ABSTRACT

The machine learning field, which can be briefly defined as enabling computers make successful predictions using past experiences, has exhibited an impressive development recently with the help of the rapid increase in the storage capacity and processing power of computers. Together with many other disciplines, machine learning methods have been widely employed in bioinformatics. The difficulties and cost of biological analyses have led to the development of sophisticated machine learning approaches for this application area. In this chapter, we first review the fundamental concepts of machine learning such as feature assessment, unsupervised versus supervised learning and types of classification. Then, we point out the main issues of designing machine learning experiments and their performance evaluation. Finally, we introduce some supervised learning methods.

When Should You Use Machine Learning?

Consider using machine learning when you have a complex task or problem involving a large amount of data and lots of variables, but no existing formula or equation. For example, machine learning is a good option if you need to handle situations like these:

Hand-written rules and equations are too complex—as in face recognition and speech recognition.

The rules of a task are constantly changing—as in fraud detection from transaction records.

The nature of the data keeps changing, and the program needs to adapt—as in automated trading, energy demand forecasting, and predicting shopping trends.

Supervised Learning

[Supervised machine learning](https://www.mathworks.com/discovery/supervised-learning.html) builds a model that makes predictions based on evidence in the presence of uncertainty. A supervised learning algorithm takes a known set of input data and known responses to the data (output) and trains a model to generate reasonable predictions for the response to new data. Use supervised learning if you have known data for the output you are trying to predict.

Supervised learning uses classification and regression techniques to develop predictive models.

**Classification techniques** predict discrete responses—for example, whether an email is genuine or spam, or whether a tumor is cancerous or benign. Classification models classify input data into categories. Typical applications include medical imaging, speech recognition, and credit scoring.

Use classification if your data can be tagged, categorized, or separated into specific groups or classes. For example, applications for hand-writing recognition use classification to recognize letters and numbers. In image processing and computer vision, [unsupervised pattern recognition](https://www.mathworks.com/discovery/pattern-recognition.html) techniques are used for object detection and image segmentation.

Common algorithms for performing classification include [support vector machine (SVM)](https://www.mathworks.com/help/stats/support-vector-machine-classification.html), [boosted](https://www.mathworks.com/help/stats/classification-ensembles.html) and [bagged](https://www.mathworks.com/help/stats/classification-ensembles.html) [decision trees](https://www.mathworks.com/help/stats/classification-trees.html), [*k*-nearest neighbor](https://www.mathworks.com/help/stats/classification-nearest-neighbors.html), [Naïve Bayes](https://www.mathworks.com/help/stats/classification-naive-bayes.html), [discriminant analysis](https://www.mathworks.com/help/stats/classification-discriminant-analysis.html), [logistic regression](https://www.mathworks.com/help/stats/generalized-linear-regression-1.html), and [neural networks](https://www.mathworks.com/help/deeplearning/pattern-recognition-and-classification.html).

**Regression techniques** predict continuous responses—for example, changes in temperature or fluctuations in power demand. Typical applications include electricity load forecasting and algorithmic trading.

Use regression techniques if you are working with a data range or if the nature of your response is a real number, such as temperature or the time until failure for a piece of equipment.

Common regression algorithms include [linear model](https://www.mathworks.com/help/stats/multiple-linear-regression-1.html), [nonlinear model](https://www.mathworks.com/help/stats/nonlinear-models.html), [regularization](https://www.mathworks.com/help/stats/regularization-1.html), [stepwise regression](https://www.mathworks.com/help/stats/stepwise-regression-1.html), [boosted](https://www.mathworks.com/help/stats/classification-ensembles.html) and [bagged](https://www.mathworks.com/help/stats/classification-ensembles.html) [decision trees](https://www.mathworks.com/help/stats/classification-trees.html), [neural networks](https://www.mathworks.com/help/deeplearning/function-approximation-and-nonlinear-regression.html), and [adaptive neuro-fuzzy learning](https://www.mathworks.com/help/fuzzy/anfis.html).

### Unsupervised Learning

[Unsupervised learning](https://www.mathworks.com/discovery/unsupervised-learning.html) finds hidden patterns or intrinsic structures in data. It is used to draw inferences from datasets consisting of input data without labeled responses.

**Clustering** is the most common unsupervised learning technique. It is used for exploratory data analysis to find hidden patterns or groupings in data. Applications for [cluster analysis](https://www.mathworks.com/discovery/cluster-analysis.html) include gene sequence analysis, market research, and object recognition.

For example, if a cell phone company wants optimize the locations where they build cell phone towers, they can use machine learning to estimate the number of clusters of people relying on their towers. A phone can only talk to one tower at a time, so the team uses clustering algorithms to design the best placement of cell towers to optimize signal reception for groups, or clusters, of their customers.

Common algorithms for performing clustering include [k-means and k-medoids](https://www.mathworks.com/help/stats/k-means-clustering-12.html), [hierarchical clustering](https://www.mathworks.com/help/stats/hierarchical-clustering-12.html), [Gaussian mixture models](https://www.mathworks.com/help/stats/gaussian-mixture-models.html), [hidden Markov models](https://www.mathworks.com/help/stats/hidden-markov-models.html), [self-organizing maps,](https://www.mathworks.com/help/deeplearning/self-organizing-maps.html) [fuzzy c-means clustering](https://www.mathworks.com/help/fuzzy/fcm.html), and [subtractive clustering](https://www.mathworks.com/help/fuzzy/subclust.html).

K-Means algorithm

K-Means algorithm based on dividing [4] [5] is a kind of cluster algorithm, and it is proposed by

J.B.MacQueen. This algorithm which is unsupervised is usually used in data mining and pattern

recognition. Aiming at minimizing cluster performance index, square-error and error criterion are

foundations of this algorithm. To seek the optimalizing outcome, this algorithm tries to find K divisions to

satisfy a certain criterion. Firstly, choose some dots to represent the initial cluster focal points(usually, we

choose the first K sample dots of income to represent the initial cluster focal point); secondly, gather the

remaining sample dots to their focal points in accordance with the criterion of minimum distance, then we

will get the initial classification, and if the classification if unreasonable, we will modify it(calculate each

cluster focal points again), iterate repetitively till we get a reasonable classification.

K-Means algorithm based on dividing is a kind of cluster algorithm, and has advantages of briefness,

efficiency and celerity.

However, this algorithm depends quite much on initial dots and the difference in choosing initial

samples which always leads to different outcomes. What’s more, this algorithm based on target function

always uses gradient method to get extremum. The direction of search in gradient method is always along

the direction in which energy decreases, which will leads to the fact that when the initial cluster focal

point is not proper, and then the whole algorithm will easily sink into local minimum point.

K-Means algorithm

K-Means algorithm based on dividing [4] [5] is a kind of cluster algorithm, and it is proposed by

J.B.MacQueen. This algorithm which is unsupervised is usually used in data mining and pattern

recognition. Aiming at minimizing cluster performance index, square-error and error criterion are

foundations of this algorithm. To seek the optimalizing outcome, this algorithm tries to find K divisions to

satisfy a certain criterion. Firstly, choose some dots to represent the initial cluster focal points(usually, we

choose the first K sample dots of income to represent the initial cluster focal point); secondly, gather the

remaining sample dots to their focal points in accordance with the criterion of minimum distance, then we

will get the initial classification, and if the classification if unreasonable, we will modify it(calculate each

cluster focal points again), iterate repetitively till we get a reasonable classification.

K-Means algorithm based on dividing is a kind of cluster algorithm, and has advantages of briefness,

efficiency and celerity.

However, this algorithm depends quite much on initial dots and the difference in choosing initial

samples which always leads to different outcomes. What’s more, this algorithm based on target function

always uses gradient method to get extremum. The direction of search in gradient method is always along

the direction in which energy decreases, which will leads to the fact that when the initial cluster focal

point is not proper, and then the whole algorithm will easily sink into local minimum point.

K-Means algorithm

K-Means algorithm based on dividing [4] [5] is a kind of cluster algorithm, and it is proposed by

J.B.MacQueen. This algorithm which is unsupervised is usually used in data mining and pattern

recognition. Aiming at minimizing cluster performance index, square-error and error criterion are

foundations of this algorithm. To seek the optimalizing outcome, this algorithm tries to find K divisions to

satisfy a certain criterion. Firstly, choose some dots to represent the initial cluster focal points(usually, we

choose the first K sample dots of income to represent the initial cluster focal point); secondly, gather the

remaining sample dots to their focal points in accordance with the criterion of minimum distance, then we

will get the initial classification, and if the classification if unreasonable, we will modify it(calculate each

cluster focal points again), iterate repetitively till we get a reasonable classification.

K-Means algorithm based on dividing is a kind of cluster algorithm, and has advantages of briefness,

efficiency and celerity.

However, this algorithm depends quite much on initial dots and the difference in choosing initial

samples which always leads to different outcomes. What’s more, this algorithm based on target function

always uses gradient method to get extremum. The direction of search in gradient method is always along

the direction in which energy decreases, which will leads to the fact that when the initial cluster focal

point is not proper, and then the whole algorithm will easily sink into local minimum point.

K-Means algorithm

K-Means algorithm based on dividing [4] [5] is a kind of cluster algorithm, and it is proposed by

J.B.MacQueen. This algorithm which is unsupervised is usually used in data mining and pattern

recognition. Aiming at minimizing cluster performance index, square-error and error criterion are

foundations of this algorithm. To seek the optimalizing outcome, this algorithm tries to find K divisions to

satisfy a certain criterion. Firstly, choose some dots to represent the initial cluster focal points(usually, we

choose the first K sample dots of income to represent the initial cluster focal point); secondly, gather the

remaining sample dots to their focal points in accordance with the criterion of minimum distance, then we

will get the initial classification, and if the classification if unreasonable, we will modify it(calculate each

cluster focal points again), iterate repetitively till we get a reasonable classification.

K-Means algorithm based on dividing is a kind of cluster algorithm, and has advantages of briefness,

efficiency and celerity.

However, this algorithm depends quite much on initial dots and the difference in choosing initial

samples which always leads to different outcomes. What’s more, this algorithm based on target function

always uses gradient method to get extremum. The direction of search in gradient method is always along

the direction in which energy decreases, which will leads to the fact that when the initial cluster focal

point is not proper, and then the whole algorithm will easily sink into local minimum point.