Pattern Recognition

Pattern recognition is the study of computational methods that can be used to improve the identification of states of nature under conditions of imperfect information.

Also known as pattern classification and shares techniques with discriminant analysis, machine learning and data mining.

For example, optical character recognition, voice recognition, document classification, spam filtering, medical diagnosis.

Issues

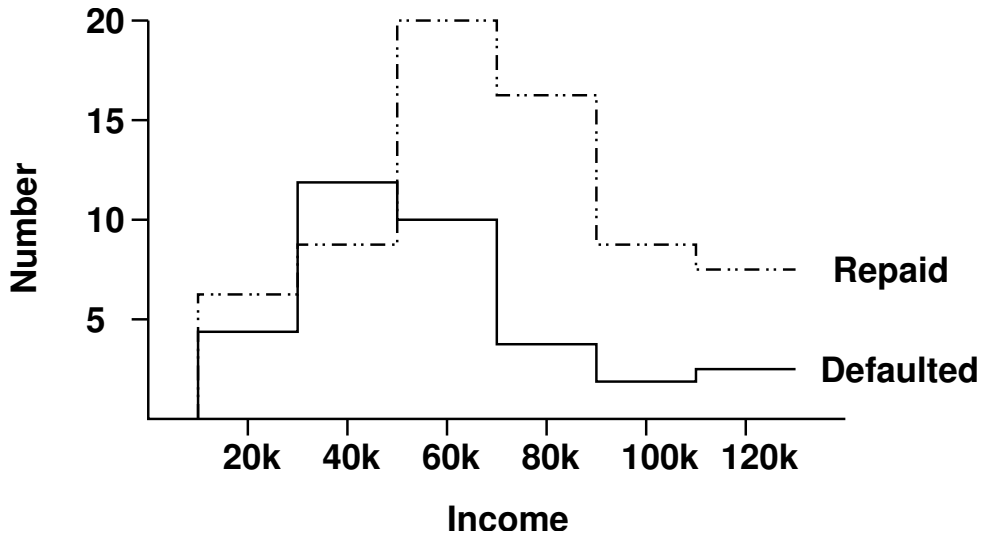
- What are the states of nature?
- Why is our information imperfect and can it be improved?
  - Unknown cause-and-effect relationships
  - Noise
  - Measurement error
- How does this relate to classification in nature (human/animal)?
- How does it relate to other disciplines - AI, ML, . . .

An Example

We want to determine if a given applicant should be given a loan or not based on their likelihood of repaying.

- A simple rule could be based on the past history of borrowers. For example, 65% of people repay, so give everyone a loan. We have 35% error.
- Can we collect data that will help us make this decision? For example, suppose we collect information on gross income and find:

Example (cont)



- We collect some design or training data and it appears that we can use income as a feature to separate the groups, but where do we put the income *threshold*?

Selecting a Threshold

Minimize the error rate?

Income	Accept Error	Reject Error	Total Error
20k	.31	.06	.36
40k	.19	.14	.33
60k	.09	.34	.43
80k	.05	.50	.55
100k	.02	.58	.61

But 33% error is still high.

Risk

What about cost? Suppose a default costs an average of $3000 and a lost loan that would be repaid costs $1000? Cost is usually called risk.

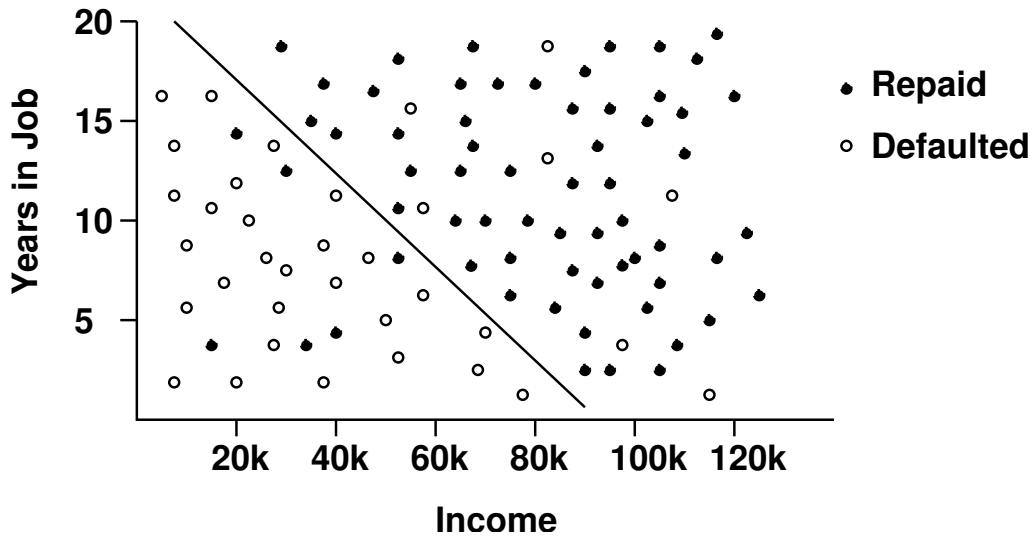
Income	Accept Risk	Reject Risk	Total Risk
20k	93,000	6,000	99,000
40k	57,000	14,000	71,000
60k	27,000	34,000	61,000
80k	15,000	50,000	65,000
100k	9,000	58,000	67,000

Risk (cont)

With a threshold of $60,000 The 100 loan applicants will result in an average of 40 acceptances with a risk of $61,000. We don't know what the income is to offset this cost.

Two-dimensional Feature Sets

Suppose we add the length of time in current job as a feature and get:



The error rate has improved to 13%.

Two-dimensional Feature Sets

Where the line is a decision boundary or showofflinear discriminant classifier. Any application falling right of the line is accepted, any to the left is rejected.

Note that a sample is now represented as a vector:

$$x = \begin{vmatrix} x_1 \\ x_2 \end{vmatrix}$$

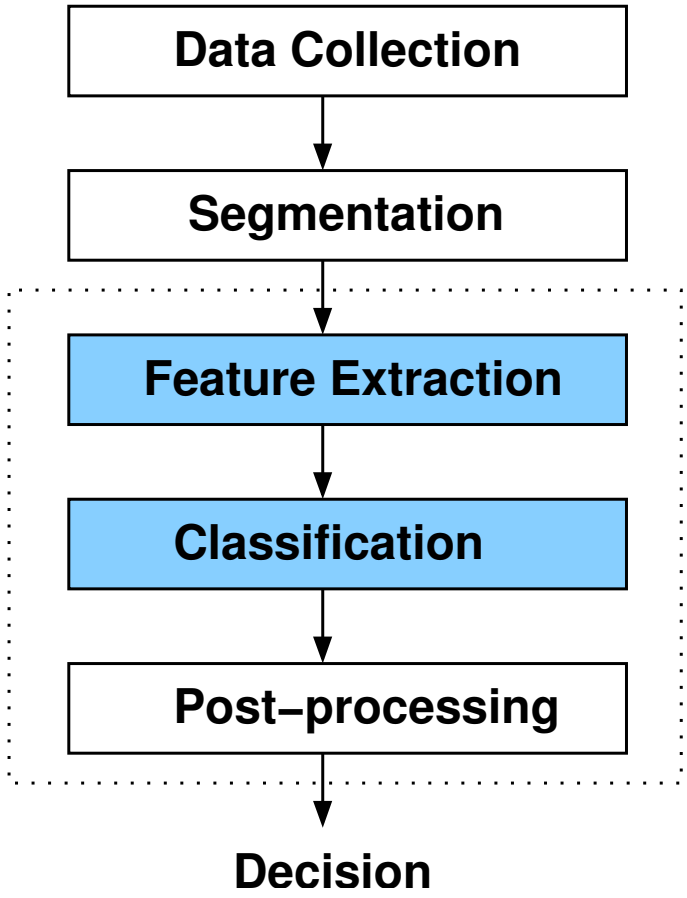
Classifiers can also be non-linear, but overly complex functions may result in unrobust behavior.

Features

Can we always improve by adding additional features?

- Additional features suffer from the curse of dimensionality which is related to the exponential increase in necessary training samples.
- Features may not be independent, resulting in larger training sets with little additional information.
- More complex models are more likely to be overly sensitive to evolution in the input stream.
- We want patterns requiring the same action to be close in feature space and those requiring different actions to be far apart/

Classification Systems



System Components

- Data collection - an experiment, sensing, collection, extraction.
- Segmentation - data processing.
- Feature Extraction - converting data into variables that have maximum impact on classification.
- Classification - selecting the class for the given experiment.
- Post-processing - error analysis; determine the performance of the classifier and determine if modifications are necessary.
- Decision - based on the results of the classifier, render a decision.

Types of Pattern Recognition

- Statistical PR is based on the probabilisitic nature of the the knowledge of the system.
- Neural network PR is based largely on the same foundations as statistical PR, but the approach is based on approximations of cognitive function.
- Syntactic PR uses logic rules.

Classifier Development