

Title: Pattern recognition

URL: [https://en.wikipedia.org/wiki/Pattern\\_recognition](https://en.wikipedia.org/wiki/Pattern_recognition)

PageID: 126706

Categories: Category:Computational fields of study, Category:Formal sciences, Category:Machine learning, Category:Pattern recognition

Source: Wikipedia (CC BY-SA 4.0).

-----

Supervised learning

Unsupervised learning

Semi-supervised learning

Self-supervised learning

Reinforcement learning

Meta-learning

Online learning

Batch learning

Curriculum learning

Rule-based learning

Neuro-symbolic AI

Neuromorphic engineering

Quantum machine learning

Classification

Generative modeling

Regression

Clustering

Dimensionality reduction

Density estimation

Anomaly detection

Data cleaning

AutoML

Association rules

Semantic analysis

Structured prediction

Feature engineering

Feature learning

Learning to rank

Grammar induction

Ontology learning

Multimodal learning

Apprenticeship learning

Decision trees

Ensembles Bagging Boosting Random forest

Bagging

Boosting

Random forest

k -NN

Linear regression

Naive Bayes

Artificial neural networks

Logistic regression

Perceptron

Relevance vector machine (RVM)

Support vector machine (SVM)

BIRCH

CURE

Hierarchical

k -means

Fuzzy

Expectation–maximization (EM)

DBSCAN

OPTICS

Mean shift

Factor analysis

CCA

ICA

LDA

NMF

PCA

PGD

t-SNE

SDL

Graphical models Bayes net Conditional random field Hidden Markov

Bayes net

Conditional random field

Hidden Markov

RANSAC

k -NN

Local outlier factor  
Isolation forest  
Autoencoder  
Deep learning  
Feedforward neural network  
Recurrent neural network LSTM GRU ESN reservoir computing  
LSTM  
GRU  
ESN  
reservoir computing  
Boltzmann machine Restricted  
Restricted  
GAN  
Diffusion model  
SOM  
Convolutional neural network U-Net LeNet AlexNet DeepDream  
U-Net  
LeNet  
AlexNet  
DeepDream  
Neural field Neural radiance field Physics-informed neural networks  
Neural radiance field  
Physics-informed neural networks  
Transformer Vision  
Vision  
Mamba  
Spiking neural network  
Memtransistor  
Electrochemical RAM (ECRAM)  
Q-learning  
Policy gradient  
SARSA  
Temporal difference (TD)  
Multi-agent Self-play  
Self-play  
Active learning  
Crowdsourcing  
Human-in-the-loop

Mechanistic interpretability

RLHF

Coefficient of determination

Confusion matrix

Learning curve

ROC curve

Kernel machines

Bias–variance tradeoff

Computational learning theory

Empirical risk minimization

Occam learning

PAC learning

Statistical learning

VC theory

Topological deep learning

AAAI

ECML PKDD

NeurIPS

ICML

ICLR

IJCAI

ML

JMLR

Glossary of artificial intelligence

List of datasets for machine-learning research List of datasets in computer vision and image processing

List of datasets in computer vision and image processing

Outline of machine learning

v

t

e

Pattern recognition is the task of assigning a class to an observation based on patterns extracted from data. While similar, pattern recognition (PR) is not to be confused with pattern machines (PM) which may possess PR capabilities but their primary function is to distinguish and create emergent patterns. PR has applications in statistical data analysis , signal processing , image analysis , information retrieval , bioinformatics , data compression , computer graphics and machine learning . Pattern recognition has its origins in statistics and engineering; some modern approaches to pattern recognition include the use of machine learning , due to the increased availability of big data and a new abundance of processing power .

Pattern recognition systems are commonly trained from labeled "training" data. When no labeled data are available, other algorithms can be used to discover previously unknown patterns. KDD and

data mining have a larger focus on unsupervised methods and stronger connection to business use. Pattern recognition focuses more on the signal and also takes acquisition and signal processing into consideration. It originated in engineering, and the term is popular in the context of computer vision: a leading computer vision conference is named Conference on Computer Vision and Pattern Recognition.

In machine learning, pattern recognition is the assignment of a label to a given input value. In statistics, discriminant analysis was introduced for this same purpose in 1936. An example of pattern recognition is classification, which attempts to assign each input value to one of a given set of classes (for example, determine whether a given email is "spam"). Pattern recognition is a more general problem that encompasses other types of output as well. Other examples are regression, which assigns a real-valued output to each input; [ 1 ] sequence labeling, which assigns a class to each member of a sequence of values [ 2 ] (for example, part of speech tagging, which assigns a part of speech to each word in an input sentence); and parsing, which assigns a parse tree to an input sentence, describing the syntactic structure of the sentence. [ 3 ]

Pattern recognition algorithms generally aim to provide a reasonable answer for all possible inputs and to perform "most likely" matching of the inputs, taking into account their statistical variation. This is opposed to pattern matching algorithms, which look for exact matches in the input with pre-existing patterns. A common example of a pattern-matching algorithm is regular expression matching, which looks for patterns of a given sort in textual data and is included in the search capabilities of many text editors and word processors.

## Overview

A modern definition of pattern recognition is:

The field of pattern recognition is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions such as classifying the data into different categories. [ 4 ]

Pattern recognition is generally categorized according to the type of learning procedure used to generate the output value. Supervised learning assumes that a set of training data (the training set) has been provided, consisting of a set of instances that have been properly labeled by hand with the correct output. A learning procedure then generates a model that attempts to meet two sometimes conflicting objectives: Perform as well as possible on the training data, and generalize as well as possible to new data (usually, this means being as simple as possible, for some technical definition of "simple", in accordance with Occam's Razor, discussed below). Unsupervised learning, on the other hand, assumes training data that has not been hand-labeled, and attempts to find inherent patterns in the data that can then be used to determine the correct output value for new data instances. [ 5 ] A combination of the two that has been explored is semi-supervised learning, which uses a combination of labeled and unlabeled data (typically a small set of labeled data combined with a large amount of unlabeled data). In cases of unsupervised learning, there may be no training data at all.

Sometimes different terms are used to describe the corresponding supervised and unsupervised learning procedures for the same type of output. The unsupervised equivalent of classification is normally known as clustering, based on the common perception of the task as involving no training data to speak of, and of grouping the input data into clusters based on some inherent similarity measure (e.g. the distance between instances, considered as vectors in a multi-dimensional vector space), rather than assigning each input instance into one of a set of pre-defined classes. In some fields, the terminology is different. In community ecology, the term classification is used to refer to what is commonly known as "clustering".

The piece of input data for which an output value is generated is formally termed an instance. The instance is formally described by a vector of features, which together constitute a description of all known characteristics of the instance. These feature vectors can be seen as defining points in an appropriate multidimensional space, and methods for manipulating vectors in vector spaces can be correspondingly applied to them, such as computing the dot product or the angle between two vectors. Features typically are either categorical (also known as nominal, i.e., consisting of one of a set of unordered items, such as a gender of "male" or "female", or a blood type of "A", "B", "AB" or

"O"), ordinal (consisting of one of a set of ordered items, e.g., "large", "medium" or "small"), integer-valued (e.g., a count of the number of occurrences of a particular word in an email) or real-valued (e.g., a measurement of blood pressure). Often, categorical and ordinal data are grouped together, and this is also the case for integer-valued and real-valued data. Many algorithms work only in terms of categorical data and require that real-valued or integer-valued data be discretized into groups (e.g., less than 5, between 5 and 10, or greater than 10).

#### Probabilistic classifiers

Many common pattern recognition algorithms are probabilistic in nature, in that they use statistical inference to find the best label for a given instance. Unlike other algorithms, which simply output a "best" label, often probabilistic algorithms also output a probability of the instance being described by the given label. In addition, many probabilistic algorithms output a list of the  $N$ -best labels with associated probabilities, for some value of  $N$ , instead of simply a single best label. When the number of possible labels is fairly small (e.g., in the case of classification),  $N$  may be set so that the probability of all possible labels is output. Probabilistic algorithms have many advantages over non-probabilistic algorithms:

They output a confidence value associated with their choice. (Note that some other algorithms may also output confidence values, but in general, only for probabilistic algorithms is this value mathematically grounded in probability theory. Non-probabilistic confidence values can in general not be given any specific meaning, and only used to compare against other confidence values output by the same algorithm.)

Correspondingly, they can abstain when the confidence of choosing any particular output is too low.

Because of the probabilities output, probabilistic pattern-recognition algorithms can be more effectively incorporated into larger machine-learning tasks, in a way that partially or completely avoids the problem of error propagation.

#### Number of important feature variables

Feature selection algorithms attempt to directly prune out redundant or irrelevant features. A general introduction to feature selection which summarizes approaches and challenges, has been given. [ 6 ] The complexity of feature-selection is, because of its non-monotonous character, an optimization problem where given a total of  $n$  features the powerset consisting of all  $2^n - 1$  subsets of features need to be explored. The Branch-and-Bound algorithm [ 7 ] does reduce this complexity but is intractable for medium to large values of the number of available features  $n$ .

Techniques to transform the raw feature vectors ( feature extraction ) are sometimes used prior to application of the pattern-matching algorithm. Feature extraction algorithms attempt to reduce a large-dimensionality feature vector into a smaller-dimensionality vector that is easier to work with and encodes less redundancy, using mathematical techniques such as principal components analysis (PCA). The distinction between feature selection and feature extraction is that the resulting features after feature extraction has taken place are of a different sort than the original features and may not easily be interpretable, while the features left after feature selection are simply a subset of the original features.

#### Problem statement

The problem of pattern recognition can be stated as follows: Given an unknown function  $g : X \rightarrow Y$  (the ground truth) that maps input instances  $x \in X$  to output labels  $y \in Y$ , along with training data  $D = \{ (x_1, y_1), \dots, (x_n, y_n) \}$  assumed to represent accurate examples of the mapping, produce a function  $h : X \rightarrow Y$  that approximates as closely as possible the correct mapping  $g$ . (For example, if the problem is filtering spam, then  $x_i$  is some representation of an email and  $y$  is either "spam" or "non-spam"). In order for this to be a well-defined problem, "approximates as closely as possible" needs to be defined rigorously. In decision theory, this is defined by specifying

a loss function or cost function that assigns a specific value to "loss" resulting from producing an incorrect label. The goal then is to minimize the expected loss, with the expectation taken over the probability distribution of  $X$ . In practice, neither the distribution of  $X$  nor the ground truth function  $g: X \rightarrow Y$  are known exactly, but can be computed only empirically by collecting a large number of samples of  $X$  and hand-labeling them using the correct value of  $Y$  (a time-consuming process, which is typically the limiting factor in the amount of data of this sort that can be collected). The particular loss function depends on the type of label being predicted. For example, in the case of classification, the simple zero-one loss function is often sufficient. This corresponds simply to assigning a loss of 1 to any incorrect labeling and implies that the optimal classifier minimizes the error rate on independent test data (i.e. counting up the fraction of instances that the learned function  $h: X \rightarrow Y$  labels wrongly, which is equivalent to maximizing the number of correctly classified instances). The goal of the learning procedure is then to minimize the error rate (maximize the correctness) on a "typical" test set.

For a probabilistic pattern recognizer, the problem is instead to estimate the probability of each possible output label given a particular input instance, i.e., to estimate a function of the form

where the feature vector input is  $x$ , and the function  $f$  is typically parameterized by some parameters  $\theta$ . [ 8 ] In a discriminative approach to the problem,  $f$  is estimated directly. In a generative approach, however, the inverse probability  $p(x | a \text{ b e l})$  is instead estimated and combined with the prior probability  $p(a \text{ b e l} | \theta)$  using Bayes' rule, as follows:

When the labels are continuously distributed (e.g., in regression analysis), the denominator involves integration rather than summation:

The value of  $\theta$  is typically learned using maximum a posteriori (MAP) estimation. This finds the best value that simultaneously meets two conflicting objects: To perform as well as possible on the training data (smallest error-rate) and to find the simplest possible model. Essentially, this combines maximum likelihood estimation with a regularization procedure that favors simpler models over more complex models. In a Bayesian context, the regularization procedure can be viewed as placing a prior probability  $p(\theta)$  on different values of  $\theta$ . Mathematically:

where  $\theta^*$  is the value used for  $\theta$  in the subsequent evaluation procedure, and  $p(\theta | D)$ , the posterior probability of  $\theta$ , is given by

In the Bayesian approach to this problem, instead of choosing a single parameter vector  $\theta^*$ , the probability of a given label for a new instance  $x$  is computed by integrating over all possible values of  $\theta$ , weighted according to the posterior probability:

Frequentist or Bayesian approach to pattern recognition

The first pattern classifier – the linear discriminant presented by Fisher – was developed in the frequentist tradition. The frequentist approach entails that the model parameters are considered unknown, but objective. The parameters are then computed (estimated) from the collected data. For the linear discriminant, these parameters are precisely the mean vectors and the covariance matrix. Also the probability of each class  $p(a \text{ b e l} | \theta)$  is estimated from the collected dataset. Note that the usage of 'Bayes rule' in a pattern classifier does not make the classification approach Bayesian.

Bayesian statistics has its origin in Greek philosophy where a distinction was already made between the 'a priori' and the 'a posteriori' knowledge. Later Kant defined his distinction between what is a priori known – before observation – and the empirical knowledge gained from

observations. In a Bayesian pattern classifier, the class probabilities  $p(\text{label} | \theta)$  can be chosen by the user, which are then a priori. Moreover, experience quantified as a priori parameter values can be weighted with empirical observations – using e.g., the Beta- (conjugate prior) and Dirichlet-distributions. The Bayesian approach facilitates a seamless intermixing between expert knowledge in the form of subjective probabilities, and objective observations.

Probabilistic pattern classifiers can be used according to a frequentist or a Bayesian approach.

#### Uses

Within medical science, pattern recognition is the basis for computer-aided diagnosis (CAD) systems. CAD describes a procedure that supports the doctor's interpretations and findings. Other typical applications of pattern recognition techniques are automatic speech recognition, speaker identification, classification of text into several categories (e.g., spam or non-spam email messages), the automatic recognition of handwriting on postal envelopes, automatic recognition of images of human faces, or handwriting image extraction from medical forms. [ 9 ] [ 10 ] The last two examples form the subtopic image analysis of pattern recognition that deals with digital images as input to pattern recognition systems. [ 11 ] [ 12 ]

Optical character recognition is an example of the application of a pattern classifier. The method of signing one's name was captured with stylus and overlay starting in 1990. [ citation needed ] The strokes, speed, relative min, relative max, acceleration and pressure is used to uniquely identify and confirm identity. Banks were first offered this technology, but were content to collect from the FDIC for any bank fraud and did not want to inconvenience customers. [ citation needed ]

Pattern recognition has many real-world applications in image processing. Some examples include:

identification and authentication: e.g., license plate recognition, [ 13 ] fingerprint analysis, face detection /verification, [ 14 ] and voice-based authentication. [ 15 ]

medical diagnosis: e.g., screening for cervical cancer (Papnet), [ 16 ] breast tumors or heart sounds;

defense: various navigation and guidance systems, target recognition systems, shape recognition technology etc.

mobility: advanced driver assistance systems, autonomous vehicle technology, etc. [ 17 ] [ 18 ] [ 19 ] [ 20 ] [ 21 ]

In psychology, pattern recognition is used to make sense of and identify objects, and is closely related to perception. This explains how the sensory inputs humans receive are made meaningful. Pattern recognition can be thought of in two different ways. The first concerns template matching and the second concerns feature detection. A template is a pattern used to produce items of the same proportions. The template-matching hypothesis suggests that incoming stimuli are compared with templates in the long-term memory. If there is a match, the stimulus is identified. Feature detection models, such as the Pandemonium system for classifying letters (Selfridge, 1959), suggest that the stimuli are broken down into their component parts for identification. One observation is a capital E having three horizontal lines and one vertical line. [ 22 ]

#### Algorithms

Algorithms for pattern recognition depend on the type of label output, on whether learning is supervised or unsupervised, and on whether the algorithm is statistical or non-statistical in nature. Statistical algorithms can further be categorized as generative or discriminative.

Classification methods (methods predicting categorical labels)

Parametric: [ 23 ]

Linear discriminant analysis

Quadratic discriminant analysis



Maximum entropy classifier (aka logistic regression , multinomial logistic regression ): Note that logistic regression is an algorithm for classification, despite its name. (The name comes from the fact that logistic regression uses an extension of a linear regression model to model the probability of an input being in a particular class.)

Nonparametric: [ 24 ]

Decision trees , decision lists

Kernel estimation and K-nearest-neighbor algorithms

Naive Bayes classifier

Neural networks (multi-layer perceptrons)

Perceptrons

Support vector machines

Gene expression programming

Clustering methods (methods for classifying and predicting categorical labels)

Categorical mixture models

Hierarchical clustering (agglomerative or divisive)

K-means clustering

Correlation clustering

Kernel principal component analysis (Kernel PCA)

Ensemble learning algorithms (supervised meta-algorithms for combining multiple learning algorithms together)

Boosting (meta-algorithm)

Bootstrap aggregating ("bagging")

Ensemble averaging

Mixture of experts , hierarchical mixture of experts

General methods for predicting arbitrarily-structured (sets of) labels

Bayesian networks

Markov random fields

Multilinear subspace learning algorithms (predicting labels of multidimensional data using tensor representations)

Unsupervised:

Multilinear principal component analysis (MPCA)

Real-valued sequence labeling methods (predicting sequences of real-valued labels)

Kalman filters

Particle filters

Regression methods (predicting real-valued labels)

Gaussian process regression (kriging)

Linear regression and extensions

Independent component analysis (ICA)

Principal components analysis (PCA)

Sequence labeling methods (predicting sequences of categorical labels)

Conditional random fields (CRFs)

Hidden Markov models (HMMs)

Maximum entropy Markov models (MEMMs)

Recurrent neural networks (RNNs)

Dynamic time warping (DTW)

See also

Adaptive resonance theory – Theory in neuropsychology

Black box – System where only the inputs and outputs can be viewed, and not its implementation

Cache language model

Compound-term processing

Computer-aided diagnosis – Type of diagnosis assisted by computers

Contextual image classification

Data mining – Process of extracting and discovering patterns in large data sets

Deep learning – Branch of machine learning

Grey box model – Mathematical data production model with limited structure

Information theory – Scientific study of digital information

List of datasets for machine learning research

List of numerical-analysis software

List of numerical libraries

Neocognitron – Type of artificial neural network

Perception – Interpretation of sensory information

Perceptual learning – Process of learning better perception skills

Predictive analytics – Statistical techniques analyzing facts to make predictions about unknown events

Prior knowledge for pattern recognition

Sequence mining – Data mining technique Pages displaying short descriptions of redirect targets

Template matching – Technique in digital image processing

References

Further reading

Fukunaga, Keinosuke (1990). Introduction to Statistical Pattern Recognition (2nd ed.). Boston: Academic Press. ISBN 978-0-12-269851-4 .

Hornegger, Joachim; Paulus, Dietrich W. R. (1999). Applied Pattern Recognition: A Practical Introduction to Image and Speech Processing in C++ (2nd ed.). San Francisco: Morgan Kaufmann Publishers. ISBN 978-3-528-15558-2 .

Schuermann, Juergen (1996). Pattern Classification: A Unified View of Statistical and Neural Approaches . New York: Wiley. ISBN 978-0-471-13534-0 .

Godfried T. Toussaint, ed. (1988). Computational Morphology . Amsterdam: North-Holland Publishing Company. ISBN 9781483296722 .

Kulikowski, Casimir A.; Weiss, Sholom M. (1991). Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems . San

Francisco: Morgan Kaufmann Publishers. ISBN 978-1-55860-065-2 .

Duda, Richard O.; Hart, Peter E.; Stork, David G. (2000). Pattern Classification (2nd ed.). Wiley-Interscience. ISBN 978-0471056690 .

Jain, Anil.K.; Duin, Robert.P.W.; Mao, Jianchang (2000). "Statistical pattern recognition: a review". IEEE Transactions on Pattern Analysis and Machine Intelligence . 22 (1): 4– 37. Bibcode : 2000ITPAM..22....4J . CiteSeerX 10.1.1.123.8151 . doi : 10.1109/34.824819 . S2CID 192934 .

An introductory tutorial to classifiers (introducing the basic terms, with numeric example)

Kovalevsky, V. A. (1980). Image Pattern Recognition . New York, NY: Springer New York. ISBN 978-1-4612-6033-2 . OCLC 852790446 .

External links

The International Association for Pattern Recognition

List of Pattern Recognition web sites

Journal of Pattern Recognition Research Archived 2008-09-08 at the Wayback Machine

Pattern Recognition Info

Pattern Recognition (Journal of the Pattern Recognition Society)

International Journal of Pattern Recognition and Artificial Intelligence Archived 2004-12-11 at the Wayback Machine

International Journal of Applied Pattern Recognition

Open Pattern Recognition Project , intended to be an open source platform for sharing algorithms of pattern recognition

Improved Fast Pattern Matching Improved Fast Pattern Matching

v

t

e

Differentiable programming

Information geometry

Statistical manifold

Automatic differentiation

Neuromorphic computing

Pattern recognition

Ricci calculus

Computational learning theory

Inductive bias

IPU

TPU

VPU

Memristor

SpiNNaker

TensorFlow

PyTorch

Keras

scikit-learn

Theano

JAX

Flux.jl

MindSpore

Portals Computer programming Technology

Computer programming

Technology

GND

United States

Japan

Czech Republic

Israel

Yale LUX