Intelligent Systems: Reasoning and Recognition

James L. Crowley

ENSIMAG 2 / MoSIG M1 Second Semester 2010/2011

Lesson 12 30 March 2011

Introduction to Bayesian Recognition

Sources Bibliographiques :
"Pattern Recognition and Machine Learning", C. M. Bishop, Springer Verlag, 2006.
"Pattern Recognition and Scene Analysis", R. E. Duda and P. E. Hart, Wiley, 1973.

Notation

x	A variable
X	A random variable (unpredictable value)
N	The number of possible values for x (Can be infinite).
$\vec{x}$	A vector of D variables.
$\vec{X}$	A vector of D random variables.
D	The number of dimensions for the vector $\vec{x}$ or $\vec{X}$
E	An observation. An event.
C_k	The class k
k	Class index
K	Total number of classes
ω_k	The statement (assertion) that E $\in T_k$
M_k	Number of examples for the class k. (think M = Mass)
M	Total number of examples. $M = \sum_{k=1}^{K} M_k$
$\{X_m^k\}$	A set of M_k examples for the class k. $\{X_m\} = \bigcup_{k=1,K} \{X_m^k\}$

Bayesian Recognition

Recognition is a fundamental ability for intelligence, and indeed for all life.
To survive, any creature must be able to recognize food, enemies and friends.

Recognize: To identify an object or an entity as known:

Two forms of recognition: Identify and Classify

Identify: To recognize an object or entity as an individual
Classify: To recognize an object or entity as a member of a class.

Categorize is sometimes used in place of classify.

A class is a form of set, defined by a membership test. The set may be defined by listing its members (closed set) or by listing properties that define members (open set). These correspond to the two methods to define a set:

Extension: Provide a list of members.
Intension: Provide a conjunction of predicates.

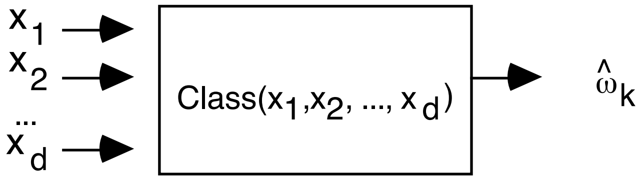
Classification is a process of associating an event to a class. The event is described by a vector of features, provided by an observation.

The event (or observation) E is described by a vector of features, $\vec{X}$
Features are provided by sensors.

<u>Features</u>: observable properties that permit assignment of events to classes.
A set of D features, x_d, are assembled into a feature vector $\vec{X}$

$$\vec{X} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_D \end{pmatrix}$$

A classifier is a process that maps an event, E, to a class label, C_k, based on features.
The result is the proposition $\omega_k = E \in \text{Class } C_k$



The techniques from pattern recognition and statistics provide a variety of methods to construct membership tests for classification of observations. The most appropriate technique depends on the number and nature of the classes and the features.

There are two families of technique: extensive and intensive

Extensive methods compare the features from an observation to a set of prototype examples, using a similarity function.

$$\hat{\omega}_k = \arg\text{-}\max_k\left\{Sim(\vec{X},\vec{X}_m^k)\right\}$$

Distance is often used to define similarity.

For an extensive class definition, we enumerate the M examples of each of K class. The estimate is the most similar, as provided by some similarity function.

Thus for an observed unknown event E, described by the features X, the "estimated" class, $\hat{\omega}_k$ is given by :

$$\hat{\omega}_k = \forall_k \forall_m : \arg\text{-}\max_k\left\{Sim(\vec{X},\vec{X}_m^k)\right\}$$

This approach is often used for generative methods for machine learning.

For an intensive class definitions, we apply a set of tests, one for each class.

$$\hat{\omega}_k = \arg\text{-}\max_k\left\{Test-for-k(\vec{X})\right\}$$

This approach is used for discriminative methods to machine learning.

Bayesian Classification

"Bayesian" refers to the 18th century mathematician and theologian Thomas Bayes (1702–1761), who provided the first mathematical treatment of a non-trivial problem of Bayesian inference. Bayesian probability was made popular by Simon Laplace in the early 19th century.

The rules of Bayesian logic can be justified by requirements of rationality and consistency and interpreted as an extension of logic. Many modern machine learning methods are based on objectivist Bayesian principles.

With a Bayesian approach, the tests are designed to minimize the number of errors. False positives and false negatives count equally as errors, but can have different costs associated.

This approach makes it possible to include the cost of error, which may not be the same for a false positive and a false negative.

Let ω_k be the proposition that the event belongs to class k: $\omega_k = E \in T_k$

ω_k Proposition that event E $\in$ the class k
$p(\omega_k) = p(E \in T_k)$ Probability that E is a member of class k

Given an observation $\vec{X}$, the decision criteria is

$$\hat{\omega}_k = \arg\text{-}\max_k \left\{ \Pr(\omega_k \mid \vec{X}) \right\}$$

where $\omega_k \equiv E \in T_k$

The meaning of "given" is provided by Bayes Rule:

$$p(\omega_k \mid \vec{X}) = \frac{P(\vec{X} \mid \omega_k) p(\omega_k)}{P(\vec{X})}$$

Applying Bayes rule for classification will require us to define probability.

Probability and Uncertainty

Bayes rule provides a method to accumulate evidence to reduce uncertainty.

The core problem of recognition is uncertainty. One could even say that recognition is a problem of assigning signals to categories in the presence of uncertainty.

We can distinguish two separate kinds of uncertainties: Confidence and Accuracy (Precision).

Confidence: Freedom from doubt, belief in the truth of a proposition.
Accuracy: Reproducibility of a measurement.

Confidence concerns the truth of a statement. The proposition is generally formalized as a predicate (truth function). Predicates are generally defined a boolean truth functions (True or false). It is possible to define probabilistic truth functions.

Accuracy concerns a selecting an entity from an ordered set. Generally there is some order between the possible values with an associated distance metric. The accuracy refers to the size of a subset of possible values or the distance spanned by possible values.

In popular language, accuracy is often confused with precision.
In informatics:

Accuracy is the degree to which a measurement can be reproduced.
Precision is the detail with which a measurement is represented.

For example, a measurement may be represented with 32 bits of precision, but be accurate to only 8 bits (1 part in 256).
In common usage, precision and accuracy are often used for the same concept.

Probability is a powerful tool for both Confidence and Accuracy.

Both confidence and precision may be addressed in using Bayesian probabilities.

Probability as Frequency of Occurence.

A frequency based definition of probability is sufficient for many practical problems.

Suppose we have M observations of random events, $\{E_m\}$, for which M_k of these events belong to the class k. The probability that one of these observed events belongs to the class k is:

$$Pr(E \in T_k) = \frac{M_k}{M}$$

If we make new observations under the same observations conditions (ergodicity), then it is reasonable to expect the fraction to be the same. However, because the observations are random, there may be differences. These differences will grow smaller as M grows larger.

The average (root-mean-square) error for

$$Pr(E \in T_k) = \frac{M_k}{M}$$

will be proportional to M_k and inversely proportional to M.