Pattern Recognition

Lecture 05. Discriminant Functions and Decision Boundaries

Dr. Shanshan ZHAO

School of AI and Advanced Computing
Xi'an Jiaotong-Liverpool University

Academic Year 2023-2024

Table of Contents

Recap&Introduction

Previously

- Combine a Gaussian probability density function with class priors probabilities using Bayes' theorem to estimate class-conditional posterior probabilities.
- Assign that point to the class with the maximum posterior probability : dividing the input space into decision regions.

We will

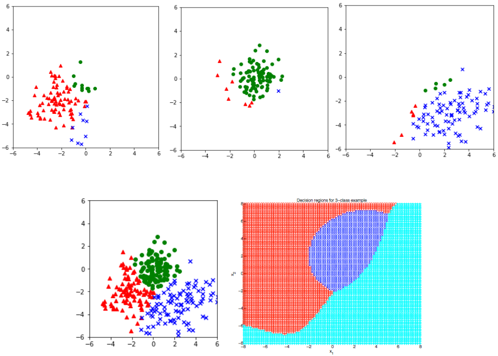
- Introduce discriminant functions which define the decision boundaries.
- Investigate the form of decision functions induced by Gaussian *pdfs* with different constraints on the covariance matrix.

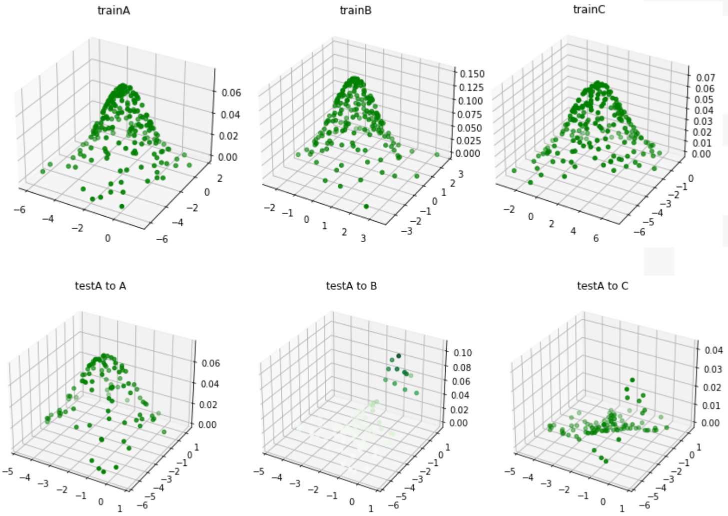
Notations

- x : a feature vector
- x: one feature, scalar

Decision Boundaries

We need to assign each point in the input space as a particular class. This divides the input space into *decision regions* $\mathscr{R}_k$, such that a point falling in $\mathscr{R}_k$ is assigned to class k.




trainA
trainB
trainC
testA to A
testA to B
testA to C

Decision Boundaries

Minimizing either the risk or the error probability is equivalent to partitioning the feature space into M regions, for a task with M classes.

For the minimum error probability case, this described by the equation

$$P(\omega_i|x) - P(\omega_j|x) = 0 \quad i, j = 1, 2, ..., M, \quad i \neq j$$

Decision Boundaries

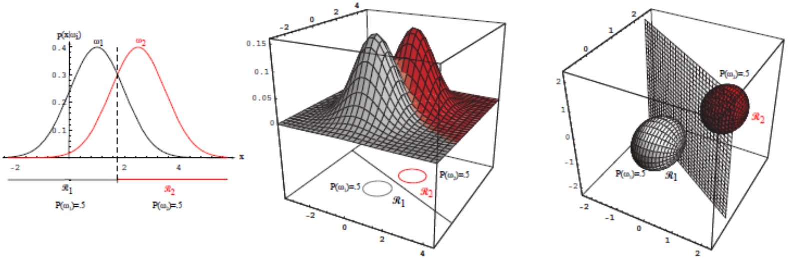


Figure 1: Decision boundaries 1D, 2D, and 3D [**duda1973pattern**].

Discriminant Functions

Sometimes, it may be more convenient to work with equivalent function of them, for example $g_i(x) \equiv f(P(\omega_i|x))$, where

- $f(\cdot)$is a monotonically increasing function.
- $g_i(x)$ is known as a **discriminant function**.

The decision rule is now stated as

$$\text{Decide } x \text{ in } \omega_i \text{ if} \quad g_i(x) > g_j(x) \quad \forall j \neq i$$

The decision boundaries, separating regions are described by

$$g_{ij}(x) \equiv g_i(x) - g_j(x) = 0, \quad i, j = 1, 2, ..., M, \quad i \neq j$$

Discriminant Functions

In terms of a set of discriminant functions $g_i(x)$, it can be viewed as a network or machine that computes c discriminant functions and select the category corresponding to the largest discriminant.

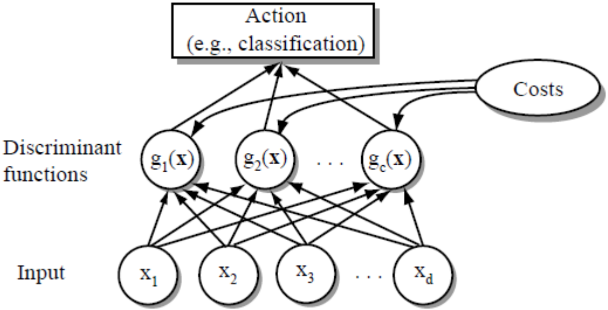


Figure 2: The functional structure of a general statistical pattern

Discriminant Functions

This is precisely what we mentioned in previous lectures when classifying based on the values of the log posterior probability. Thus the log posterior probability of class ω_k given a data point x is a possible discriminant function.

$$g_k(x) = \ln P(\omega_k|x) = \ln p(x|\omega_k) + \ln P(\omega_k) + const.$$

Decision boundaries are not changed by monotonic transformation(such as taking the log) of the discriminant functions.

recall : Gaussian distribution

The **Gaussian(or Normal)** distribution is the most commonly encountered (and easily analysed) continuous distribution.

- The one-dimensional or the univariate Gaussian or defined by

$$p(x|\mu,\sigma^2) = \mathcal{N}(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(\frac{-(x-\mu)^2}{2\sigma^2})$$

- The multivariate generalization of a Gaussian pdf in the D dimensional space is defined by

$$p(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))$$

Discriminant Functions for Normal density

What is the form of the discriminant function when using a Gaussian pdf?

As before, we take the discriminant function as the log posterior probability:

$$
\begin{aligned}
g_k(x) &= \ln P(\omega_k|x) \equiv \ln p(x|\omega_k) + \ln P(\omega_k) \\
&= -\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k) - \frac{1}{2}\ln|\Sigma_k| + \ln P(\omega_k) + c_k
\end{aligned}
$$

- We can drop the term $c_k = -1/2\ln(2\pi)$, since it is a constant that occurs in the discriminant function for each class.
- The first term on the right hand side of the equation is quadratic in the elements of x(i.e., if you multiply out the elements, there will be some terms containing x_i^2 or $x_i x_j$).

Discriminant Functions for Normal density

"We strongly encourage the reader to become proficient in manipulating Gaussian distributions using the techniques presented here as this will prove invaluable in understanding the more complex models".
– Pattern Recognition and Machine Learning, Bishop

Discriminant Functions for Normal density

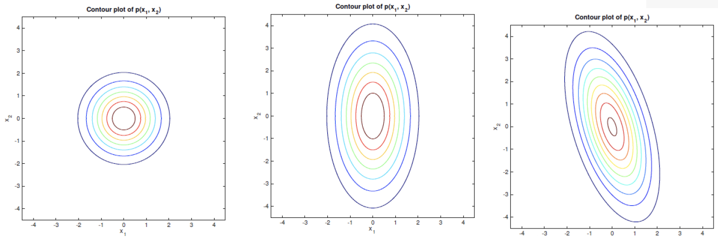
Exercise

Considering a two-class classification problem, where each class is modelled with a D-dimensional Gaussian distribution. Derive the formula for the decision boundary, and show that it is quadratic in x.

Recall

The covariance matrix Σ determines the shape of the density.

- case 1: $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
- case 2: $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$
- case 3: $\Sigma = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$



Recall

The covariance matrix Σ determines the shape of the density. For simplicity, let's focus on cases in 2-dimensitional space. We have

$$\Sigma = E\left[\begin{bmatrix} x_1-\mu_1, \\ x_2-\mu_2 \end{bmatrix}\begin{bmatrix} x_1-\mu_1 & x_2-\mu_2 \end{bmatrix}\right]$$
$$= \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

Where $E[x_i] = \mu_i, i = 1, 2$, and by definition $\sigma_{12} = E[(x_1-\mu_1)(x_2-\mu_2)]$, which is known as the covariance between the random variables x_1 and x_2 and it is a measure of their mutual statistical correlation.

Recall

The isovalue curves are ellipses of different orientation and with different ratios of major to minor axis lengths. To compute the isovalue curves is equivalent to computing the curves of constant values for the exponent, that is,

$$x^T\Sigma^{-1}x = [x_1-\mu_1, x_2-\mu_2]\begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} = C$$

or

$$\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} = C$$

For some constant C, This is the equation of an ellipse whose axes are determined by the variances of the involved features.

Discriminant Functions for Normal density

We assume that

$$\begin{bmatrix} \sigma_{i1}^2 & 0 \\ 0 & \sigma_{i2}^2 \end{bmatrix}$$

Then the Discriminant functions for normal density becomes,

$$g_i(x) = -\frac{1}{2}\left[\frac{(x_1-\mu_1)^2}{\sigma_{i1}^2} + \frac{(x_2-\mu_2)^2}{\sigma_{i2}^2}\right] - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i) + c_i$$

Obviously the associated decision curves $g_i(x) - g_j(x) = 0$ are quadric(i.e., ellipsoids, parabolas, hyperbolas, pairs of lines).

Discriminant Functions for Normal Density

Examples of quadric decision curvesplaying with covariance matrices of the Gaussian functions, different decision result, which can be *ellipsoids*, *parabolas*, *hyperbolas*, *pair of lines*.

$$\Sigma_1 = \begin{bmatrix} 0.3 & 0.0 \\ 0.0 & 0.35 \end{bmatrix} \Sigma_2 = \begin{bmatrix} 1.2 & 0.0 \\ 0.0 & 1.85 \end{bmatrix}$$

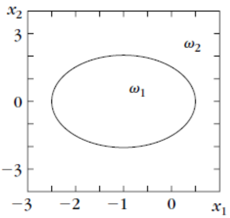


Figure 3: Ellipsoid decision boundary

$$\Sigma_1 = \begin{bmatrix} 0.1 & 0.0 \\ 0.0 & 0.75 \end{bmatrix} \Sigma_2 = \begin{bmatrix} 0.75 & 0.0 \\ 0.0 & 0.1 \end{bmatrix}$$

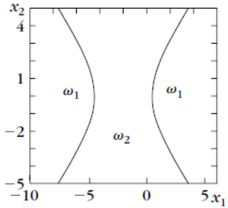


Figure 4: Hyperbolas decision boundary

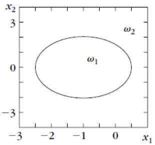


Figure 5: Ellipse decision boundary

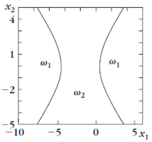


Figure 7: Hyperbolas decision boundary

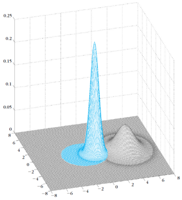


Figure 6: Ellipse decision boundary

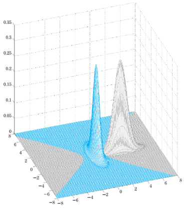
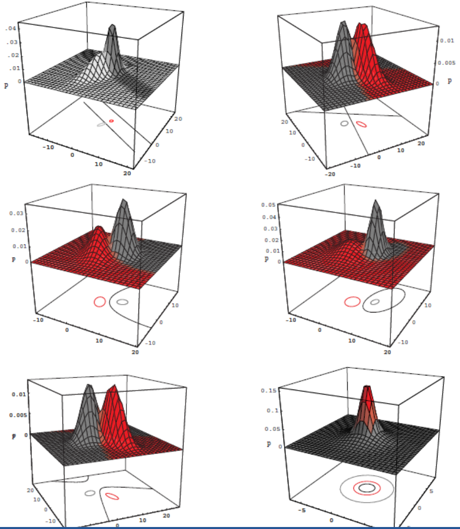
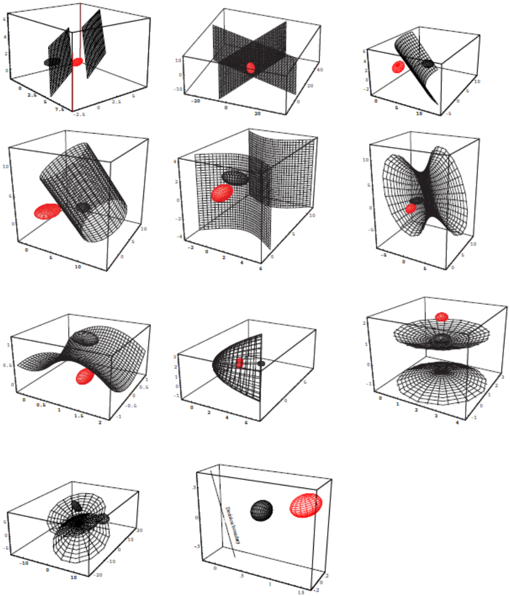


Figure 8: Hyperbolas decision boundary

Arbitrary Gaussian distributions lead to Bayes decision boundaries that are general hyperquadrics.





1. Recap & Introduction
2. Discriminant Functions
3. Linear Discriminants

Linear Discriminants

Take the discriminant function as the log posteriori probability

$$g_k(x) = \ln P(\omega_k|x) = \ln p(x|\omega_k) + \ln P(\omega_k) + const.$$
$$= -\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k) - \frac{1}{2}\ln|\Sigma_k| + \ln P(\omega_k)$$

Let's consider the case in which the Gaussian pdfs for each class all share the same covariance matrix.

That is, for all classes ω_k, $\Sigma_k = \Sigma$. In this case Σ is class-independent (since it is equal for all classes). Therefore, the discriminant function can be:

$$g_k(x) = -\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k) + \ln P(\omega_k)$$

Linear Discriminants

Now that we have,

$$g_k(x) = -\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k) + \ln P(\omega_k)$$

If we explicitly expand the quadratic matrix-vector expression we obtain the following:

$$g_k(x) = -\frac{1}{2}(x^T \Sigma^{-1} x - x^T \Sigma^{-1} \mu_k - \mu_k^T \Sigma^{-1} x + \mu_k^T \Sigma^{-1} \mu_k) + \ln P(\omega_k)$$

Can we drop $x^T \Sigma^{-1} x$ from the discriminant?

Linear Discriminants

Now since the covariance matrix Σ is symmetric, it follows that Σ^{-1} is also symmetric. Therefore,

$$x^T \Sigma^{-1} \mu_k = \mu_k^T \Sigma^{-1} x$$

The discriminant function can be simplified as:

$$g_k(x) = \mu_k^T \Sigma^{-1} x - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \ln P(\omega_k)$$

Linear Discriminants

We can define two new variables w_k(D-dimension vector) and w_{k0},

$$w_k^T = \mu_k^T \Sigma^{-1}$$
$$w_{k0} = -\frac{1}{2}\mu_k^T \Sigma^{-1} \mu_k + \ln P(\omega_k)$$

We obtain:

$$g_k(x) = w_k^T x + w_{k0}$$

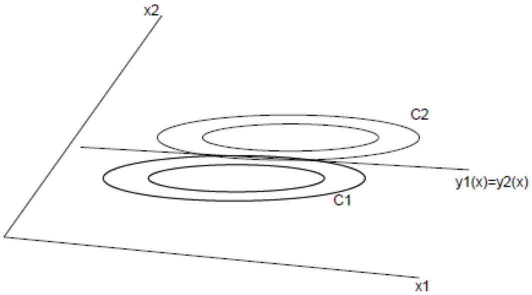
This is a linear equation in D dimensions, We refer to w_k as the **weight vector** and w_{k0} as the **bias** for class ω_k.

Linear Discriminants

Example

Considering a classification problem of two classes, whose discriminant function takes the form, $y(x) = w_k^T x + w_{k0}$.
When $D = 2$, the decision boundary is a straight line

- the decision boundary is a straight line
- the weight vector w is normal to the decision boundary.



Example

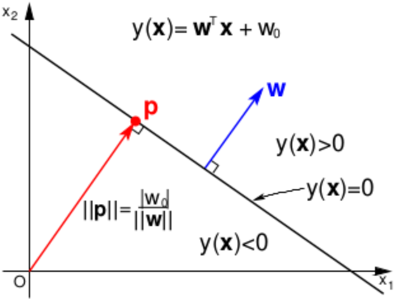


Figure 12: Geometry of a two-class linear discriminant.

Thank You !

$\mathcal{Q}$ & $\mathcal{A}$