بازشناسی الگو

درس ۴

طبقه‌بندی مبتنی بر نظریه‌ی تصمیم بیز

طبقه‌بندی بیزی برای توزیع‌های نرمال

Classification Based on Bayes Decision Theory

Bayesian Classification for Normal Densities

کاظم فولادی
دانشکده مهندسی برق و کامپیوتر
دانشگاه تهران

http://courses.fouladi.ir/pr

توزیع نرمال

تابع چگالی گاوسی

THE NORMAL DENSITY

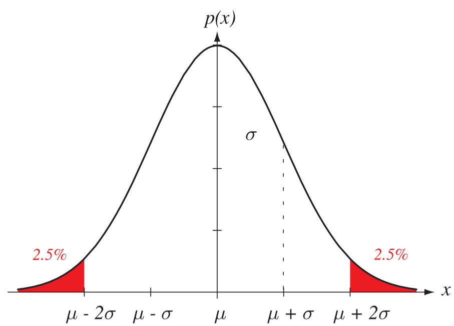
توزیع نرمال (گاوسی): مهم‌ترین و پرکاربردترین تابع توزیع در بازشناسی الگو

دلایل اهمیت و کاربرد متعدد:

۱) قضیه‌ی حد مرکزی

توزیع مجموع N متغیر تصادفی (با هر توزیعی) برای $N \to \infty$ به توزیع نرمال میل می‌کند.

۲) محاسبات اندک و سادگی فرمول‌بندی مسائل



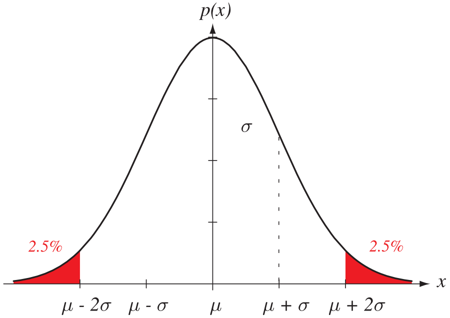
توزیع نرمال

تابع چگالی گاوسی : مزایا

THE NORMAL DENSITY

- توزیع گاوسی قابل بررسی به‌صورت تحلیلی است.
- توزیع گاوسی با گشتاورهای اول و دوم (μ و σ) کاملاً مشخص می‌شود.
- توزیه گاوسی با میانگین و واریانس داده شده، دارای حداکثر آنتروپی ممکن در میان همه‌ی توزیع‌هاست.
- بسیاری از فرآیندهای تصادفی در حالت مجانبی، گاوسی هستند.
- عدم همبستگی در توزیع گاوسی، استقلال را نتیجه می‌دهد.
- ۹۵% مساحت توزیع گاوسی در بازه‌ی $|x-\mu| \le 2\sigma$ قرار دارد.

$$P(|x-\mu| < \sigma) \cong 0.68$$
$$P(|x-\mu| < 2\sigma) \cong 0.95$$
$$P(|x-\mu| < 3\sigma) \cong 0.997$$



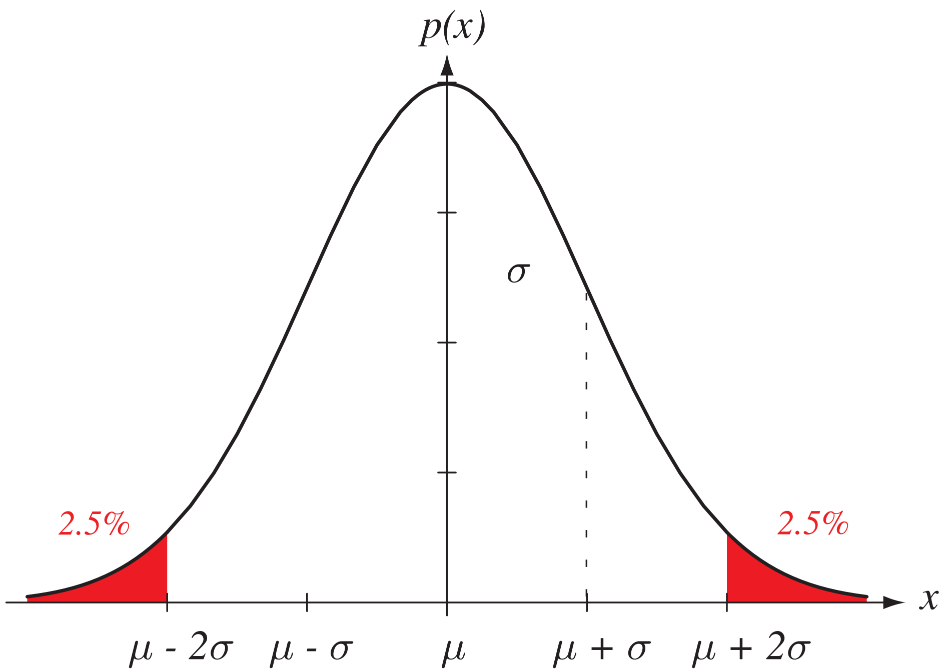
طبقه‌بندی مبتنی بر نظریه‌ی تصمیم بیز:
طبقه‌بندی بیزی برای توزیع‌های نرمال

۱

چگالی تک‌متغیره

توزیع نرمال

تابع چگالی گاوسی تکمتغیره



From: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification, John Wiley & Sons, Inc., 2001.

توزیع نرمال

تابع چگالی گاوسی تک‌متغیره

For $x \in \mathbb{R}$:

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

where

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

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

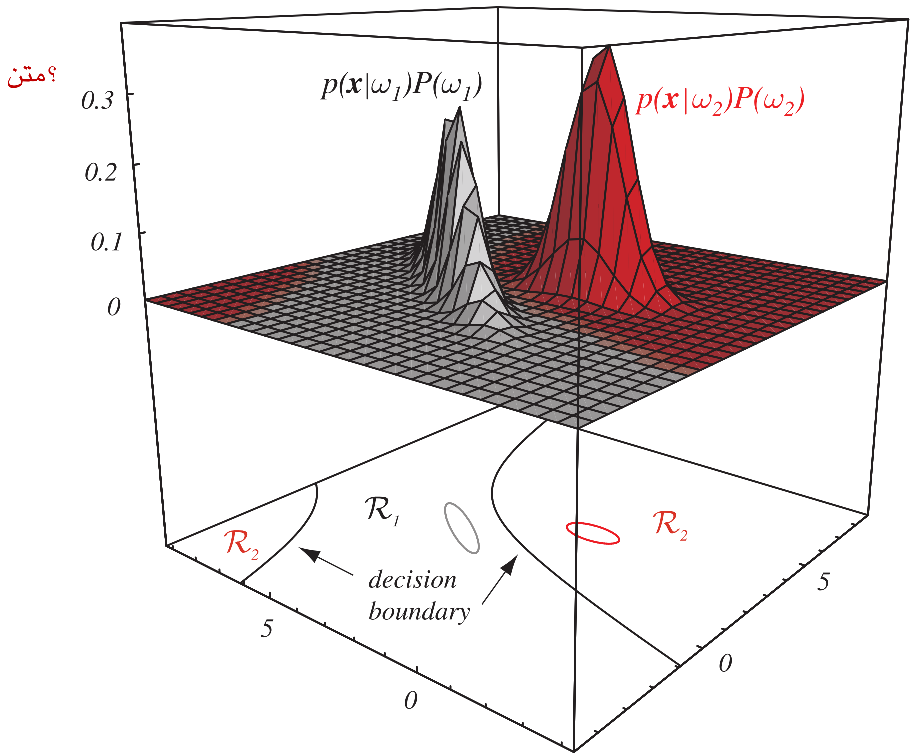
طبقه‌بندی مبتنی بر نظریه‌ی تصمیم بیز:
طبقه‌بندی بیزی برای توزیع‌های نرمال

۲

چگالی چندمتغیره

توزیع نرمال

تابع چگالی گاوسی چندمتغیره



From: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification, John Wiley & Sons, Inc., 2001.

توزیع نرمال

تابع چگالی گاوسی چندمتغیره

For $\mathbf{x} \in \mathbb{R}^d$:

$$p(\mathbf{x}) = N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

where

$$\boldsymbol{\mu} = E[\mathbf{x}] = \int \mathbf{x}\, p(\mathbf{x})\, d\mathbf{x}$$

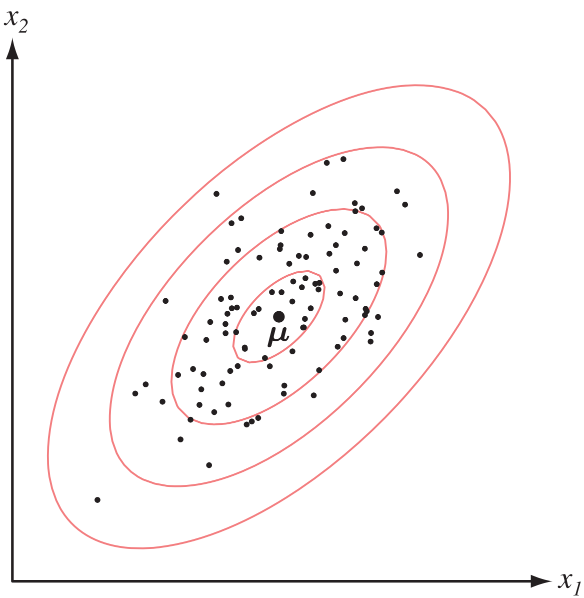
$$\boldsymbol{\Sigma} = E[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T] = \int (\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T p(\mathbf{x})\, d\mathbf{x}$$

- متقارن و معین مثبت (positive-definite) [دترمینان مثبت]
- عناصر قطر اصلی σ_{ii} = واریانس متغیر x_i
- عناصر خارج از قطر اصلی σ_{ij} = کوواریانس متغیرهای x_i و x_j

Prepared by Kazim Fouladi | Spring 2017 | 2nd Edition
کاظم فولادی

توزیع نرمال

تابع چگالی گاوسی چندمتغیره: شکل داده‌های دارای توزیع نرمال



Samples drawn from a two-dimensional Gaussian lie in a cloud centered on the mean $\boldsymbol{\mu}$. The loci of points of constant density are the ellipses for which $(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ is constant, where the eigenvectors of $\boldsymbol{\Sigma}$ determine the direction and the corresponding eigenvalues determine the length of the principal axes. The quantity $r^2 = (\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ is called the squared *Mahalanobis distance* from $\mathbf{x}$ to $\boldsymbol{\mu}$.

شکل داده‌های دارای توزیع نرمال

- داده‌های دارای یک توزیع نرمال:
  داخل یک **خوشه** قرار می‌گیرند.
  - مرکز خوشه: میانگین
  - شکل خوشه: بیضی
    (میزان کشیدگی/فشردگی با
    واریانس/ماتریس کوواریانس)
  - بردارهای ویژه ماتریس کوواریانس:
    محورهای اصلی بیضی
  - مقادیر ویژه ماتریس کوواریانس:
    طول محورهای اصلی بیضی

From: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification, John Wiley & Sons, Inc., 2001.

توزیع نرمال

تابع چگالی گاوسی چندمتغیره: فاصله‌ی ماهالونوبیس

MAHALONOBIS DISTANCE

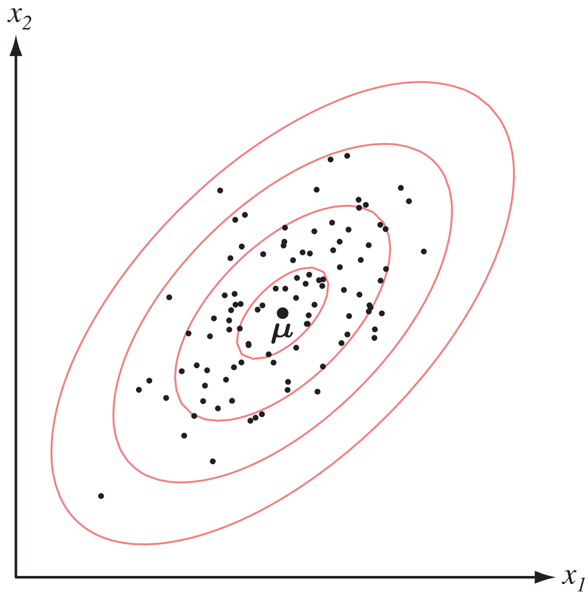
فاصله‌ی ماهالونوبیس:

فاصله‌ی میان یک مجموعه داده‌ی معین (با پارامترهای میانگین $\boldsymbol{\mu}$ و واریانس $\boldsymbol{\Sigma}$) و یک نمونه داده $\boldsymbol{x}$.

$$r^2 = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

وابستگی میان داده‌ها در نظر گرفته می‌شود (بر خلاف فاصله‌ی اقلیدسی).

در برابر مقیاس، تغییرناپذیر است (scale-invariant).



توزیع نرمال

تابع چگالی گاوسی چندمتغیره: اعمال تبدیل خطی

Linear Transformation

حاصل اعمال یک تبدیل خطی بر یک متغیر تصادفی گاوسی، یک متغیر تصادفی گاوسی خواهد بود.

- Recall that, given $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{A} \in \mathbb{R}^{d \times k}$, $\mathbf{y} = \mathbf{A}^T\mathbf{x} \in \mathbb{R}^k$, if $x \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $y \sim N(\mathbf{A}^T\boldsymbol{\mu}, \mathbf{A}^T\boldsymbol{\Sigma}\mathbf{A})$.
- As a special case, the *whitening transform*

$$\mathbf{A}_\mathbf{w} = \boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2}$$

where

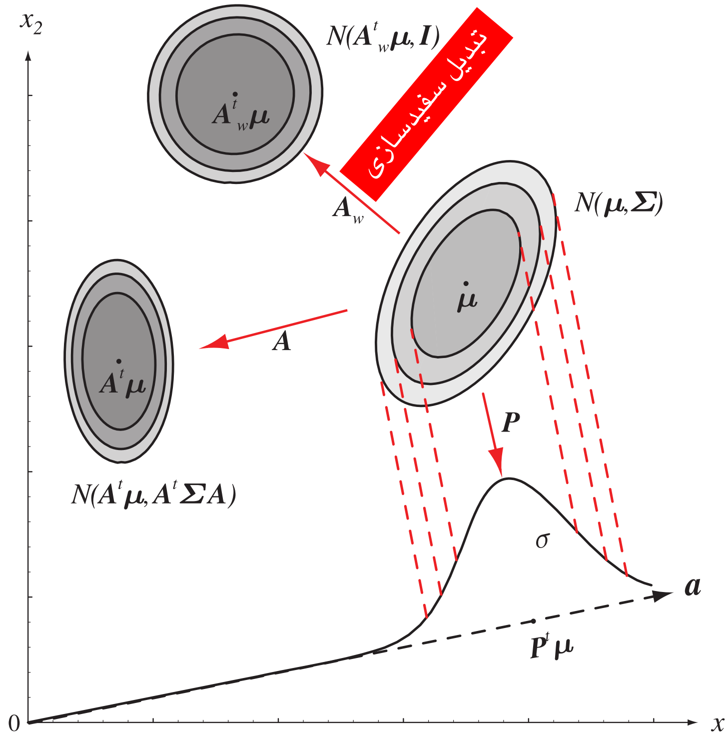
$\boldsymbol{\Phi}$ is the matrix whose columns are the orthonormal eigenvectors of $\boldsymbol{\Sigma}$,

$\boldsymbol{\Lambda}$ is the diagonal matrix of the corresponding eigenvalues,

gives a covariance matrix equal to the identity matrix $\mathbf{I}$.

توزیع نرمال

تابع چگالی گاوسی چندمتغیره : اعمال تبدیل خطی : مثال



From: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification, John Wiley & Sons, Inc., 2001.

طبقه‌بندی مبتنی بر نظریه‌ی تصمیم بیز:
طبقه‌بندی بیزی برای توزیع‌های نرمال

۳

توابع تفکیک برای چگالی نرمال

قاعده‌ی تصمیم بیزی

تابع چگالی گاوسی چندمتغیره: توابع تفکیک

توابع تفکیک برای طبقه‌بندی با نرخ خطای می‌نیمم، و برای توزیع گاوسی:

- Discriminant functions for minimum-error-rate classification can be written as

$$g_i(\mathbf{x}) = \ln p(\mathbf{x}|w_i) + \ln P(w_i)$$

- For $p(\mathbf{x}|w_i) = N(\boldsymbol{\mu_i}, \boldsymbol{\Sigma}_i)$

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{\Sigma}_i| + \ln P(w_i)$$

قاعده‌ی تصمیم بیزی

تابع چگالی گاوسی چندمتغیره: توابع تفکیک: حالت کوواریانس مساوی و متناسب با ماتریس همانی

$$g_i(\mathbf{x}) = -\frac{\|\mathbf{x}-\boldsymbol{\mu}_i\|^2}{2\sigma^2} + \ln P(\omega_i),$$

$$\|\mathbf{x}-\boldsymbol{\mu}_i\|^2 = (\mathbf{x}-\boldsymbol{\mu}_i)^t(\mathbf{x}-\boldsymbol{\mu}_i).$$

$$g_i(\mathbf{x}) = -\frac{1}{2\sigma^2}[\mathbf{x}^t\mathbf{x} - 2\boldsymbol{\mu}_i^t\mathbf{x} + \boldsymbol{\mu}_i^t\boldsymbol{\mu}_i] + \ln P(\omega_i),$$

$$g_i(\mathbf{x}) = \mathbf{w}_i^t\mathbf{x} + w_{i0},$$

Case 1: $\Sigma_i = \sigma^2\mathbf{I}$

Discriminant functions are

$$g_i(\mathbf{x}) = \mathbf{w}_i^T\mathbf{x} + w_{i0} \quad \text{(linear discriminant)}$$

where

$$\mathbf{w}_i = \frac{1}{\sigma^2}\boldsymbol{\mu}_i$$

$$w_{i0} = -\frac{1}{2\sigma^2}\boldsymbol{\mu}_i^T\boldsymbol{\mu}_i + \ln P(w_i)$$

(w_{i0} is the threshold or bias for the i'th category)

Case 1: $\Sigma_i = \sigma^2 \mathbf{I}$

- Decision boundaries are the hyperplanes $g_i(\mathbf{x}) = g_j(\mathbf{x})$, and can be written as

$$\mathbf{w}^T(\mathbf{x} - \mathbf{x_0}) = 0$$

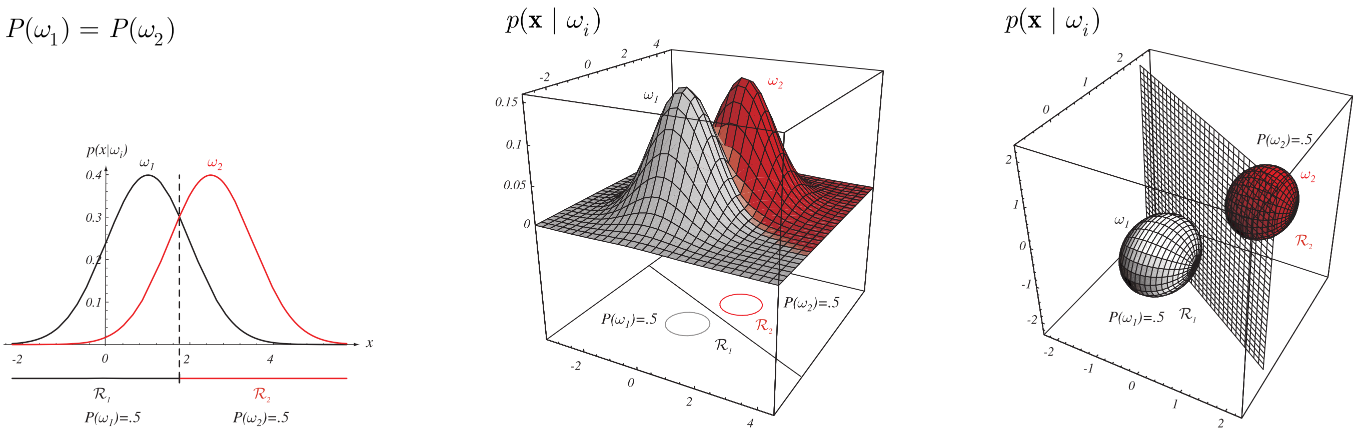
where

$$\mathbf{w} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j$$
$$\mathbf{x_0} = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{\sigma^2}{\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2} \ln \frac{P(w_i)}{P(w_j)} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

- Hyperplane separating $\mathcal{R}_i$ and $\mathcal{R}_j$ passes through the point $\mathbf{x_0}$ and is orthogonal to the vector $\mathbf{w}$.

قاعده‌ی تصمیم بیزی

تابع چگالی گاوسی چندمتغیره: توابع تفکیک: حالت کوواریانس مساوی و متناسب با ماتریس همانی



اگر ماتریس کوواریانس برای دو توزیع مساوی بوده و متناسب با ماتریس همانی باشد،
آن‌گاه توزیع‌ها کره‌های d–بعدی خواهند بود و
مرز یک ابرصفحه‌ی تعمیم‌یافته با $d-1$ بعد (عمود بر خط جداکننده‌ی میانگین‌ها) خواهد بود.

From: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification, John Wiley & Sons, Inc., 2001.

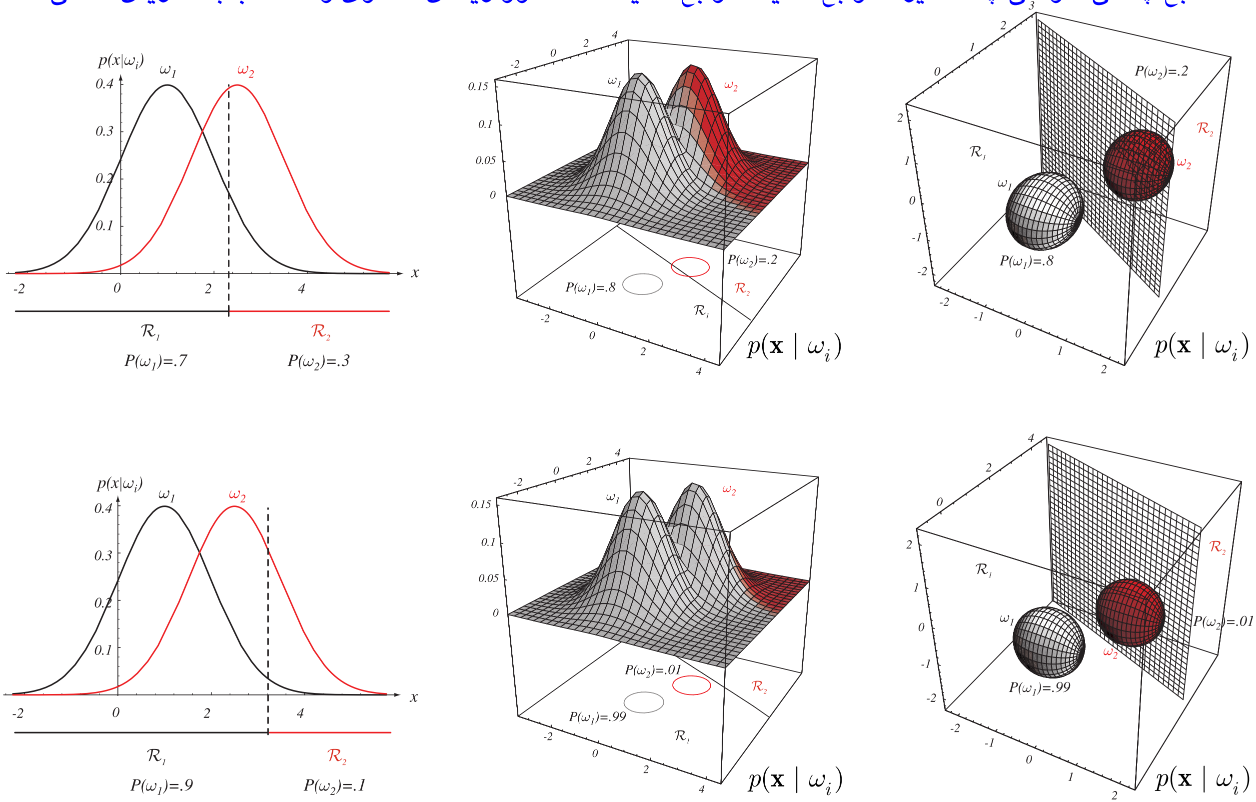
Case 1: $\mathbf{\Sigma}_i = \sigma^2 \mathbf{I}$

Special case when $P(w_i)$ are the same for $i = 1, \ldots, c$ is the *minimum-distance classifier* that uses the decision rule

$$\text{assign } \mathbf{x} \text{ to } w_{i^*} \text{ where } i^* = \arg\min_{i=1,\ldots,c} \|\mathbf{x} - \boldsymbol{\mu}_i\|$$

قاعده‌ی تصمیم بیزی

تابع چگالی گاوسی چندمتغیره: توابع تفکیک: توابع تفکیک حالت کوواریانس مساوی و متناسب با ماتریس همانی



با تغییر احتمال پیشین طبقه‌ها، مرز تصمیم شیفت پیدا می‌کند.

From: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification, John Wiley & Sons, Inc., 2001.

تابع چگالی گاوسی چندمتغیره: توابع تفکیک: حالت کوواریانس مساوی

Case 2: $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$

Discriminant functions are

$$g_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + w_{i0} \quad \text{(linear discriminant)}$$

where

$$\mathbf{w}_i = \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i$$
$$w_{i0} = -\frac{1}{2} \boldsymbol{\mu}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i + \ln P(w_i)$$

From: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification, John Wiley & Sons, Inc., 2001.

Case 2: $\Sigma_i = \Sigma$

- Decision boundaries can be written as

$$\mathbf{w}^T(\mathbf{x} - \mathbf{x_0}) = 0$$

where

$$\mathbf{w} = \mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

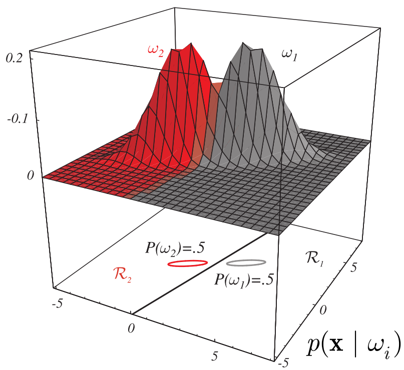
$$\mathbf{x_0} = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{\ln(P(w_i)/P(w_j))}{(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

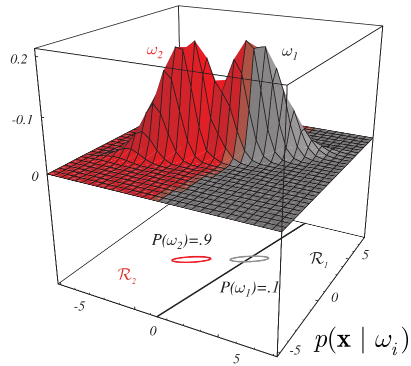
- Hyperplane passes through $\mathbf{x_0}$ but is not necessarily orthogonal to the line between the means.

From: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification, John Wiley & Sons, Inc., 2001.

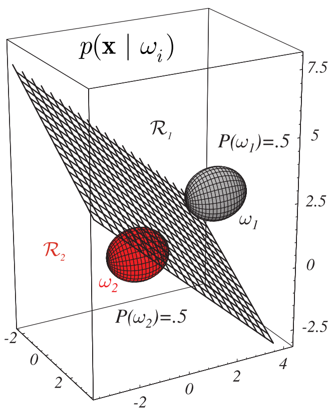
قاعده‌ی تصمیم بیزی

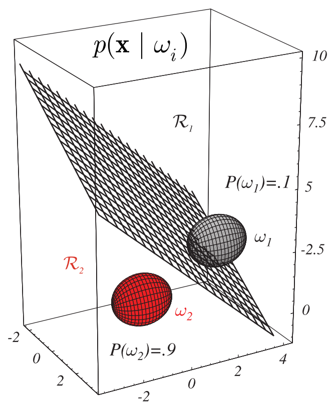
تابع چگالی گاوسی چندمتغیره: توابع تفکیک: حالت کوواریانس مساوی





حالت دوبعدی:
چگالی‌های احتمال و نواحی تصمیم متناظر برای
توزیع‌های گاوسی مساوی اما نامتقارن





حالت سه‌بعدی:
چگالی‌های احتمال و نواحی تصمیم متناظر برای
توزیع‌های گاوسی مساوی اما نامتقارن

ابرصفحه‌های تصمیم لزوماً بر خط واصل بین میانگین‌ها عمود نیست.

From: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification, John Wiley & Sons, Inc., 2001.

Case 3: Σ_i = arbitrary

- Discriminant functions are

$$g_i(\mathbf{x}) = \mathbf{x}^T \mathbf{W}_i \mathbf{x} + \mathbf{w}_i^T \mathbf{x} + w_{i0} \quad \text{(quadratic discriminant)}$$

where

$$\mathbf{W}_i = -\frac{1}{2}\boldsymbol{\Sigma}_i^{-1}$$

$$\mathbf{w}_i = \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i$$

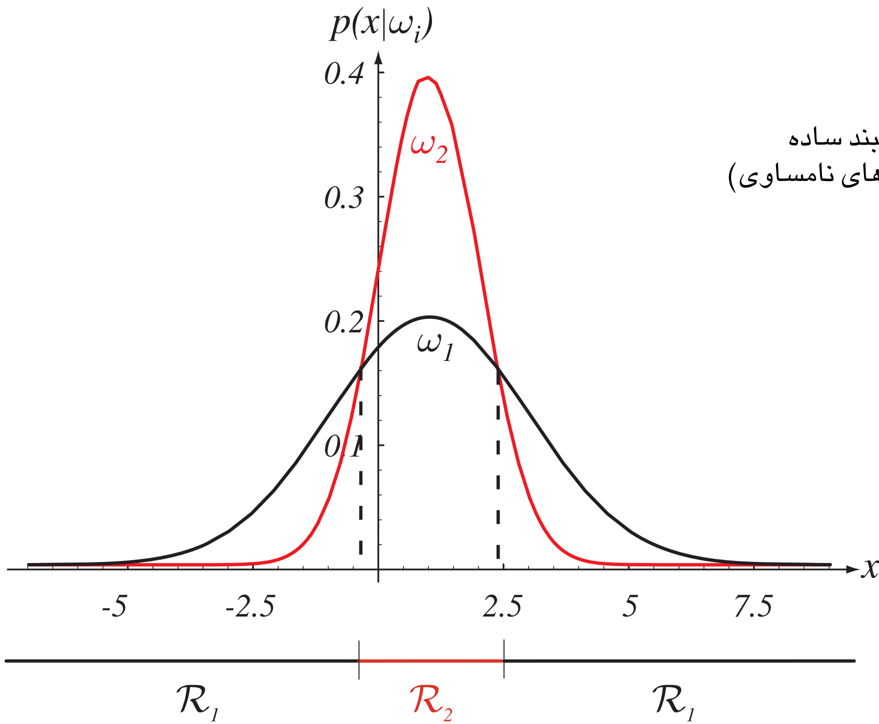
$$w_{i0} = -\frac{1}{2}\boldsymbol{\mu}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(w_i)$$

- Decision boundaries are hyperquadrics.

From: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification, John Wiley & Sons, Inc., 2001.

قاعده‌ی تصمیم بیزی

تابع چگالی گاوسی چندمتغیره: توابع تفکیک: حالت کلی

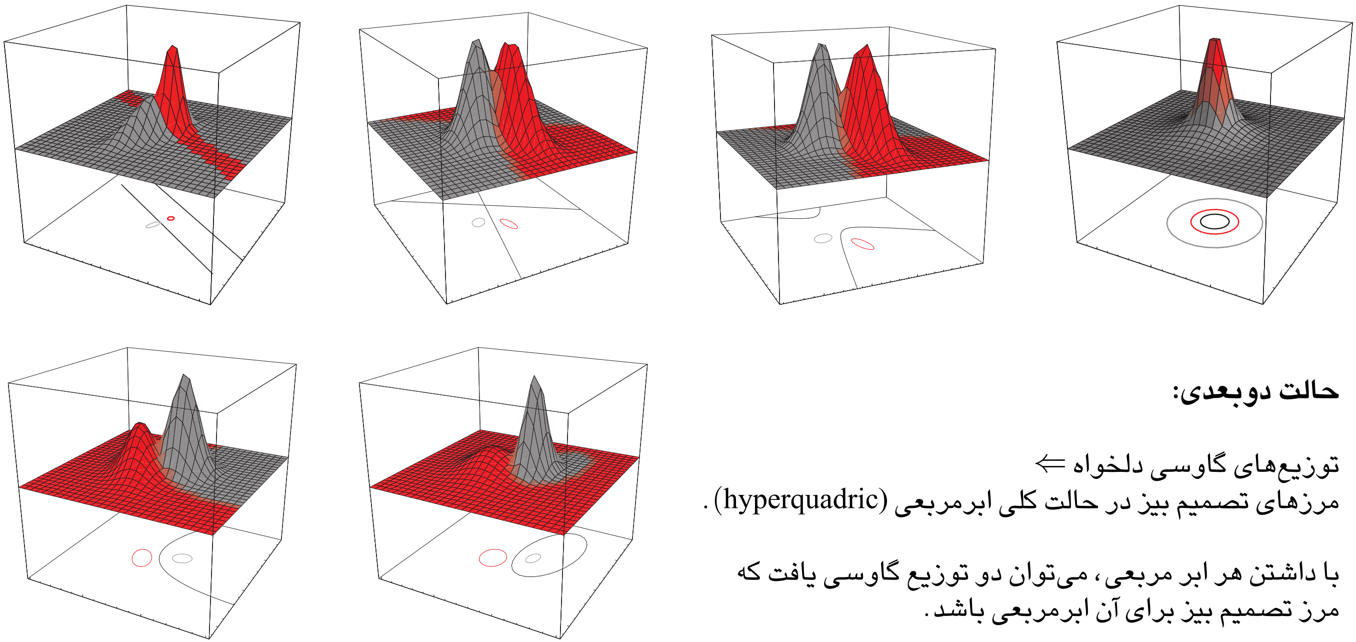


نواحی تصمیم غیر همبند ساده
(برای حالت واریانس‌های نامساوی)

From: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification, John Wiley & Sons, Inc., 2001.

قاعده‌ی تصمیم بیزی

تابع چگالی گاوسی چندمتغیره: توابع تفکیک: حالت کلی



حالت دوبعدی:

توزیع‌های گاوسی دلخواه $\Leftarrow$
مرزهای تصمیم بیز در حالت کلی ابرمربعی (hyperquadric).

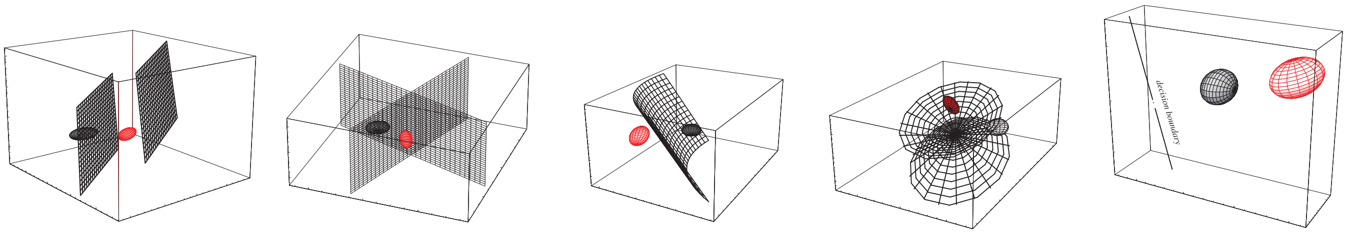
با داشتن هر ابر مربعی، می‌توان دو توزیع گاوسی یافت که
مرز تصمیم بیز برای آن ابرمربعی باشد.

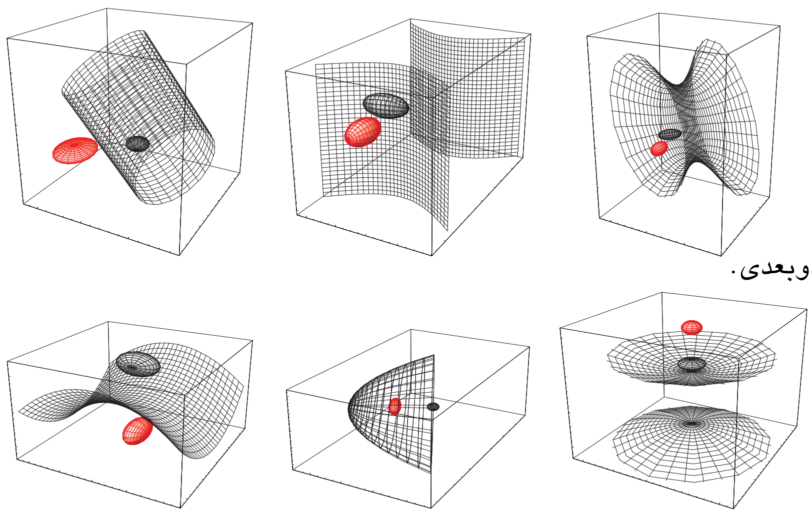
(واریانس‌ها توسط کانتورهای چگالی احتمال ثابت نشان داده شده‌اند.)

From: Richard O. Duda, Peter E. Hart, and David G. Stork,
Pattern Classification, John Wiley & Sons, Inc., 2001.

قاعده‌ی تصمیم بیزی

تابع چگالی گاوسی چندمتغیره: توابع تفکیک: حالت کلی





حالت سه‌بعدی:

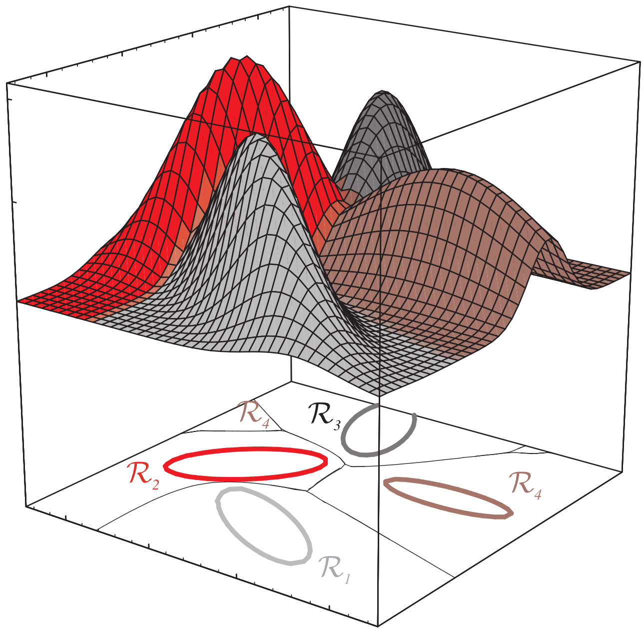
توزیع‌های گاوسی سه‌بعدی دلخواه ⇐
مرزهای تصمیم بیز: ابرمربعی (hyperquadric) دوبعدی.

حالت‌های تباهیده (degenerate cases) نیز
وجود دارد که مرز تصمیم یک **خط** است.

From: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification, John Wiley & Sons, Inc., 2001.

قاعده‌ی تصمیم بیزی

تابع چگالی گاوسی چندمتغیره: توابع تفکیک: حالت کلی



ناحیه‌های تصمیم برای چهار توزیع نرمال دوبعدی:
با وجود تعداد کم دسته‌ها، شکل نواحی مرز می‌تواند نسبتاً پیچیده باشد.

From: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification, John Wiley & Sons, Inc., 2001.

❖ **Multivariate Gaussian pdf**

$$p(\underline{x}|\omega_i) = \frac{1}{(2\pi)^{\frac{\ell}{2}}|\Sigma_i|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\underline{x}-\underline{\mu}_i)^{\mathrm{T}}\Sigma_i^{-1}(\underline{x}-\underline{\mu}_i)\right)$$

$$\underline{\mu}_i = E[\underline{x}] \qquad \ell \times \ell \text{ matrix in } \omega_i$$

$$\Sigma_i = E\left[(\underline{x}-\underline{\mu}_i)(\underline{x}-\underline{\mu}_i)^{\mathrm{T}}\right]$$

called covariance matrix

- ln(.) is monotonic. Define:

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

  - $$g_i(\underline{x}) = -\frac{1}{2}(\underline{x} - \underline{\mu}_i)^T \Sigma_i^{-1}(\underline{x} - \underline{\mu}_i) + \ln P(\omega_i) + C_i$$
    $$C_i = -(\frac{\ell}{2})\ln 2\pi - (\frac{1}{2})\ln|\Sigma_i|$$

  - Example: $$\Sigma_i = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

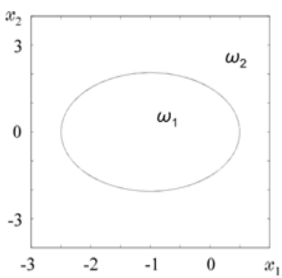
➢ $$g_i(\underline{x}) = -\frac{1}{2\sigma^2}(x_1^2 + x_2^2) + \frac{1}{\sigma^2}(\mu_{i1}x_1 + \mu_{i2}x_2) - \frac{1}{2\sigma^2}(\mu_{i1}^2 + \mu_{i2}^2) + \ln P(\omega_i) + C_i$$

That is, $g_i(x)$ is quadratic and the surfaces

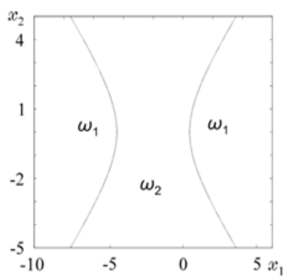
$$g_i(\underline{x}) - g_j(\underline{x}) = 0$$

quadrics, ellipsoids, parabolas, hyperbolas, pairs of lines.

For example:



(a)



(b)

❖ Decision Hyperplanes

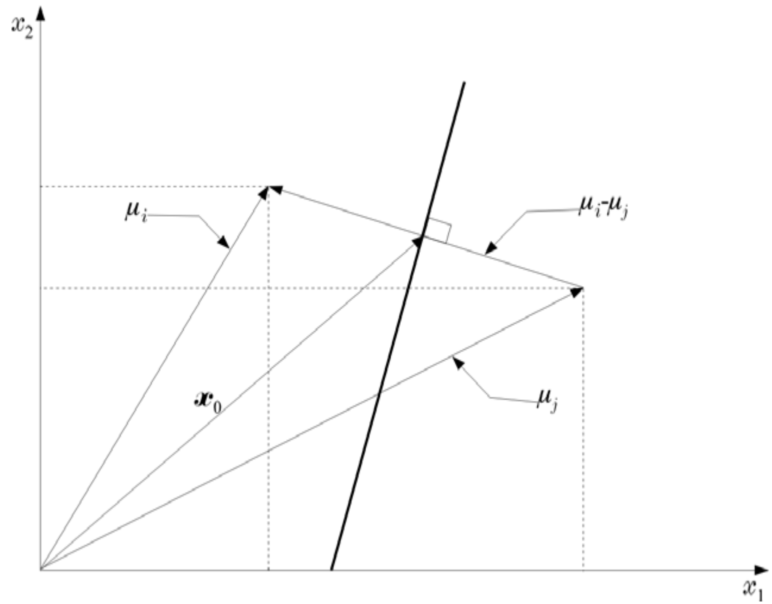
➢ Quadratic terms: $\underline{x}^T \Sigma_i^{-1} \underline{x}$

If ALL $\Sigma_i = \Sigma$ (the same) the quadratic terms are not of interest. They are not involved in comparisons. Then, equivalently, we can write:

$$\boxed{\begin{aligned} g_i(\underline{x}) &= \underline{w}_i^T \underline{x} + w_{io} \\ \underline{w}_i &= \Sigma^{-1} \underline{\mu}_i \\ w_{i0} &= \ln P(\omega_i) - \frac{1}{2} \underline{\mu}^T{}_i \Sigma^{-1} \underline{\mu}_i \end{aligned}}$$

Discriminant functions are LINEAR

- Let in addition:



$$\Sigma = \sigma^2 I. \text{ Then}$$

$$g_i(\underline{x}) = \frac{1}{\sigma^2} \underline{\mu}_i^T \underline{x} + w_{i0}$$

$$g_{ij}(\underline{x}) = g_i(\underline{x}) - g_j(\underline{x}) = 0$$

$$= \underline{w}^T (\underline{x} - \underline{x}_0)$$

$$\underline{w} = \underline{\mu}_i - \underline{\mu}_j,$$

$$\underline{x}_0 = \frac{1}{2}(\underline{\mu}_i + \underline{\mu}_j) - \sigma^2 \ln \frac{P(\omega_i)}{P(\omega_j)} \frac{\underline{\mu}_i - \underline{\mu}_j}{\left\| \underline{\mu}_i - \underline{\mu}_j \right\|^2}$$

- Nondiagonal: $\Sigma \neq \sigma^2 I$

$$\boxed{g_{ij}(\underline{x}) = \underline{w}^T(\underline{x} - \underline{x}_0) = 0}$$

$$\underline{w} = \Sigma^{-1}(\underline{\mu}_i - \underline{\mu}_j)$$

$$\underline{x}_0 = \frac{1}{2}(\underline{\mu}_i + \underline{\mu}_j) - \ell n\left(\frac{P(\omega_i)}{P(\omega_j)}\right)\frac{\underline{\mu}_i - \underline{\mu}_j}{\left\|\underline{\mu}_i - \underline{\mu}_j\right\|^2_{\Sigma^{-1}}}$$

$$\left\|\underline{x}\right\|_{\Sigma^{-1}} \equiv (\underline{x}^T \Sigma^{-1} \underline{x})^{\frac{1}{2}}$$

- Decision hyperplane
  - not normal to $\underline{\mu}_i - \underline{\mu}_j$
  - normal to $\Sigma^{-1}(\underline{\mu}_i - \underline{\mu}_j)$

❖ Minimum Distance Classifiers

- $P(\omega_i) = \frac{1}{M}$ **equiprobable**

- $g_i(\underline{x}) = -\frac{1}{2}(\underline{x} - \underline{\mu}_i)^T \Sigma^{-1} (\underline{x} - \underline{\mu}_i)$

- $\Sigma = \sigma^2 I :$ Assign $\underline{x} \rightarrow \omega_i :$

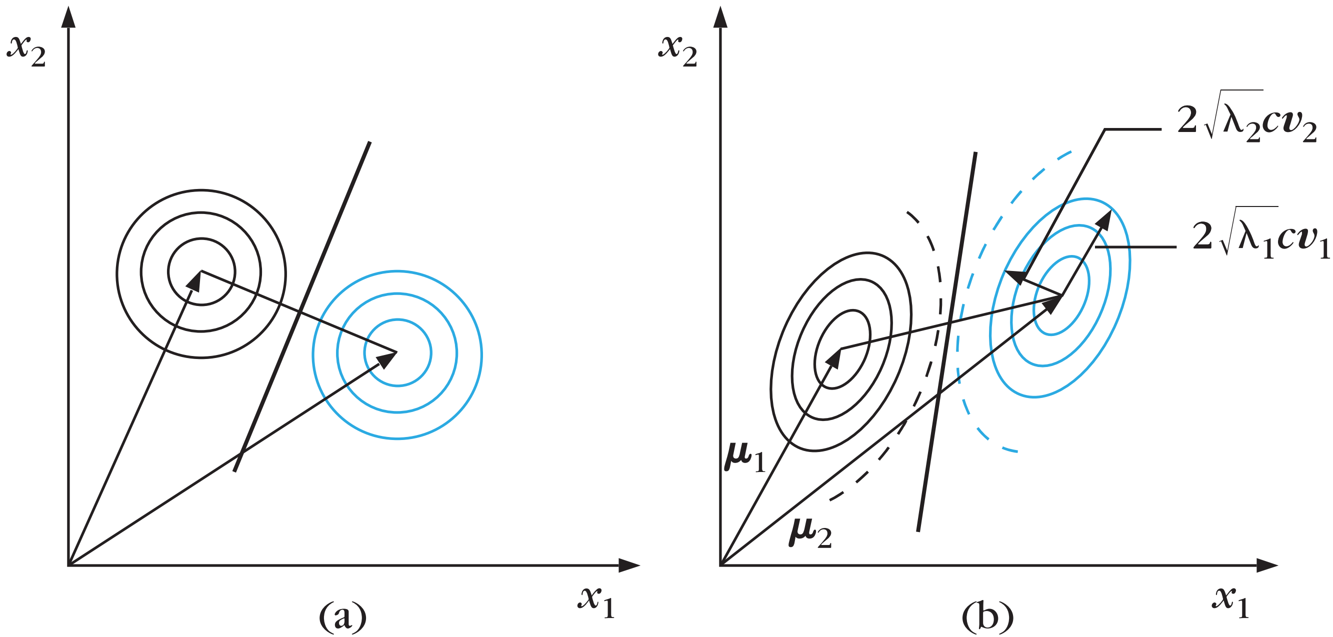
  Euclidean Distance: $d_E \equiv \left\| \underline{x} - \underline{\mu}_i \right\|$

  smaller

- $\Sigma \neq \sigma^2 I :$ Assign $\underline{x} \rightarrow \omega_i :$

  Mahalanobis Distance: $d_m = ((\underline{x} - \underline{\mu}_i)^T \Sigma^{-1} (\underline{x} - \underline{\mu}_i))^{\frac{1}{2}}$

  smaller



equal Euclidean distance

equal Mahalanobis distance

from the mean points of each class.

In the two-dimensional space, they are circles in the case of Euclidean distance and ellipses in the case of Mahalanobis distance. Observe that in the latter case the decision line is no longer orthogonal to the line segment joining the mean values. It turns according to the shape of the ellipses.

❖ Example:

Given ω_1, ω_2 : $P(\omega_1) = P(\omega_2)$ and $p(\underline{x}|\omega_1) = N(\underline{\mu}_1, \Sigma)$,

$$p(\underline{x}|\omega_2) = N(\underline{\mu}_2, \Sigma),\ \underline{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix},\ \underline{\mu}_2 = \begin{bmatrix} 3 \\ 3 \end{bmatrix},\ \Sigma = \begin{bmatrix} 1.1 & 0.3 \\ 0.3 & 1.9 \end{bmatrix}$$

classify the vector $\underline{x} = \begin{bmatrix} 1.0 \\ 2.2 \end{bmatrix}$ using Bayesian classification :

- $\Sigma^{-1} = \begin{bmatrix} 0.95 & -0.15 \\ -0.15 & 0.55 \end{bmatrix}$
  - Compute Mahalanobis d_m from μ_1, μ_2 : $d^2_{m,1} = [1.0, \quad 2.2]$

$$\Sigma^{-1} \begin{bmatrix} 1.0 \\ 2.2 \end{bmatrix} = 2.952,\ d^2_{m,2} = [-2.0, \quad -0.8]\, \Sigma^{-1} \begin{bmatrix} -2.0 \\ -0.8 \end{bmatrix} = 3.672$$

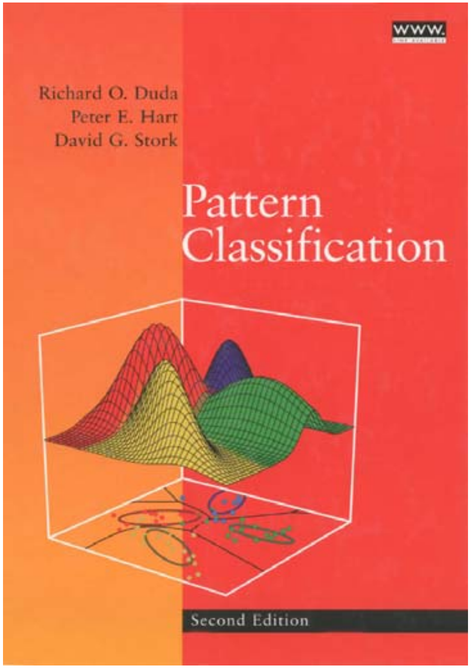
- Classify $\underline{x} \rightarrow \omega_1$. Observe that $d_{E,2} < d_{E,1}$

طبقه‌بندی مبتنی بر نظریه‌ی تصمیم بیز:
طبقه‌بندی بیزی برای توزیع‌های نرمال

۶

منابع

منبع اصلی



R.O. Duda, P.E. Hart, and D.G. Stork,
Pattern Classification,
Second Edition, John Wiley & Sons, Inc., 2001.

Chapter 2

CHAPTER 2

BAYESIAN DECISION THEORY

2.1 INTRODUCTION

Bayesian decision theory is a fundamental statistical approach to the problem of pattern classification. This approach is based on quantifying the tradeoffs between various classification decisions using probability and the costs that accompany such decisions. It makes the assumption that the decision problem is posed in probabilistic terms, and that all of the relevant probability values are known. In this chapter we develop the fundamentals of this theory and we show how it can be viewed as being simply a formalization of common-sense procedures; in subsequent chapters we will consider the problems that arise when the probabilistic structure is not completely known.

While we will give a quite general, abstract development of Bayesian decision theory in Section 2.2, we begin our discussion with a specific example. Let us reconsider the hypothetical problem posed in Chapter 1 of designing a classifier to separate two kinds of fish: sea bass and salmon. Suppose that an observer watching fish arrive along the conveyor belt finds it hard to predict what type will emerge next and that the sequence of types of fish appears to be random. In decision-theoretic terminology we would say that as each fish emerges nature is in one or the other of the two possible states: Either the fish is a sea bass or the fish is a salmon. We let ω denote the *state of nature*, with $\omega = \omega_1$ for sea bass and $\omega = \omega_2$ for salmon. Because the state of nature is so unpredictable, we consider ω to be a variable that must be described probabilistically.

STATE OF NATURE

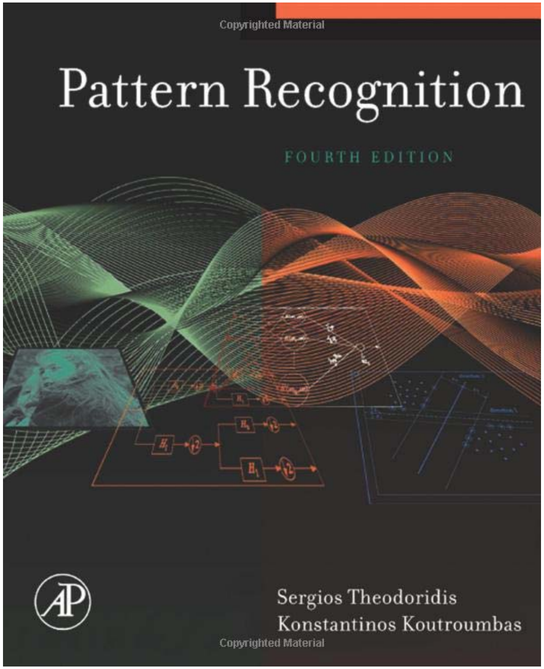
If the catch produced as much sea bass as salmon, we would say that the next fish is equally likely to be sea bass or salmon. More generally, we assume that there is some *a priori probability* (or simply *prior*) $P(\omega_1)$ that the next fish is sea bass, and some prior probability $P(\omega_2)$ that it is salmon. If we assume there are no other types of fish relevant here, then $P(\omega_1)$ and $P(\omega_2)$ sum to one. These prior probabilities reflect our prior knowledge of how likely we are to get a sea bass or salmon before the fish actually appears. It might, for instance, depend upon the time of year or the choice of fishing area.

PRIOR

Suppose for a moment that we were forced to make a decision about the type of fish that will appear next without being allowed to see it. For the moment, we shall assume that any incorrect classification entails the same cost or consequence, and

20

منبع اصلی



S. Theodoridis, K. Koutroumbas,
Pattern Recognition,
Fourth Edition, Academic Press, 2009.

Chapter 2

CHAPTER

Classifiers Based on Bayes Decision Theory 2

2.1 INTRODUCTION

This is the first chapter, out of three, dealing with the design of the classifier in a pattern recognition system. The approach to be followed builds upon probabilistic arguments stemming from the statistical nature of the generated features. As has already been pointed out in the introductory chapter, this is due to the statistical variation of the patterns as well as to the noise in the measuring sensors. Adopting this reasoning as our kickoff point, we will design classifiers that classify an unknown pattern in the most probable of the classes. Thus, our task now becomes that of defining what "most probable" means.

Given a classification task of M classes, $\omega_1, \omega_2, \ldots, \omega_M$, and an unknown pattern, which is represented by a feature vector $\boldsymbol{x}$, we form the M conditional probabilities $P(\omega_i|\boldsymbol{x}), i = 1, 2, \ldots, M$. Sometimes, these are also referred to as *a posteriori probabilities*. In words, each of them represents the probability that the unknown pattern belongs to the respective class ω_i, given that the corresponding feature vector takes the value $\boldsymbol{x}$. Who could then argue that these conditional probabilities are not sensible choices to quantify the term *most probable*? Indeed, the classifiers to be considered in this chapter compute either the maximum of these M values or, equivalently, the maximum of an appropriately defined function of them. The unknown pattern is then assigned to the class corresponding to this maximum.

The first task we are faced with is the computation of the conditional probabilities. The Bayes rule will once more prove its usefulness! A major effort in this chapter will be devoted to techniques for estimating probability density functions (pdf), based on the available experimental evidence, that is, the feature vectors corresponding to the patterns of the training set.

2.2 BAYES DECISION THEORY

We will initially focus on the two-class case. Let ω_1, ω_2 be the two classes in which our patterns belong. In the sequel, we assume that the *a priori probabilities*

13