Homework 2

Linear Models for Classification

Deadline: 2017.4.4 Tuesday (p.m.23:59)

Data

Database of Faces (AT&T Laboratories Cambridge)

Reference : http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

Each class has 1000 30x30 images.

You need to use them to build the classification models.

Models

Multi-class Probabilistic Generative Model

$$p(C_k|x) = \frac{p(x|C_k)p(C_k)}{\sum_j p(x|C_j)p(C_j)}$$

$$p(x|C_k) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(x - \mu_k)^T \Sigma^{-1}(x - \mu_k)\}$$

Multi-class Probabilistic Discriminative Model

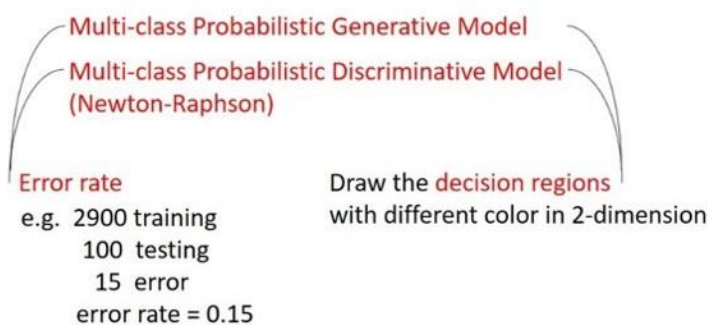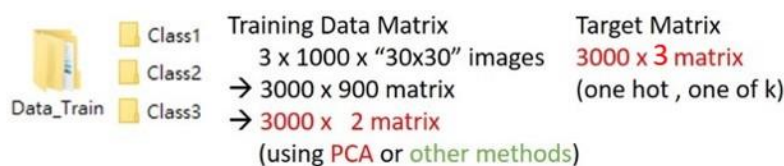$$p(C_k|x) = \frac{\exp(a_k)}{\sum_j \exp(a_j)}$$

$$a_k = w_k^T x$$

To optimize w_k , please use the Newton-Raphson iterative optimization and define your own way to judge whether w_k is converged or not.

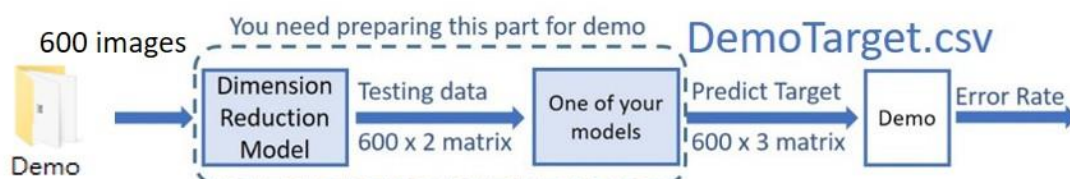(You can find the details in the textbook.)

Tasks

1. Use the training images to build "Multi-class Probabilistic Generative Model" and "Multi-class Probabilistic Discriminative Model".
2. Use Newton-Raphson iterative optimization in "Multi-class Probabilistic Discriminative Model".
3. You can separate some images for testing your models and note the error rate in the report.
   Choosing how many data for testing is up to you.
4. Please use Principal component analysis (PCA) to map data down to 2 dimensions.
   You can use other dimension reduction methods to make your report be better.
   This part is bonus.
5. Draw the pictures of your decision region.
6. Test unbalanced data how to affect your models.
   You should decrease one certain class and retry the task 1 – 4 again.

Hints and Demo

Class1
Class2
Data_Train Class3

Training Data Matrix
3 x 1000 x "30x30" images
→ 3000 x 900 matrix
→ 3000 x 2 matrix
(using PCA or other methods)

Target Matrix
3000 x 3 matrix
(one hot , one of k)

Multi-class Probabilistic Generative Model

Multi-class Probabilistic Discriminative Model
(Newton-Raphson)

Error rate
e.g. 2900 training
     100 testing
      15 error
error rate = 0.15

Draw the decision regions
with different color in 2-dimension

Test unbalanced data

600 images

You need preparing this part for demo

DemoTarget.csv

Demo → Dimension Reduction Model → Testing data 600 x 2 matrix → One of your models → Predict Target 600 x 3 matrix → Demo → Error Rate

<u>Reminders</u>

1. Report within 12 pages
2. Using Python is encouraged for you.
3. Don't use high level functions and tool boxes.
   Use functions and tool boxes for reading and writing files are allowed.
4. Do not copy!
   (When using some reference materials, please give credit to them.)

Machine Learning 2017

Grading Policy & Homework Rules

● <u>Homework will be graded by</u>

Completeness

Correctness

Algorithm description

Discussion

● <u>You should upload homework files to E3</u>

● <u>Homework Rules</u>

File Name: hw2_StudentID.zip/rar (e.g. hw2_1234567.zip)

Code with comments

You can use any programing language to finish your homework

Report (.pdf format)

ReadMe.txt (describes how to run your code)

Hand in a hardcopy report on the due day.

● <u>Deadline</u>

Late Submission (1-7 days): 70% score

Don't accept after 7 days.