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

-- Vin Scully

Terminology

- Feature
- Sample
- Dimension
- Supervised
- Unsupervised
- Training set
- Test set
- Validation set

- Pattern classification
- Pattern recognition (PR)
- Machine learning (ML)
- Data mining (DM)
- Artificial intelligence (AI)

- ECE-471/571
- CS
PR = Feature Extraction + Pattern Classification

- Feature extraction
- Pattern classification

Input media → Feature extraction → Feature vector → Recognition result

Need domain knowledge

Different Approaches - Overview

- Supervised
  - Derive decision rule
    - Known classification (Training set)
    - Unknown classification (Testing set)

- Unsupervised
  - Data set

Pattern Classification

- Statistical Approach
  - Supervised
    - Basic concepts
    - Bayesian decision rule
    - Discriminative methods (LDA, FDA)
  - Unsupervised
    - Basic concepts
    - Distance estimation methods
    - Clustering methods

- Non-Statistical Approach
  - Supervised
    - Basic concepts
    - Distance estimation methods
    - Clustering methods
  - Unsupervised
    - Basic concepts
    - Distance estimation methods
    - Clustering methods

Dimensionality Reduction: FDA, PCA

Classifier Fusion: majority voting, SCR, BBS

Performance Evaluation: ROC curve, TN, TP, FPR, FNR

Stochastic Methods: local opt. (LDA, FDA), global opt. (SA, GA)

Syntactic Approach: Decision tree

Deep Learning (DL): Mean shift, Support Vector Machine, NN (BP)
Strong Emphasis on

- Why
- How
- Implementation
- Project

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  4.4  1.05  71.87 0.06 0.00 0.00  1
  - 1.51761 13.89  3.6  0.1  36.73 0.48 0.00 0.00  1

Example – Face Recognition
Landmark file structure

<table>
<thead>
<tr>
<th>column 1</th>
<th>column 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lm1:</td>
<td>row-of-Lm1</td>
</tr>
<tr>
<td>Lm2:</td>
<td>row-of-Lm2</td>
</tr>
<tr>
<td></td>
<td>...</td>
</tr>
<tr>
<td>Lm35:</td>
<td>row-of-Lm35</td>
</tr>
<tr>
<td>Line36</td>
<td>row-of-image</td>
</tr>
</tbody>
</table>

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
Homework and Projects

- Programming
  - Regular projects
    - C/C++ (EECS Major)
    - Matlab or C/C++ (non-major)
  - Final project (C/C++ or Matlab)
  - Homework (C/C++ or Matlab)

Programming environment

- OS – Linux
- Compiler – g++
- Editor – Emacs
- Language – C++

Computing resources

- Hydra Linux Lab
  - hydra01-25.eecs.utk.edu
Programming style

- Indenting
- Comment
- Spacing
- Simple
- Flexible

Step 1 – Get an account

- Get an account on the student server
- Login and change password
  - passwd

Step 2 – Get familiar with Linux commands

- ls – list files
- less xxx – view a readable file (used to quickly browse the file)
- rm xxx – delete a file
- mv xxx yyy – rename xxx to yyy
- emacs – a nice file editor
- mkdir – create a new directory
- cd xxx – go into a new directory
- Access floppy disk
  - mcopy source dest
  - mdtr a:
  - mdel a:\xxx
Step 3 – Learn to use g++
- `g++ -o xxx xxx.cpp Matrix.cpp`
- `xxx.cpp` is the application source code you wrote
- `xxx` is the executable application
- Learn to use Makefile

Step 4 – Test a sample code
- Download the dataset from the “testing datasets” webpage and uncompress it
  - Use “unzip” to uncompress the .zip file
  - Use “gunzip” to uncompress the .gz file
- Download the sample source code (read.cpp) from the course website
- Download the Matrix library from the course website (Matrix_v1.tar)
- Test a sample code
  - Run "make" under the directory "lib"
  - Run "make" under the directory "example"
  - Under "example", type ./testMatrix

Step 5 – Understanding the matrix library
- `Matrix.h` declares the private data and defines the member functions
- `Matrix.cpp` implements the member functions
Step 6 - How to organize the source files

- \lib
  - With the .cpp source code with function implementations
  - Use Makefile to generate a library
- \include
  - Header files
- \example
  - Test data
  - Use Makefile to generate executables
- \data