Linear Discriminant Functions

- Linear Discriminant Functions and Decisions Surfaces
- Generalized Linear Discriminant Functions

Introduction

• Parametric Methods

- Underlying pdfs are known
- Training samples used to estimate pdf parameters

• Linear Discriminant Functions

- Forms of discriminant functions are known
- Similar to non-parametric techniques
- Sub-optimal, but simple to use

Linear discriminant function definition

A linear combination of components of *x*

$$g(x) = w^t x + w_0 \tag{1}$$

where

w is the weight vector w_0 the bias or threshold weight

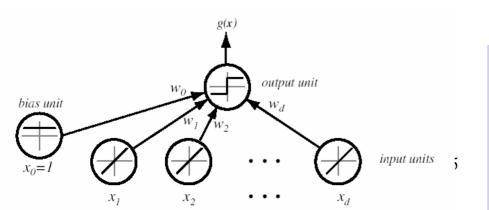
In general there are c such functions

Srihari: CSE 555

Two-category case

- With a discriminant function of the form (1) use:
- Decide ω_1 if $g(x) = w^t x + w_0 > 0$ and ω_2 if $g(x) = w^t x + w_0 < 0$
 - Decide ω_1 if $w^t x > -w_0$ and ω_2 otherwise If g(x) = 0 then x is assigned to either class

Simple Linear Two Category Classifier



 \Leftrightarrow

Each unit is shown as having inputs and outputs. The input units exactly output the same values as the inputs (except the bias unit which outputs a constant 1). The output unit emits a 1 if the sum of its weighted inputs is greater than zero and -1 otherwise

Hyperplane

- The equation g(x) = 0 defines the decision surface that separates points assigned to the category ω_1 from points assigned to the category ω_2
- When g(x) is linear, the decision surface is a *hyperplane*
- If x_1 and x_2 are both on the hyperplane then

$$\mathbf{w}^{t}\mathbf{x}_{1} + \mathbf{w}_{0} = \mathbf{w}^{t}\mathbf{x}_{2} + \mathbf{w}_{0}$$

or
$$\mathbf{w}^{t}(\mathbf{x}_{1} - \mathbf{x}_{2}) = 0$$

• Or w is normal to any vector lying on the hyperplane

Hyperplane Geometry

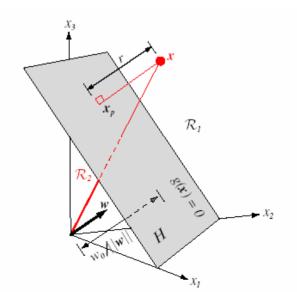
- Positive side, R_1 , if g(x) > 0
- Negative side, R_2 , if g(x) < 0
- <u>Algebraic measure of distance</u> of *x* to hyperplane
 - Define

 $\mathbf{x} = \mathbf{x}_{p} + \frac{\mathbf{r.w}}{\|\mathbf{w}\|}$ $\mathbf{g}(\mathbf{x}) = \mathbf{w}^{t}.\mathbf{x} + \mathbf{w}_{0} = \mathbf{w}^{t}.\mathbf{x}_{p} + \mathbf{w}^{t}\frac{\mathbf{r.w}}{\|\mathbf{w}\|} + \mathbf{w}_{0}$ $\mathbf{Because} \quad \mathbf{g}(\mathbf{x}_{p}) = \mathbf{w}^{t}.\mathbf{x}_{p} + \mathbf{w}_{0} = \mathbf{0} \quad \text{and } \mathbf{w}^{t}.\mathbf{w} = \|\mathbf{w}\|^{2}$ $\mathbf{r} = \frac{\mathbf{g}(\mathbf{x})}{\|\mathbf{w}\|}$

• Distance of origin to hyperplane

$$\mathbf{r} = \frac{\mathbf{g}(0)}{\|\mathbf{w}\|} = \frac{\mathbf{w}_0}{\|\mathbf{w}\|}$$

- Origin is on positive side if $w_0 > 0$ otherwise on negative side
- Orientation of hyperplane is determined by normal vector w
- If $w_0 = 0$ then g(x) has homogeneous form $w^t x$ and hyperplane passes through the origin
- Location of hyperplane is determined by the bias



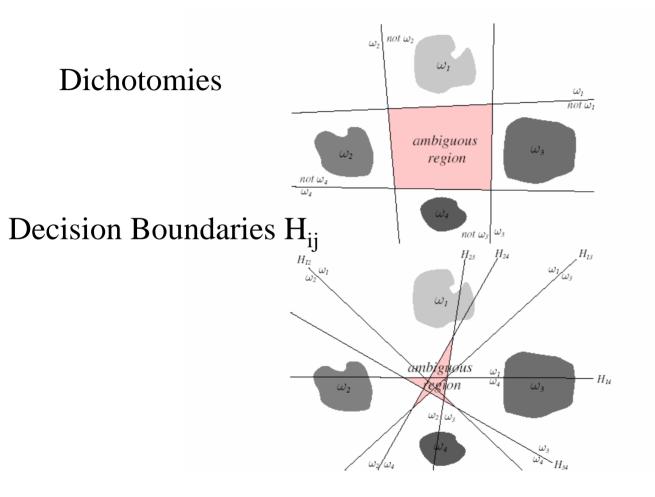
 W_0

Srihari: CSE 555

Multicategory Case

- Regard the problem as c two-class problems
- Two methods
 - 1. Separate points assigned to ω_i from those not assigned to ω_i
 - Use one hyperplane for each pair for classes
 Will need c(c-1)/2 discriminant functions

Using dichotomies for four-class problem



Multi-category case: Linear Machine

• We define c linear discriminant functions

$$g_i(x) = w_i^t x + w_{i0}$$
 $i = 1,...,c$

- and assign x to ωi if gi(x) > gj(x) ∀ j ≠ i; in case of ties, the classification is undefined
- In this case, the classifier is a "linear machine"
- A linear machine divides the feature space into c decision regions, with gi(x) being the largest discriminant if x is in the region Ri

Two class case of Linear Machine

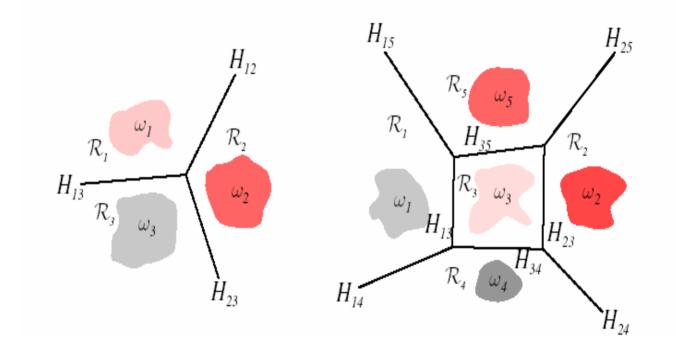
- For two contiguous regions R_i and R_j
- Boundary that separates them is a portion of hyperplane H_{ij} defined by:

 $g_i(x) = g_j(x)$

- $(w_i w_j)^t x + (w_{i0} w_{j0}) = 0$
- $w_i w_j$ is normal to H_{ij}
- Distance from x to H_{ij} is

$$\frac{\mathbf{g_i}(\mathbf{x}) - \mathbf{g_j}(\mathbf{x})}{\left\| \mathbf{w_i} - \mathbf{w_j} \right\|}$$

Linear Machine Boundaries for 3 and 5-class problems



Linear Machines and Unimodal Distributions

- It is easy to show that the decision regions for a linear machine are convex, this restriction limits the flexibility and accuracy of the classifier
- Every decision region is singly connected
- Suitable for Unimodal distributions
- Sometimes can be made to give good results for multimodal distributions as well

Generalized Linear Discriminant Functions

• Linear Discriminant functions can be written as

$$g(\mathbf{x}) = w_0 + \sum_{i=1}^d w_i x_i,$$

• Adding additonal terms we get quadratic discriminant functions

$$g(\mathbf{x}) = w_0 + \sum_{i=1}^d w_i x_i + \sum_{i=1}^d \sum_{j=1}^d w_{ij} x_i x_j.$$

Polynomial Discriminant Functions

- Continue to add terms such as
 - $W_{ijk} x_i x_j x_k$
- Leads to generalized linear discriminant function

$$g(\mathbf{x}) = \sum_{i=1}^{\hat{d}} a_i y_i(\mathbf{x})$$
$$g(\mathbf{x}) = \mathbf{a}^t \mathbf{y},$$

• a is a d^ dimensional weight vector

- Functions $y_i(x)$ are called ϕ -functions
- They map points in d-dimensional space into points in d^dimensional space
- Homogeneous discriminant function *a*^{*t*}*y* separates points by a hyperplane that passes through origin in transformed space
- Problem is one of finding a *Homogeneous linear* discriminant function

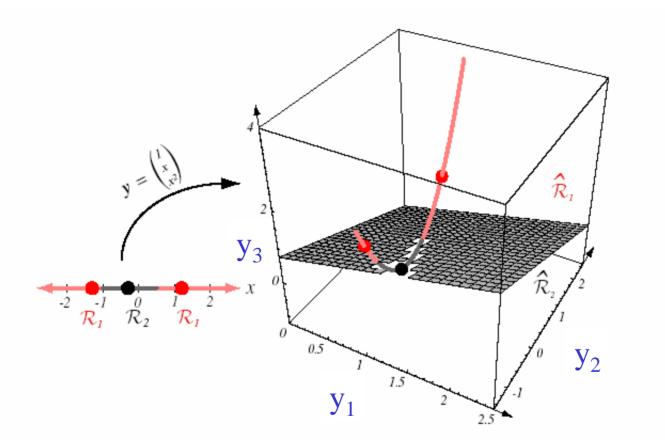
Quadratic as Generalized LDF

• Example

$$g(x) = a_1 + a_2 x + a_3 x^2,$$

$$\mathbf{y} = \left(\begin{array}{c} 1\\x\\x^2\end{array}\right).$$

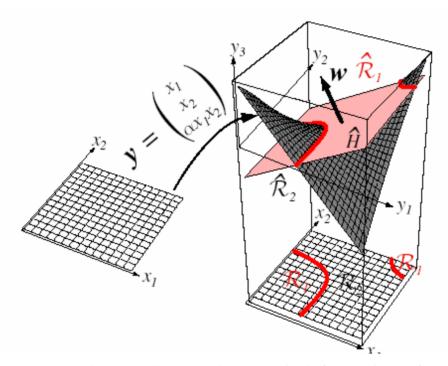
Transforming 1-d to 3-d



Error in figure: Plane should pass through origin

Mapping $\mathbf{y} = (1, x, x^2)^t$ takes a line and transforms it to a parabola in 3-d. A plane splits the resulting **y**-space into regions corresponding to two categories which gives a nonsimply connected decision region in **x**-space

Case of a=(-1,1,2)



2-d input space x is mapped through a polynomial function f to y. Here the mapping is $y_1 = x_1$, $y_2 = x_2$ and $y_3 \alpha x_1 x_2$ A linear discriminant in this transformed space is a hyperplane which cuts the surface. Points on the positive side of H correspond to ω_1 and those beneath it correspond to ω_2 .

Augmented Feature Vector

$$\mathbf{y} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_d \end{bmatrix} = \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix},$$

$$g(\mathbf{x}) = w_0 + \sum_{i=1}^d w_i x_i = \sum_{i=0}^d w_i x_i$$

$$\mathbf{a} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} = \begin{bmatrix} w_0 \\ \mathbf{w} \\ \mathbf{w} \end{bmatrix}.$$

3-d augmented feature space **V** \mathcal{R}_{i} \mathcal{R}_2 $v_o =$ $y_0=0$

Augmented weight vector **a** (at the origin).

Set of points for which $\mathbf{a}^t y = 0$ is a plane perpendicular to \mathbf{a} and passing through the origin of y-space, as indicated by red disk. Such a plane need not pass through origin of 2-d feature space as illustrated by dashed decision boundary at top of box.

Thus there exists an augmented weight vector \mathbf{a} that will lead to any straight decision line in \mathbf{x} -space.