Pattern Recognition

Guide to AI and Data Courses · Part II: Branches of AI

Feature extraction | Bayes decisions | Classifiers | Clustering | Measuring performance | Neighboring courses

Sercan KÜLCÜ · Giresun University · Computer Engineering

© Sercan Külcü. All rights reserved. · Interactive HTML version of these slides: https://sercankulcu.github.io/files/ai_courses/slides/Chapter_08_Pattern_Recognition.html

What Question Does It Ask?

> **How do we decide from measurements which class an object belongs to?**

Pattern recognition studies how to decide which class an object belongs to by looking at a **feature vector** extracted from measurements. The course treats the chain from sensor to decision as a whole: preprocessing, feature selection, classifier design and error analysis.

At a glance

APPROACH	INPUT	OUTPUT
statistical decision theory	measurement or feature vector	class label

SUCCESS MEASURE	MATH LOAD	DATA NEEDS
classification error	medium-high: probability, linear algebra	labeled examples

A Short History

1936
Fisher publishes linear discriminant analysis; the Iris data from that paper is still the first example.

1962
Hu introduces moment features invariant to translation, scale and rotation.

1967
Cover and Hart show that the nearest neighbor rule's error is at most twice the Bayes error.

1973
Duda and Hart publish the field's core textbook, "Pattern Classification and Scene Analysis".

1991
Turk and Pentland build face recognition on PCA with eigenfaces.

1998
LeCun et al. publish LeNet-5 and the MNIST handwritten digit dataset.

THE COURSE

Core Concepts

Feature vector
Measurements that represent an object: length, color, texture.

Feature space
Each example is a point; classes are regions in this space.

Bayesian decision theory
Choose the class with the largest posterior; minimum error.

Decision boundary
The surface that splits feature space into class regions.

Density estimation
Parametric (Gaussian) or nonparametric (Parzen).

Nearest neighbor
Assign a new example to the class of the nearest labeled ones.

Dimensionality reduction
Few but discriminative features with PCA and LDA.

Confusion matrix
Shows which class is confused with which.

A Typical Workflow



Example: A mail sorting machine scans the postcode on an envelope, binarizes the image and splits it into digits, extracts shape features from each digit, decides which digit it is with nearest neighbor or an SVM, and sends misread envelopes to manual sorting.

THE COURSE

What Does the Course Cover?

1. Recognition systems: sensing, features, decision
2. Bayesian decision theory, risk and loss
3. The normal distribution and discriminant functions
4. Parameter estimation: maximum likelihood, Bayesian
5. Nonparametric density estimation: Parzen windows
6. Nearest-neighbour rule and distance measures
7. Linear discriminants and the perceptron
8. Support vector machines
9. Feature extraction, selection, curse of dimensionality
10. PCA and Fisher linear discriminant analysis (LDA)
11. Evaluating and comparing classifiers
12. Clustering and Gaussian mixture models

Skills you will gain

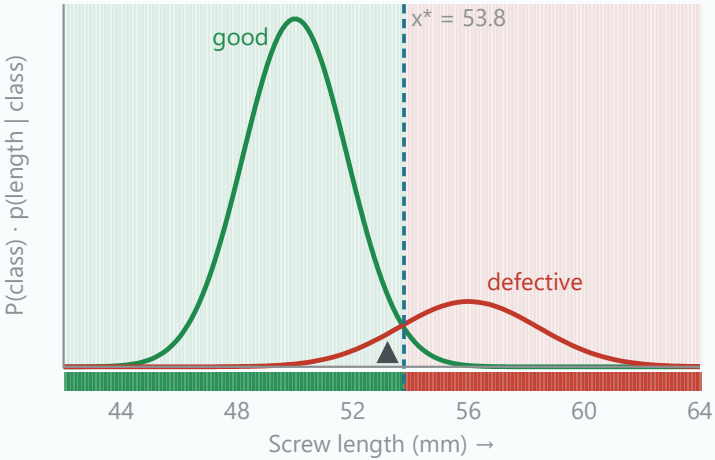
designing features from measurements · building a Bayes classifier · k-NN and SVM with scikit-learn · extracting image features with OpenCV · error analysis with a confusion matrix

A typical semester; the order and coverage vary between universities.

TRY IT

Reject the Defective Part

Every screw on the line is measured; defective ones are longer on average, but the distributions overlap. Change the defect rate and error cost to move the Bayes boundary; click the chart to measure a part.



Defect rate (prior) P(D) = **20%**

Missing a defect costs **1**× scrapping a good one

Reset

Decision boundary: **53.8 mm**
Missed defects 17.1% · good parts scrapped 1.9%
Total error **5.0%** · cost per 1000 parts 50

Measured part (▲) 53.2 mm: P(D | length) = 31.6% → **accepted**. As the cost rises the boundary moves left: total error grows, the costly error shrinks.

Pattern recognition decides with **probability and cost**: the same measurement changes class when the defect rate or the cost of an error changes. The Bayes rule gives the boundary that minimizes expected loss.

CONTEXT

Prerequisites

Courses worth knowing before this one, and why they are needed.

Course	Why it is needed
Probability and Statistics	Bayesian decision theory, density estimation
Programming	arrays, loops and functions in Python or Java
Linear Algebra	vectors, distances, eigenvalues: the language of PCA and LDA

CONTEXT

Are You Ready for This Course?

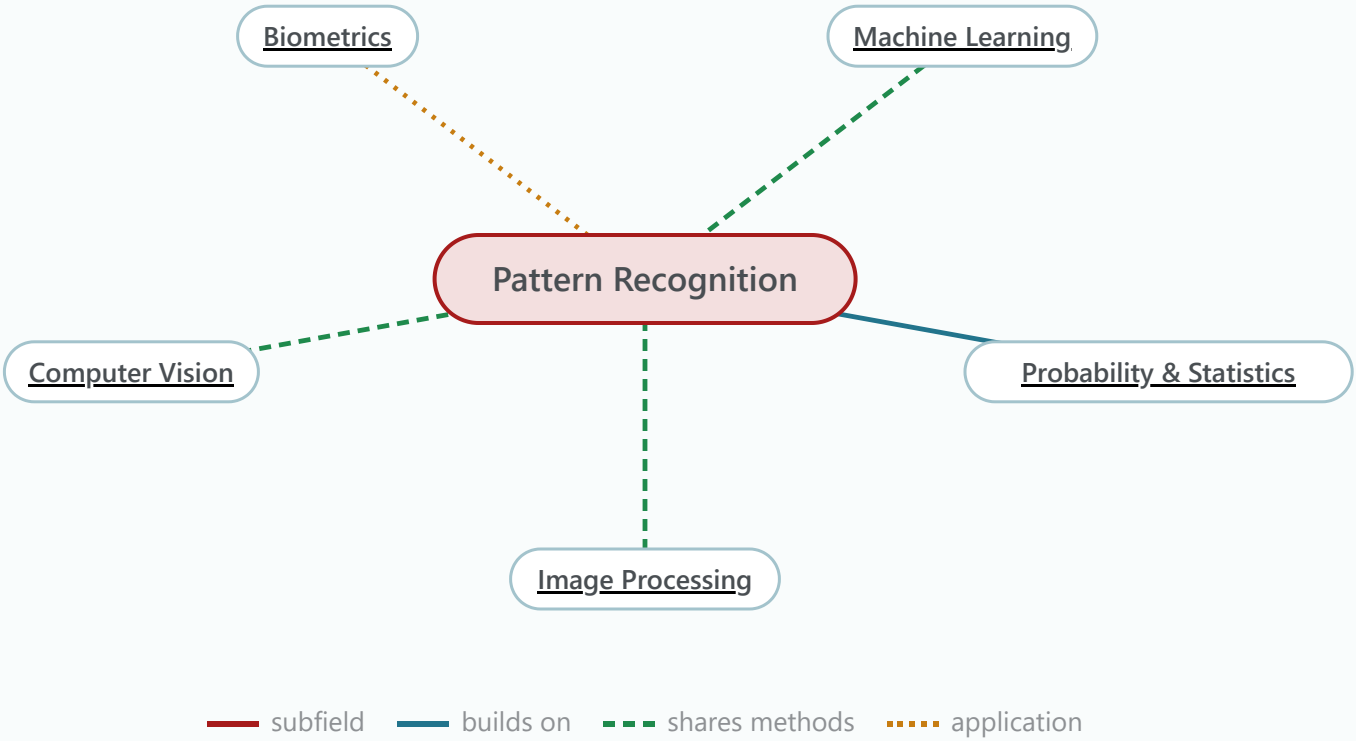
Ask yourself; for every "no", the right-hand column shows where to start.

Ask yourself	If not, start with
Can you explain the normal distribution, mean and covariance?	Probability and Statistics
Can you explain Euclidean distance, eigenvalues and eigenvectors?	Linear Algebra
Can you find the minimum of a function by taking its derivative?	Derivatives and extrema from calculus
Can you load and process a dataset with NumPy arrays in Python?	An intro programming course, then a short NumPy tutorial

If you can say "yes" to most of these, you are ready; one or two gaps can be closed during the semester.

CONTEXT

Neighbouring Courses



OVERLAP

Overlapping Courses

Relation	Course	Shared ground
builds on	This course builds on Probability and Statistics	Bayesian decision theory, density estimation
application	Biometric Systems applies the methods of this course	matching, verification and identification
shares methods	Shares methods with Machine Learning	classifiers, decision boundaries
shares methods	Shares methods with Image Processing	feature extraction, template matching
shares methods	Shares methods with Computer Vision	object classification

Where It Differs (1/2)

Criterion	Pattern Recognition	Machine Learning	Computer Vision
Core question	How do we decide from measurements which class an object belongs to?	How does a program generalize from examples?	How does a computer understand what is in an image or a video?
Approach	statistical decision theory	statistical learning	extracting meaning from images
Input	measurement or feature vector	labeled or unlabeled examples	image or video
Output	class label	a model that makes predictions	label, box, depth, 3D structure
Success measure	classification error	error on unseen data	accuracy, mAP, IoU
Math load	medium-high: probability, linear algebra	high: linear algebra, probability, optimization	high: geometry, linear algebra
Data needs	labeled examples	moderate to large	many labeled images

- **Machine Learning:** ML studies learning algorithms in general; this course focuses on the design from measurement to class decision and on the Bayes error.
- **Computer Vision:** Computer vision tries to understand the scene in an image; this course treats class decisions from any kind of measurement.

DIFFERENCES

Where It Differs (2/2)

Criterion	Pattern Recognition	Image Processing	Biometrics
Core question	How do we decide from measurements which class an object belongs to?	How is an image enhanced, transformed and divided into regions?	How is a person recognized from their face, fingerprint or voice?
Approach	statistical decision theory	pixel-level transformations	matching and verification
Input	measurement or feature vector	image	face, fingerprint, iris, voice
Output	class label	processed image or regions	identity decision
Success measure	classification error	visual quality, PSNR	FAR, FRR, EER
Math load	medium-high: probability, linear algebra	medium: convolution, Fourier	medium
Data needs	labeled examples	little; learning not required	few samples per person

- **Image Processing:** Image processing enhances and transforms images, and the result is still an image; this course outputs a class label.
- **Biometrics:** Biometrics is an application of pattern recognition; this course teaches general decision methods, not person-specific ones.

PRACTICE

Applications

Handwriting recognition
Reading postcodes, bank checks and forms.

Optical character recognition
Turning a scanned document into editable text.

Face and fingerprint
Identity verification and identification.

Industrial inspection
Rejecting defective products on the line.

Medical diagnosis
Classifying ECGs, blood tests or tissue samples.

Speaker recognition
Finding out who is speaking from the voice.

PRACTICE

Risks and Ethics

Face recognition without consent

Face recognition enables covert tracking; Clearview AI was fined in several European countries for faces scraped from the internet.

Error gaps between groups

In the Gender Shades study, commercial gender classifiers erred far more often on darker-skinned women than on lighter-skinned men.

Treating all errors alike

If a missed cancer weighs the same as a false alarm, total error looks low while outcomes are dangerous; choose the loss matrix with care.

Leakage in feature selection

Selecting features on the full dataset and then cross-validating makes the error look far lower than it really is.

In depth: Explainable AI and AI Ethics.

PRACTICE

Tools and Datasets

Tools and libraries

scikit-learn | NumPy | OpenCV | MATLAB Statistics and ML Toolbox | Weka | PRTools

Well-known datasets

Iris | MNIST | Optical Recognition of Handwritten Digits (UCI) | Wisconsin Breast Cancer | AT&T (ORL) face dataset | EMNIST

PRACTICE

Common Misconceptions

Common belief	In fact
Pattern recognition is the old name for machine learning.	The two overlap heavily; pattern recognition is an engineering view that also covers sensing and feature design.
Choosing the classifier is the key step.	In most problems **feature selection** decides success; a good classifier cannot rescue bad features.
With the right features the error drops to zero.	If classes overlap, even the best decision errs as much as the Bayes error.
The more features, the better.	As dimension grows, examples become sparse (the curse of dimensionality); useless features hurt performance.

NEXT STEPS

After This Course

Courses that build on or extend this one

Biometric Systems

Related careers

Image and signal processing engineer

Machine learning engineer

Biometrics specialist

Quality inspection systems engineer

Pattern recognition is the shared core of computer vision, speech recognition and biometrics. After this course you can move to a sensing area (vision, audio, biometrics) or go on to deep learning, which also learns the features from data.

NEXT STEPS

Further Reading

- Duda, Hart and Stork, *Pattern Classification*, 2nd ed., Wiley, 2001.
- Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- Theodoridis and Koutroumbas, *Pattern Recognition*, 4th ed., Academic Press, 2009.
- Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed., Academic Press, 1990.

The End

Chapter 8: Pattern Recognition

← Previous chapter · Map of the field · Next chapter →