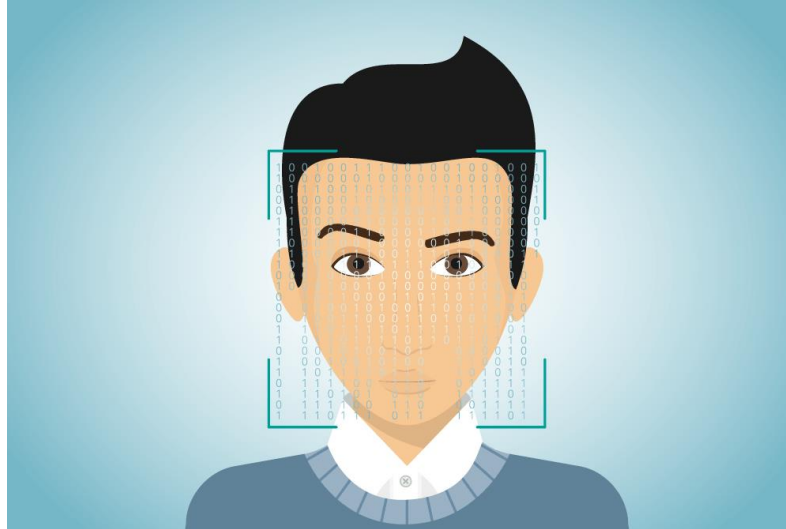
- class: numerals, English (Chinese, etc.) characters

Examples

- Speech recognition

- raw data: speech signal
- class: spoken words

Examples

- Face detection
  - raw data: image
  - class: face/non-face

Examples

- Fingerprint identification
 - raw data: fingerprint image
- class: known/unknown person

Examples

- Document classification
  - raw data: (web) document
  - class: semantic categories

User Guide

The User Guide covers all of pandas by topic area. Each of the subsections introduces a topic (such as "working with missing data"), and discusses how pandas approaches the problem, with many examples throughout.
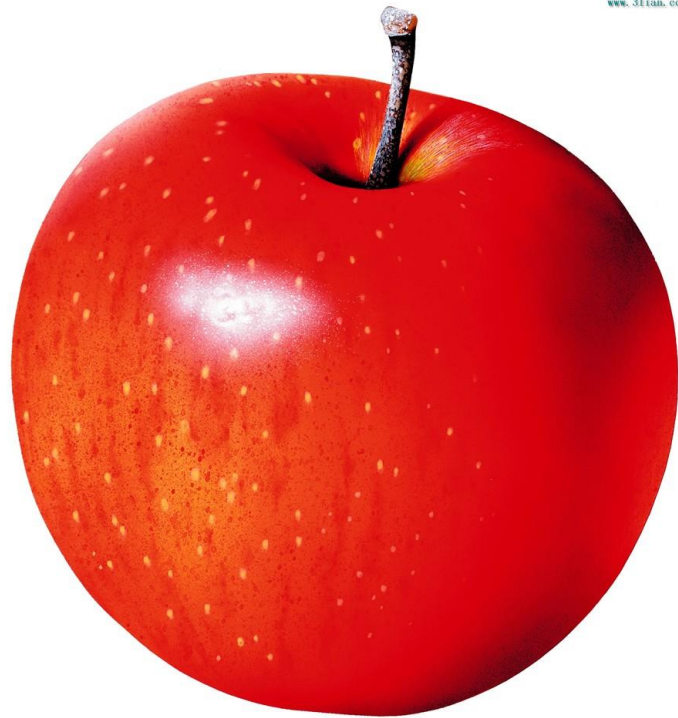
Users brand-new to pandas should start with 10 minutes to pandas.

Further information on any specific method can be obtained in the API reference.

- IO tools (text, CSV, HDF5, …)
  - CSV & text files
  - JSON
  - HTML
  - Excel files
  - OpenDocument Spreadsheets
  - Clipboard

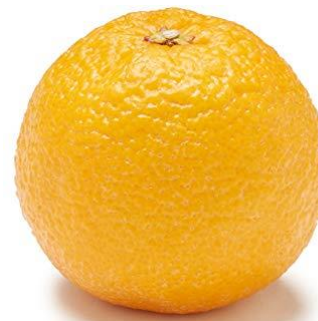# Why can this be solved?

- Is this an apple?

Training examples

Apples

Not Apples

Some terminologies

- **Model:** the collection of parameters you are trying to fit
- **Data:** what you are using to fit the model
- **Target:** the value you are trying to predict with your model
- **Features:** attributes of your data that will be used in prediction
- **Methods:** algorithms that will use your data to fit a model

Features

Automatic fish sorting

Preprocessing

Feature extraction

Classification

"salmon" "sea bass"
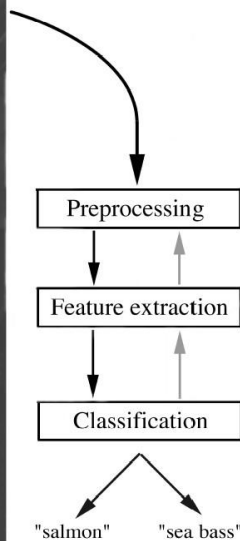
Different types of fish differ in length, lightness, width, number and shape of fins, position of the mouth, etc

Also there are variations in lighting, position of the fish on the conveyer, etc
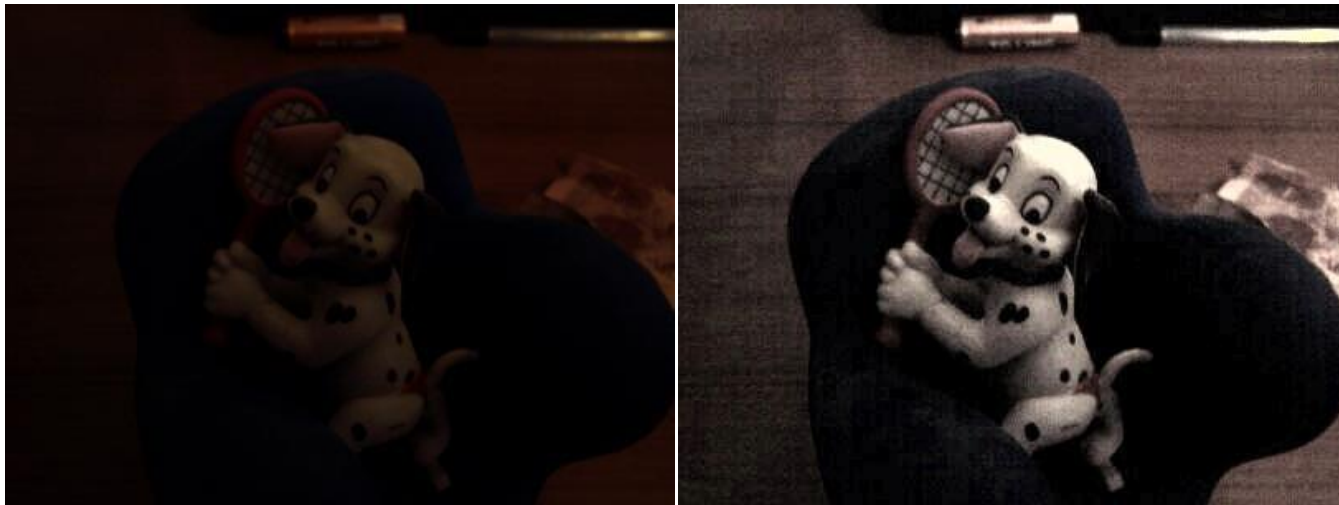
- noise
- needs pre-processing

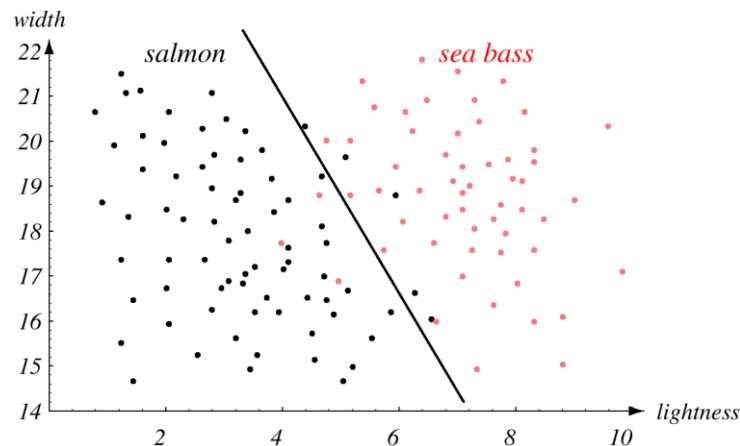# Example of Pre-processing

- Remove background

Example of Pre-processing

• Adjust the light level

Feature extraction

Consider each fish as a point in some feature space
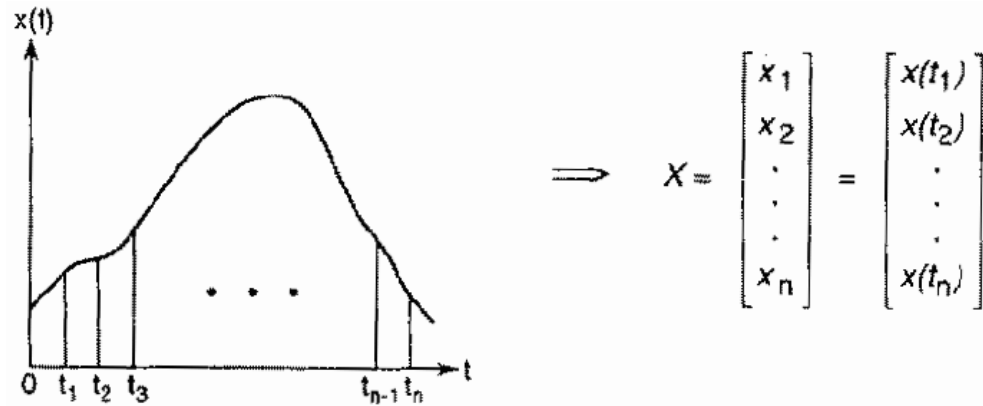

2-dimensional feature space: x = [x1,x2]T
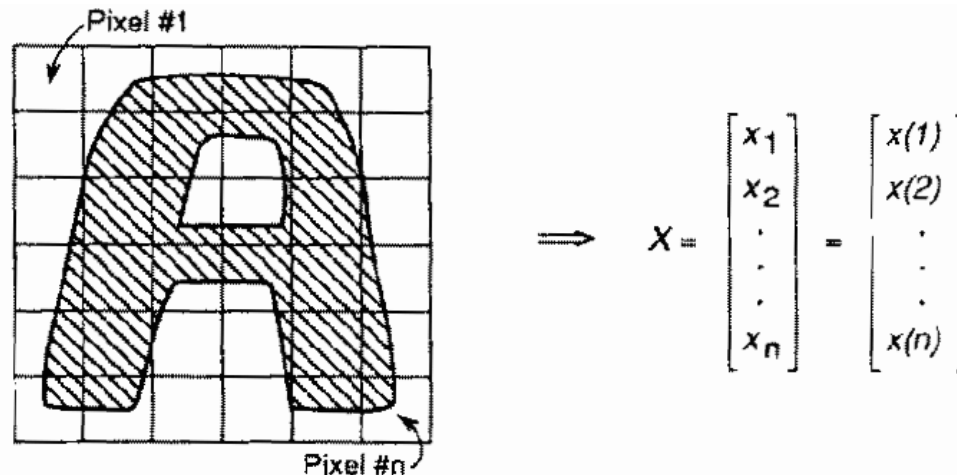Make sure you have a good set of features to work with!

Example (bad feature)

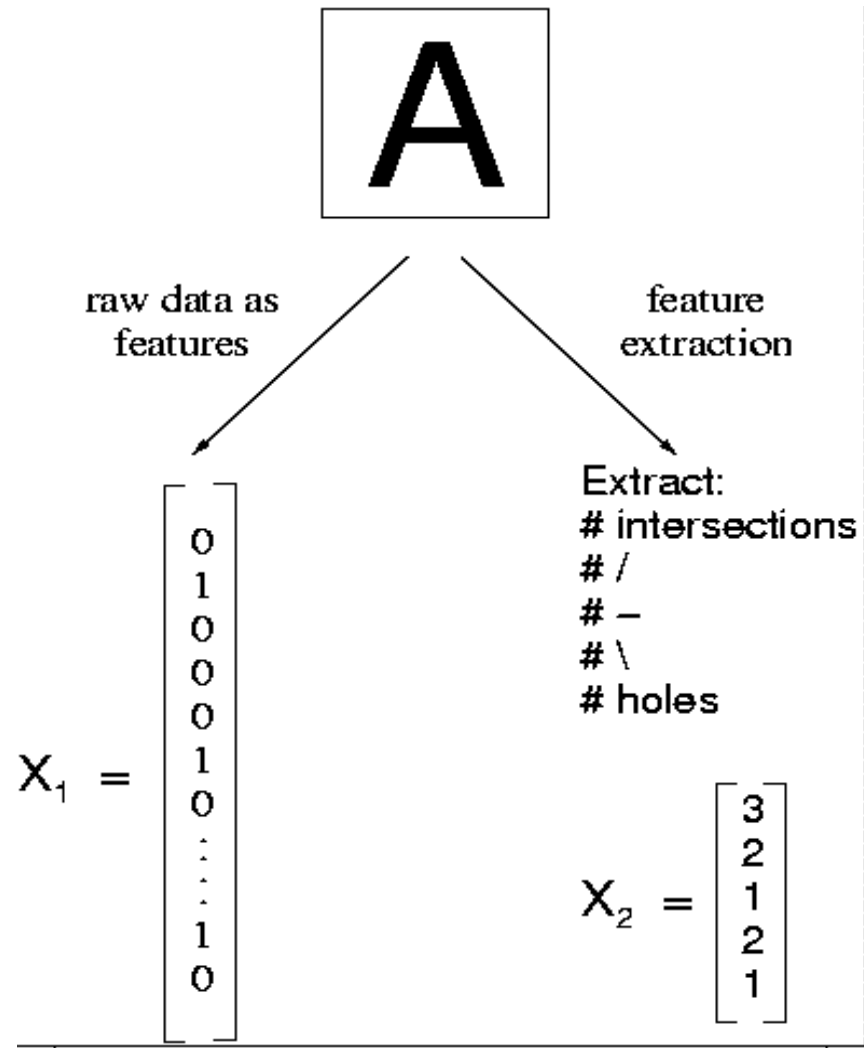Location of a fish on the conveyer belt

Examples of features

waveform

$$\implies X = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = \begin{bmatrix} x(t_1) \\ x(t_2) \\ \cdot \\ \cdot \\ \cdot \\ x(t_n) \end{bmatrix}$$

character

$$\implies X = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = \begin{bmatrix} x(1) \\ x(2) \\ \cdot \\ \cdot \\ \cdot \\ x(n) \end{bmatrix}$$
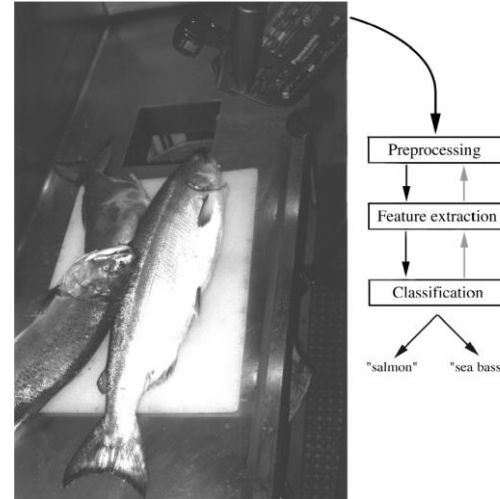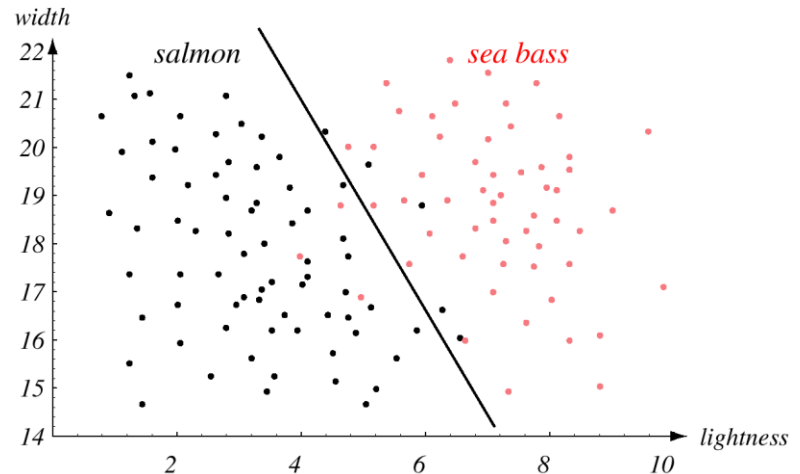
Examples of features

Learning

Training sample
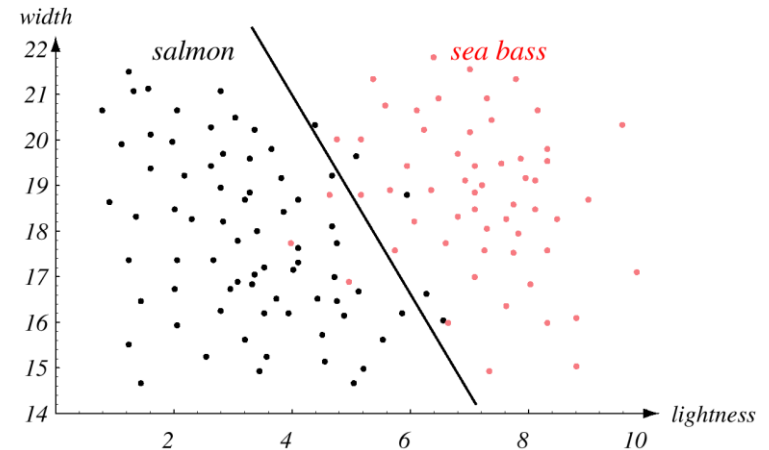- tell me which species a particular fish belongs to (data collection)

Partition the feature space into 2 regions, one for each type of fish
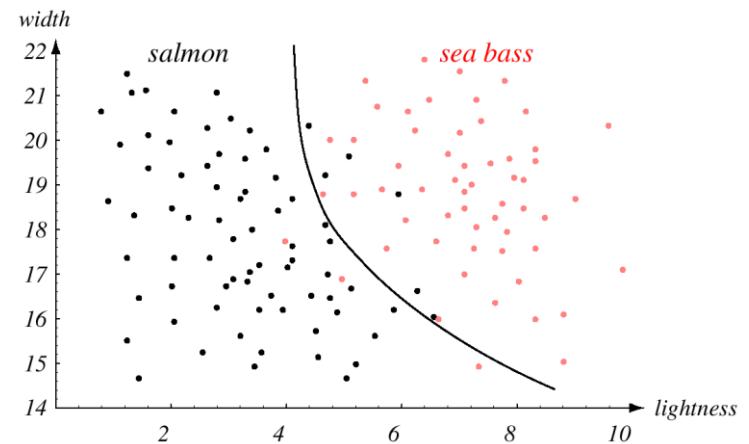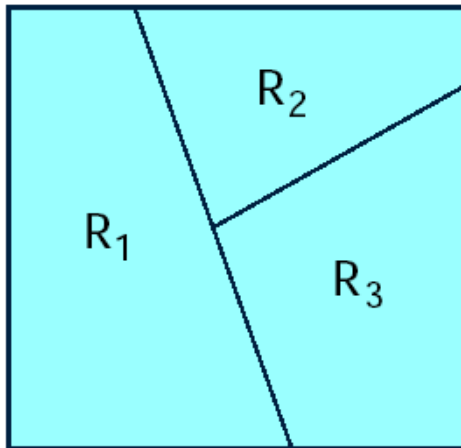- decision boundary

Classifier

Linear classifier

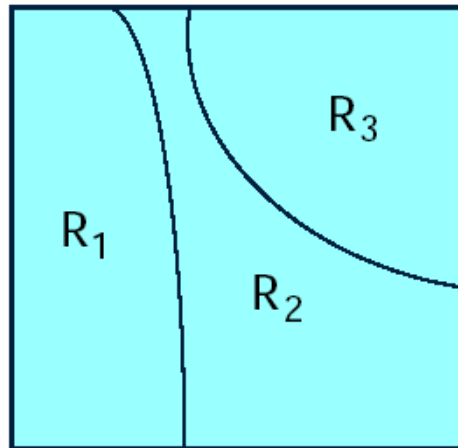Can also use a more complicated decision boundary
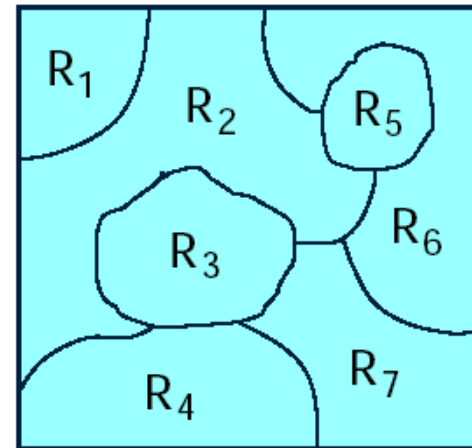   e.g. quadratic classifier

Classifier

can be even more complicated

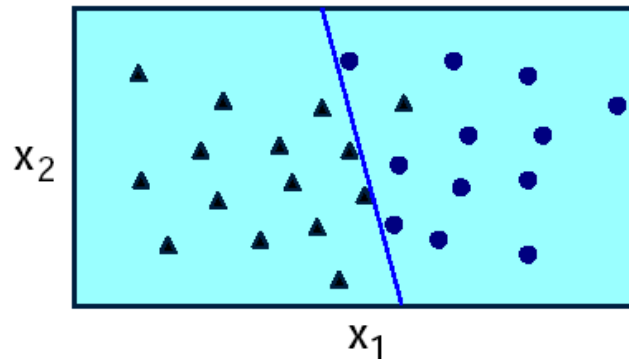linear boundaries   quadratic boundaries   more complicated boundaries

How to handle new images?

- Use the trained classifier to classify the new image

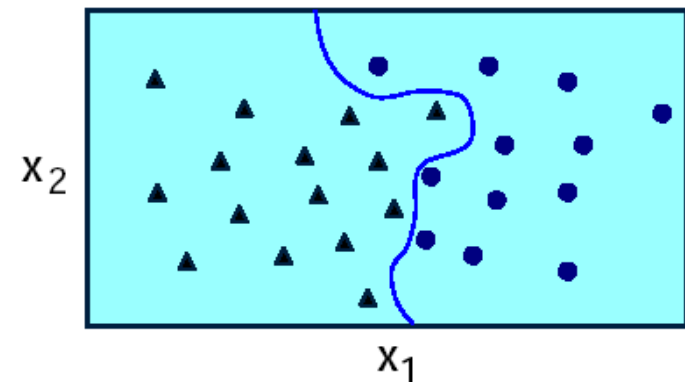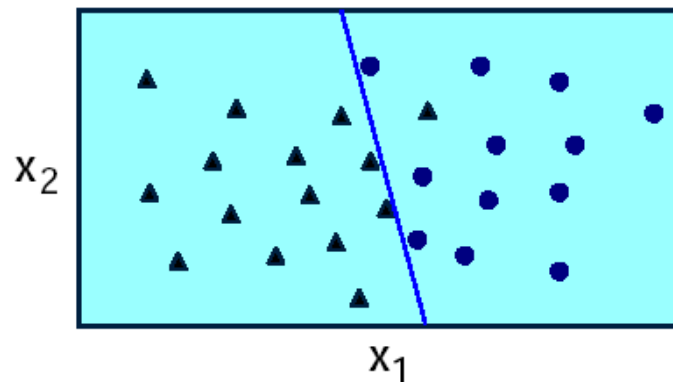Question: How to measure the classifier performance?
- classification error rate
  % patterns that are assigned to the wrong category
- other aspects may be important too
  e.g. computational complexity
  e.g. user-friendliness

How to handle new images?

- Question: will the classifier work for this unseen fish?
  - Issue of generalization

- Example:
  - Which model is better?

Classifier design
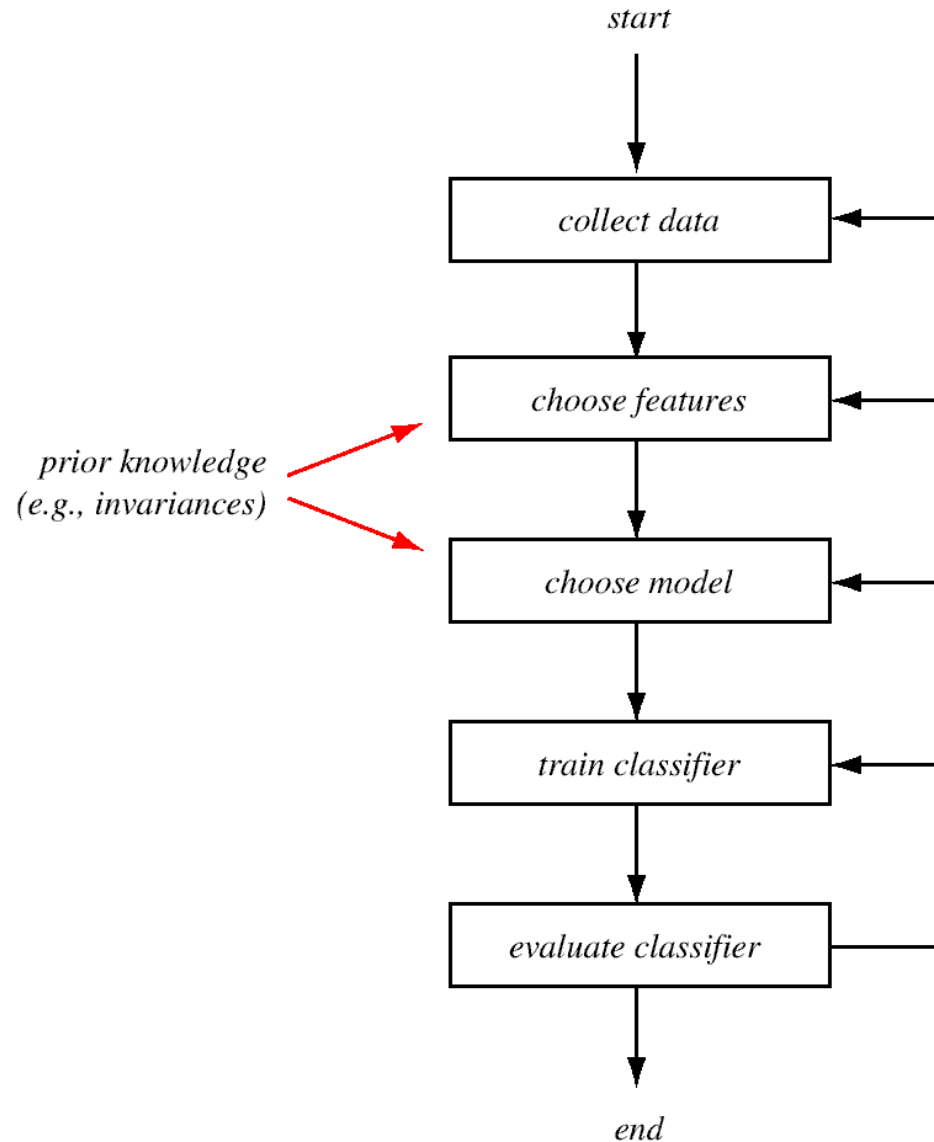
Example: OCR
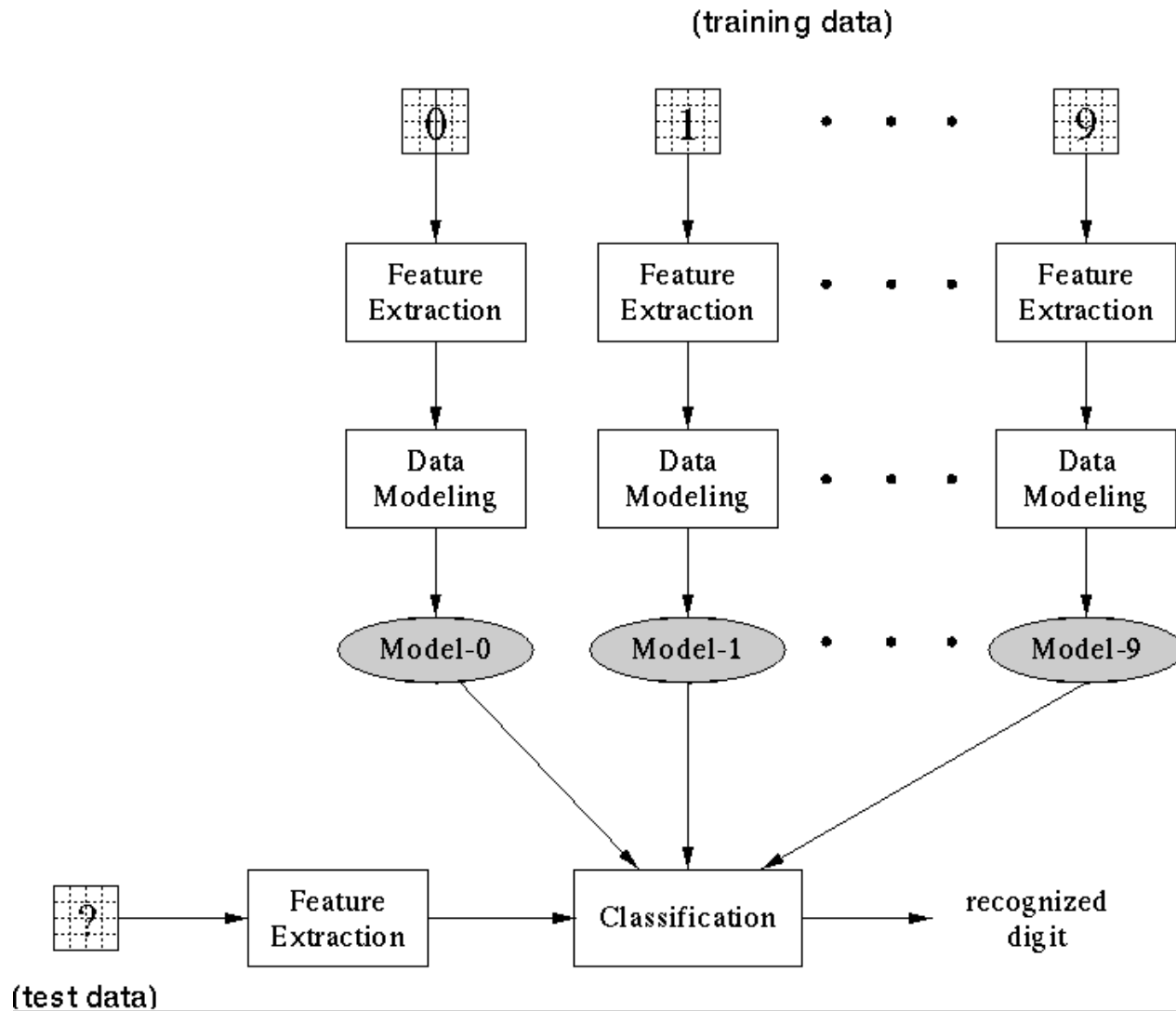
Learning paradigm: supervised learning

- The learner is provided with a set of inputs together with the corresponding desired outputs
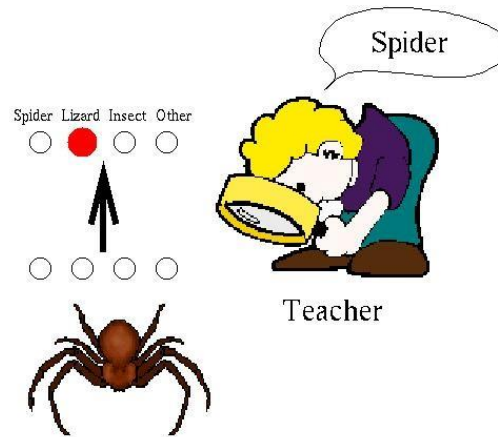
- Has a teacher
- Example:
  - teaching kids to recognize different animals
  - graded examinations with correct answers provided

Example: Recommender System
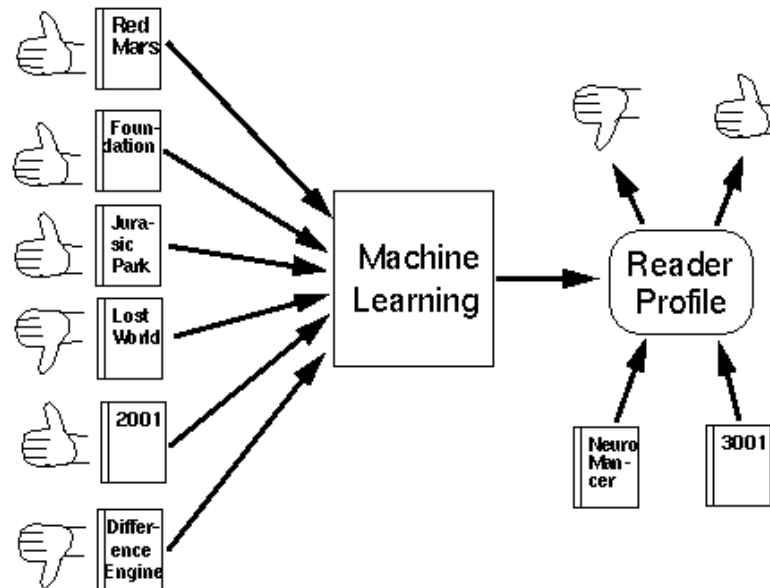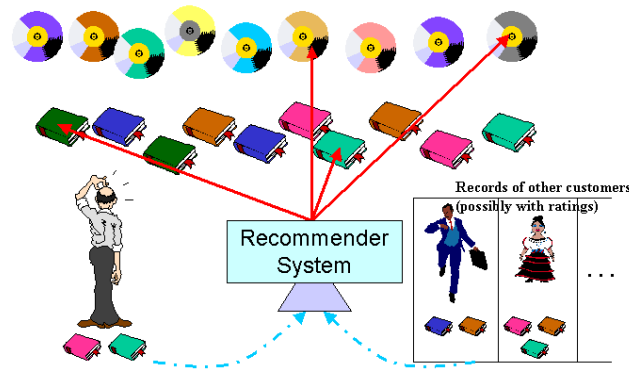
Learning paradigm: unsupervised learning

- Training examples as input patterns, with no associated output

- No teacher

- similarity measure exists to detect groupings / clusterings

Example: Novelty Detection

- Network intrusion detection

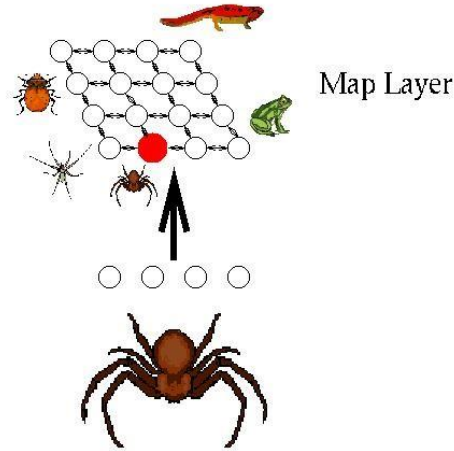| Application	Protocol	Local Address	Remote Address	State	Creation Time	Rx [Bytes]	Type	Tx [Bytes]
SVCHOST.EXE	TCP	all:135	--------------	Listening	27/Feb/2007 09:41:27	0	0	0
SVCHOST.EXE	TCP	all:3389	--------------	Listening	27/Feb/2007 09:41:33	0	0	0
SVCHOST.EXE	TCP	all:1025	--------------	Listening	27/Feb/2007 09:41:29	0	0	0
SYSTEM	TCP	192.168.123.128:139	--------------	Listening	27/Feb/2007 09:41:30	0	0	0
PERSFW.EXE	TCP	all:44334	--------------	Listening	27/Feb/2007 09:41:43	0	0	0
PERSFW.EXE	TCP	all:44334	localhost:1986	Connected In	27/Feb/2007 16:45:42	2931	<attack>	152516
SVCHOST.EXE	TCP	all:5000	--------------	Listening	27/Feb/2007 09:42:15	0	0	0
LSASS.EXE	TCP	all:27155	--------------	Listening	27/Feb/2007 09:42:17	0	0	0
PUTTY.EXE	TCP	all:1898	uststu1.ust.hk:22	Connected Out	27/Feb/2007 15:54:54	445204	<normal>	141398
CCAPP.EXE	TCP	localhost:1035	--------------	Listening	27/Feb/2007 09:42:23	0	0	0
SSHCLIENT.EXE	TCP	all:1639	lcpu2.cse.ust.hk:22	Connected Out	27/Feb/2007 11:43:48	135198	<normal>	3764
ICQ.EXE	TCP	all:21470	--------------	Listening	27/Feb/2007 14:41:02	0	0	0
ICQ.EXE	TCP	all:1759	64.12.24.205:5190	Connected Out	27/Feb/2007 14:41:04	59285	<normal>	12191
MSNMSGR.EXE	TCP	all:1722	by2msg1104003.ph...	Connected Out	27/Feb/2007 14:38:33	176880	<normal>	129012
PFWADMIN.EXE	TCP	all:1986	localhost:44334	Connected Out	27/Feb/2007 16:45:42	278215	<normal>	2931
LSASS.EXE	UDP	all:500	--------------	Listening	27/Feb/2007 09:41:40	0	0	0
SYSTEM	UDP	192.168.123.128:138	--------------	Listening	27/Feb/2007 09:41:30	205	0	40995
LSASS.EXE	UDP	localhost:1033	--------------	Listening	27/Feb/2007 09:42:23	0	0	2
SYSTEM	UDP	192.168.123.128:137	--------------	Listening	27/Feb/2007 09:41:30	1088	0	17982
PERSFW.EXE	UDP	all:44334	--------------	Listening	27/Feb/2007 09:41:43	0	0	0
IEXPLORE.EXE	UDP	localhost:1600	--------------	Listening	27/Feb/2007 11:17:52	1173	0	1173
IEXPLORE.EXE	UDP	localhost:1697	--------------	Listening	27/Feb/2007 14:09:48	15	0	15

Learning Paradigm: Reinforcement Learning
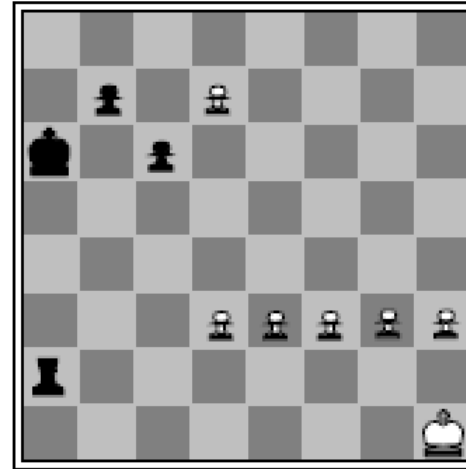
- Training examples as input-output pairs, with evaluative output only
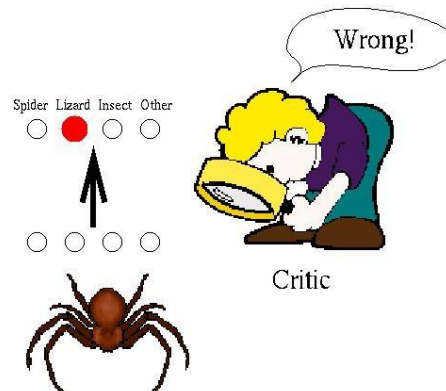
- try to increase the reinforcement it receives
- Example:
  - graded examinations with only overall scores but no correct answers

Example

Game playing

Black to move

Pole balancing