CS4640 Week 14

Classification

medical images : healthy – unhealthy

fruit : high quality/not

pills : looks OK / not

lumber : type

image → features → classes

variables/ parameters

what classes?

Supervised : set of labeled samples

images → [choose features] → classifier → class
labels ↗

↑ new image

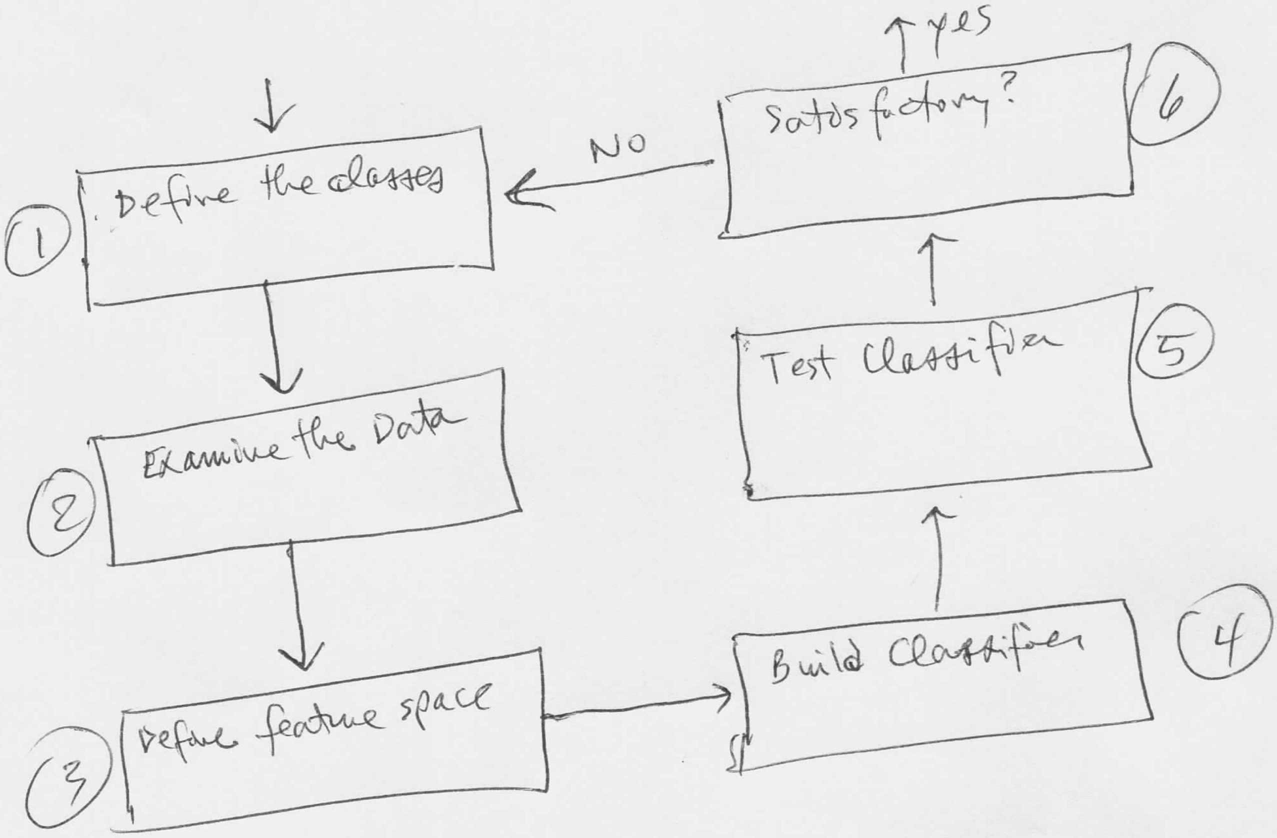
Text p. 292-293 example

see CS4640_week14

Trouble! [e.g., how many seeds in Fig 11.1
how many items in Fig 11.2]

see separation lines between classes.

Design of Classification System



1. distinct upper + lower case characters & digits
   punctuation : ( ) .) ' ; : -
   other symbols : + * ÷ / ...

2. See what kinds of noise or other pre-processing is required, scope of problem

3. geometric, moments, Fourier, ...
   discriminatory power ; complementary
   [e.g., a basis!]

(4) Classifier

Pick type of classifier

Train

(5) Test classifier

need new samples

(6) Satisfactory if meets specification

e.g., 90% correct recognition

Simple Classifiers

Given a set of vectors for a class,
calculate the mean as the model

Then, for a new sample,
pick class whose model is closest to new sample

e.g., try lentils, ... data!
pumpkin does well; lentils & pist.
need more!

<u>Distance Measures</u>

Euclidean distance $\quad L(\bar{x},\bar{y}) = \left[\sum_{i=1}^{N} (x_i - y_i)^2\right]^{1/2}$

Matlab: norm

Manhattan distance $\quad L(\bar{x},\bar{y}) = \sum_{i=1}^{N} |x_i - y_i|$

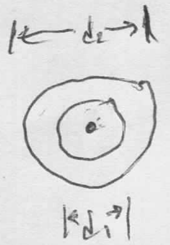
Minkowski distance $\quad L(\bar{x},\bar{y}) = \left[\sum_{i=1}^{N} |x_i - y_i|^K\right]^{1/K}$

Mahalanobis distance $\quad L(\bar{x},\bar{y}) = (\bar{x}-\bar{y})^T \Sigma^{-1} (\bar{x}-\bar{y})]^{1/2}$

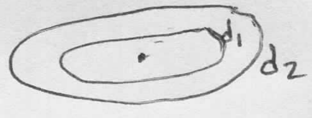
where $\Sigma = (\bar{x} - \mu_x)(\bar{y} - \mu_y)^T$

* <u>confusing</u>

This is supposed to tell how far a point is from a distribution where level sets may be ellipses



Euclidean



distribution characterized by $\bar{\mu}$ + Σ

$$L(\bar{x}, (\bar{\mu}, \Sigma)) = \sqrt{(\bar{x}-\bar{\mu})^T \Sigma^{-1} (\bar{x}-\bar{\mu})}$$

if $\bar{x}$ an $\bar{y}$ are both from same distribution,

i.e., $\bar{x}, \bar{y} \sim \mathcal{N}(\bar{\mu}, \Sigma)$

then $L(\bar{x}, \bar{y}) = \sqrt{(\bar{x}-\bar{y})^T \Sigma^{-1} (\bar{x}-\bar{y})}$

note: if $\Sigma = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}$ identity

then $L(\bar{x}, \bar{y}) = \sqrt{(\bar{x}-\bar{y})^T (\bar{x}-\bar{y})}$

= Euclidean distance

i.e., $\sqrt{(\bar{x}-\bar{y})^T (\bar{x}-\bar{y})}$

$= \sqrt{\sum (x_i - y_i)^2}$

Linear Discriminant Function

Given 2 classes represented by an average vector

$\cdot\ \bar{c}_1$

$\bar{c}_2\ \cdot$

Determine class of new point by closer center

$\cdot\ \bar{c}_1$

$\bar{c}_2\ \cdot$ ——— $* \ \bar{x}$

$$d_1 = \underbrace{\sqrt{(\bar{x}-\bar{c}_1)^T(\bar{x}-c_1)}}_{\text{quadratic and a sqrt}}$$

$$\sqrt{(\bar{x}-\bar{c}_2)^T(\bar{x}-\bar{c}_2)} = d_2$$

Use a <u>line</u> as follows

$$l(x,y) = ax + by + c = 0$$

$=0$

$l(x,y) > 0$ $\quad$ $l(x,y) < 0$

$\cdot\ \bar{c}_1$

$\bar{c}_2\ \cdot$

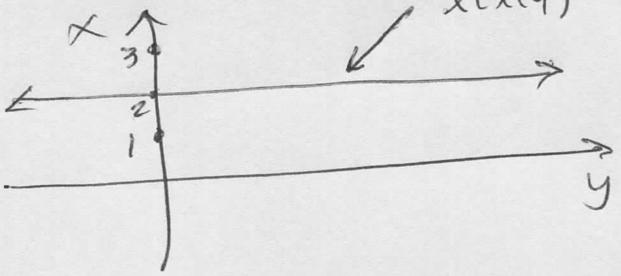
(1) points on the line yield 0 in eqn

(2) points closer to $\bar{c}_1$ yield < 0 in eqn

(3) points closer to $\bar{c}_2$ yield > 0 in eqn

Some special lines

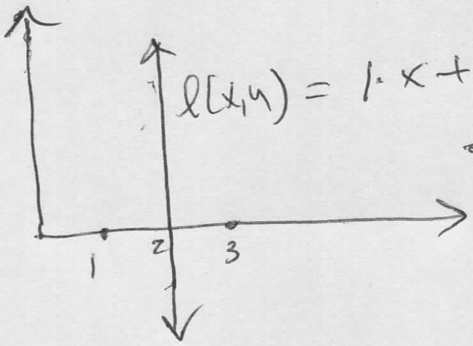
(1)



$l(x,y) = 0x + 1 \cdot y - 2 = 0$

$\Rightarrow y = 2$

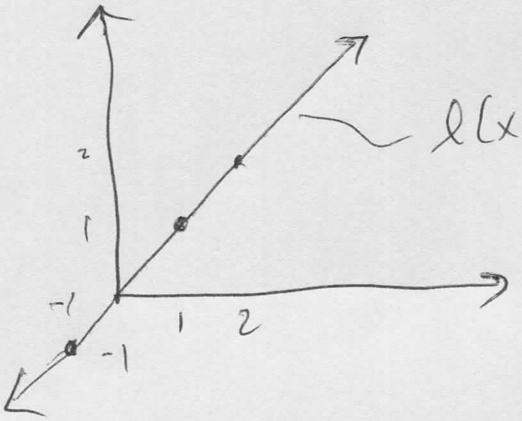
(2)



$l(x,y) = 1 \cdot x + 0 \cdot y - 2 = 0$

$\Rightarrow x = 2$

(3)



$l(x,y) = -0.7071x + 0.7071y + 0 = 0$

$\Rightarrow y = x$

Now, take a look at the vector $[a; b]$

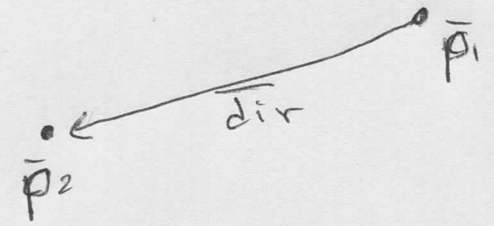
(1) $[0; 1]$ $\uparrow$ (2) $[1; 0]$ $\rightarrow$ (3) $[-0.7071; 0.7071]$ $\nwarrow$

all normal to their line

14/6

How to get the discriminat line?

(1) Find vector from $\bar{p}_2$ to $\bar{p}_1$: $\overline{dir} = \bar{p}_2 - \bar{p}_1$



(2) Determine unit vector : $\overline{dir_u} = \dfrac{\overline{dir}}{|\overline{dir}|}$

$|\overline{dir}| = \text{norm}(dir)$

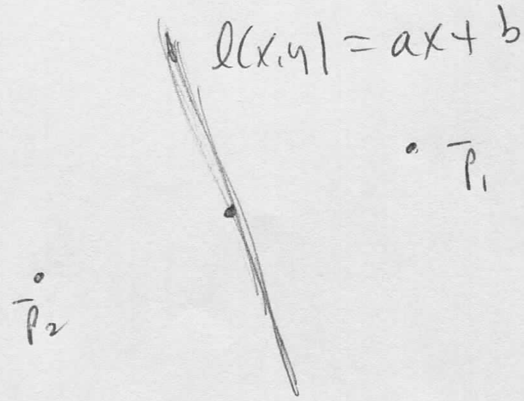
(3) So, $l(x,y) = dir_u(1)x + dir_u(2)y + c$

Find c : want line mid-way between $\bar{p}_1$ and $\bar{p}_2$: $\bar{P}_m = \dfrac{(\bar{P}_1 + \bar{P}_2)}{2}$

$a = dir_u(1)$
$b = dir_u(2)$

so, $C = -dir_u(1) * P_m(1) - dir_u(2) * P_m(2)$

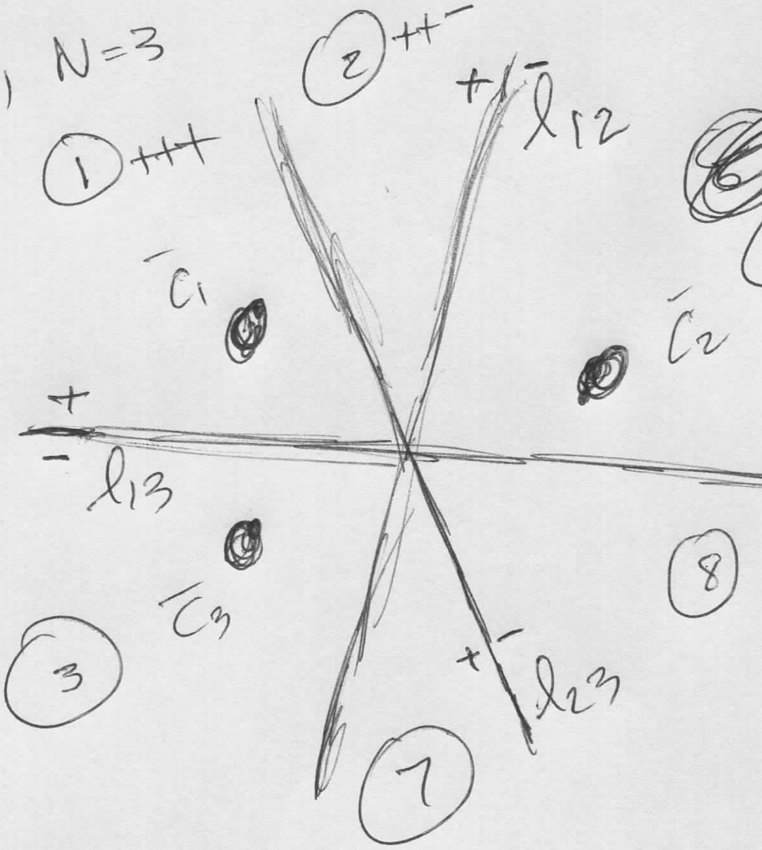
$l(x,y) = ax + by + c = 0$

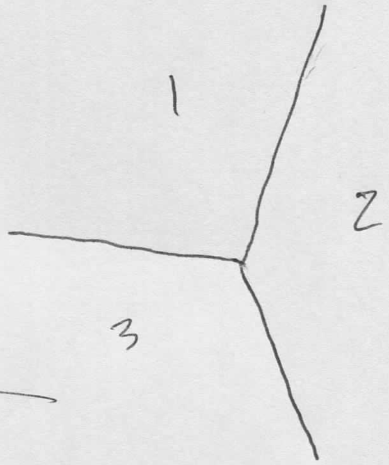


(14|7)

For more than 2 classes, there are 2N regions for N classes

E.g., N=3





4 = $\emptyset$

5 = $\emptyset$

Region	l_{12}	l_{13}	l_{23}	class
1	+	+	+	1
2	+	+	−	1
3	+	−	+	3
4	+	−	−	not possible
5	−	+	+	not possible
6	−	+	−	2
7	−	−	+	3
8	−	−	−	2

Same approach works in n-dimensions

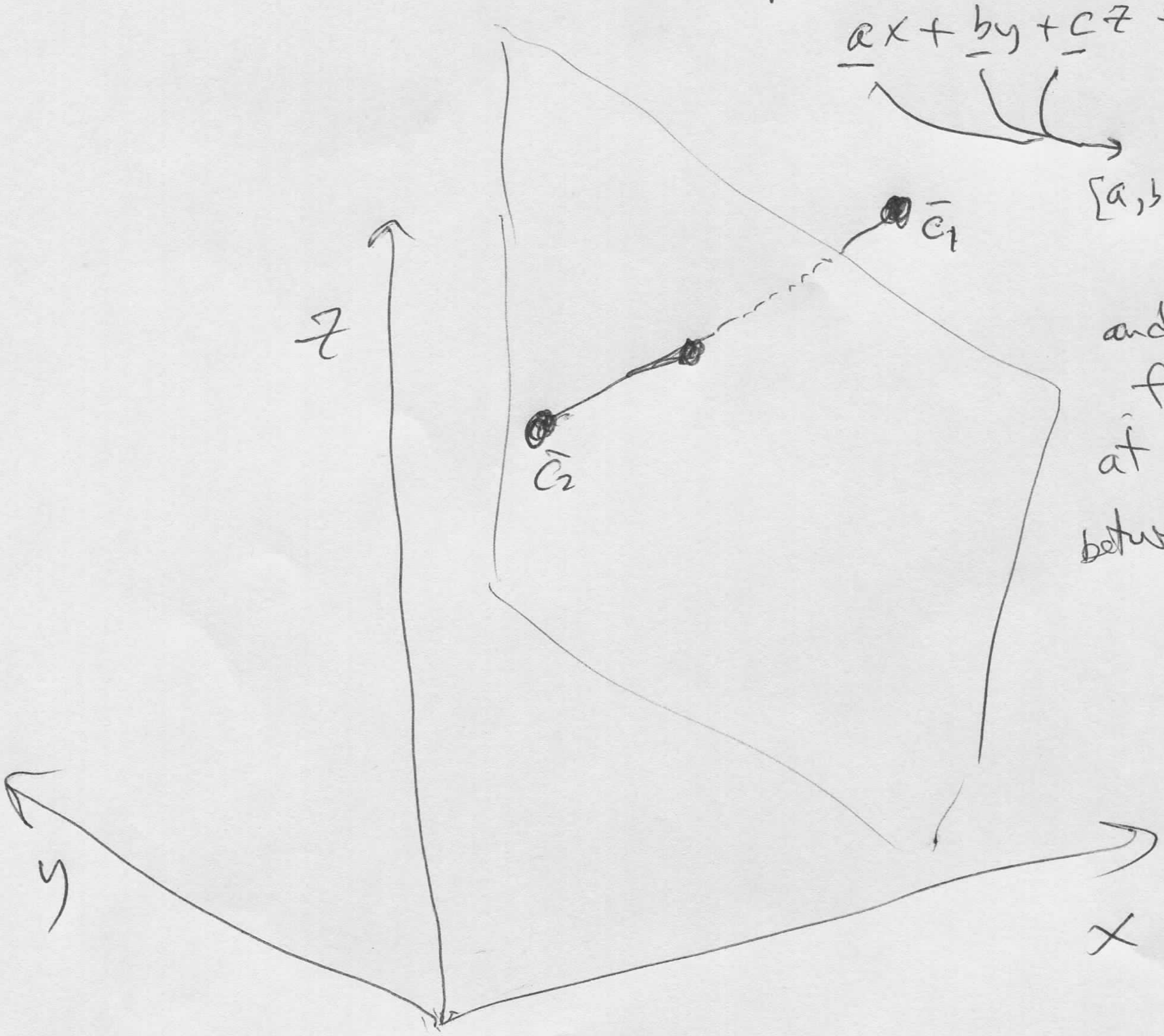
E.g., 3D

equation of plane

$\underline{a}x + \underline{b}y + \underline{c}z + d = 0$

$[a, b, c] = \dfrac{\bar{c}_2 - \bar{c}_1}{|\bar{c}_2 - \bar{c}_1|}$

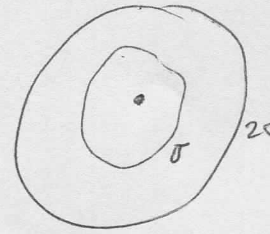
and solve for d at mid-point between $\bar{c}_1$ + $\bar{c}_2$



14|9

Mahalanobis distance

scales according to ellipse level sets



$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$\Rightarrow$ Euclidean

scales by standard deviation

$\Sigma = \begin{bmatrix} 5 & \\ & 1 \end{bmatrix}$

Bayesian Classification

assign sample to most probable class

general definitions:

patterns (features): $\bar{x} = [x_1, x_2, \ldots x_N]$

measures on shape or object

classes: C classes $C = \{\omega_1, \omega_2, \ldots, \omega_C\}$

priors: $p(\omega_i)$: probability of drawing a feature vector from class ω_i

training set $S = \{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M\}$

M feature vectors whose classes are known

Bayes Decision Rule

if $\quad p(w_j \mid \bar{x}) > p(w_k \mid \bar{x}) \quad \forall k \neq j$

then assign $\bar{x}$ to w_j

where $p(w_j \mid \bar{x})$ probability of class w_j given $\bar{x}$

Use Bayes theorem:

$$p(w_j \mid \bar{x}) = \frac{p(\bar{x} \mid w_j)\, p(w_j)}{p(\bar{x})}$$

---

in the rule:

if $p(\bar{x} \mid w_j)\, p(w_j) > p(\bar{x} \mid w_k)\, P(w_k) \quad \forall k \neq j$

then assign $\bar{x}$ to w_j

don't need division by $p(\bar{x})$ since it's in both sides

$p(\bar{x} \mid \omega_j)$: probability of vector $\bar{x}$ given class ω_j

$p(\omega_j)$: priors are determined by frequency
e.g., 'e' is most probable letter in english 12.7 %

$p(\bar{x})$: $p(\bar{x}) = \sum_j p(\bar{x} \mid \omega_j)\, p(\omega_j)$

Multi-variate normal distribution

mvnrnd (mu, Sigma, n)

e.g., [0;0], eye(2), 1000

$$p(\bar{x}) = \frac{1}{2\pi^{\frac{N}{2}} |C|^{1/2}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu})^T C^{-1} (\bar{x}-\bar{\mu})}$$

$\bar{\mu}$ is distribution mean C is covariance matrix

Bayesian classifiers : multivariate normal distribution

if $p(\bar{x}|w_j)\,p(w_j) > p(\bar{x}|w_k)\,p(w_k) \quad \forall k \neq j$

then assign $\bar{x}$ to w_j

taking log:

if $\log(p(\bar{x}|w_j)) + \log(p(w_j)) > \log(p(\bar{x}|w_k)) + \log(p(w_j))$

then ~~a~~ssign $\bar{x}$ to w_j

$$\log(p(\bar{x}|w_j)) + \log(p(w_j))$$

$$= -\frac{1}{2}(\bar{x}-\bar{\mu})^T C_j^{-1}(\bar{x}-\bar{\mu}) - \frac{N}{2}\log 2\pi - \frac{1}{2}\log|C_j|$$

$$+ \log(p(w_j))$$

Simpler cases $\quad C_j = \sigma^2 I$

then
$$= -\frac{\|\bar{x}-\bar{\mu}\|^2}{2\sigma^2} + \log(p(w_j))$$

$$= \bar{w}_j \bar{x} + v_j \qquad \bar{w}_j = \frac{\bar{\mu}_j}{\sigma^2}$$

$$v_j = \frac{1}{2\sigma^2}\bar{\mu}_j^T \bar{\mu}_i + \log(p(w_j))$$

If $C_j = C \quad \forall j$

$$= \bar{\omega}_j^T \bar{x} + v_j$$

where $\bar{\omega}_j = C^{-1} \bar{\mu}_j$

$$v_j = -\frac{1}{2} \bar{\mu}_j^T C \bar{\mu}_j + \log(p(\omega_j))$$

Fisher linear discriminant from N-D to 1D

Given feature vectors, $\{\bar{x}_i\}$ $\quad \bar{x}_i \in \mathbb{R}^N$

project them onto an axis

so they are maximally separated.

$\bar{y} = \bar{\omega}^T \bar{x}$ $\quad \bar{y}$ is new vector

criterion : $J(\bar{w}) = \dfrac{|\bar{m}_A - \bar{m}_B|^2}{s_A^2 + s_B^2}$ (2 classes)

$\bar{m}_A, \bar{m}_B$ are means ; s_A^2, s_B^2 sample variances

then: $\bar{w} = S_W^{-1} [\bar{M}_A - \bar{M}_B]$ original class means

$S_W = S_A + S_B$