

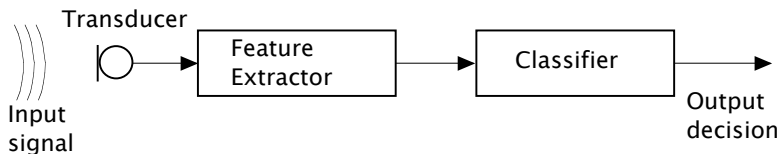
Pattern Recognition 1: Introduction

Arne Leijon

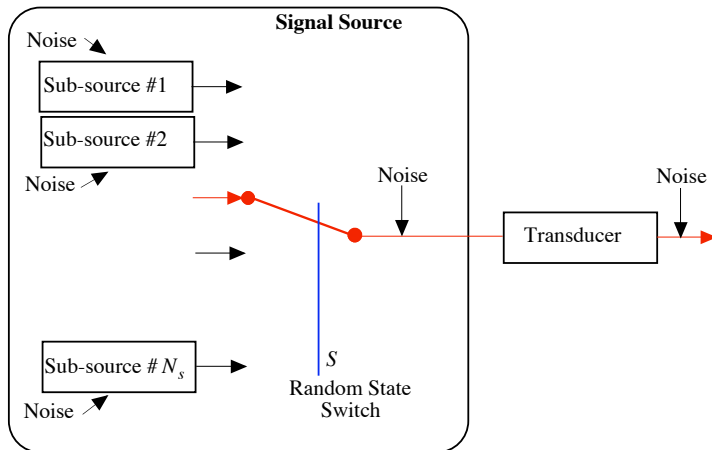
KTH Sound and Image Processing

Aug 27, 2012

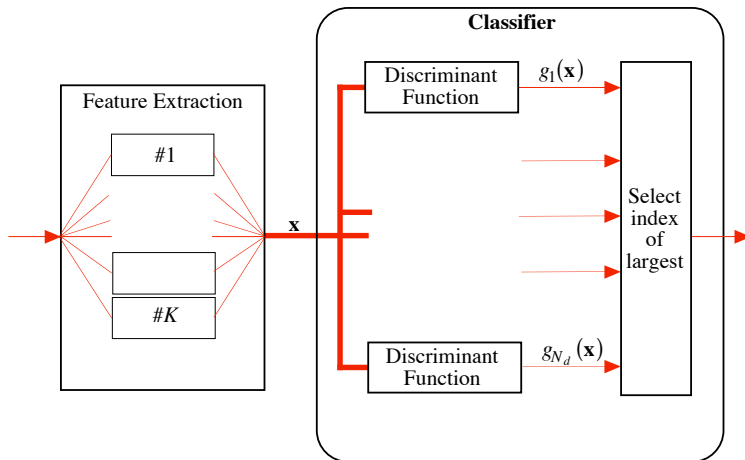
A Pattern-Classification System



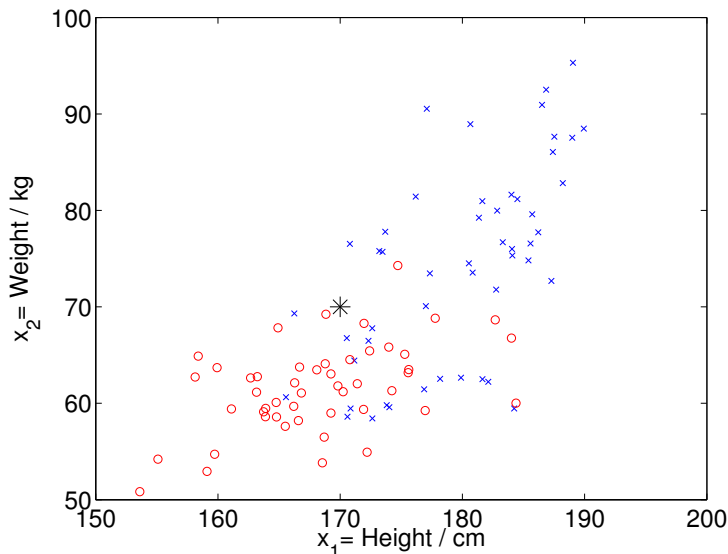
Source Categories



Optimal Classification Mechanism

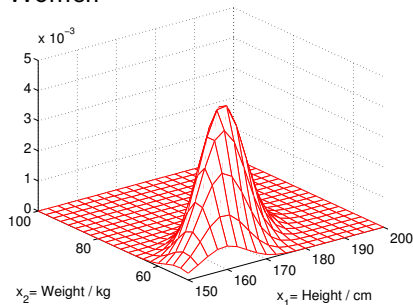


First Classifier Example: Man or Woman?

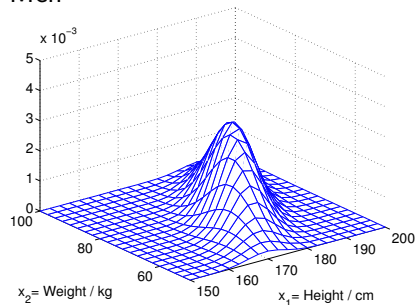


Two Gaussian Feature Probability Density Functions

Women

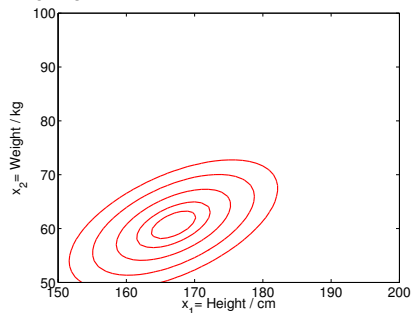


Men

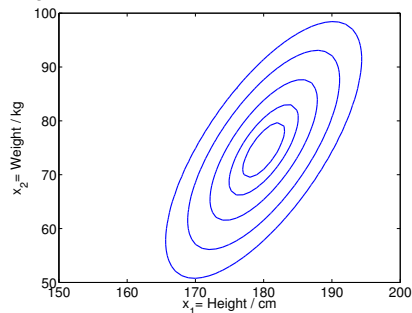


Contours of Equal Probability Density

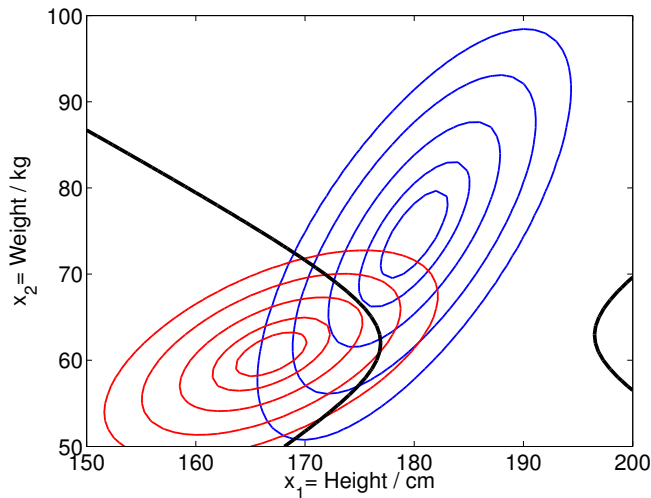
Women



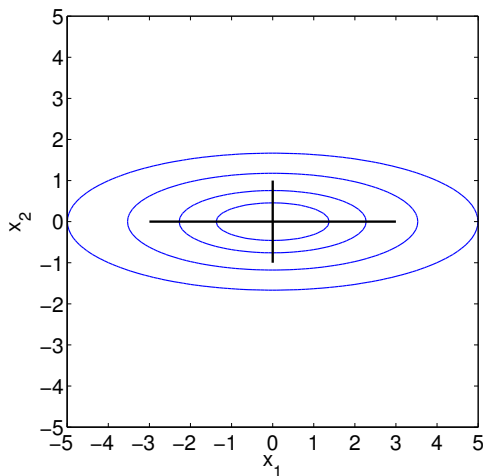
Men



Decision Boundary and Decision Regions

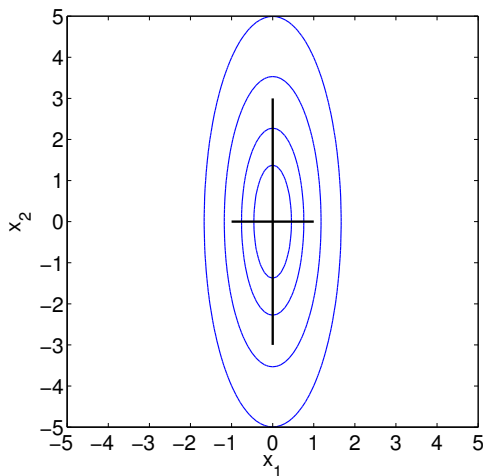


Gaussian Probability Density (2-dim)



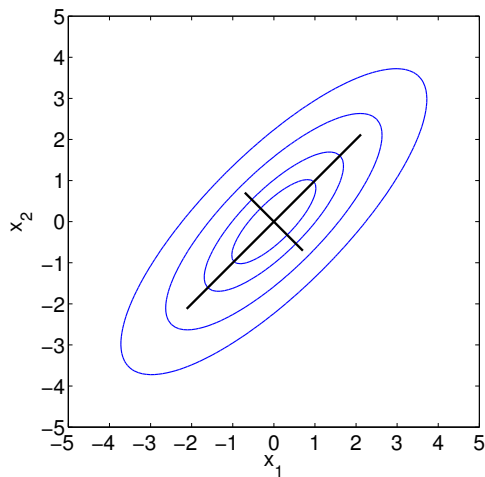
$$\begin{aligned}\text{cov}[X] &= \begin{pmatrix} 3^2 & 0 \\ 0 & 1^2 \end{pmatrix} = \\ &= \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}\end{aligned}$$

Gaussian Probability Density (2-dim)



$$\begin{aligned}\text{cov}[X] &= \begin{pmatrix} 1^2 & 0 \\ 0 & 3^2 \end{pmatrix} = \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}\end{aligned}$$

Gaussian Probability Density (2-dim)



$$\begin{aligned}\text{cov}[X] &= P \begin{pmatrix} 3^2 & 0 \\ 0 & 1^2 \end{pmatrix} P^T = \\ &= \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}\end{aligned}$$

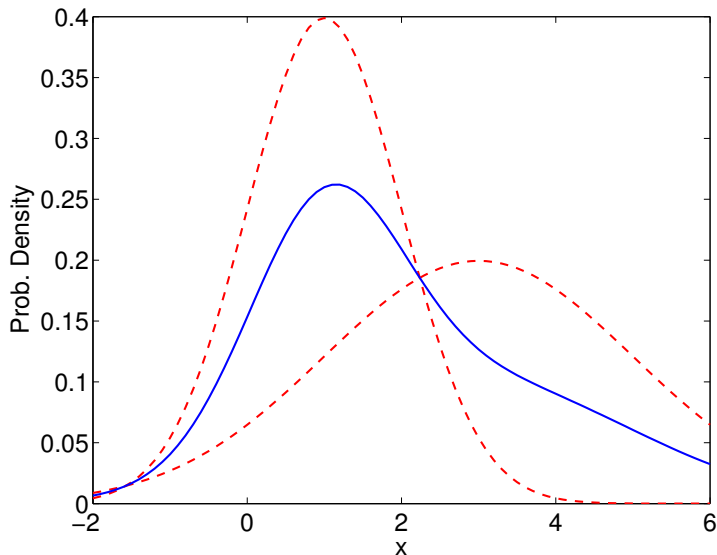
where

$$P = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

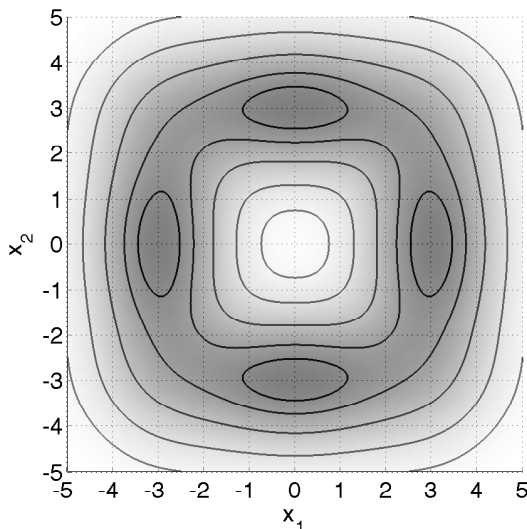
Gaussian Probability Density (K -dim)

$$f_X(x) = \frac{1}{(2\pi)^{K/2} \sqrt{\det C}} e^{-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)}$$

General Probability Density: Gaussian Mixture (1-dim)



General Probability Density: Gaussian Mixture (2-dim)



General Probability Density: Gaussian Mixture (K -dim)

$$f_X(x) = \sum_{m=1}^M w_m \frac{1}{(2\pi)^{K/2} \sqrt{\det C_m}} e^{-\frac{1}{2}(x-\mu_m)^T C_m^{-1}(x-\mu_m)}$$

Summary: Important Concepts

Feature Vector: vector x containing all input data to the classifier

Observation Space: set of all possible feature-vector values

Decision Function: $d(x)$ maps feature vector x into discrete decision output, coded as integer

Decision Region: set of feature-vector values giving same decision:
 $\Omega_i = \{x : d(x) = i\}$

Discriminant Function: $g_i(x)$ maps any feature vector x into real number, used for classification

General Optimal Classifier: $d(x) = \operatorname{argmax}_k g_k(x)$

Two-category Classifier: can be implemented using a single
Discriminant Function: $d(x) = \operatorname{sgn} g(x)$

Gaussian Mixture Model: can approximate any distribution