Machine Intelligence (SE2MI11)

Pattern Recognition

(Spring term, 2011-2012)

Xia Hong

Recommended books:

1. R. O. Duda and P. E. Hart: Pattern Classification and Scene Analysis (1973) J. Wiley

2. C. M. Bishop: Neural Networks for Pattern Recognition (1995) Oxford University Press.

3. S. Haykin: Neural Networks (1999) Prentice Hall.

Introduction to Pattern Recognition

Pattern recognition is the discipline of building machines to perform perceptual tasks which we humans are particularly good at. e.g. recognize faces, voice, identify species of flowers, spot an approaching storm ....

There are practical needs to find more efficient ways of doing things. e.g. read hand-written symbols, diagnose diseases, identify incoming missiles from radar or sonar signals. The machines may perform these tasks faster, more accurately and cheaply.

The goal of pattern recognition research is to clarify complicated mechanisms of decision making processes and automatic these function using computers.

Pattern classification

Although we humans can perform some of the perceptual tasks with ease, there is not sufficient understanding to duplicate the performance with a computer.

Because the complex nature of the problems, many pattern recognition research has been concerned with more moderate problems of pattern classification — the assignment of a physical object or event to one of several prespecified categories.

Example: A lumber mill producing assorted hardwoods wants to automate the process of sorting finished lumber according to the species of trees.

Optical sensing is used to distinguish birch lumber from ash lumber. A camera takes a pictures of the lumber and passes to on to a feature extractor.

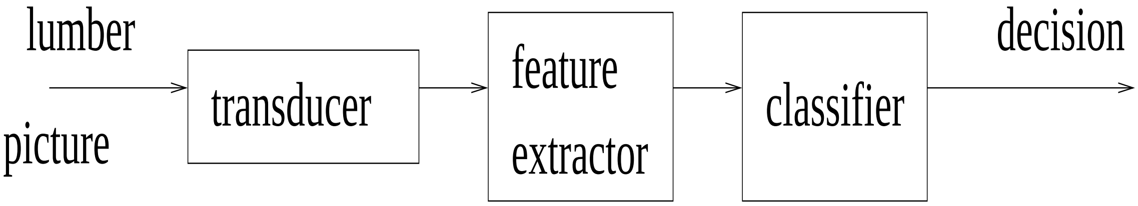

Figure: A pattern classification system

The feature extractor reduces the data by measuring certain "properties" that distinguish pictures of birch lumber from pictures of ash lumber.

These features are then passed to a classifier that evaluates the evidence presented and makes a final decision about the lumber type.

Suppose that somebody at the lumber mill tells us that birth is often lighter colored than ash. Then brightness becomes an obvious feature. We might attempt to classify the lumber merely by seeing whether or not the average brightness x exceeds some critical value.

One characteristic of human pattern recognition is that it involves a teacher. Similarly a machine pattern recognition system needs to be trained. A common mode of learning is to be given a collection of labelled examples, known as training data set. From the training data set, structure information is distilled and used for classifying new inputs.

In this case, we would obtain samples of different types of wood, make brightness measurements, and inspect the results. Suppose that we obtain the following histogram based on these data samples.

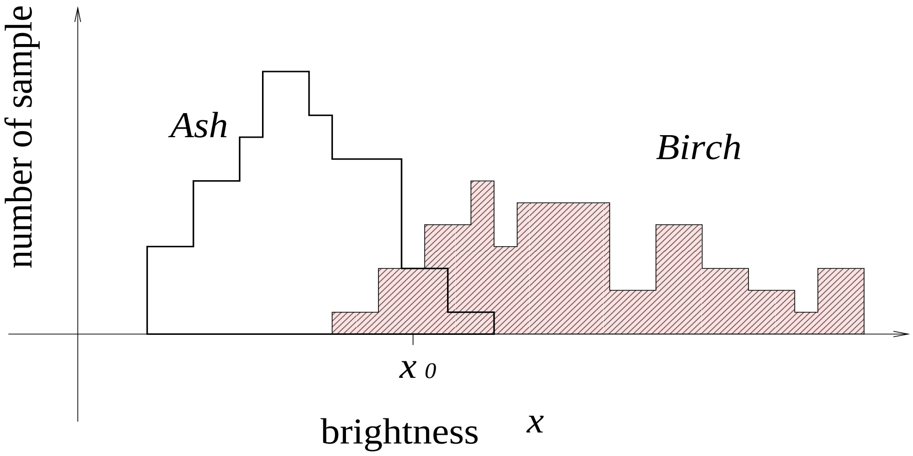

Figure: Histogram for the brightness feature.

The histogram bears out the statement that birch is usually lighter than ash, but it is clear that this single criterion is not infallible. No matter how we choose x_0, we cannot reliably separate birch from ash by brightness alone.

The second feature is based on the observation that ash typically has a more prominent grain pattern than birch. It is reasonable to assume that we can obtain a measure of this feature from the magnitude and frequency of occurrence of light-to-dark transitions in the picture.

The feature extractor has thus reduced each picture to a point or a feature vector $\mathbf{x}$ in a two dimensional space, where,

$$\mathbf{x} = [x_1, x_2]^T$$

where x_1 denotes the brightness, x_2 denotes the grain prominence.

Our problem now is to partition the feature space into two regions for birth and ash. Suppose that we measure the feature vectors for our training data samples and obtain the following scatter diagram.

This plot suggests the rule for classifying the data: Classify the lumber as ash if its feature vector falls above the line AB, and as birch otherwise.

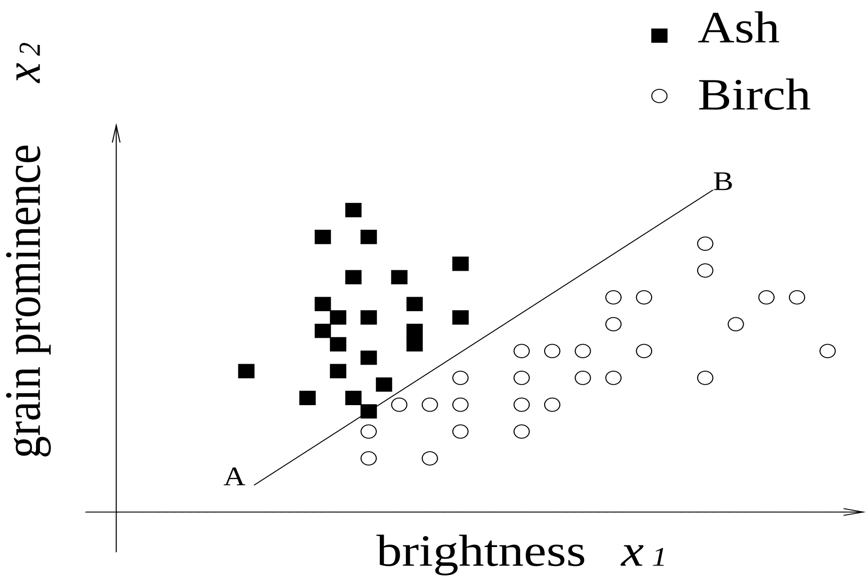

Figure: Scatter diagram for the feature vector.

In order to make sure that this rule performs well, we obtain more samples and adjust the position of line AB in order to minimize the probability of error. This suggests that pattern recognition problems has a statistical nature.

We also see from above example that pattern recognition involves transducer, feature extractor and classifier. In this course we will focus on *introducing different types of classifiers*.

Concepts in pattern recognition

Nonparametric: Nonparametric techniques do not rely on a set of parameters/weights.

Parametric: These models are parameterized, with its parameters/weights to be determined through some parameter optimization algorithm, which are then determined by fitting the model to the training data set.

Supervised: The training samples are given as some input/output pairs. The output is the desired response for the input. The parameters/weights are adjusted so as to minimize the errors between the response of the networks and the desired response.

Unsupervised: Suppose that we are given data samples without being told which classes they belong to. There are schemes that are aimed to discover significant patterns in the input data without a teacher (labelled data samples).

Overview of the course

Nonparametric, supervised

1. Parzen windows

This lecture deals with the problem of modelling the probability distribution of a set of data.

2. Probabilistic neural network (PNN)

PNN is a nonparametric classier based on Parzen windows density estimator by comparing probabilities of different classes. It can be used for multi class classification.

3. The nearest neighbor classifiers

The nearest neighbor classifiers bypass the probability distribution estimation and go directly to classification. It can be used for multi class classification.

Parametric, supervised

4. Linear discriminant analysis

All parametric methods involve a specific functional form. This lecture introduces the linear discriminant function. It is assumed that the decision function has linear relationship with the feature vector.

5. Radial basis functions (RBF) neural networks

Neural networks have the capability of modelling unknown nonlinear function well. RBF neural networks is a popular neural network. This lecture introduces the structure of RBF. and illustrates its use of modelling unknown nonlinear function.

6. RBF classifier

In this lecture we see how RBF neural networks can be used as a nonlinear discriminant function.

Unsupervised

7. k-means clustering

The k-means clustering itself is a type of unsupervised algorithm. In the lecture, its application in RBF center selection is introduced. The RBF neural network using k-means clustering for center selection is a hybrid algorithm of unsupervised /supervised (semi-supervised).

8. Kohonen's self-organizing feature (SOM) map

The best known unsupervised is the Kohonen's self-organizing feature (SOM) map. Finally we introduce SOM.