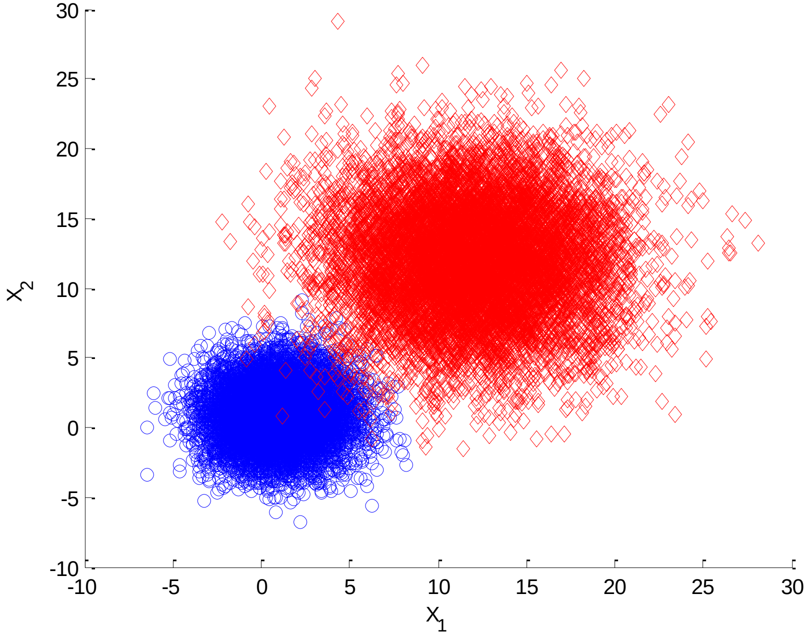
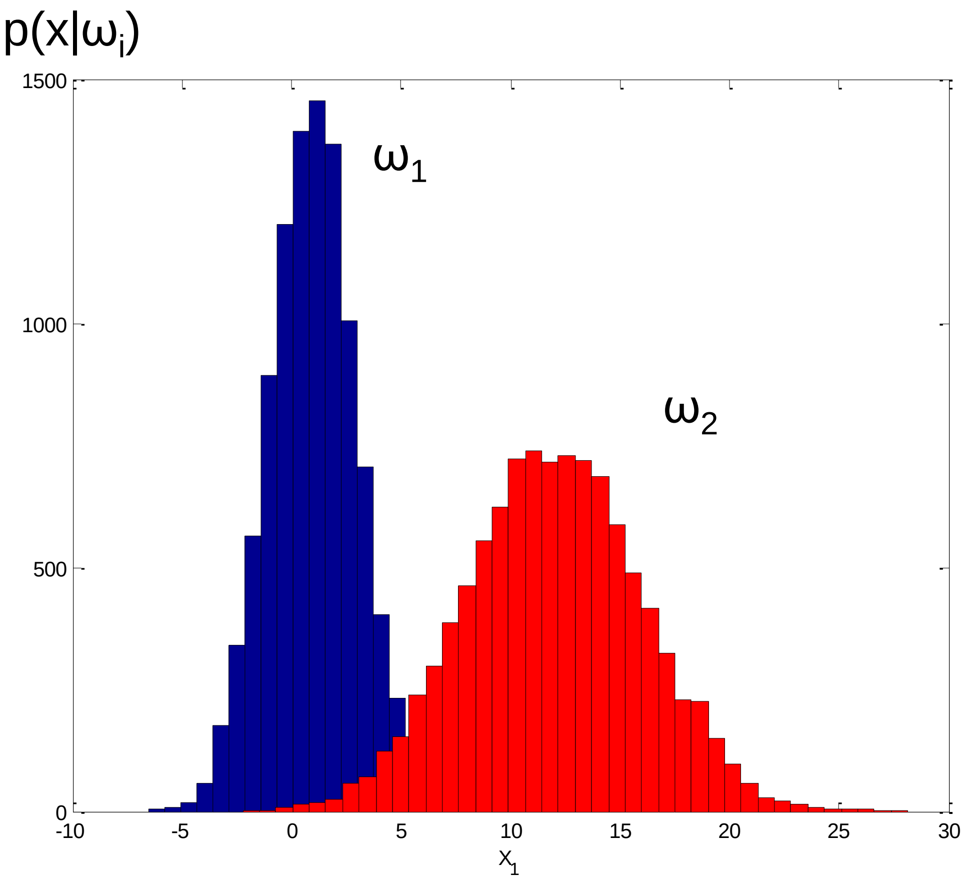
Machine Learning (521289S) Multidimensional Classification and Discriminant Functions

M.Sc. Antti Isosalo

Physiological Signal Analysis Team
Center for Machine Vision and Signal Analysis (CMVS)
University of Oulu

Multidimensional Classification: Two class scenario



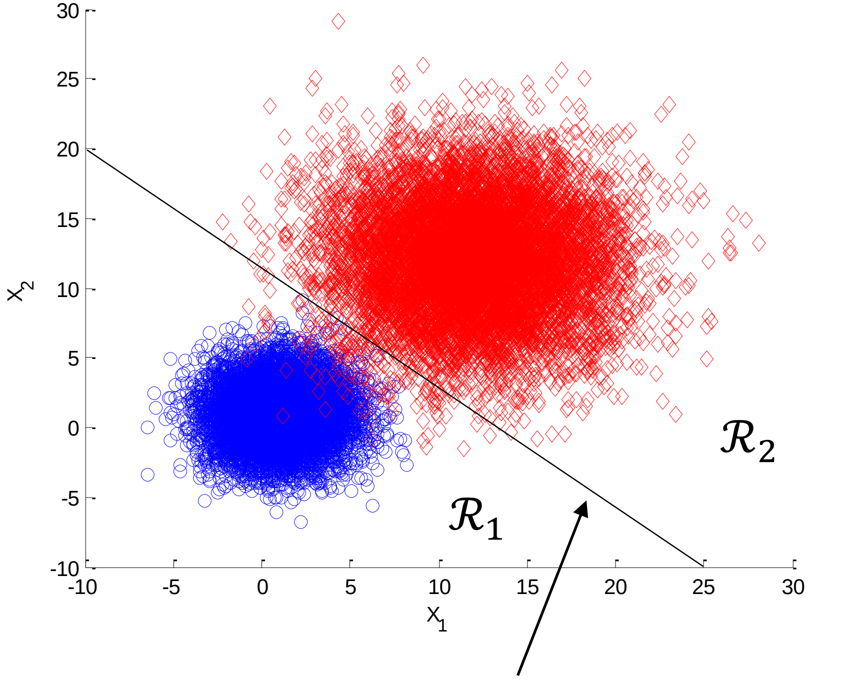


Two features, namely X_1 and X_2, have been measured. The points in the dataset are pairs **x** = (x_1, x_2), realizations of the 2-dimensional random variable **X** = (X_1, X_2).

Here we have marginal distribution of X_1 without reference to the values of the other feature X_2.

Multidimensional Classification: Two class scenario (cont.)

- This time we want to work in the *feature space*.
- We wish to set a *decision boundary* which divides the feature space into *decision regions*.
- It is clear that we can place the line describing the decision boundary in an *infinite amount* of equally *correct ways*.
- Now we need a way to determine on which side of the boundary we are (decision boundary is actually a surface, hyperplane).

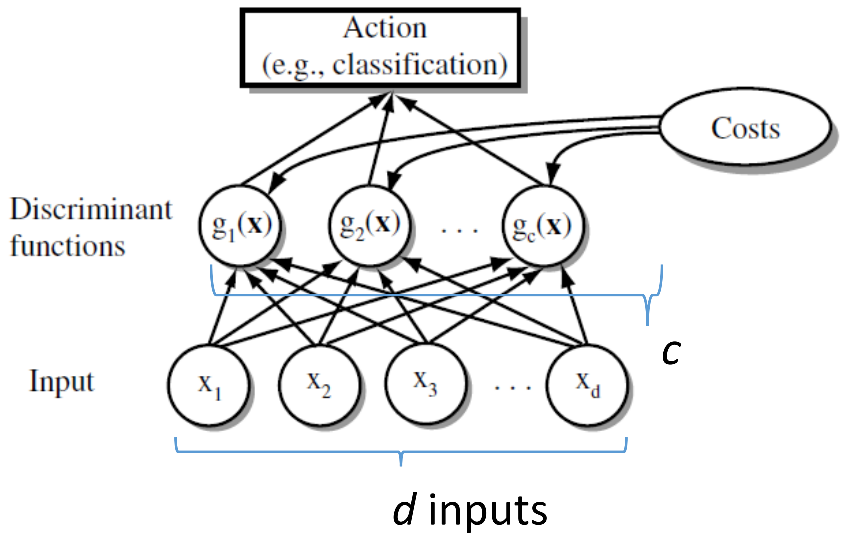


decision boundary

Discriminant Functions

- One way of representing a classifier is by using a set of discriminant functions $g_i(x),\ i = 1, \ldots, c$.
- "The classifier is said to assign a feature vector x to class ω_i if

  $g_i(\boldsymbol{x}) > g_j(\boldsymbol{x})$ for all $j \neq i$."
- The classifier can be seen as "a network that computes c discriminant functions and selects the category corresponding to the largest discriminant".
  - "A subsequent step determines which of the discriminant values is the maximum, and categorizes the input pattern accordingly."
- "The effect of any decision rule is to divide the feature space into c **decision regions**, $\mathcal{R}_1, \ldots, \mathcal{R}_c$."
  - "If $g_i(\boldsymbol{x}) > g_j(\boldsymbol{x})$ for all $j \neq i$, then $\boldsymbol{x}$ is in region $\mathcal{R}_i$, and the decision rule calls for us to assign $\boldsymbol{x}$ to ω_i."



The functional structure of a general statistical pattern classifier. From Duda R., Hart P., Stork D., "Pattern Classification," John Wiley & Sons Inc., 2nd ed., 2001

Discriminant Functions (cont.)

- Discriminant functions can be simplified to make them easier to understand or simpler to compute.
- In the exercises you will come accross the following:

$$g_i(\boldsymbol{x}) = P(\omega_i|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\omega_i)P(\omega_i)}{\sum_{j=1}^{c} p(\boldsymbol{x}|\omega_j)P(\omega_j)}$$

$$g_i(\boldsymbol{x}) = p(\boldsymbol{x}|\omega_i)P(\omega_i)$$

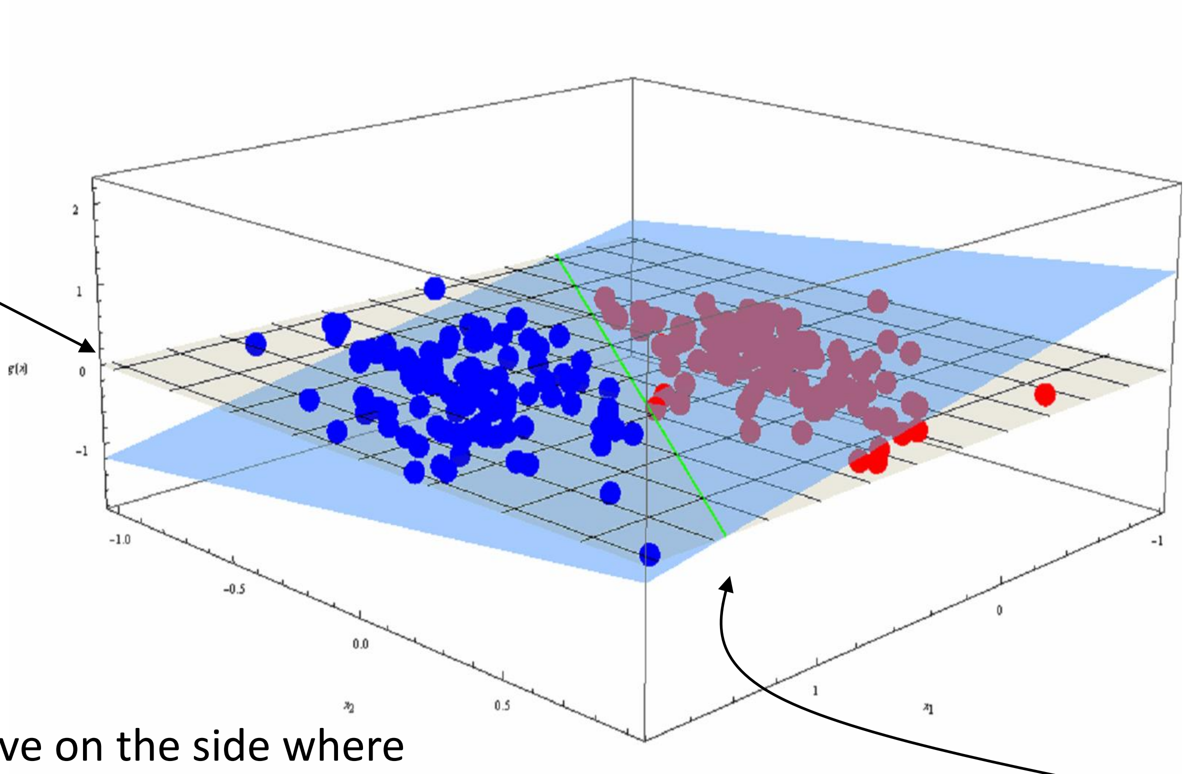
$$g_i(\boldsymbol{x}) = \ln p(\boldsymbol{x}|\omega_i) + \ln P(\omega_i)$$

$\ln(x)$ logarithm base e, or natural logarithm of x; $\ln(e) = \log_e(e) = 1$

Discriminant Functions: Task 1a

Figure shows the three-dimensional space in which the two-dimensional feature space is embedded as a plane.

Feature space,
here it is a plane



The discriminant function is a (hyper)plane that intersects the feature space plane forming the desired line as the decision boundary

The discriminant is positive on the side where all the red dots are and negative on the side where all the blue dots are.

On the decision boundary

$$g(x_1, x_2) = 0$$

$$\Leftrightarrow ax_1 + bx_2 + c = 0$$