ECE471-571 – Lecture 1

Introduction

Statistics are used much like a drunk uses a lamppost: for support, not illumination

-- Vin Scully

Terminology

- Feature
- Sample
- Dimension
- Pattern classification (PC)
- Pattern recognition (PR)
PR = Feature Extraction + Pattern Classification

Need domain knowledge

An Example

fglass.dat
- forensic testing of glass collected by German on 214 fragments of glass
- Data file has 10 columns
  - RI – refractive index
  - Na – weight of sodium oxide(s)
  - ...
  - Type
RI  Na  Mg  Al  Si  K  Ca  Ba  Fe  type
1.52101 13.64 4.49 1.10 71.78 0.06 8.75 0.00 0.00 1
1.51761 13.89 3.60 1.36 72.73 0.48 7.83 0.00 0.00 1

On the Different Courses at UT
- Machine Learning (ML) (CS525/528)
- Pattern Recognition (PR) (ECE471/571)
- Data Mining (DM) (Big Data) (CS526)
- Deep Learning (DL) (ECE599/592)
- From Google Trends
- Digital Image Processing (DIP) (ECE472/572)
- Computer Vision (CV) (ECE573)
A Graphical Illustration

Input (e.g., Images) \[\xrightarrow{\text{DIP}}\] A Better Input (e.g., less storage, less noisy, less blurred) \[\xrightarrow{\text{CV}}\] Features

Knowledge/Inference \[\xrightarrow{\text{AI}}\] Objects Recognized

Conferences

- Computer Vision and Pattern Recognition (CVPR)
- International Conference on Pattern Recognition (ICPR)

Different Approaches - Overview

- Supervised classification
  - Parametric
  - Non-parametric
- Unsupervised classification
  - clustering

Decision rule
- derive
- apply

Training set
- Known classification

Testing set
- Unknown classification

Data set
Pattern Classification

Statistical Approach

Non-Statistical Approach

Supervised

Unsupervised

Decision tree

Basic concepts:

Bayesian decision rule (MPP, LR, Discri.)

Parameter estimate (ML, BL)

Non-parametric learning (kNN)

LDF (Perceptron)

k-means

Winner-take-all

Kohonen maps

Dimensionality Reduction:

FLD, PCA

Performance Evaluation:

ROC curve (TP, TN, FN, FP)

cross validation

Stochastic Methods:

Local opt (GD)

Global opt (SA, GA)

Classifier Fusion:

Majority voting

NB, BKS

Stochastic Methods

Example – Face Recognition

Landmark file structure

column 1 column 2

Lm1: col-of-lm1 row-of-lm1

Lm2: col-of-lm2 row-of-lm2

Lm35: col-of-lm35 row-of-lm35

Line36: col-of-image row-of-image
Example - Network Intrusion Detection

- **KDD Cup 99**
- **Features**
  - basic features of an individual TCP connection, such as its duration, protocol type, number of bytes transferred and the flag indicating the normal or error status of the connection
  - domain knowledge
  - 2-sec window statistics
  - 100-connection window statistics

Example - Gene Analysis for Tumor Classification

- **Early detection of cancer**
- **Tumor classification**
  - Observation of abnormal consequences of tumor development
  - Physical examination (X-rays)
  - Molecular marker detection
- **Tumor gene expression profiles: molecular fingerprint**
- **Challenge: high dimensionality (in the order of thousands)**
  - 16,063 known human genes and expressed sequence tags

Example - Color Image Compression
Example - Automatic Target Recognition

Example – Bio/chemical Agent Detection in Drinking Water

x-axis: time (seconds)
y-axis: relative fluorescence induction

Assignment and Tests

Programming and Reporting
- 4 Regular projects
  - C/C++/Python (major)
  - Matlab or C/C++/Python (non-major)
- 1 Final project (C/C++/Python or Matlab)
  - With milestone deliverables
  - With final presentation
- 4~5 Homework (C/C++/Python or Matlab)
- 2 Tests