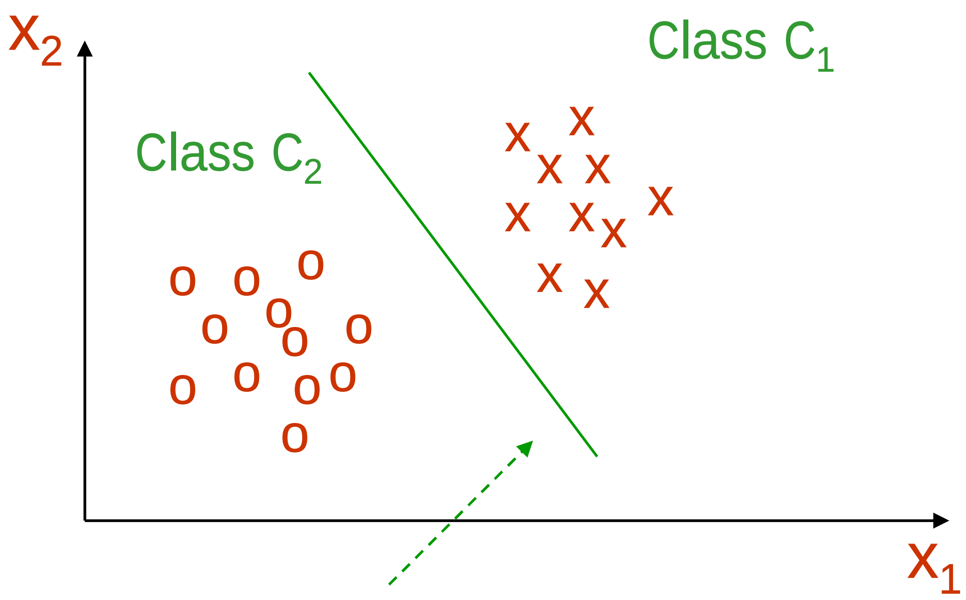
Pattern Classification

Hsiao-Lung Chan, Ph.D.
Dept Electrical Engineering
Chang Gung University, Taiwan
chanhl@mail.cgu.edu.tw

Linear discriminate function



Discriminate function

$g(\mathbf{x})=0 \quad w_0 + w_1x_1 + w_2x_2 = 0$

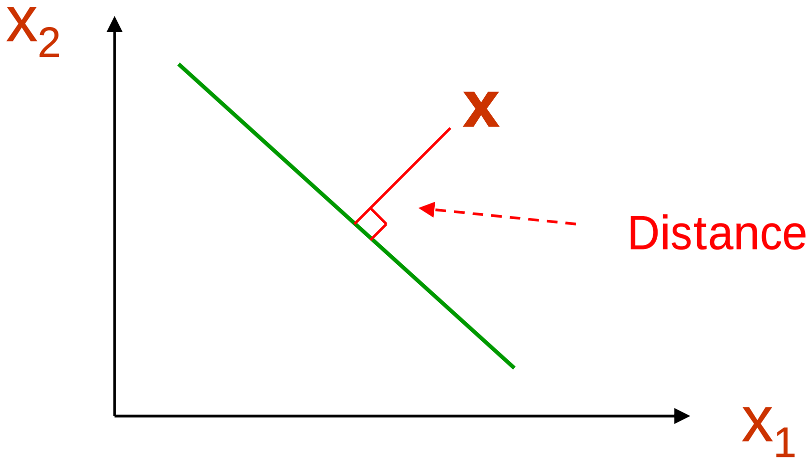
Linear discriminate function (cont.)

A two-class pattern classifier

$$g(\mathbf{x}) = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n = \mathbf{w}^T \mathbf{x}$$

$$\text{where } \mathbf{w} = [w_0, w_1, w_2, \cdots, w_n] \text{ and } \mathbf{x} = [1, x_1, x_2, \cdots, x_n]$$

$$g(\mathbf{x}) = \begin{cases} \mathbf{w}^T \mathbf{x} > 0 & \text{if } \mathbf{x} \in C_1 \\ \mathbf{w}^T \mathbf{x} \le 0 & \text{if } \mathbf{x} \in C_2 \end{cases}$$

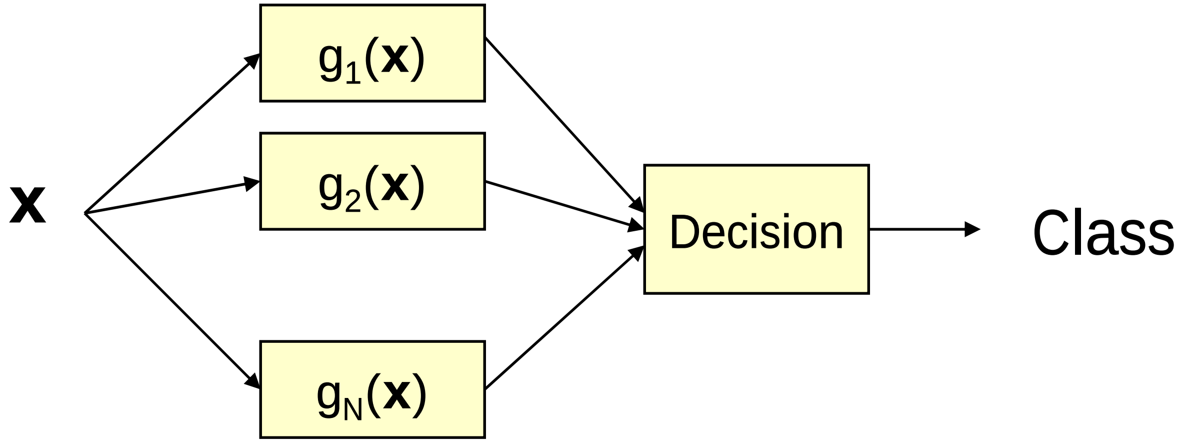


Multi-category classifier

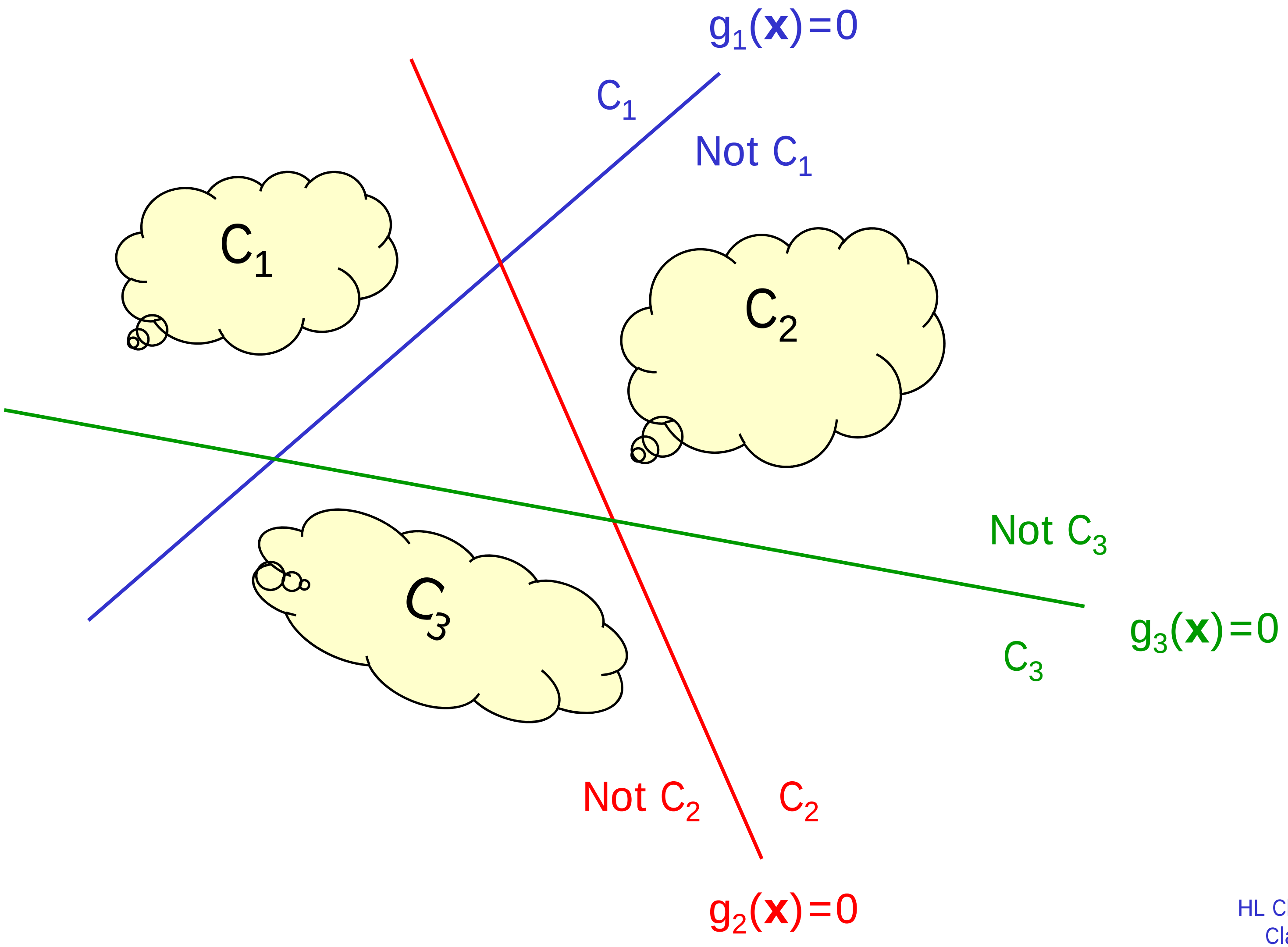
- Build M weight vectors and M discriminate functions

$$g_i(\mathbf{x}) = \begin{cases} \mathbf{w}_i^T \mathbf{x} > 0 & \text{if } \ \mathbf{x} \in C_i \\ \mathbf{w}_i^T \mathbf{x} \le 0 & \text{otherwise} \end{cases}, \ i = 1, 2, \cdots, M$$

- Make decision



Build M discriminate functions



Make decisions based on M discriminate functions

- One-to-rest separable
  - Each class is separable from the rest by a single discriminate function

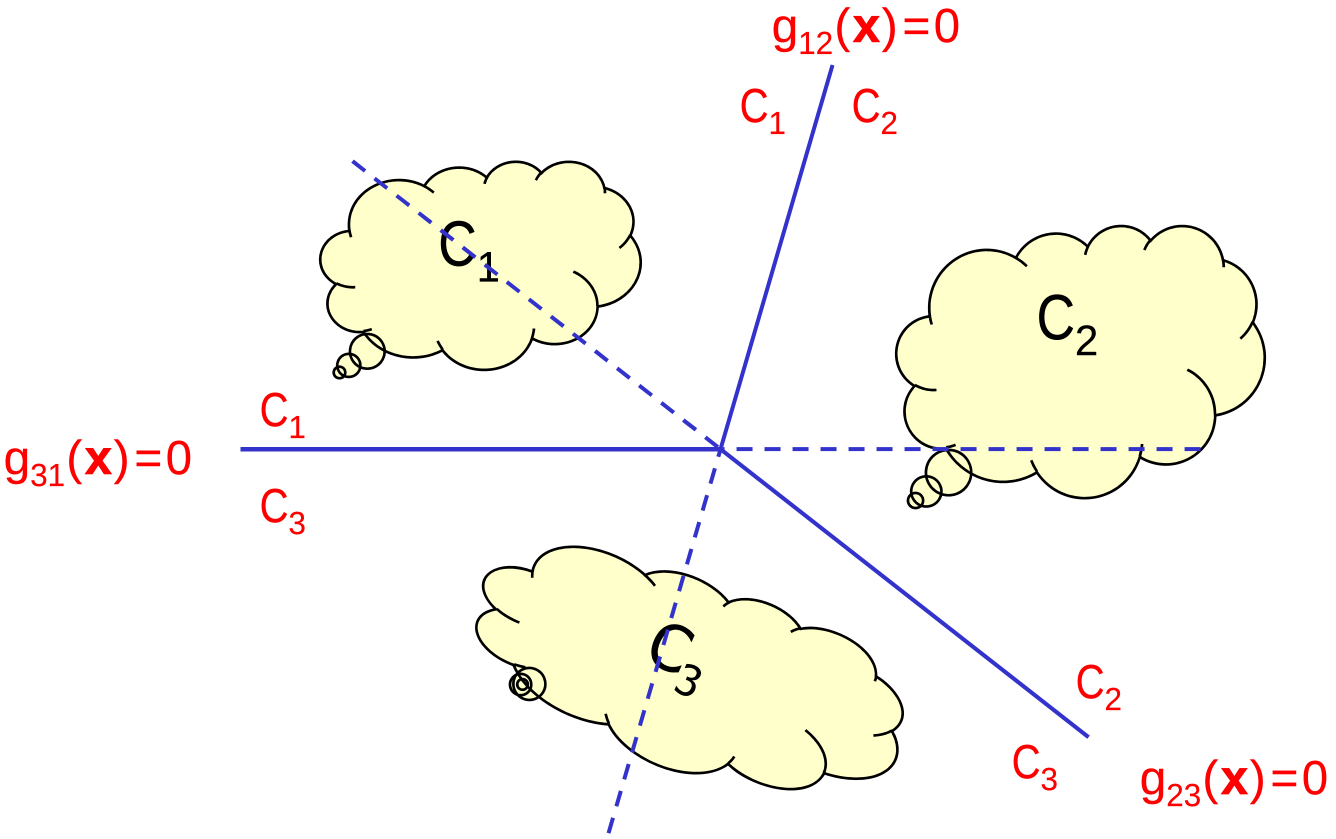
$$\text{if } g_i(\mathbf{x}) > 0 \text{ then } \mathbf{x} \in C_i$$

- Pair-wise separable
  - There exists M discriminate functions with the property that

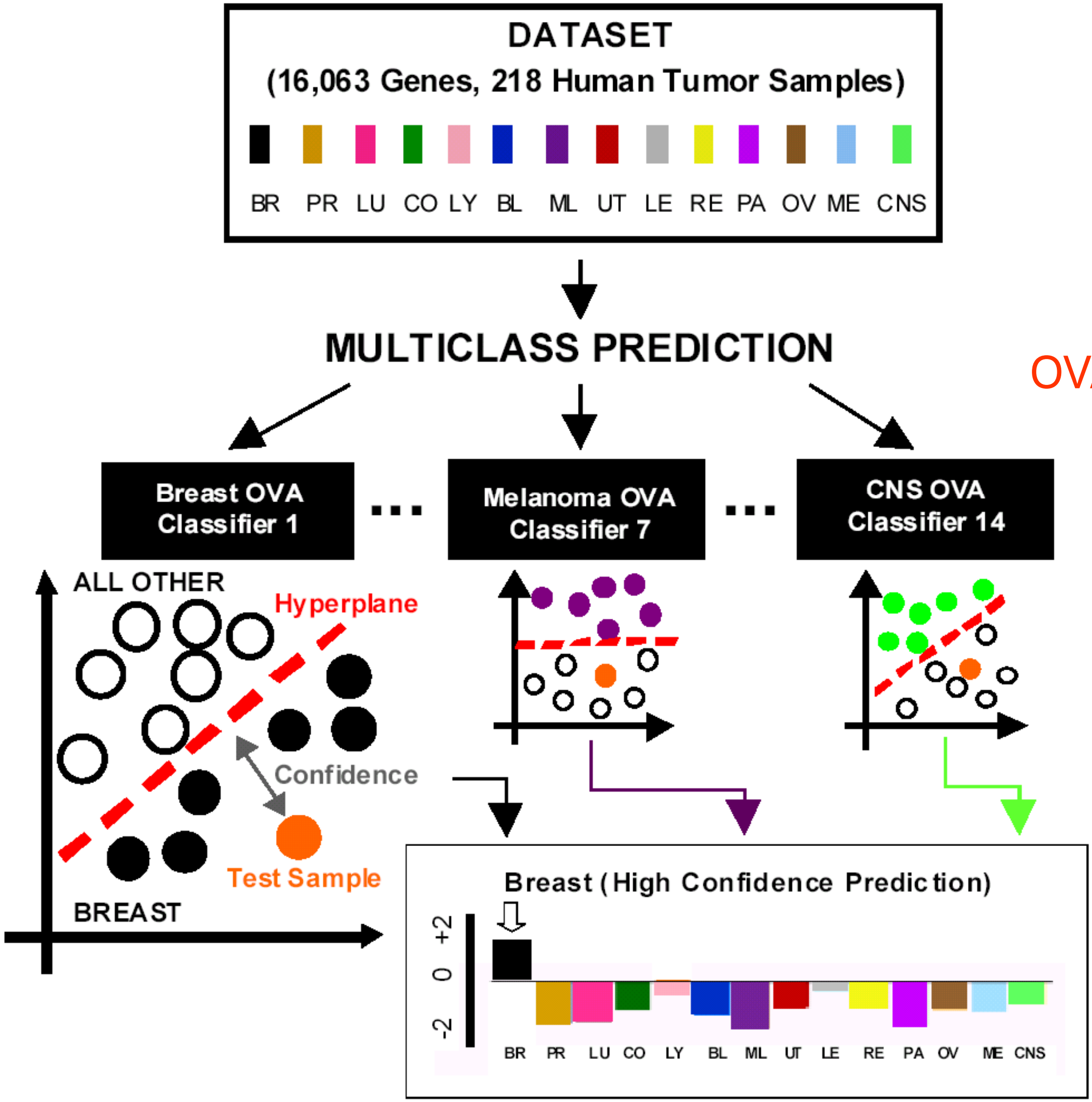
$$\text{if } g_i(\mathbf{x}) > g_j(\mathbf{x}) \ \forall j \neq i \text{ then } \mathbf{x} \in C_i$$

$$g_{ij}(\mathbf{x}) = g_i(\mathbf{x}) - g_j(\mathbf{x}) = \mathbf{w}_i\,\mathbf{x} - \mathbf{w}_j\,\mathbf{x} = (\mathbf{w}_i - \mathbf{w}_j)\,\mathbf{x}$$

Pair-wise separable: M(M-1)/2 two-class classifier



Minimum distance classifier



OVA (One-versus-all) classifier

Probability model for statistical decision

- Assumes a Gaussian probability density function (PDF) to represent the features for each class
  - Represents the statistical attribute for each class using test samples

Probability model

The a prior probability

$$P(C_i) \quad i = 1,2,\cdots,M$$

The likelihood function of class C_i

$$p(\mathbf{x} \mid C_i)$$

PDF of **x** regardless of class membership

$$p(\mathbf{x}) = \sum_{i=1}^{M} P(C_i)\, p(\mathbf{x} \mid C_i)$$

The a posterior probability

$$P(C_i \mid \mathbf{x}) = \frac{P(C_i)\, p(\mathbf{x} \mid C_i)}{p(\mathbf{x})}$$

The probability that observed **x** comes from C_i

Bayes classifier for Gaussian PDF

- Gaussian PDF of a single random variable x

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-m)^2}{2\sigma^2} \right]$$

where the mean an the variance

$$m = E[x] = \int_{-\infty}^{\infty} x p(x) dx$$

$$\sigma^2 = E[(x-m)^2] = \int_{-\infty}^{\infty} (x-m)^2 p(x) dx$$

Bayes classifier for Gaussian PDF (cont.)

- In the case of M classes and P pattern vector, $\mathbf{x}=[x_1, x_2, \ldots, x_n]$

$$p(x \mid C_i) = \frac{1}{(2\pi)^{n/2}\left|\boldsymbol{\Sigma}_i\right|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x}-\mathbf{m}_i) \right]$$

where the mean vector an the covariance matrix

$$\mathbf{m}_i = E_i[\mathbf{x}]$$

$$\boldsymbol{\Sigma}_i = E_i[(\mathbf{x}-\mathbf{m}_i)(\mathbf{x}-\mathbf{m}_i)^T]$$

Bayes classifier for Gaussian PDF (cont.)

- Discriminate functions

$$g_i(\mathbf{x}) = p(\mathbf{x} \mid C_i)P(C_i), \; i = 1,2,\cdots,M$$

$$g_i(\mathbf{x}) = \ln\left[p(\mathbf{x} \mid C_i)P(C_i)\right] = \ln p(\mathbf{x}|C_i) + \ln P(C_i)$$

$$g_i(\mathbf{x}) = \ln P(C_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| - \frac{1}{2}\left[(\mathbf{x}-\mathbf{m}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\mathbf{m}_i)\right]$$

Bayes classifier for Gaussian PDF (cont.)

- If the features are statistically independent $\mathbf{\Sigma}_i = \sigma^2 \mathbf{I}$ the bayes classifier is equivalent to linear discriminate function

$$g_i(\mathbf{x}) = \ln P(C_i) - \frac{\|\mathbf{x} - \mathbf{m}_i\|^2}{2\sigma^2}$$

$$= \ln P(C_i) - \frac{1}{2\sigma^2}\left[\mathbf{x}^T\mathbf{x} - 2\mathbf{m}_i^T\mathbf{x} - \mathbf{m}_i^T\mathbf{m}_i\right]$$

$$g_i(\mathbf{x}) = \mathbf{w}_i^T\mathbf{x} + \mathbf{w}_{i0}$$

Bayes classifier for Gaussian PDF (cont.)

- If all covariance matrix are equal, the Bayses classifier is also equivalent to linear discriminate functions

$$g_i(\mathbf{x}) = \ln P(C_i) + \mathbf{x}^T \mathbf{\Sigma}^{-1} \mathbf{m}_i - \frac{1}{2} \mathbf{m}_i^T \mathbf{\Sigma}^{-1} \mathbf{m}_i$$

$$= \mathbf{w}_i^T \mathbf{x} + \mathbf{w}_{i0}$$

Matlab implemenation

- class = classify(sample,training,group,type)
  - 'linear' — Fits a multivariate normal density to each group, with a pooled estimate of covariance
  - 'diaglinear' — Similar to 'linear', but with a diagonal covariance matrix estimate
  - 'quadratic' — Fits multivariate normal densities with covariance estimates stratified by group
  - 'diagquadratic' — Similar to 'quadratic',but with a diagonal covariance matrix estimate
  - 'mahalanobis' — Uses Mahalanobis distances with stratified covariance estimates

Fisher's sepal measurements for iris versicolor and virginica

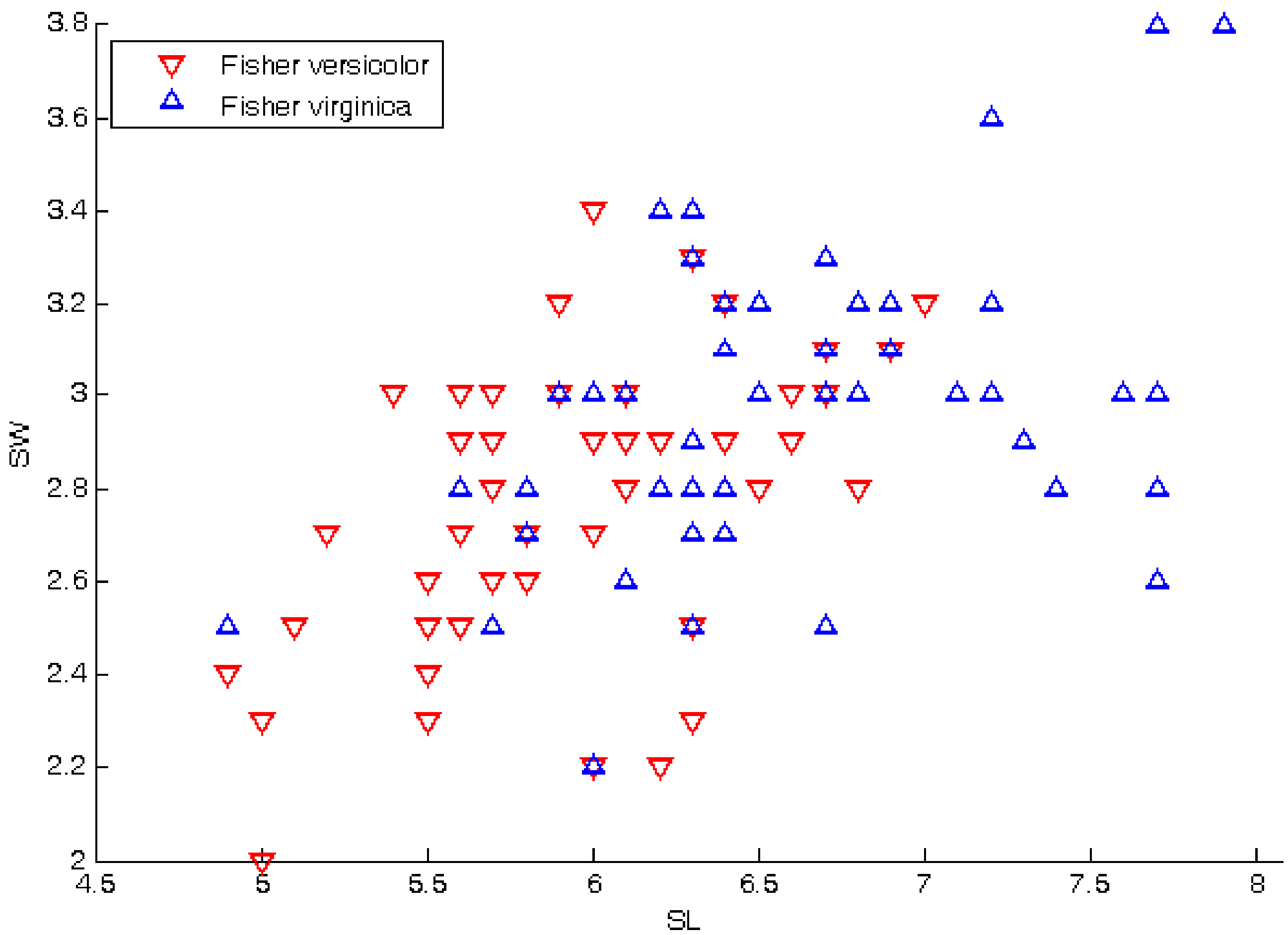
```
% Training data
load fisheriris
SL = meas(51:end,1);
SW = meas(51:end,2);
group = species(51:end);

h1 = gscatter(SL,SW,group,'rb','v^',[],'off');
set(h1,'LineWidth',2)
legend('Fisher versicolor','Fisher virginica', 'Location','NW')

% Classify a grid of measurements on the same scale:
[X,Y] = meshgrid(linspace(4.5,8),linspace(2,4));
X = X(:); Y = Y(:);
[C,err,P,logp,coeff] = classify([X Y],[SL SW],group,'quadratic');
```

Fisher's sepal measurements for iris versicolor and virginica

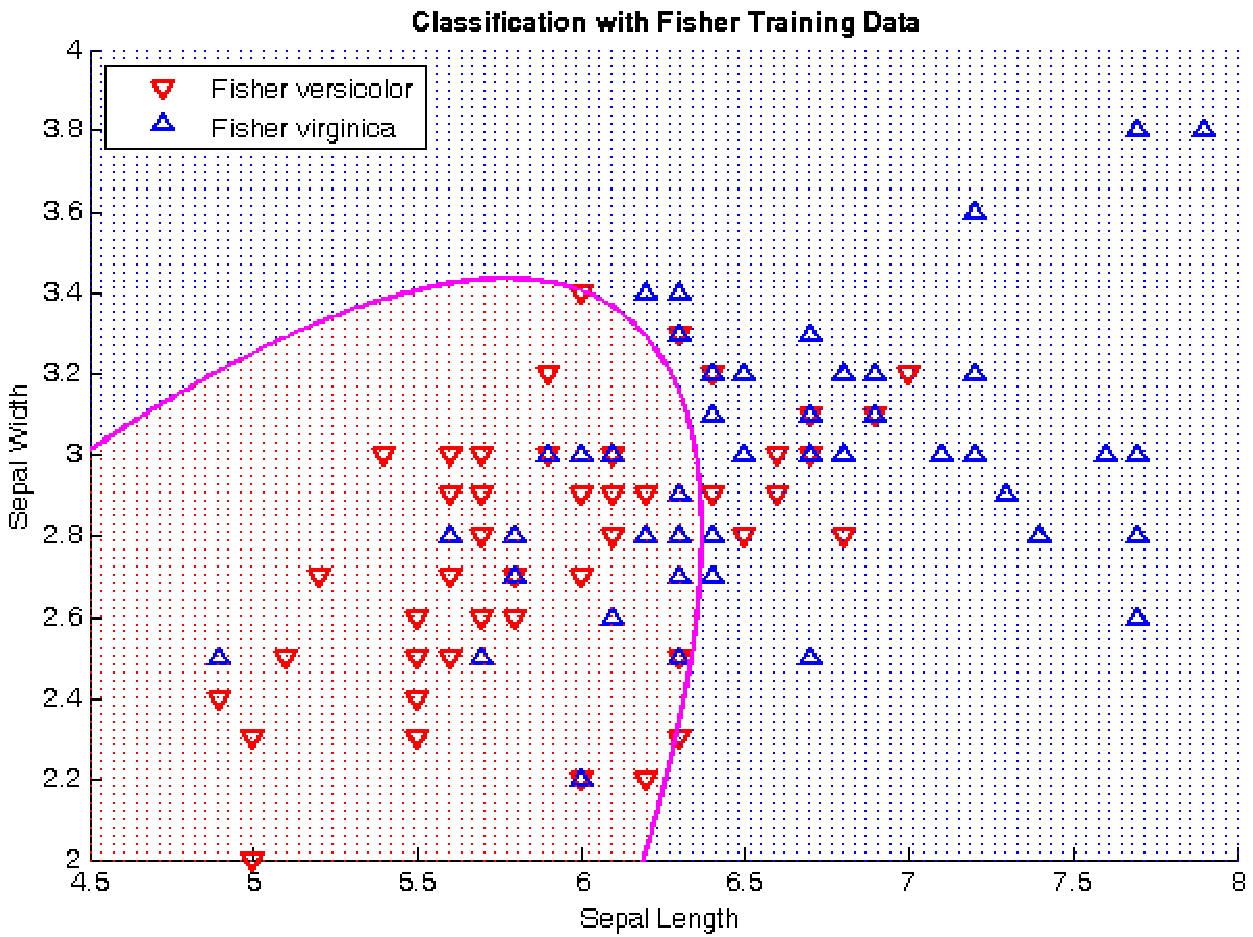
% Visualize the classification

```
hold on;
gscatter(X,Y,C,'rb','.',1,'off');
K = coeff(1,2).const;
L = coeff(1,2).linear;
Q = coeff(1,2).quadratic;
f = sprintf('0 = %g+%g*x+%g*y+%g*x^2+%g*x.*y+%g*y.^2',
    K,L,Q(1,1),Q(1,2)+Q(2,1),Q(2,2));
h2 = ezplot(f,[4.5 8 2 4]);
set(h2,'Color','m','LineWidth',2)
axis([4.5 8 2 4])
xlabel('Sepal Length')
ylabel('Sepal Width')
title('{\bf Classification with Fisher Training Data}')
```

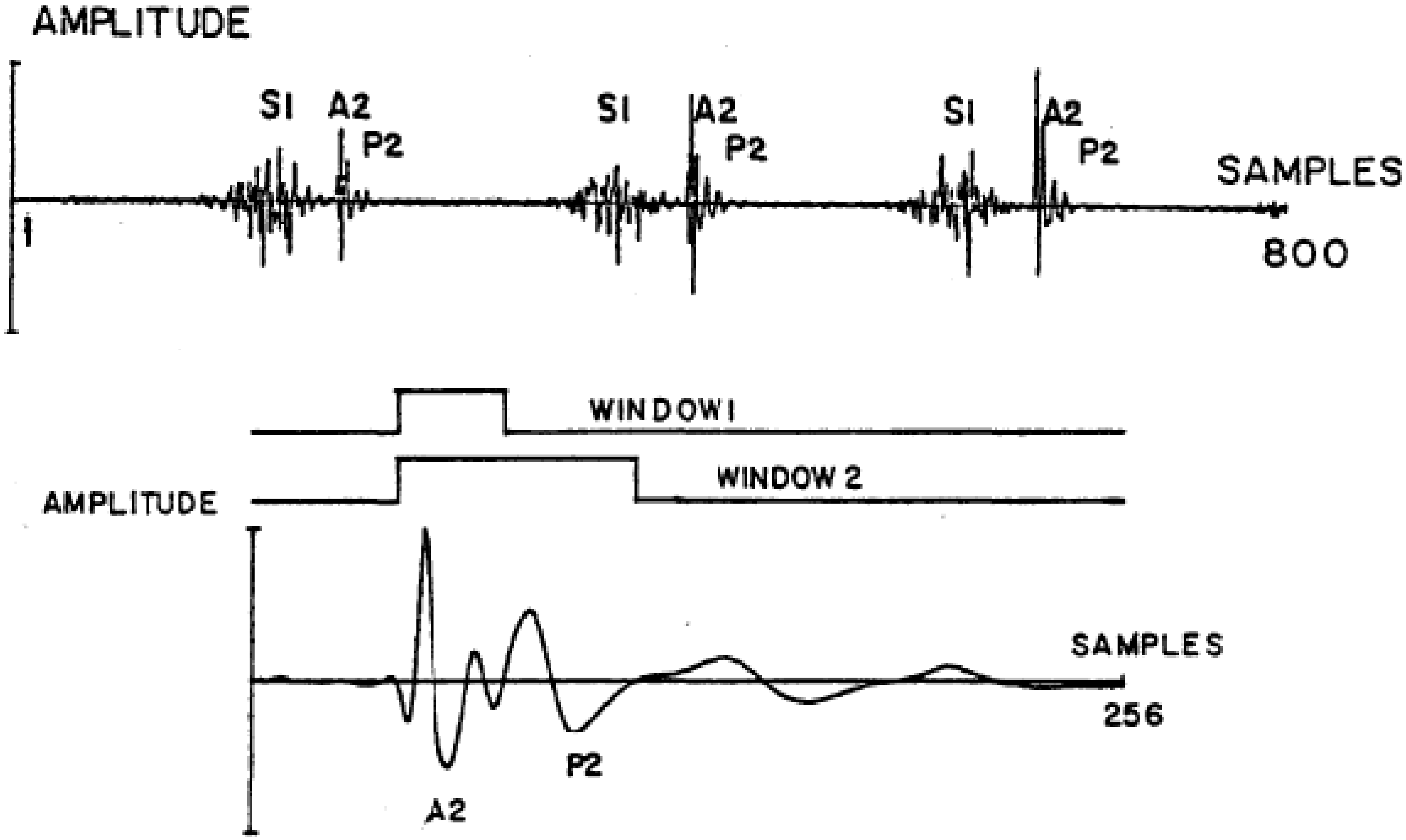
Classifying patterns



Classification



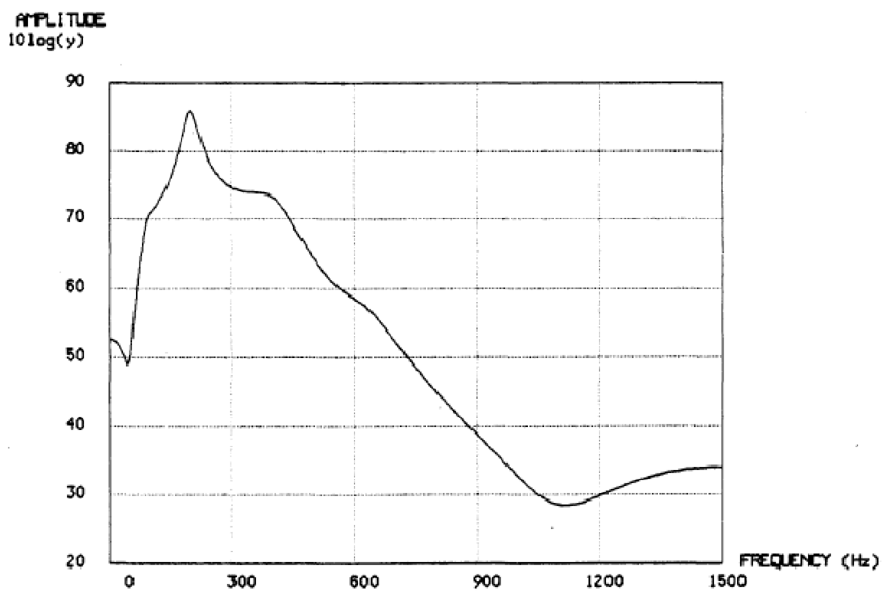
Application in PCG diagnosis



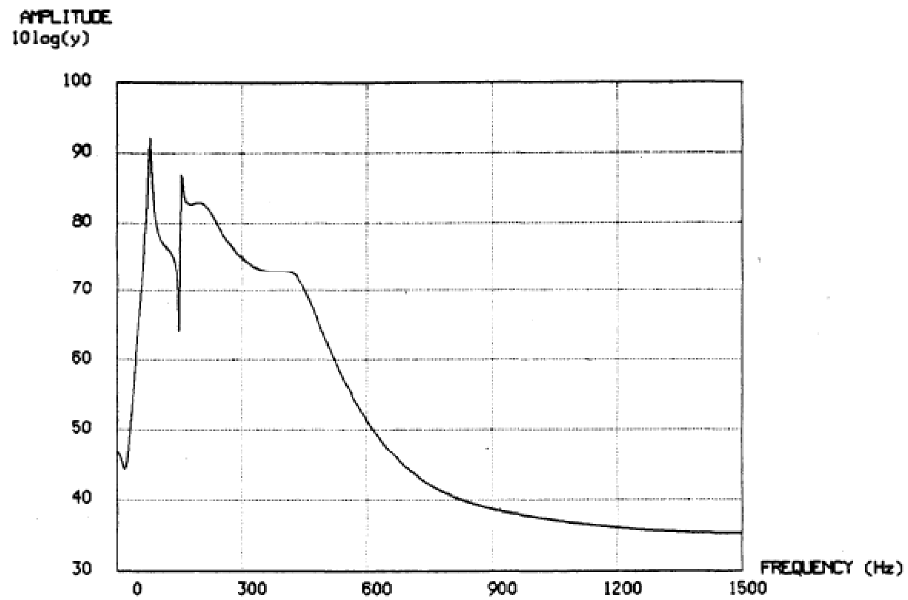
TH. Joo, et al IEEE Trans Biomed Eng 1983

High-resolution spectral analysis by pole-zero modeling

Using window 1



Using window 2



Training sets

TABLE I
THE TRAINING SET USED TO DESIGN THE CLASSIFIER

Training Set		
Valve	f1(Hz)	f2(Hz)
normal#1	172.	79.
normal#2	140.	99.
normal#3	82.	114.
normal#4	84.	123.
normal#5	123.	155.
normal#6	90.	131.
normal#7	79.	137.
normal#8	134.	99.
normal#9	146.	102.
normal#10	111.	73.
normal#11	93.	73.
normal#12	96.	131.
normal#13	87.	143.
normal#14	125.	70.
average	121.	109.
abnormal#1	184.	257.
abnormal#2	158.	216.
abnormal#3	213.	87.
abnormal#4	193.	117.
abnormal#5	178.	290.
abnormal#6	131.	193.
average	176.	193.

Classifier design

$$g(x) = g_n(x) - g_a(x)$$

$$g(x) = (x-\mu_a)^T \Sigma_a^{-1}(x-\mu_a) - (x-\mu_n)^T \Sigma_n^{-1}(x-\mu_n) + \ln\frac{|\Sigma_a|}{|\Sigma_n|} - 2\ln\frac{P(C_a)}{P(C_n)}$$

$$g(x) = \begin{bmatrix} f_1 - 176 \\ f_2 - 193 \end{bmatrix}^T \begin{bmatrix} 1.81e-3 & 2.88e-4 \\ 2.88e-4 & 5.82e-10 \end{bmatrix} \begin{bmatrix} f_1 - 176 \\ f_2 - 193 \end{bmatrix}$$

$$- \begin{bmatrix} f_1 - 121 \\ f_2 - 109 \end{bmatrix}^T \begin{bmatrix} 1.65e-3 & 7.85e-4 \\ 7.85e-4 & 1.69e-3 \end{bmatrix} \begin{bmatrix} f_1 - 121 \\ f_2 - 109 \end{bmatrix} + 1.82$$

Result

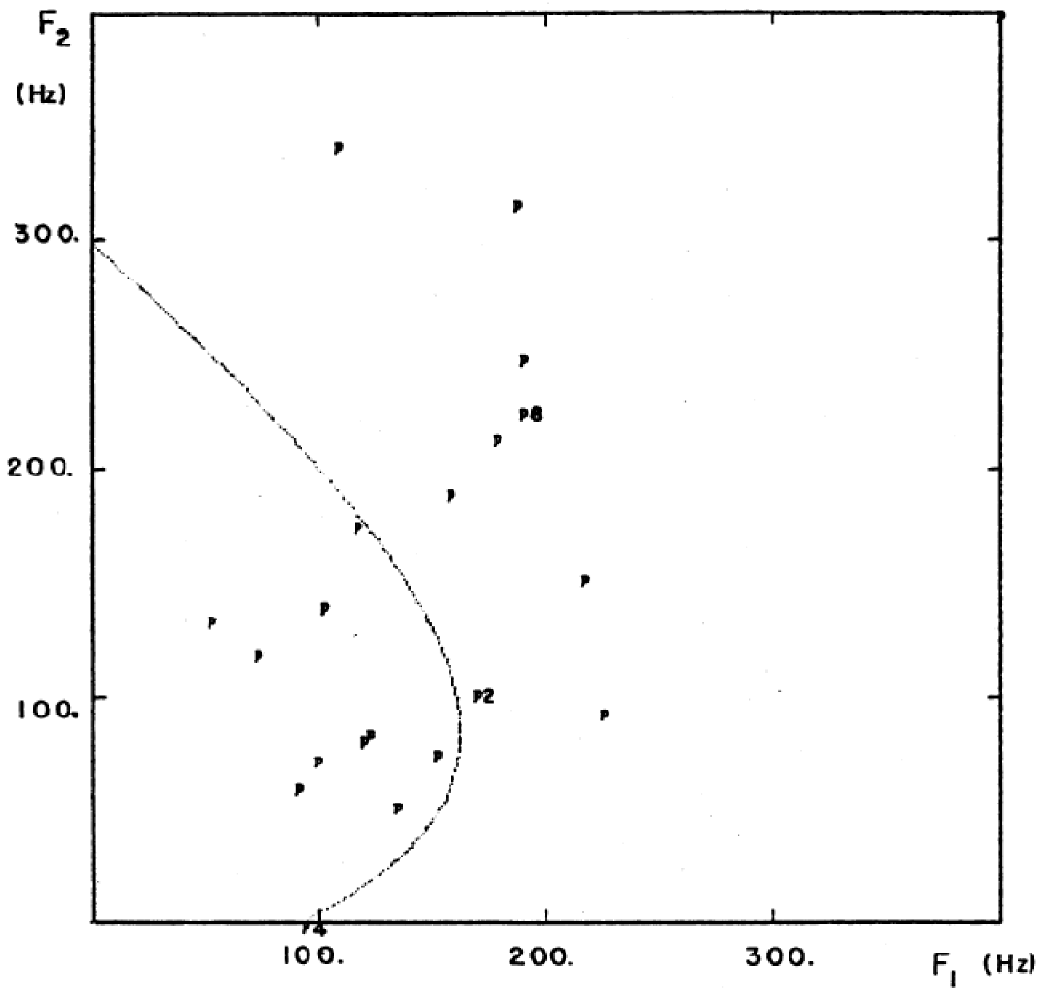


Fig. 6. Test set and the decision boundary in the feature space. Misclassified patients are Patients 2, 4, and 8. Refer to Table II for numerical values.

Test sets

TABLE II

Test Set Outcome of the Classifier and the Clinical Diagnosis

Test Set					
patient#	f1	f2	g(x)	classification	diagnosis
1	73.	120.	23.6	normal	normal
2	169.	102.	-2.68	abnormal	normal
3	120.	82.	10.2	normal	normal
4	93.	0.0	-0.45	abnormal	normal
5	52.	134.	29.4	normal	normal
6	123.	84.	9.64	normal	normal
7	134.	52.	4.14	normal	normal
8	190.	225.	-44.7	abnormal	normal
9	99.	73.	14.8	normal	normal
10	90.	61.	15.5	normal	normal
11	102.	140.	12.8	normal	normal
12	117.	175.	0.9	normal	normal
13	152.	76.	2.1	normal	normal
14	158.	190.	-17.9	abnormal	abnormal
15	216.	152.	-24.4	abnormal	abnormal
16	178.	213.	-34.7	abnormal	abnormal
17	225.	93.	-15.5	abnormal	abnormal
18	108.	342.	-85.8	abnormal	abnormal
19	190.	249.	-57.5	abnormal	abnormal
20	187.	316.	-103.	abnormal	abnormal

Measure of decision accuracy

- Let A be the class of abnormal and N the class of normal.
- True positive (TP)
  - The case when test is positive for a subject with disease

$$\text{Sensitivity} = \text{TP fraction} = P(T^{+} \mid A) = \frac{\text{number of TP decision}}{\text{number of subjects with disease}}$$

- True negative (TN)
  - The case when test is negative for a subject without disease

$$\text{Specificty} = \text{TN fraction} = P(T^{-} \mid N) = \frac{\text{number of TN decision}}{\text{number of subjects without disease}}$$

Measure of decision accuracy (cont.)

- False negative (FN)

$$\text{FN fraction} = P(T^{-} \mid A) = \frac{\text{number of negative decision who has disease}}{\text{number of subjects with disease}}$$

- False positive (FP)

$$\text{FP fraction} = P(T^{+} \mid N) = \frac{\text{number of positive decision who has not disease}}{\text{number of subjects without disease}}$$

Measure of decision accuracy (cont.)

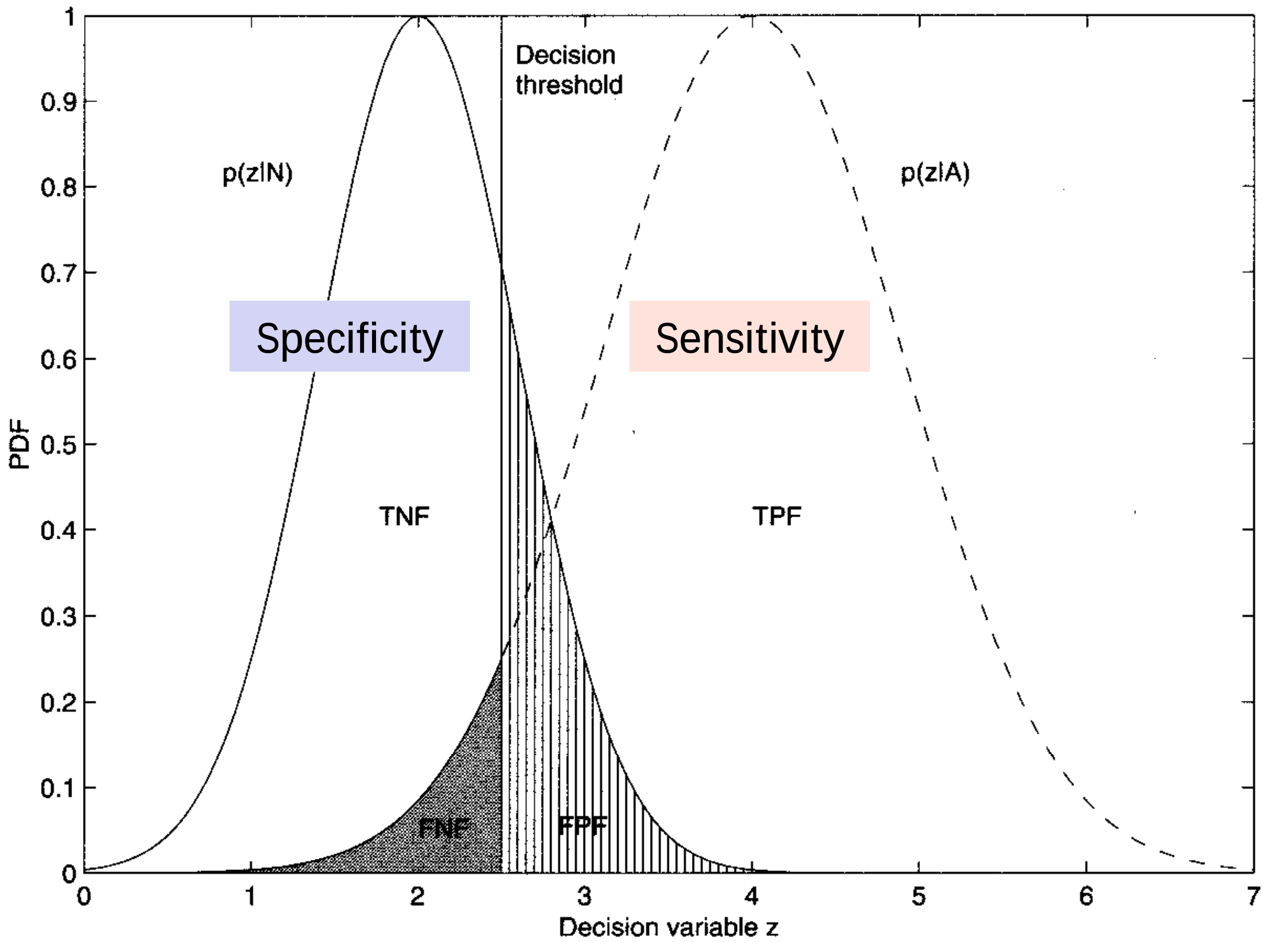
Actual Group	Predicted Group	
	Normal	Abnormal
Normal	TN	FP
Abnormal	FN	TP

TPF + FNF = 1

TNF + FPF = 1

Accuracy = $P(A)\,P(T^{+}|A) + P(N)\,P(T^{-}|N)$

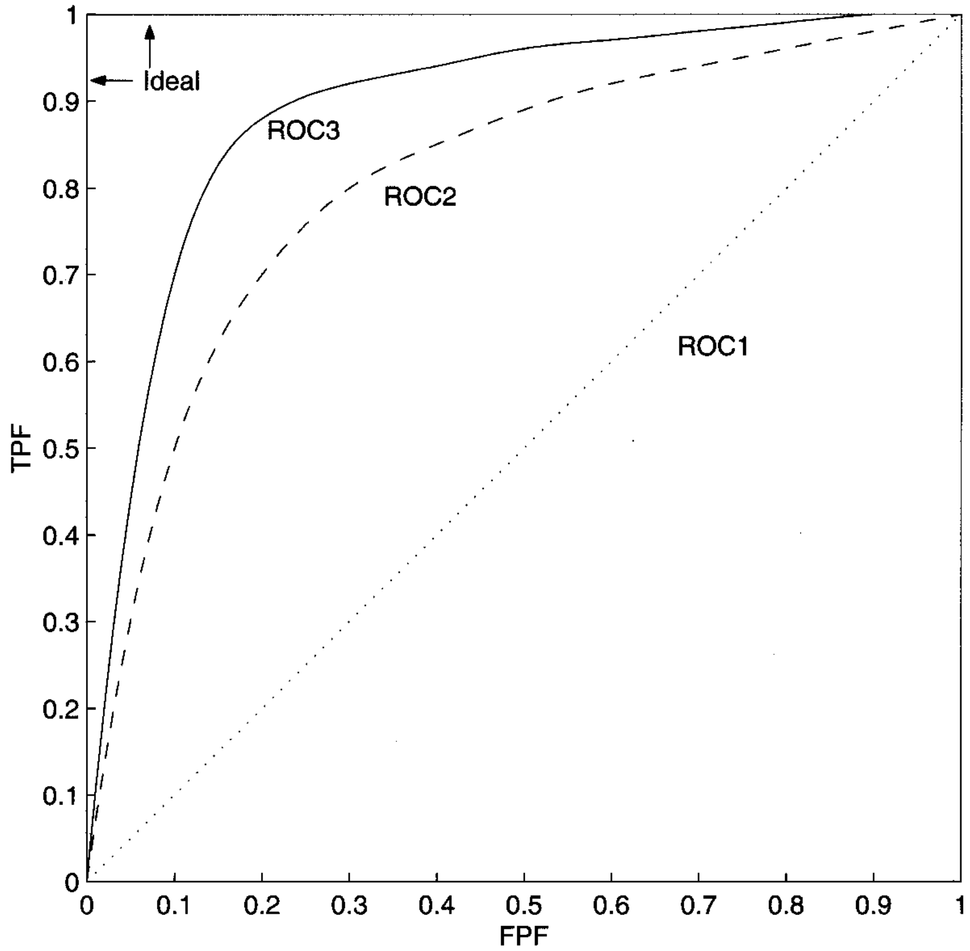
Measure of decision accuracy (cont.)



Decision threshold is the tradeoff between sensitivity and specificity

Receiver operating characteristic (ROC)

- Curve indicate the relationship between sensitivity and (1-specificity)
- A receiver (user) may choose to operate at any point along the curve



Reference

- R. Rangayyan, Biomedical Signal Analysis, John Wiely & Sons, 2002.
- TH. Joo et al, Pole-Zero Modeling and Classification of Phonocardiograms, IEEE Trans. on Biomedical Engineering, 30(2): 110-117, 1983.