CSCE 970 Lecture 0: Administrivia

Stephen D. Scott

Welcome to CSCE 970!

You should have the following handouts:

1. Syllabus

2. Copies of slides (also on web page)

Please check off or write your name on the roster (if you write your name, indicate if you plan to register for the course)

CSCE 970 Lecture 1: Introduction

Stephen D. Scott

What is Pattern Recognition?

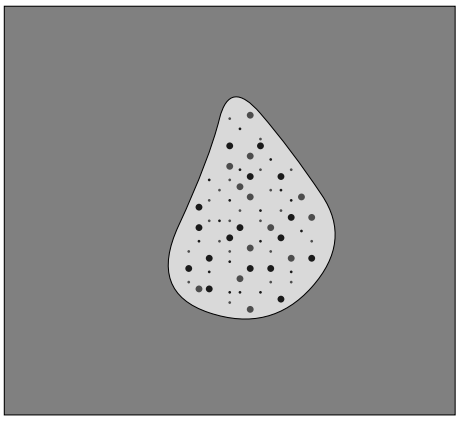
- Pattern reocognition: classify objects (instances, examples) into categories (classes, labels)

- Has deep roots in probability theory, statistics, machine learning, linear algebra, image processing, algorithms

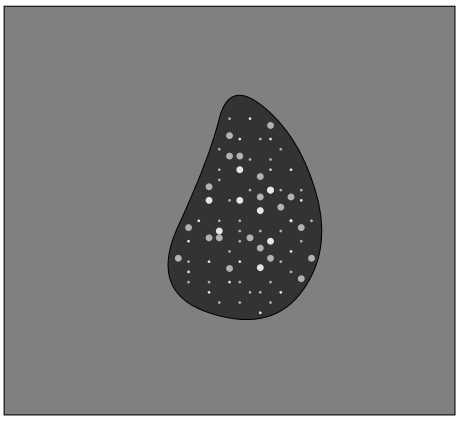
- Applications: Machine vision, OCR, handwriting recognition, computer-aided diagnosis, speech recognition

An Example

Class A



Class B

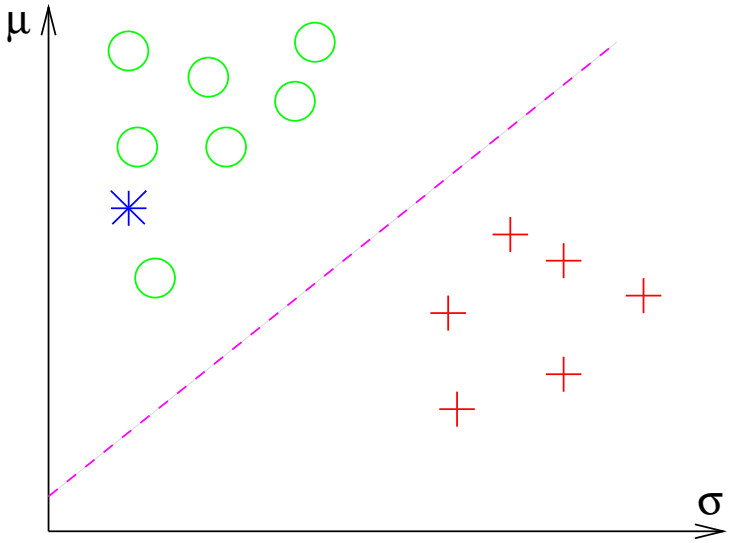


○ = Class A

\+ = Class B

✳ = unclassified

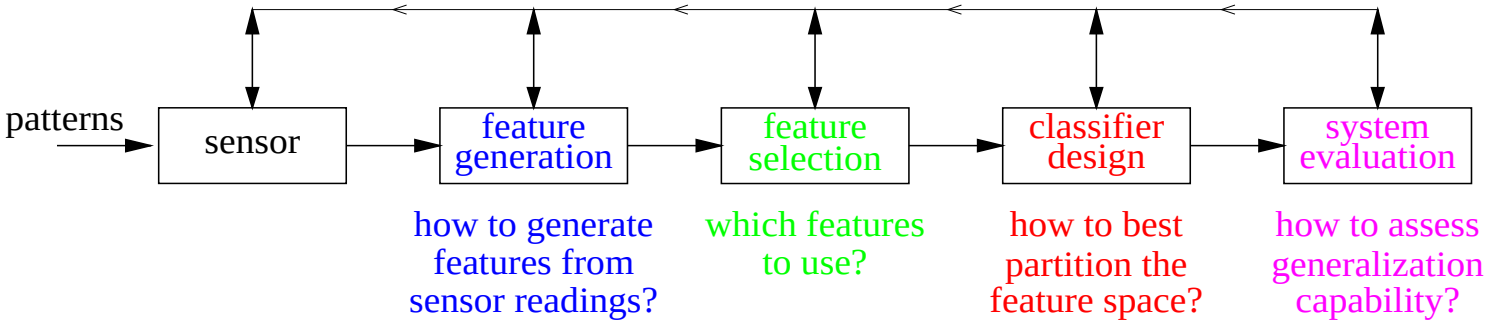
- - - - = decision line



Features, Feat. Vectors, Classifiers

- $\mathbf{x} = [x_1, \ldots, x_\ell]^T$ is a feature vector of ℓ features
  - E.g. $\mathbf{x} = [\mu, \sigma]^T$ from prev. ex.
  - Will consider features to be random variables
  - Feature vectors also known as instances or examples
- A classifier separates the feature space into regions corresponding to two or more classes (also known as labels)
  - Use to classify new, unlabeled instances
  - E.g. decision line from prev. ex.
- Classifier built by training (learning) using a training set of labeled instances
- Can also use labeled instances as a testing set to evaluate the classifier

Solving a PR Problem



- Feat. Gen.: Want to reduce sensitivity to noise and reduce complexity but retain important info

  "Pack" sensor info into small number of features

- Feat. Sel.: Want to reduce complexity and reduce redundancy but retain important info

  Select small set of features that separates classes

- Classif. Des.: Want small generalization error and fast training and classification (i.e. low complexity)

- Sys. Eval.: Want to accurately estimate classifier's generalization error

- Some stages might be combined

- Feedback loops

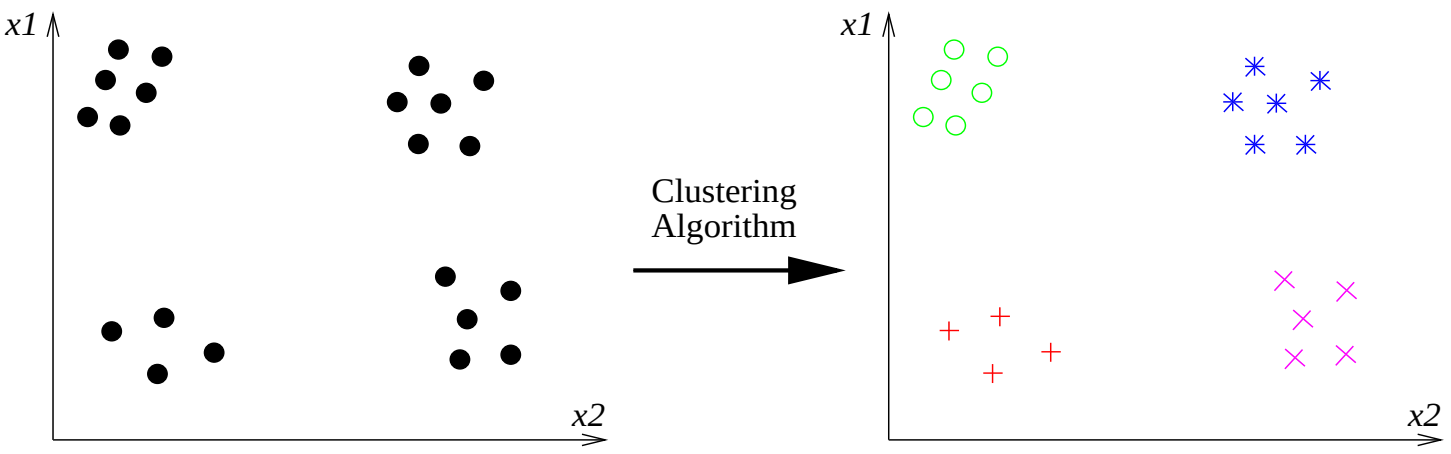
Unsupervised PR

- What if labels unavailable?

- E.g. feat. vectors are measurements of electromag. energy reflected from remote parts of Earth, can't afford to visit each area to determine labels

- Clustering algorithms group similar feature vectors together based on a similarity measure

- If clustering is good, then can find label for one of each group & use it as label for entire group



Graphical Models

- In this course, we'll focus on probabilistic models
  - Formal means to model a probability distribution over the data
- We'll study the use of directed graphs to represent the models
  - Hidden Markov models model sequential data, e.g. biological sequences and human speech
    - HMM is a generative model (can "push a button" and get a sequence
    - Algorithms determine the probability that some sequence comes from the model
  - Bayesian networks model distributions over fixed-length feature vectors
    - Model conditional probabilities
    - Algorithms can determine the probability of certain attribute values of a feature vector given others