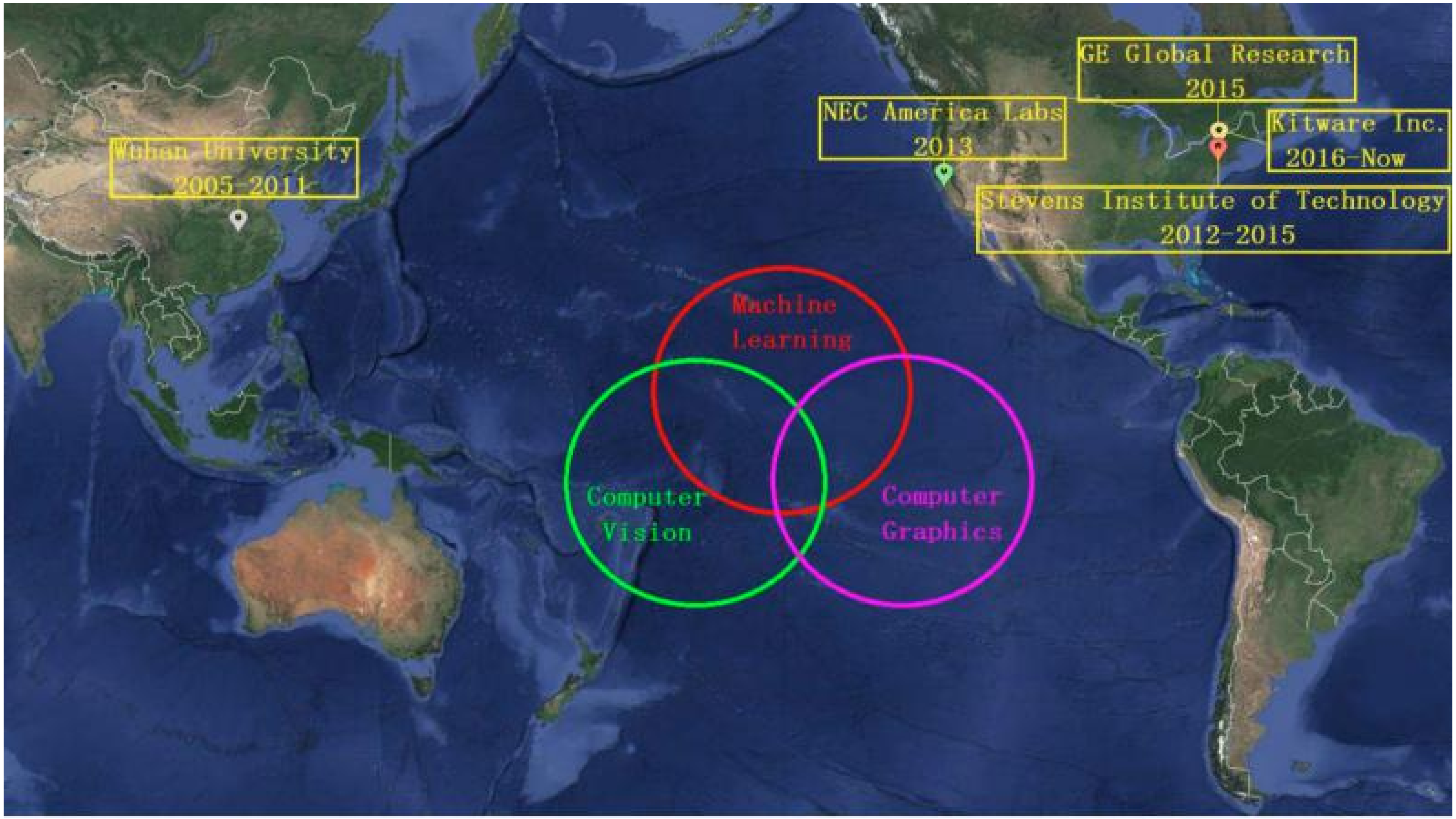
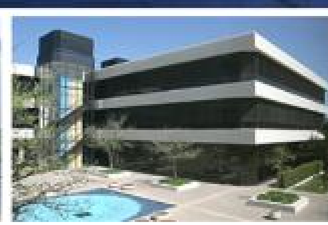
Rensselaer

Lecture 1: Introduction to Pattern Recognition

Dr. Chengjiang Long
Computer Vision Researcher at Kitware Inc.
Adjunct Professor at RPI.
Email: **longc3@rpi.edu**

Self-introduction





Outline

- Course Information
- What is Pattern Recognition?
- Components of a Pattern Recognition System
- Pattern Recognition Design Cycle

Outline

- **Course Information**
- What is Pattern Recognition?
- Components of a Pattern Recognition System
- Pattern Recognition Design Cycle

Course information

- **ECSE 6610** Pattern Recognition
- **Term**: Spring 2018
- **Instructor**: Dr. Chengjiang Long
- **Email:** cjfykx@gmail.com
- **Class time**: 2:00 pm—3:20 pm, Tueseday & Friday
- **Location**: JEC 4107
- **Office Hour**: 3:20 pm—4:00 pm, Tuesday & Friday
- **Office**: JEC 6045.
- **Course Assistant**: Il-Young Son
- **Course Website**: www.chengjianglong.com/teachings.html

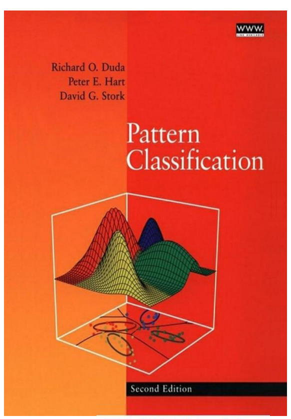
Home Research/Publications Services Teachings Honors About Me

ECSE-6110 Pattern Recognition

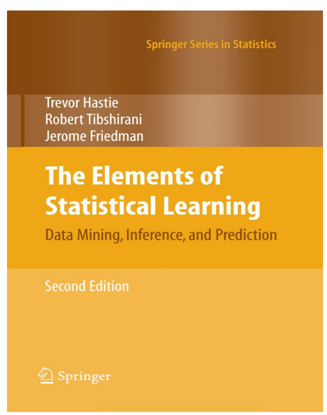
Term: Spring 2018
Instructor: Dr. Chengjiang Long
Time: Tuesday and Friday, 2:00pm – 3:20pm
Building/Room: JONSSN 4107
Office Hour: Tuesday and Friday 3:20pm—4:00pm by appointment
Office Hour Location: TBA
Course Assistant: Il-Young Son
Course Website: www.chengjianglong.com/teachings.html

Topics and textbooks

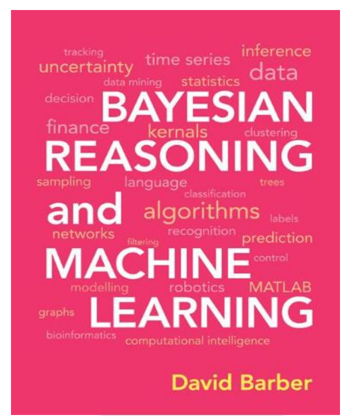
- Introduction to Pattern Recognition.
- Basic graph concepts and Belief network.
- Bayesian decision theory. Maximum-likelihood estimation. Bayesian methods.
- Naive Bayes, Nonparametric techniques.
- Dimension reduction: Principal Component Analysis; Fisher Discrimination.
- Linear models for regression and classification.
- Support Vector Machines.
- Bagging, Random Forests and Boosting.
- Introduction to neural networks and multilayer neural networks.
- Introduction to deep learning: Deep feedforward Network and Convolutional Neural Network.
- Unsupervised learning and clustering.



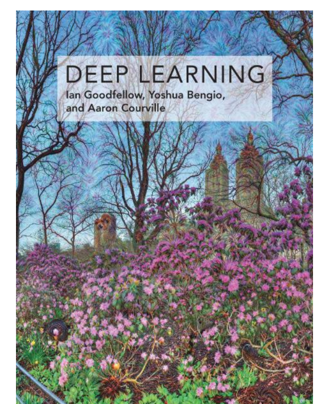
2000 (Duda)



2009 (HTF)



2012 (Barber)



2016 (Ian)

Prerequisites

- Introduction to Pattern Recognition.
- Basic graph concepts and Belief network.
- Bayesian decision theory. Maximum-likelihood estimation. Bayesian methods.
- Naive Bayes, Nonparametric techniques.
- Dimension reduction: Principal Component Analysis; Fisher Discrimination.
- Linear models for regression and classification.
- Support Vector Machines.
- Bagging, Random Forests and Boosting.
- Introduction to neural networks and multilayer neural networks.
- Introduction to deep learning: Deep feedforward Network and Convolutional Neural Network.
- Unsupervised learning and clustering.

- Probability and statistics theory
- Some linear algebra
  - – Must not be afraid of eigenvalues
- Matlab, python, Java or C/C++ programming
  - – This could be "language of your choice", but then you are responsible for debugging etc.
  - – I suggest Matlab or python for short development time.
- Your grade will be affected by any weaknesses in these.

Grading

- Introduction to Pattern Recognition.
- Basic graph concepts and Belief network.
- Bayesian decision theory. Maximum-likelihood estimation. Bayesian methods. **(HW1)**
- Naive Bayes, Nonparametric techniques.
- Dimension reduction: Principal Component Analysis; Fisher Discrimination. **(HW2)**
- Linear models for regression and classification.
- Support Vector Machines. **(HW3)**
- Bagging, Random Forests and Boosting.
- Introduction to neural networks and multilayer neural networks.
- Introduction to deep learning: Deep feedforward Network and Convolutional Neural Network. **(HW4)**
- Unsupervised learning and clustering.

- ❑ Class participation: 5%
- ❑ 4 homework assignments: 20%
- ❑ Midterm exam & Final exam: 40%
- ❑ Final project & Presentation: 35% (5% proposal, 10% presentation, 10% implementation, 10% final report)
- ❑ Final grade: A+(⩾97), A(⩾92), A-(⩾90), B+(⩾87), B(⩾82), B-(⩾80), C+(⩾77), C(⩾72), C-(⩾70), D+(⩾67), D(⩾62), D-(⩾60), F(< 60).

Schedule

Class	Date	Topic	Reading	Homework & Project
0	1/16/2018	Short class		
1	1/19/2018	Introduction to Pattern Recognition	Duda Ch 1	
2	1/23/2018	Probability theory and Linear algebra review	Duda Ch 2	
3	1/26/2018	Bayesian decision theory	Duda Ch 2.3-4	
4	1/30/2018	Basic graph concepts and Belief Network	Duda Ch. 2.5-6, 3.10	
5	2/2/2018	Parameter Estimation – Max Likelihood	Duda Ch 3.1-2, Barber Ch 8	
6	2/6/2018	Parameter Estimation – Bayesian Methods	Duda Ch 3.3-5, Barber Ch 13	HW1 assigned
7	2/9/2018	Non-Parametric Methods – Parzen Estimation	Duda Ch 4.1-3	
8	2/13/2018	Non-Parametric Methods – KNN	Duda Ch 4.4-5	HW1 due
9	2/16/2018	Dimensionality – PCA and Fisher Discrimination	Duda Ch 3.7-8, Barber Ch 15-16	HW2 assigned
10	2/20/2018	Linear Discriminant Functions (1)	Duda Ch 5.1-2, Barber Ch 17	
11	2/23/2018	Linear Discriminant Functions (2)	Duda Ch 5.3-4, Barber Ch 17, HTF Ch 12	HW3 assigned, HW2 due
12	2/27/2018	Perceptron Classification and Learning	Duda Ch 5.5-7, HTF Ch 4.	
13	3/2/2018	Minimum Squared-Error Classification	Duda Ch 5.8-9	HW3 due
14	3/6/2018	Bagging; Random Forests; Boosting (1)	HTF Ch. 10 and 15	
15	3/9/2018	Midterm exam		Midterm exam
	3/13/2018	Spring break		
	3/16/2018	Spring break		
16	3/20/2018	Presentation of Project Proposals to Class		Final Project Proposal
17	3/23/2018	Bagging; Random Forests; Boosting (2)	HTF Ch. 10 and 15	
18	3/27/2018	Introduction to Nueral Networks	Duda Ch 6.1	
19	3/30/2018	Multilayer Neural Networks	Duda Ch 6.2	
20	4/3/2018	Multilayer Neural Networks: Backpropagation	Duda Ch 6.3-4	
21	4/6/2018	Multilayer Neural Networks: Implementation	Duda Ch 6.5-8	
22	4/10/2018	Introduction to Deep Feedforward Nework	Ian Ch 6	
23	4/13/2018	Introduction to Regularization and Optimization	Ian Ch 7-8	
24	4/17/2018	Introduction to Convolutional Neural Network	Ian Ch 9	Final HW4 assigned
25	4/20/2018	Unsupervised Learning	Duda Ch 10.1-5, HTF Ch 14	
26	4/24/2018	Clustering Algorithms	Duda Ch 10.6-7	HW4 due
27	4/27/2018	Hierarchical and On-Line Clustering	Duda Ch 10.9-14	
28	5/1/2018	Presentation of Project Reports to Class		Project Presentation
29	TBA	Final exam		Final exam

Course objective

On completion of the course,

- You should be sufficiently familiar with the formal theoretical structure, notation, and vocabulary of pattern recognition to be able to read and understand current technical literature.

- You will also have experience in the design and implementation of pattern recognition systems and be able to use those methods to program and solve practical problems.

Rules

- **Need to be absent from class?**
  - 1 point per class: please send notification and justification at least 2 days before the class
- **Late submission of homework?**
  - The maximum grade you can get from your late homework decreases 50% per day
- **Zero tolerance on plagiarism**!!
  - The first time you receive zero grade for the assignment
  - The second time you get “F” in your final grade
  - Refer to Rensselaer honor system for your behavior

Outline

- Course Informatition
- What is Pattern Recognition?
- Components of a Pattern Recognition System
- Pattern Recognition Desgin Cycle
- Summary

Human Pattern

- Humans have developed highly sophisticated skills for sensing their environment and taking actions according to
  - what they observe, e.g.,
  - recognizing a face,
  - understanding spoken words,
  - reading handwriting,
  - distinguishing fresh food from its smell.

- **We would like to give similar capabilities to machines.**

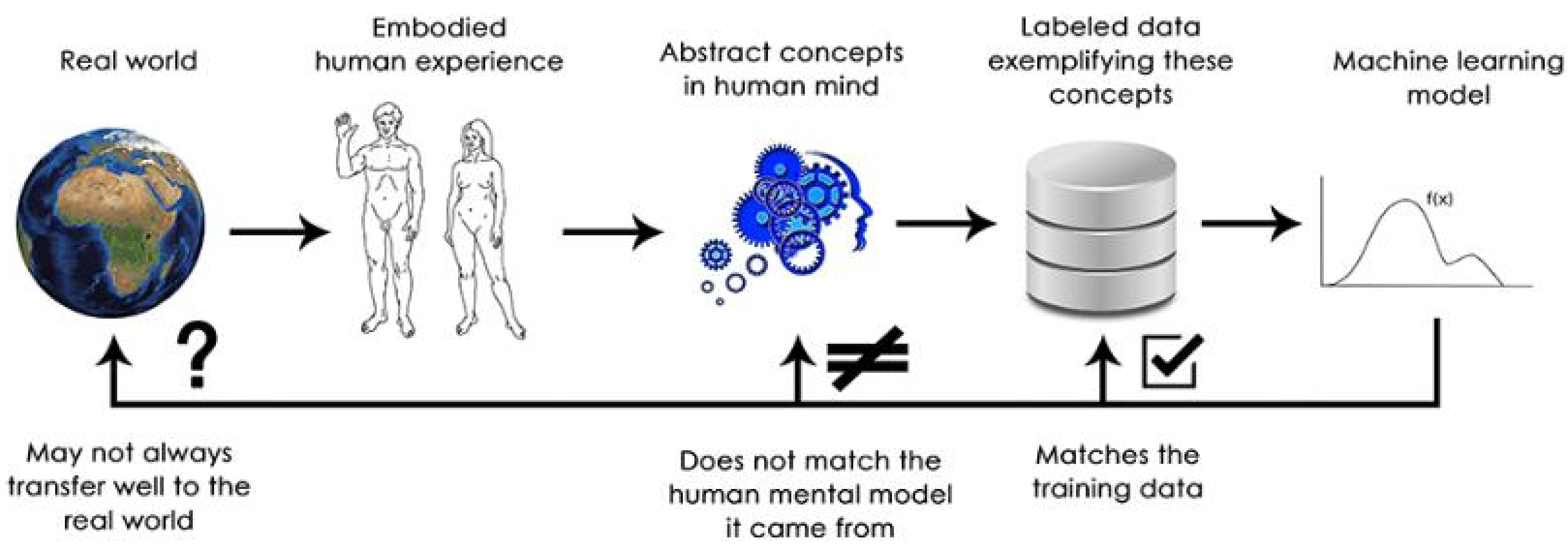
What is Pattern Recognition?

"The assignment of a physical object or event to one of several prespecified categeries"
--Duda & Hart

- A pattern is an entity, vaguely defined, that could be given a name, e.g.,
  - fingerprint image,
  - handwritten word,
  - human face,
  - speech signal,
  - DNA sequence,
  - . . .
- Pattern recognition is the study of how machines can
  - observe the environment,
  - learn to distinguish patterns of interest,
  - make sound and reasonable decisions about the categories of the patterns.

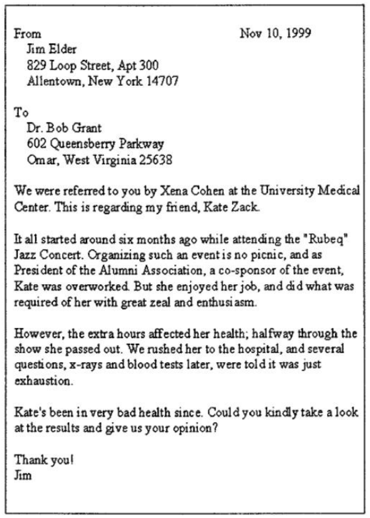
Human and Machine Pattern

- We are often influenced by the knowledge of how patterns are modeled and recognized in nature when we develop pattern recognition algorithms.
- Research on machine perception also helps us gain deeper understanding and appreciation for pattern recognition systems in nature.
- Yet, we also apply many techniques that are purely numerical and do not have any correspondence in natural systems.



Application: Speech recognition

Application: English handwriting recognition


From Nov 10, 1999
Jim Elder
829 Loop Street, Apt 300
Allentown, New York 14707

To
Dr. Bob Grant
602 Queensberry Parkway
Omar, West Virginia 25638

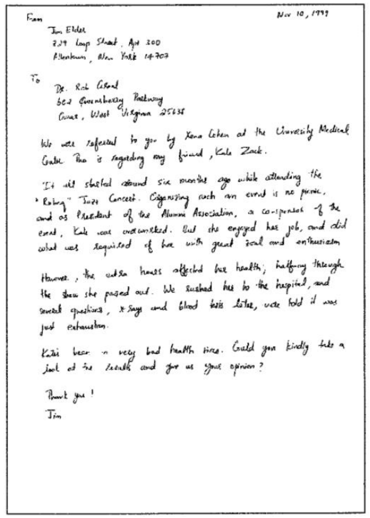
We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

It all started around six months ago while attending the "Rubeq" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event, Kate was overworked. But she enjoyed her job, and did what was required of her with great zeal and enthusiasm.

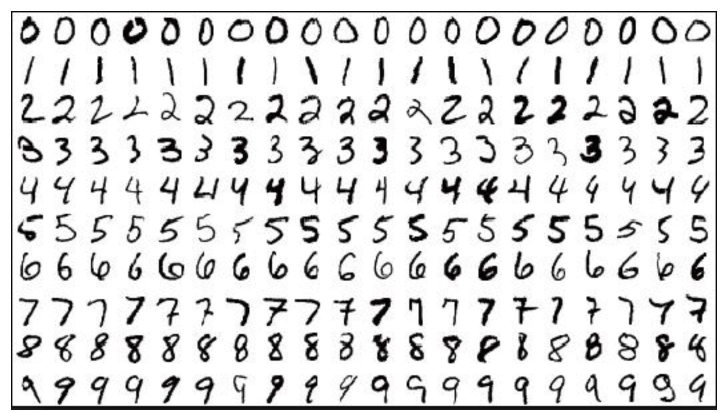
However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, were told it was just exhaustion.

Kate's been in very bad health since. Could you kindly take a look at the results and give us your opinion?

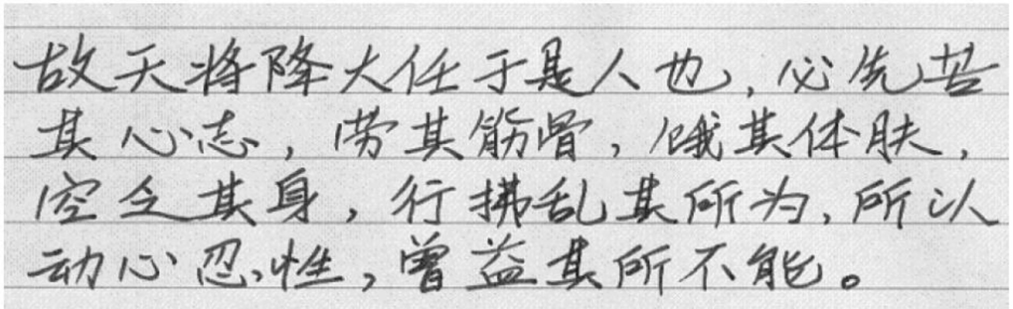
Thank you!
Jim




Letter Recognition [Peter 1991]


MINST Dataset

Application: Chinese handwriting recognition



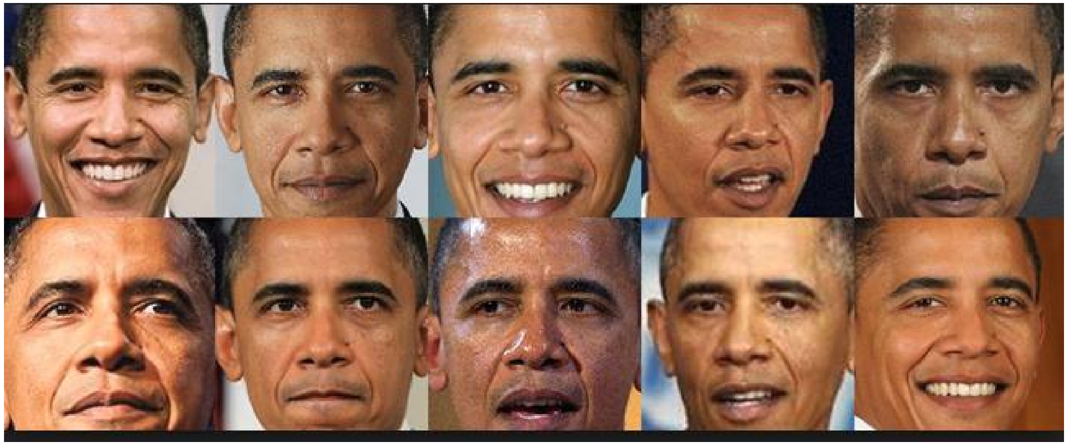
(a) Handwriting

故天将降大任于是人也，必先苦
其心志，劳其筋骨，饿其体肤，
空乏其身，行拂乱其所为，所以
动心忍性，曾益其所不能。

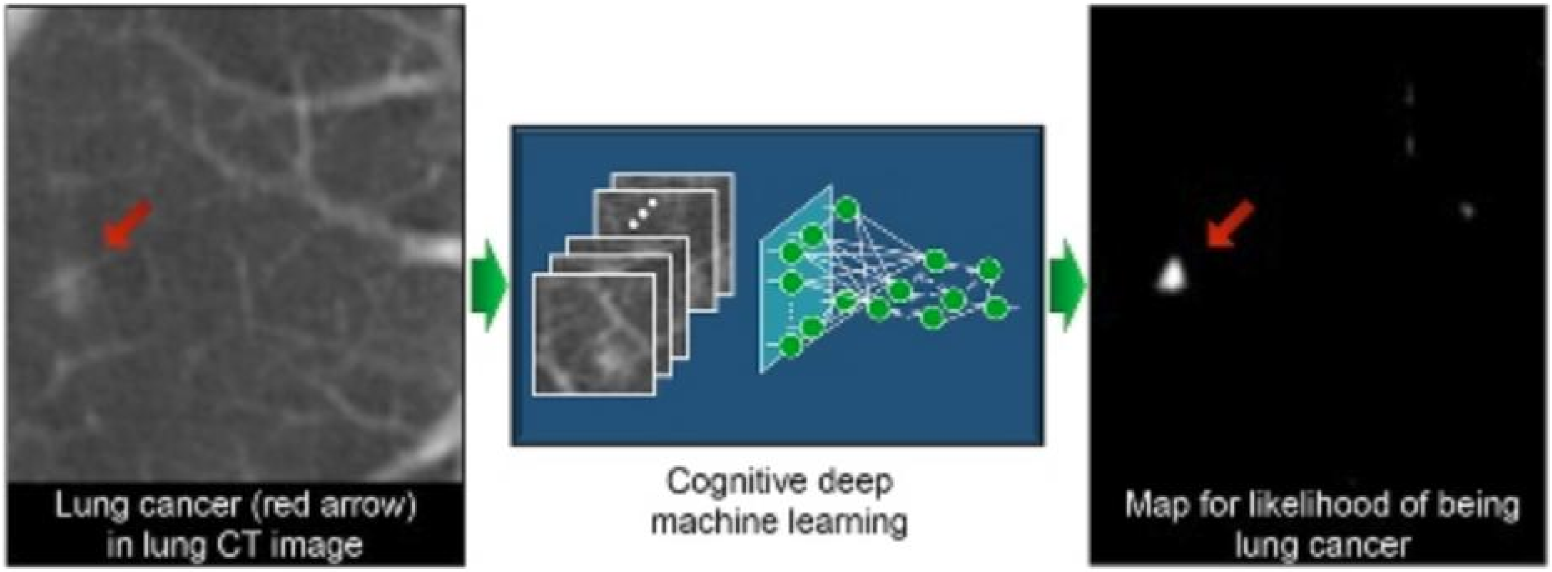
(b) Corresponding Machine Print

[Ming-KeZhou et al. *Discriminative quadratic feature learning for handwritten Chinese character recognition*. Pattern Recognition, 2016]

Application: Face recognition



Application: Cancer detection

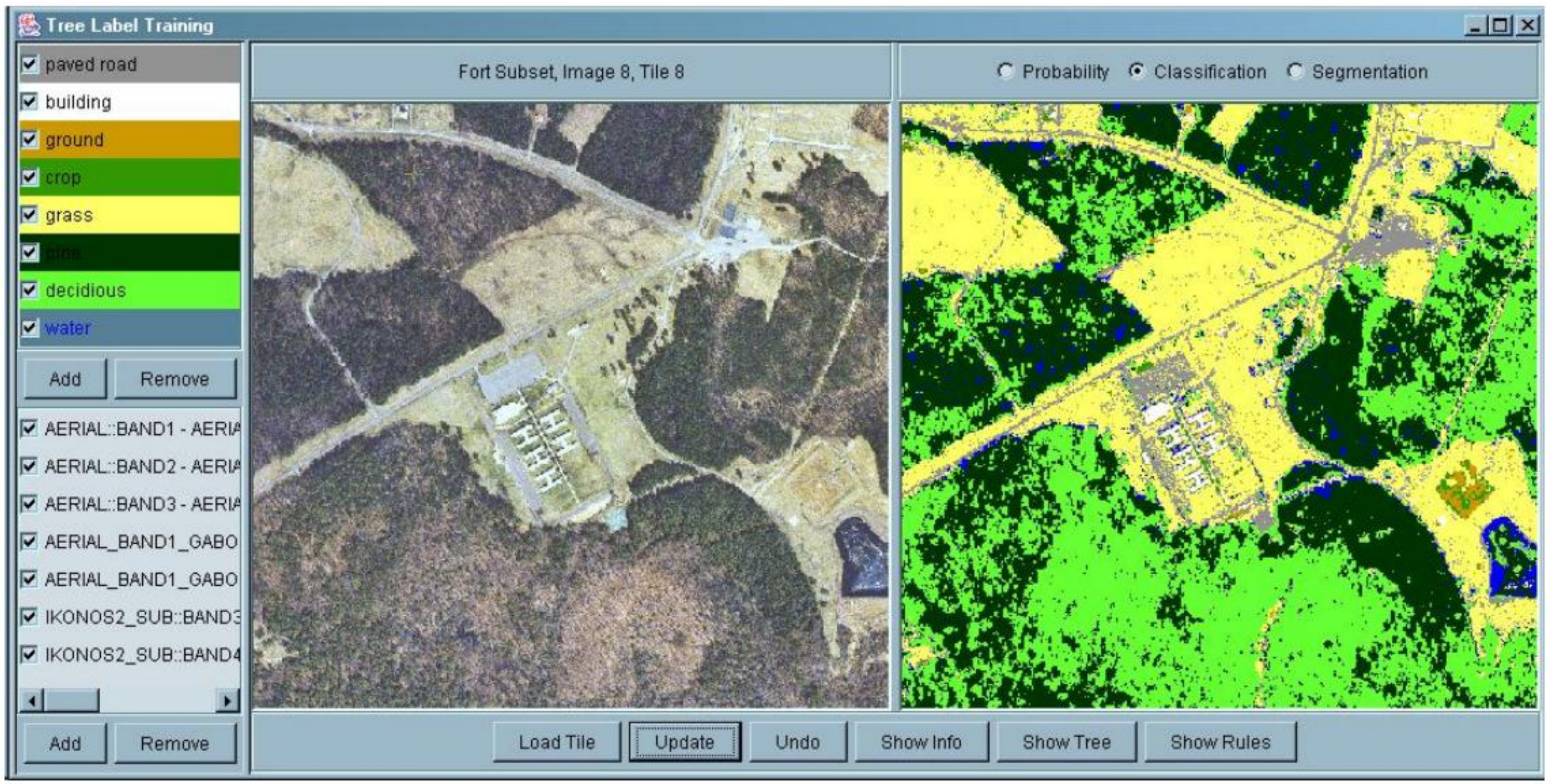


Cognitive Machine Learning for Estimating Likelihood of Being Lung Cancer in CT

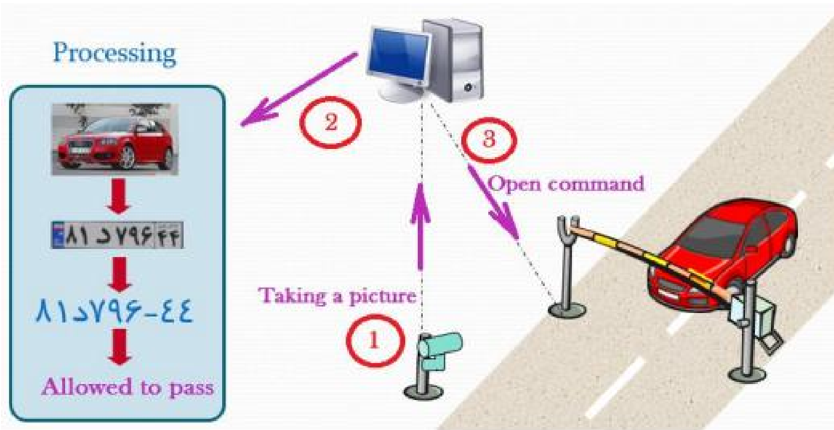
Application: Building and building grouping using satellite image

SpaceNet Dataset

Application: Land classification using satellite image



Application: License plate recognition: US license plates.



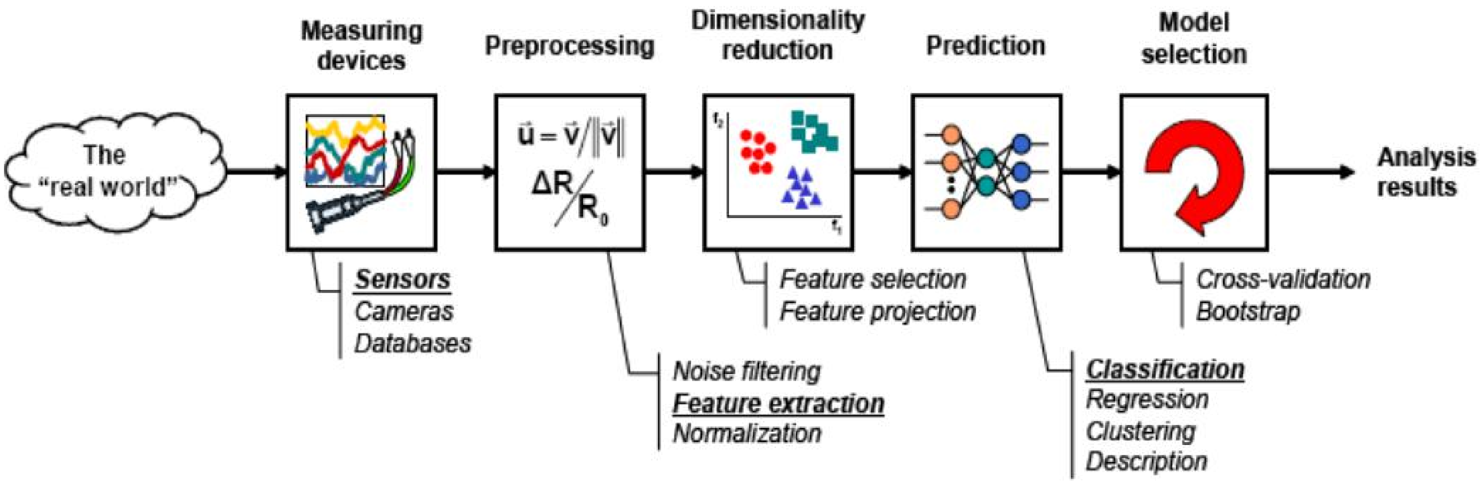
Application: Automatic navigation

Outline

- Course Informatition
- What is Pattern Recognition?
- Components of a Pattern Recognition System
- Pattern Recognition Desgin Cycle
- Summary

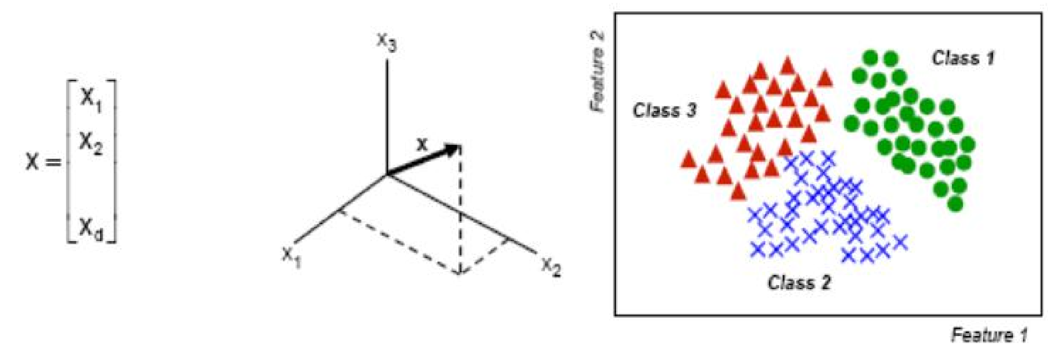
Components of a Pattern Recognition System

- A sensor
- A preprocessing mechanism
- A feature extraction mechanism (manual or automatic)
- A classification algorithm
- A set of example (training set) already classified or describe



Feature

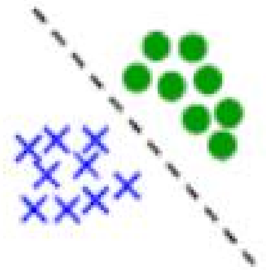
- Feature is any distinctive aspect, quality or characteristic
  - Features may be symbolic (i.e., color) or numeric (i.e., height)
- Definitions
  - The combination of d features is represented as a d-dimensional column vector called a ***feature vector***
  - The d-dimensional space defined by the feature vector is called the ***feature space***
  - Objects are represented as points in feature space. The representation is called a ***scatter plot***.



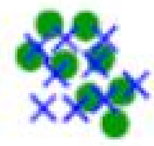
Feature vector | Feature space (3D) | Scatter plot (2D)

What's a "good" feature vector?

- The quality of a feature vector is related to its ability to discriminate examples from different classes.
  - Examples from the same class should have similar feature values.
  - Examples from different classes have different feature values.

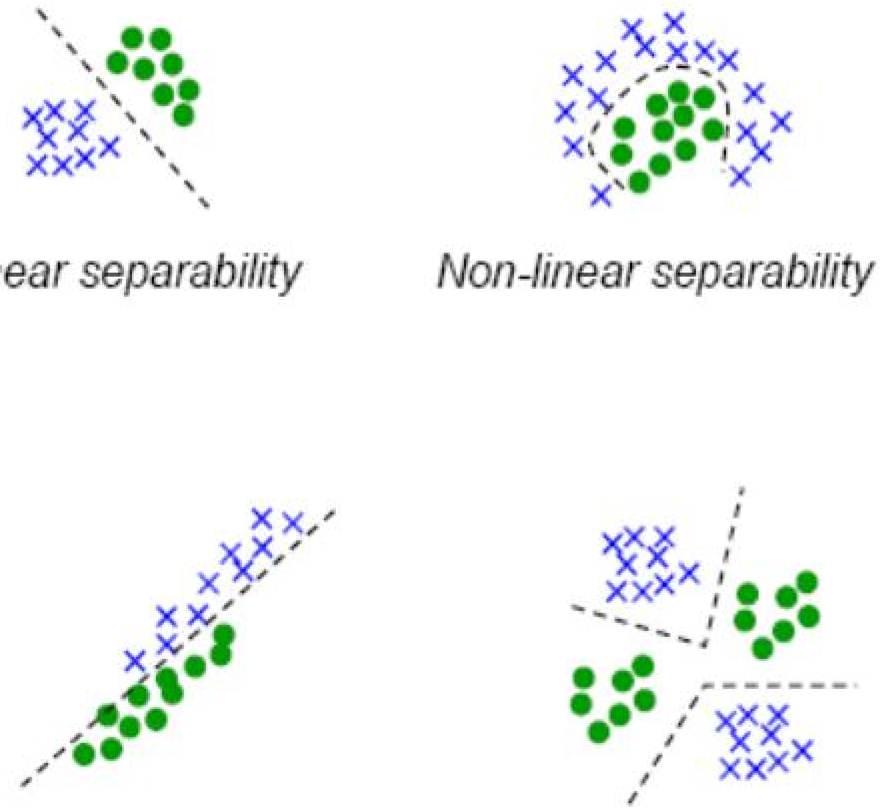


"Good" features



"Bad" features

More feature properties

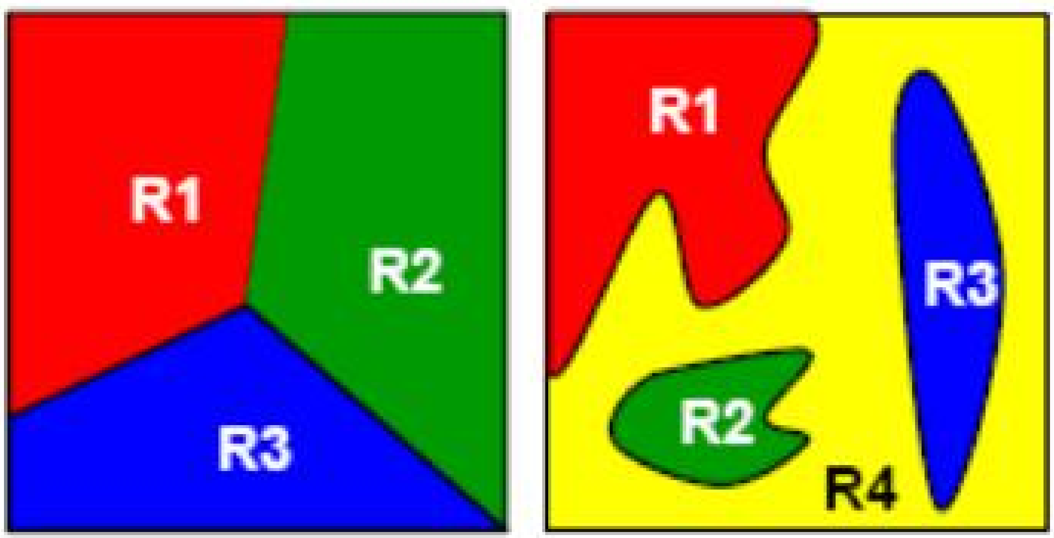


Linear separability *Non-linear separability*

Highly correlated features *Multi-modal*

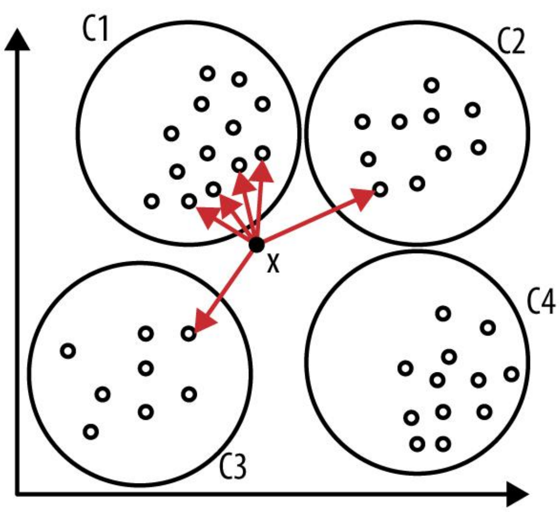
Classifier

- The task of a classifier is to partition feature space into class/ labeled decision region
  - Borders between decision regions are called decision boundaries
  - The classification of feature vector x consists of determining which decision region it belongs to, and assign x to this class.

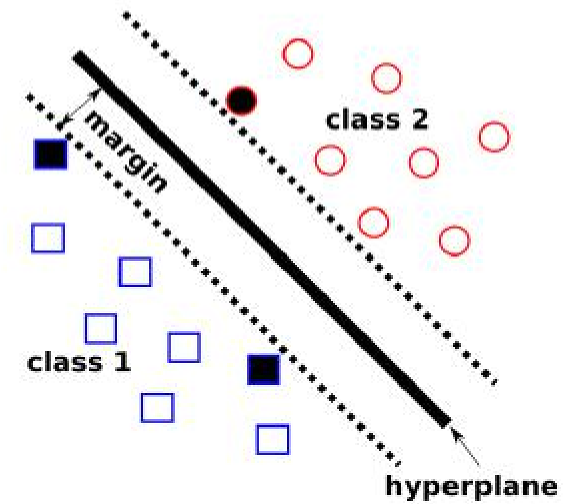


Classifier: Statistical approaches

- Patterns classified based on an underlying statistical model of the features
  - The statistical model is defined by a family of class/ conditional probability density function P(x|c) (Probability of feature vector x given class c)



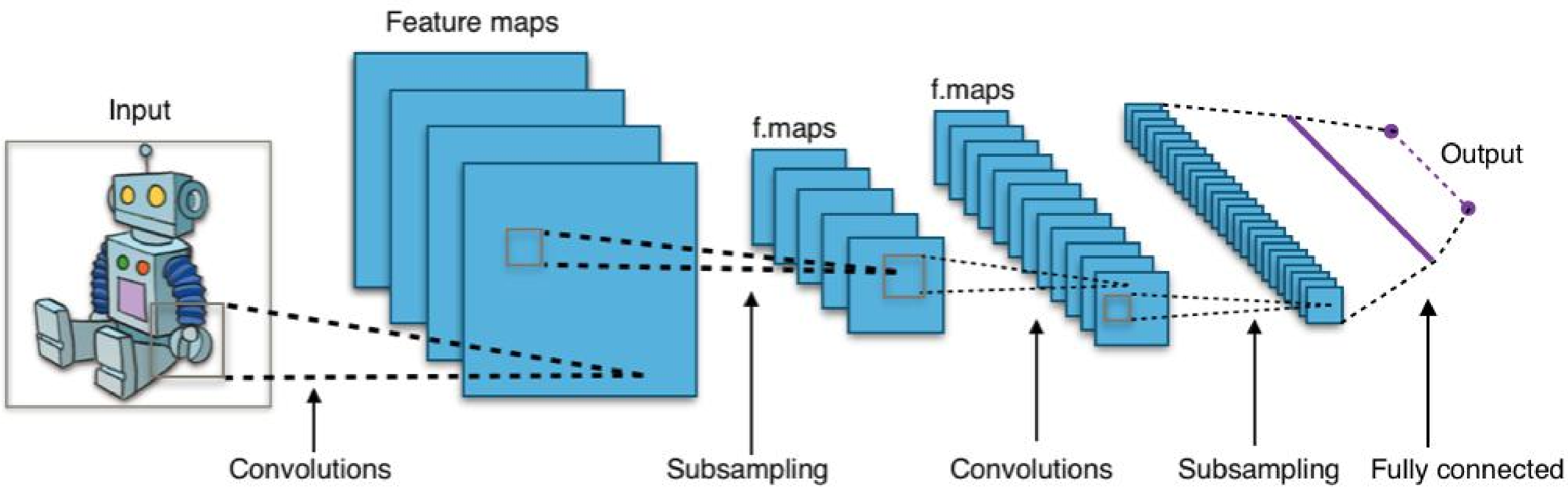
KNN classification



SVM

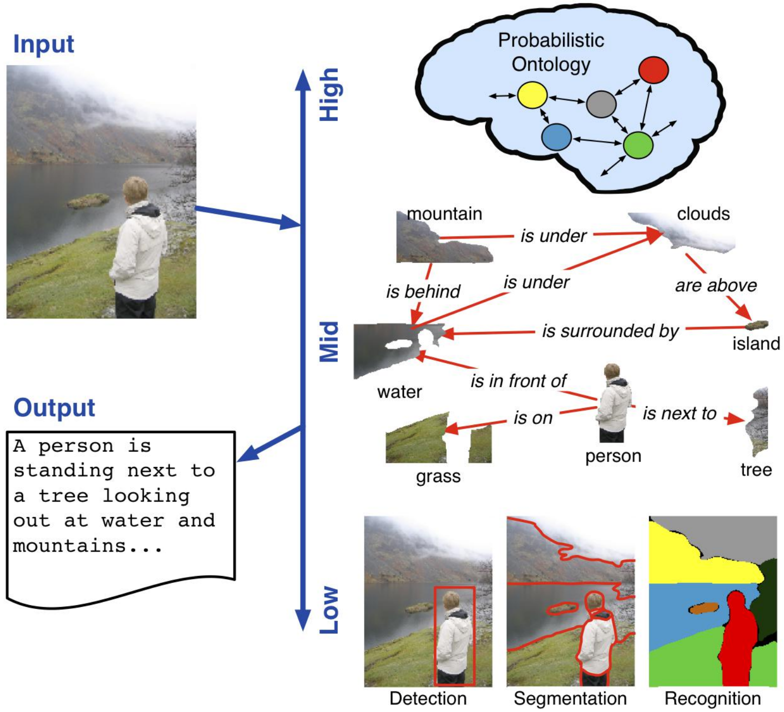
Classifier: Neural networks

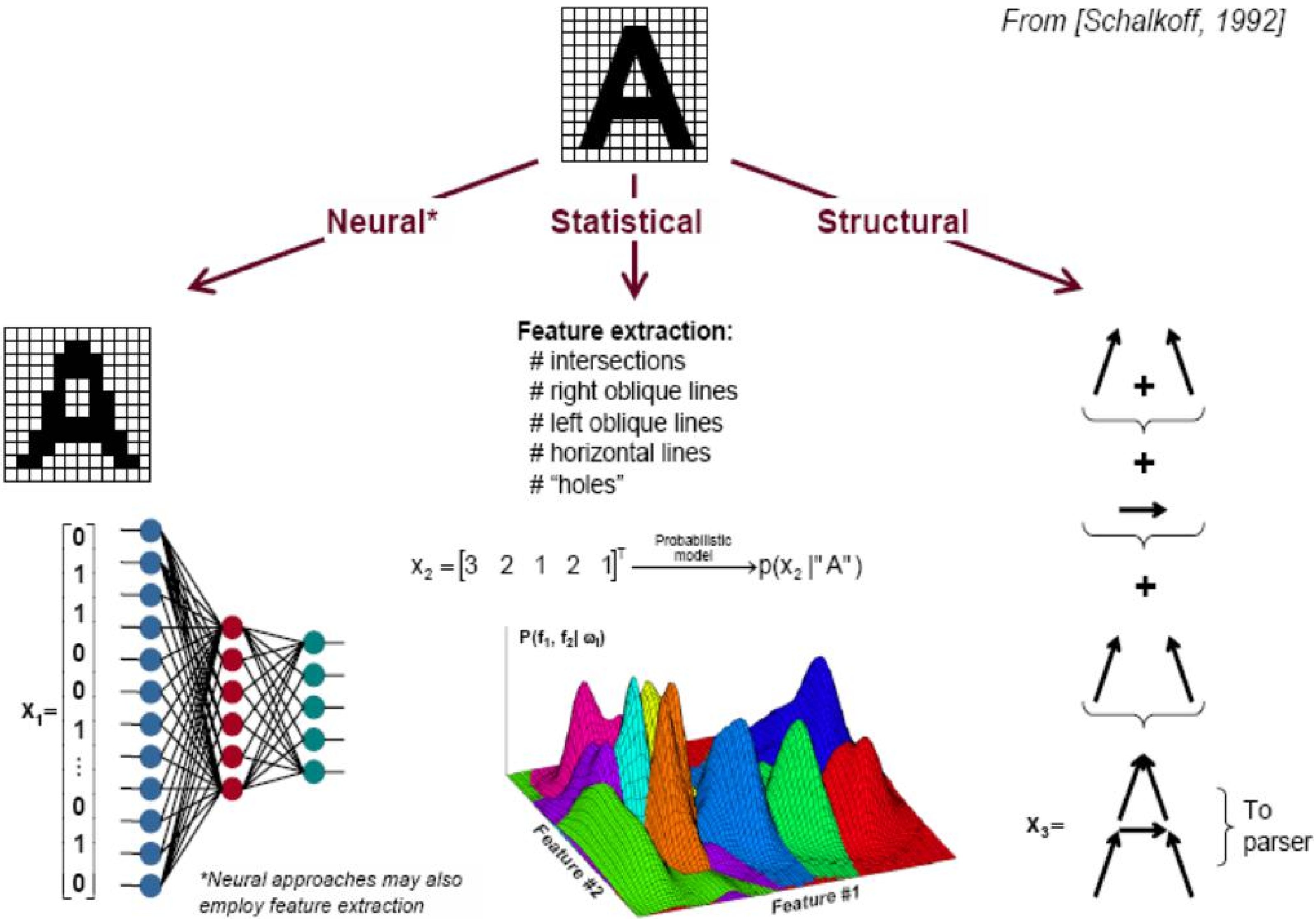
- Classification is based on the response of a network of processing units (neurons) to an input stimuli (pattern)
  - Knowledge is stored in the connectivity and strength of the synaptic weights.
- Trainable, non–algorithmic, black–box strategy.
- Very attractive since
  - it requires minimum a priori knowledge
  - with enough layers and neurons, an ANN can create any complex decision region.



Classifier: Structural approaches

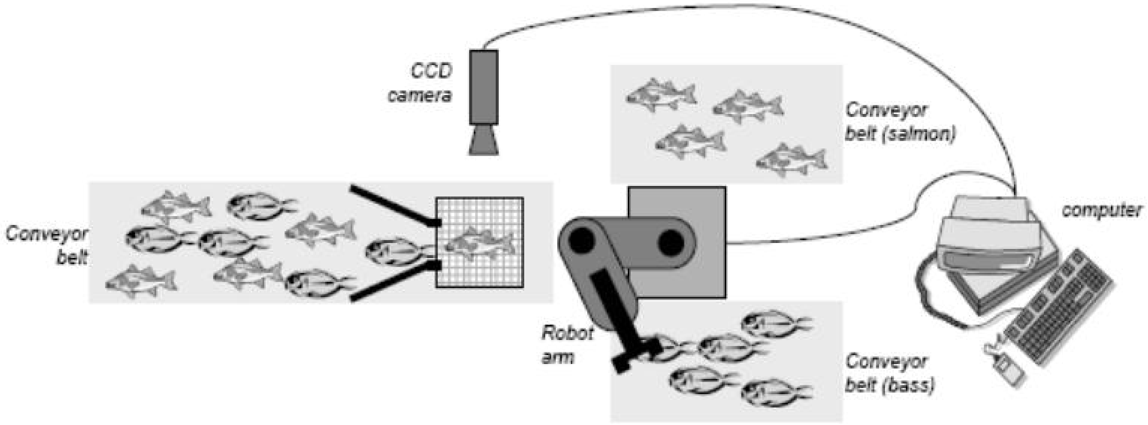
- Patterns classified based on measures of structural similarity.
  - "Knowledge" is represented by means of formal grammars or relational descriptions (graph).
- Used not only for classification, but also for description
  - Typically, structural approaches formulate hierarchical descriptions of complex patterns built up from simple sub patterns.





From [Schalkoff, 1992]
Neural*
Statistical
Structural
Feature extraction:
intersections
right oblique lines
left oblique lines
horizontal lines
"holes"
$x_1=$
0
1
1
0
0
1
0
1
0
$x_2 = [3 \quad 2 \quad 1 \quad 2 \quad 1]^T$
Probabilistic model
$p(x_2 \mid "A")$
$P(f_1, f_2 \mid \omega_1)$
Feature #2
Feature #1
*Neural approaches may also employ feature extraction
$x_3=$
To parser

An Example



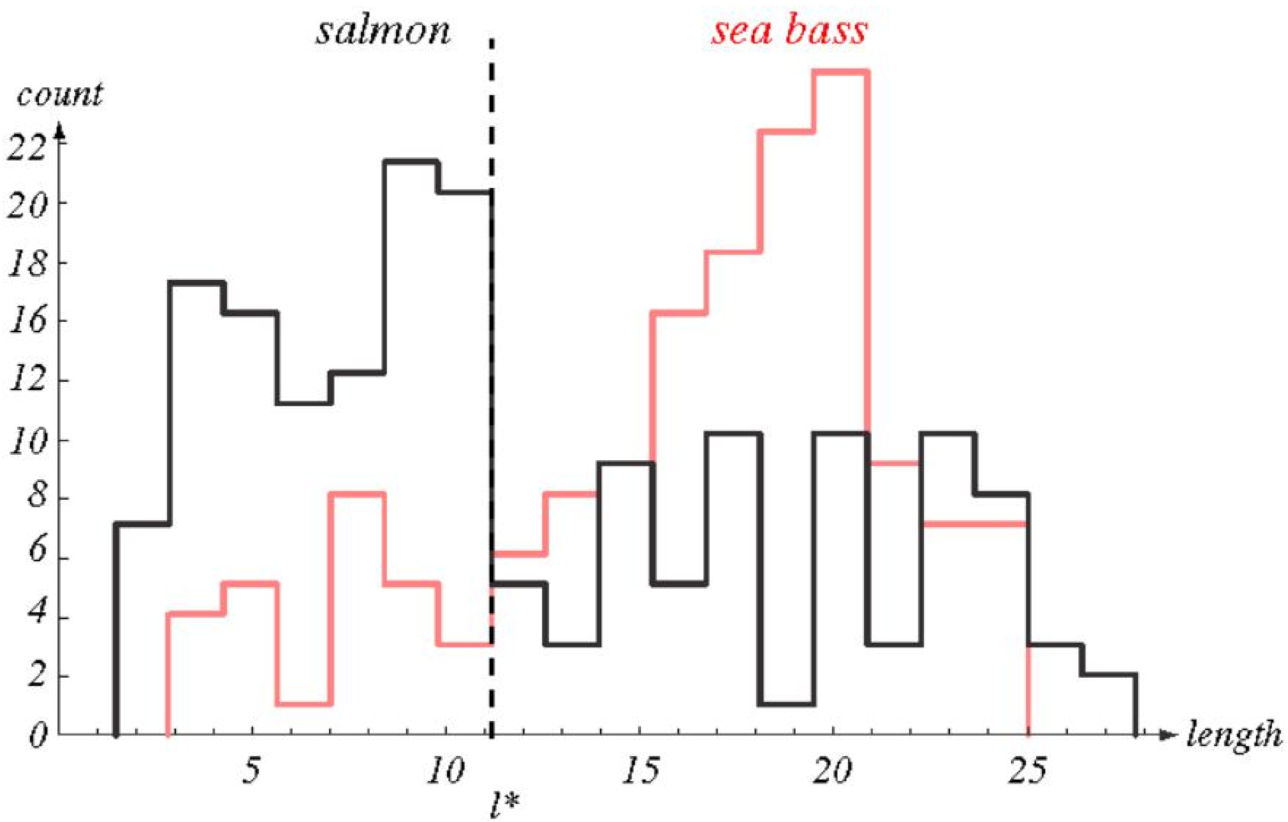
From [Duda, Hart and Stork, 2001]

- ❑ Problem: Sorting incoming fish on a conveyor belt according to species.
- ❑ Assume that we have only two kinds of fish:
  - sea bass,
  - salmon.

An Example: Selected Feature

- Assume a fisherman told us that a sea bass is generally longer than a salmon.
- We can use length as a feature and decide between sea bass and salmon according to a threshold on length.
- How can we choose this threshold？

An Example: Selected Feature

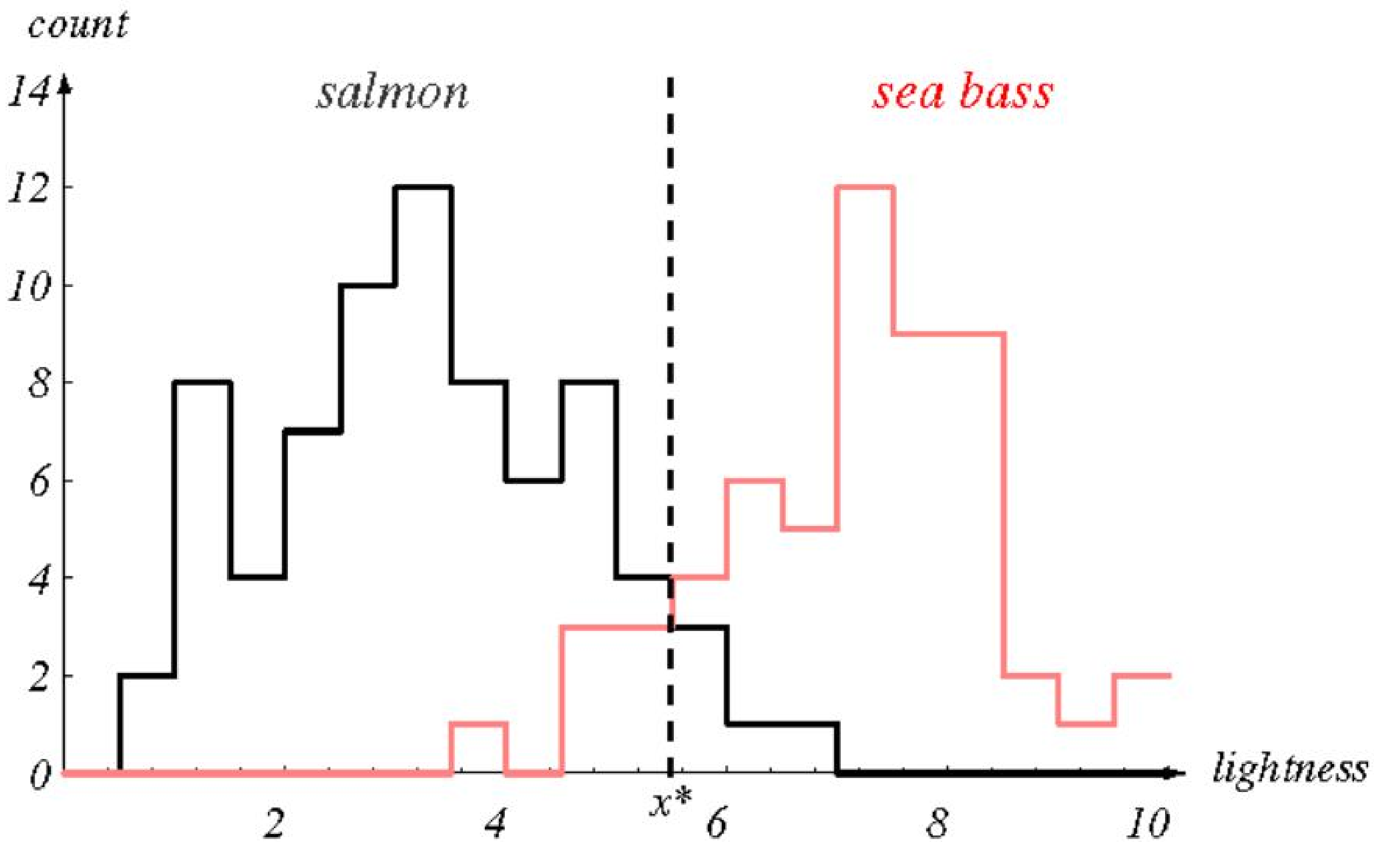


Histograms of the length feature for two types of fish in trainingsamples. How can we choose the threshold l^* to make a reliable decision?

An Example: Selected Feature

- Even though sea bass is longer than salmon on the average, there are many examples of fish where this observation does not hold.
- Try another feature: average lightness of the fish scales.

An Example: Selected Feature



Histograms of the lightness feature for two types of fish in training samples. It looks easier to choose the threshold x^* but we still cannot make a perfect decision.

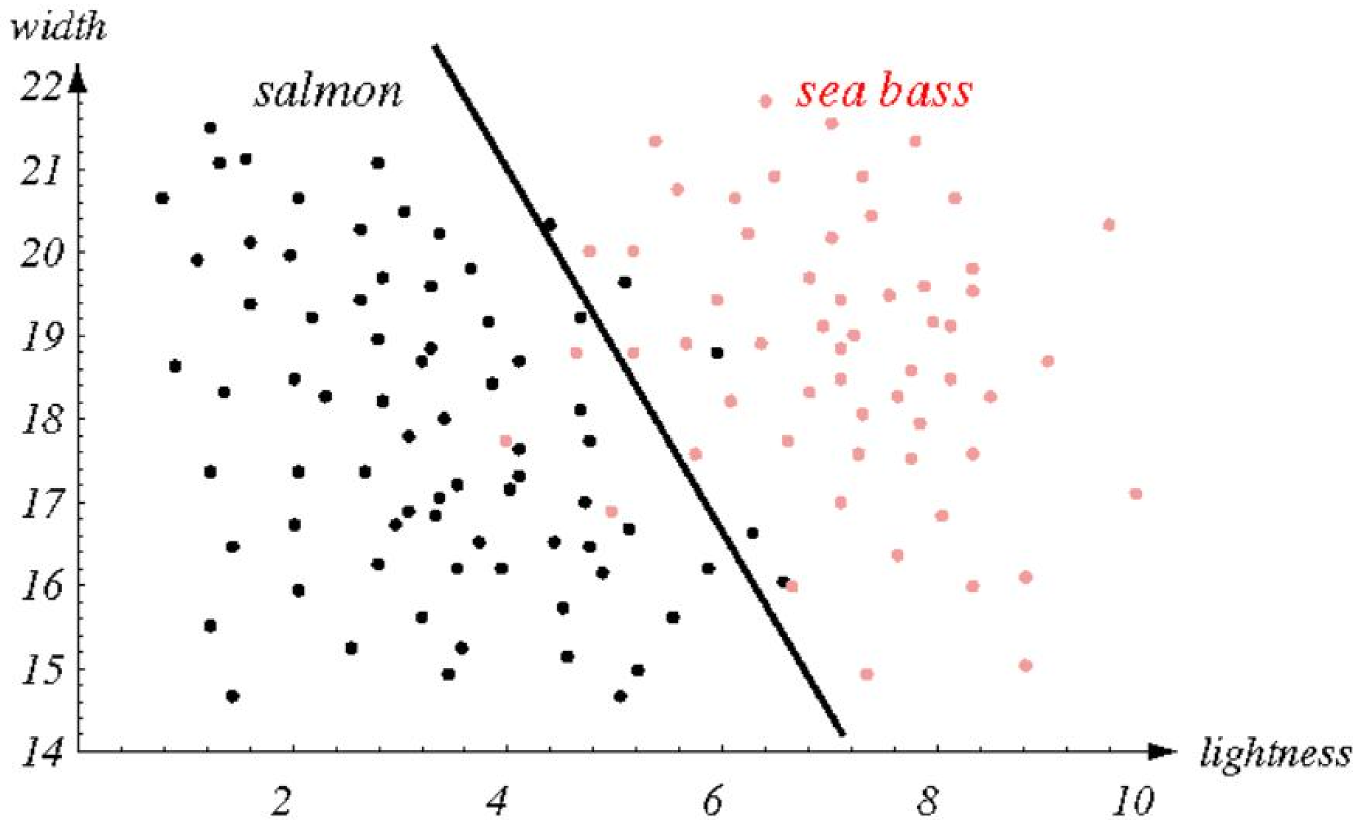
An Example: Multiple Features

- Assume we also observed that sea bass are typically wider than salmon.
- We can use two features in our decision:
  - lightness: x_1
  - width: x_2
- Each fish image is now represented as a point (***feature vector***)

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

in a two–dimensional ***feature space***.

An Example: Multiple Features



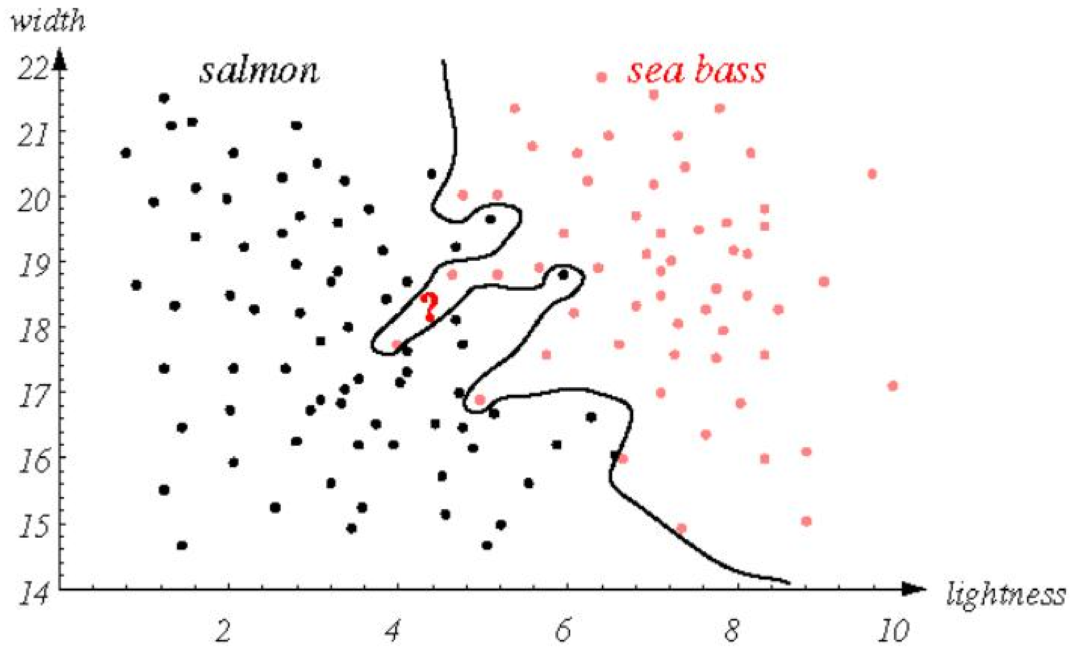
Scatter plot of lightness and width features for training samples.
We can draw a decision boundary to divide the feature space into two regions. Does it look better than using only lightness?

An Example: Multiple Features

- Does adding more features always improve the results？
  - Avoid unreliable features.
  - Be careful about correlations with existing features.
  - Be careful about measurement costs.
  - Be careful about noise in the measurements.
- Is there some curse for working in very high dimensions？

An Example: Decision Boundaries

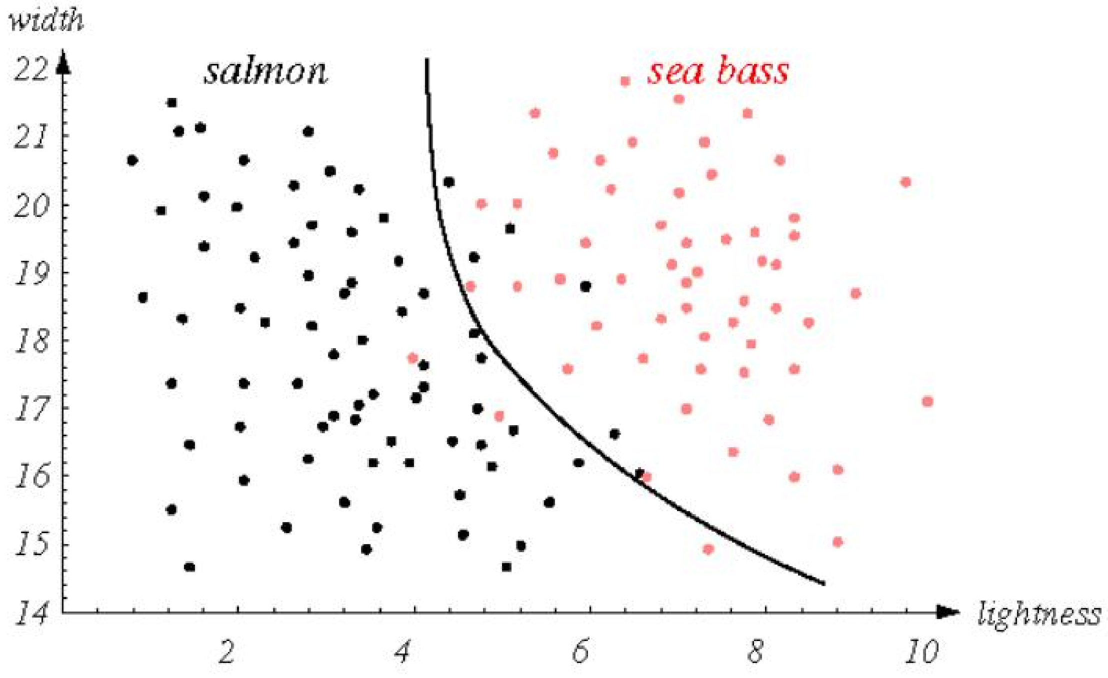
- Can we do better with another decision rule？
- More complex models result in more complex boundaries.



We may distinguish training samples perfectly but how can we predict how well we can **generalize** to unknown samples?

An Example: Decision Boundaries

- How can we manage the trade–off between complexity of decision rules and their performance to unknown samples？

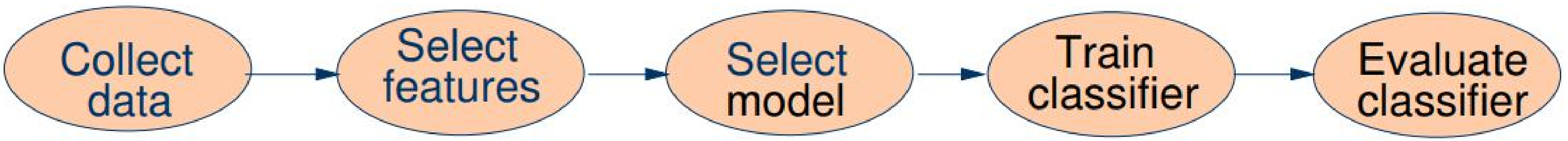


Different criteria lead to different decision boundaries.

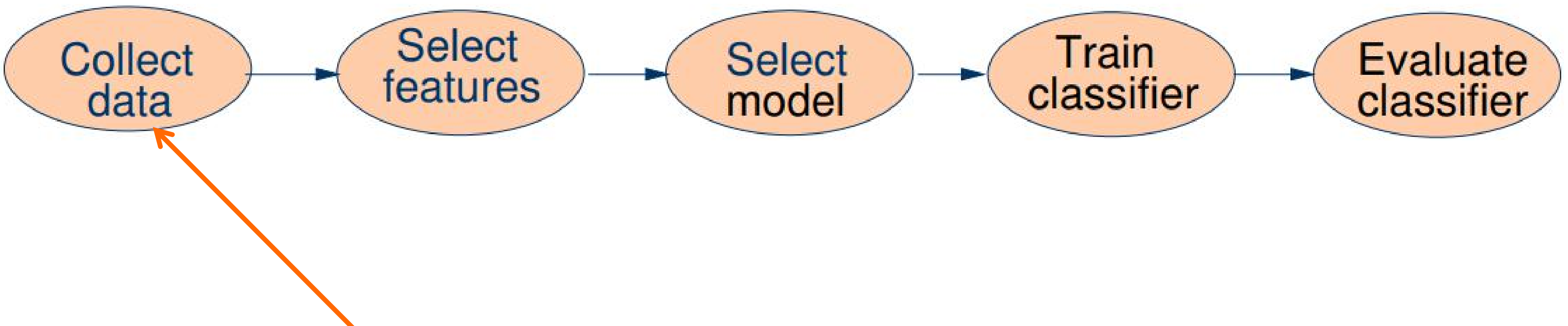
Outline

- Course Information
- What is Pattern Recognition?
- Components of a Pattern Recognition System
- Pattern Recognition Design Cycle
- Summary

Pattern recognition design cycle

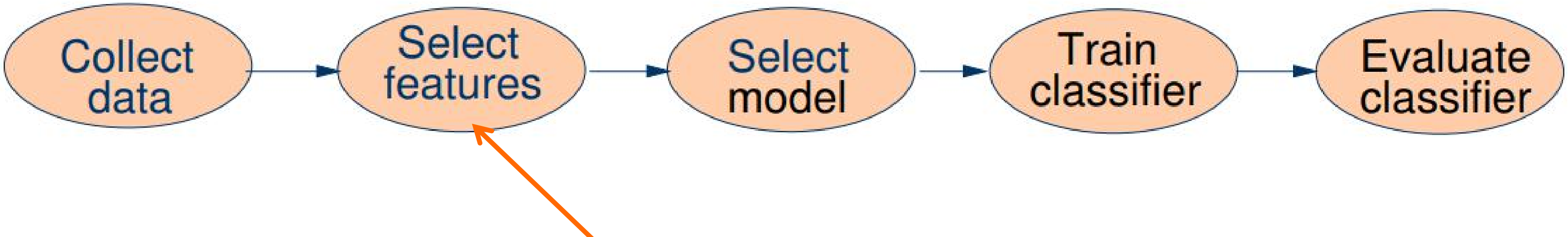


Pattern recognition design cycle



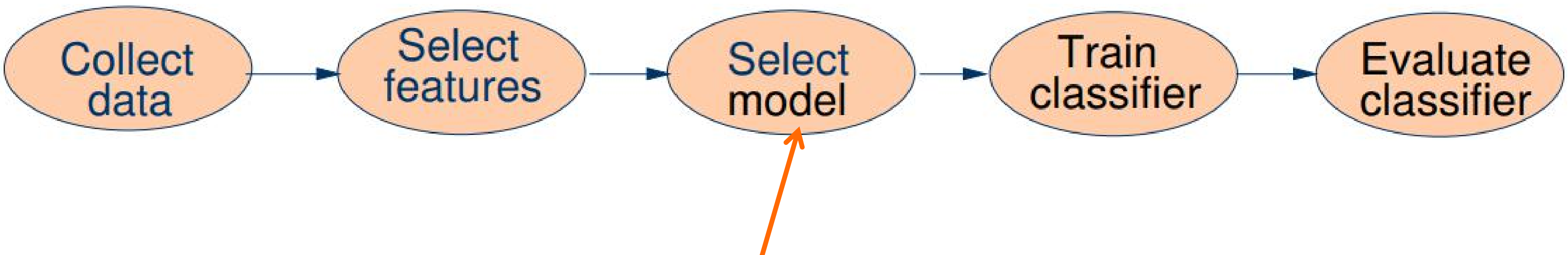
- Collecting training and testing data.
- How can we know when we have adequately large and representative set of samples?

Pattern recognition design cycle



- Domain dependence and prior information.
- Computational cost and feasibility.
- Discriminative features, i.e., similar values for similar patterns, and different values for different patterns.
- Invariant features with respect to translation, rotation and scale.
- Robust features with respect to occlusion, distortion,
- deformation, and variations in environment.

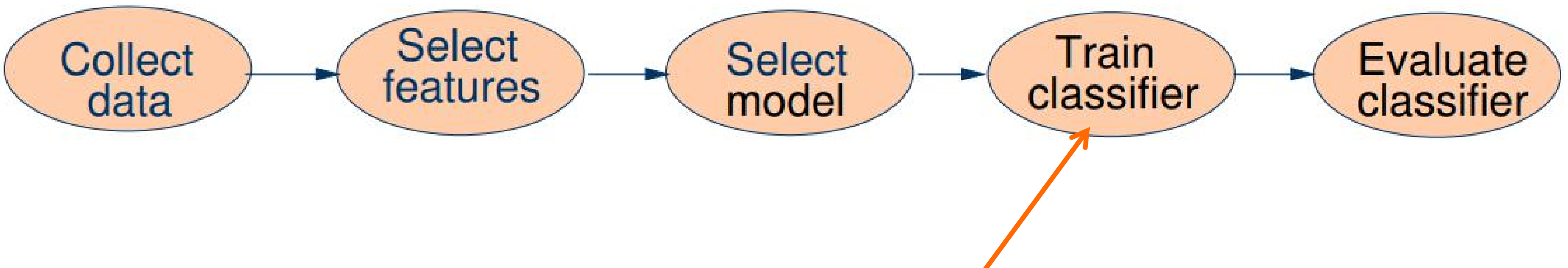
Pattern recognition design cycle



How can we know how close we are to the true model underlying the patterns?

- Domain dependence and prior information.
- Definition of design criteria.
- Parametric vs. non-parametric models.
- Handling of missing features.
- Computational complexity.
- Types of models: templates, decision-theoretic or statistical,syntactic or structural, neural, and hybrid.

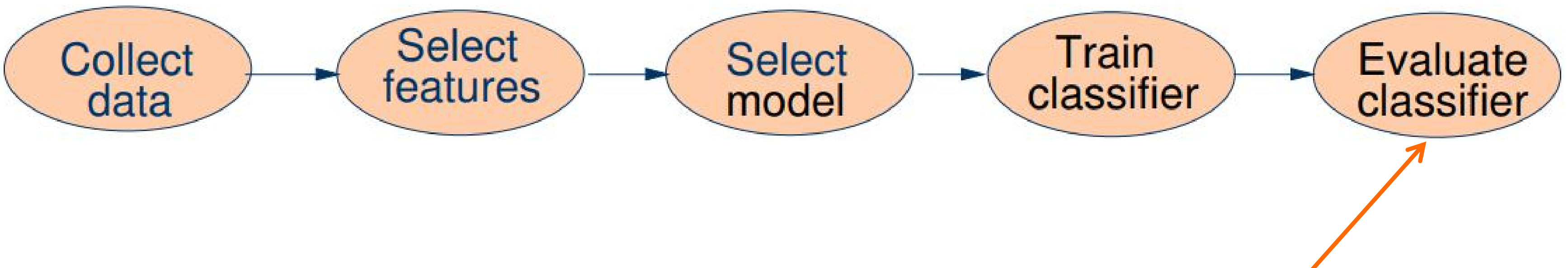
Pattern recognition design cycle



How can we learn the rule from data?

- **Supervised learning**: a teacher provides a category label or cost for each pattern in the training set.
- **Unsupervised learning**: the system forms clusters or natural groupings of the input patterns.
- **Reinforcement learning**: no desired category is given but the teacher provides feedback to the system such as the decision is right or wrong.

Pattern recognition design cycle



- How can we estimate the performance with training samples?
- How can we predict the performance with future data?
- Problems of overfitting and generalization.

Q & A