Pr is clsfctn of data based on statistical info extracted from patterns

Pr also involves clustering of data

In pr we have to process data and convert into form that a machine could process

For example we may measure feature of pattern and save them as a vector

In classification we use meaningful names for classes but in clustering it is not important

Representation and choice of attributes plays important role in classification

A pattern is represented as vector of feature values

**Paradigms for pattern recognition**

Statistical pattern recognition (for noise env)

Syntactic pattern recognition

Features are called attributes

**Data structures for pr**

Patterns as vectors

As strings ex dna sequence

Logical description

**Fuzzy and rough pattern sets**

Fuzziness is used when it is not possible to make precise stmnts. It is useful to model subjective,imprecise and incomplete data

in fuzziness objects belong to set depending on membership value which is between 0 and 1

The fundamental concept behind Rough Set Theory is the approximation of lower and

upper spaces of a set, the approximation of spaces being the formal classification of

knowledge regarding the interest domain.

The subset generated by lower approximations is characterized by objects that will

definitely form part of an interest subset, whereas the upper approximation is characterized

by objects that will possibly form part of an interest subset. Every subset defined through

upper and lower approximation is known as Rough Set

**patterns as trees and graphs**

**minimum spanning tree**

this can be used for clustering patterns

**frequent pattern tree**

used in transaction DB’s

generated from transactions in DB’s

it is a compressed tree useful for finding associations btw items in tr DB

prescne of some items imply the presence of other items

**representation of clusters**

there are 2 data structures here

one is partition p of patterns and other is set of cluster representatives c

**proximity measures**

**distance measure**

following properties hold for a metric

**minkowski**

**Euclidean distance**

If we are using distance measure we shud make sure that all the features or attributes have same range of values. Else it would appear as if we are giving more weightage to features of higher values. To do this we should perform normalization

**Mahalnobis**

**Weighted distance measure**

If we want to give importance to particular attribute we can use a weight to it

**Hausdroff distance**

H dist is the max dis between a point in one shape and the point that is nearest to it in other

**Non metric similarity fn**

Dese fn doesn’t obey triangular inequality and symmetry

Useful for string or images

Robust to extremely noisy data

**Ex:**squared Euclidean dist,k median dist

**Edit distance(levenshtein dist)**

Measures the dist btw 2 strings

Min no of changes needed to change s1 to s2

Mutations involve

Changing letter

Inserting letter

Deleting letter

**Mutual neighborhood dist**

**Conceptual cohesiveness**

Concept is a description of class of objects sharing some attribute value

CC uses domain knowledge in the form of set of concepts to characterize the similarity

Conceptual dist combines both symbolic and numerical methods

**Kernel funs**

**Polynomial kernelns**

**Radial basis fn kernel**

**Size of patterns**

**Normalization of data**

To make all patterns have same dimensions

Can also be done to give same imp to every feature

**Use of appropriate similarity measures**

**Abstraction of data set**

If we use the whole set of training patters the processing may take a very long time so we use abstraction

Depending on abstraction different classifiers are used

**No abstraction**

**Single representative per class**

**Multiple representatives per class**

**Cluster representatives as abstractions**

**Support vectors as representatives**

**Frequent item set based abstraction**

In transaction DB’s each pattern represents a trstn.

**Feature extraction**

Involved detecting and isolating features

**Fishers linear discriminant**

**Principal component analysis**

Transforms number of correlated variables into small number of uncorrelated variables called principal components

The first pc accounts for as much variation in data as possible the succeeding pc accounts for as much of the remaining variability as psbl

Pca find the most accurate data repre in lower dimensional space

Data is projected in the direction of max variance

**Feature selection**

Feature selection ensures

**Reduction in cost of pattern classification and design of the classifier**

**Improvement of classification accuracy**

Performance of classifier depends on interrelation between sample sizes, number of features and classifier complexity

Peaking phenomena

Feature selection algos search through different feature sub-sets

**Evaluation of classifiers**

*Accuracy of classifier*

*Design time and classification time*

*Space required*

*Explanation ability*

*Noise tolerance*

Resubstitution estimate

**Holdout method**

**Random sub-sampling**

**Cross-validation**

**Bootstrap procedure**