Feature (machine learning)

In machine learning and pattern recognition, a **feature** is an individual measurable property of a phenomenon being observed.[1] Choosing informative, discriminating and independent features is a crucial step for effective algorithms in pattern recognition, classification and regression. Features are usually numeric, but structural features such as strings and graphs are used in syntactic pattern recognition. The concept of "feature" is related to that of explanatory variable used in statistical techniques such as linear regression.

The initial set of raw features can be redundant and too large to be managed. Therefore, a preliminary step in many applications of machine learning and pattern recognition consists of selecting a subset of features, or constructing a new and reduced set of features to facilitate learning, and to improve generalization and interpretability.

Extracting or selecting features is a combination of art and science; developing systems to do so is known as feature engineering. It requires the experimentation of multiple possibilities and the combination of automated techniques with the intuition and knowledge of the domain expert. Automating this process is feature learning, where a machine not only uses features for learning, but learns the features itself.

1 Classification

A set of numeric features can be conveniently described by a feature vector. An example of reaching a two way classification from a feature vector (related to the perceptron) consists of calculating the scalar product between the feature vector and a vector of weights, comparing the result with a threshold, and deciding the class based on the comparison.

Algorithms for classification from a feature vector include nearest neighbor classification, neural networks, and statistical techniques such as Bayesian approaches.

2 Examples

See also: Feature (computer vision)

In character recognition, features may include histograms counting the number of black pixels along horizontal and vertical directions, number of internal holes, stroke detection and many others.

In speech recognition, features for recognizing phonemes can include noise ratios, length of sounds, relative power, filter matches and many others.

In spam detection algorithms, features may include the presence or absence of certain email headers, the email structure, the language, the frequency of specific terms, the grammatical correctness of the text.

In computer vision, there are a large number of possible features, such as edges and objects.

3 References

[1] Bishop, Christopher (2006). *Pattern recognition and machine learning*. Berlin: Springer. ISBN 0-387-31073-8.

4 See also

- Covariate

- Hashing trick

5   Text and image sources, contributors, and licenses

5.1   Text

- **Feature (machine learning)** *Source:* https://en.wikipedia.org/wiki/Feature_(machine_learning)?oldid=702634191 *Contributors:* Edward, Kku, Benwing, SchmuckyTheCat, KellyCoinGuy, D6, 3mta3, Brookie, BD2412, Qwertyus, YurikBot, TimBentley, Nbarth, DHN-bot~enwiki, Renne, Thijs!bot, Escarbot, AnAj, Rettetast, SieBot, Paux, Ecov, MystBot, Addbot, Buster7, Yobot, Erik9bot, HRoestBot, Leonardo61, Widr and Anonymous: 11

5.2   Images

- **File:Question_book-new.svg** *Source:* https://upload.wikimedia.org/wikipedia/en/9/99/Question_book-new.svg *License:* Cc-by-sa-3.0 *Contributors:*
  Created from scratch in Adobe Illustrator. Based on Image:Question book.png created by User:Equazcion *Original artist:*
  Tkgd2007

5.3   Content license

- Creative Commons Attribution-Share Alike 3.0