بازشناسی آماری الگو

جلسه سوم

Bayesian Decision Theory

فرآیند طراحی کلاسیفایر

توابع جدا ساز

فهرست مطالب

- مقدمه و یاد آوری
- رویکرد های اصلی طبقه بندی
  - بیشترین نمره (Bayesian)
  - کمترین فاصله
- لحاظ کردن ریسک و خطا
  - تفاوت اقدام و تصمیم
- طراحی کلاسیفایر به شکل یک فرآیند بهینه سازی
- معرفی توابع جدا ساز
  - چند نونه از توابع جدا ساز خطی
- خلاصه بحث

مقدمه

- در جلسات گذشته با مفاهیم اصلی SPR و مبانی طراحی کلاسیفایر (نوعی فرآیند بهینه سازی) آشنا شدیم.
- از مهمترین نکات ذکر شده آن بود که در رویکرد آماری، قرار است تمام اطلاعات اساسی کلاس ها به صورت چگونگی توزیع اعضای آنها(توابع توزیع) داده شود.
- در این جلسه، هدف آن است که ساختاری کلی برای طراحی کلاسیفایر با داشتن این اطلاعات آماری ارائه شود.
- به تعبیر دیگر، در این جلسه فرض می شود که اعضای هر کلاس، متغیر های اتفاقی هستند که توزیع احتمال آنها مشخص است. حال بر این اساس می خواهیم کلاسیفایر را با هدف مشخص(مثلا حد اقل کردن احتمال وقوع خطای انتساب) طرح کنیم

معیار های اساسی طراحی کلاسیفایر

- عموما دو نوع رویکرد برای معیارهای طراحی کلاسیفایر وجود دارد که عبارتند از رویکرد **کسب امتیاز** و رویکرد **کسب فاصله**
- رویکرد کسب امتیاز: در این رویکرد با توجه به موارد مشاهده شده وخصوصیات کلاس ها، هر کلاس برای داشتن هر نمونه امتیاز یا نمره ای کسب می کند.
  - این نمره در SPR مرتبط با احتمال عضویت در کلاس خواهد بود.
- رویکرد کسب فاصله: در این رویکرد، هر کلاس شاخص ها و نمایندگانی دارد و فاصله هر نمونه باید با آنها سنجیده شده و نمونه به کلاسی نسبت داده می شود که کمترین فاصله با نماینده آن کلاس را دارد
- هر دوی این رویکرد ها باید به نوعی فرآیند بهینه سازی تبدیل شودکه طی آن پارامتر های کلاسیفایر به منظور بهینه کردن معیار های یاد شده تنظیم گردند.

تئوری طبقه بندی Bayes

- در کلاسیفایر Bayes معیار مورد توجه، احتمال وقوع خطای انتساب یک نمونه است
- ایده طبقه بندی آماری بر این قانون بنا می شود که باید یک نمونه را به کلاسی نسبت دهیم که احتمال وقوع خطا(یا ریسک) حداقل شود.
- با توجه به موارد ذکر شده، بدیهی است احتمال وقوع خطا، بستگی به توابع احتمال عضویت نمونه ها در هر یک ازکلاس ها دارد
- لذا نیاز مندیم احتمال عضویت نمونه در هر کلاس را داشته باشیم.
- در تئوری Bayes همه این اطلاعات، تبدیل به توابع احتما ل شده و در قالب یک ساختار ریاضی با هم تر کیب می شوند.
- این ساختار ریاضی، فرمول Bayes برای یافتن احتمال شرطی یا مرکب است.

فرمول Bayes:

- در بحث احتمال عضویت، دو نوع احتمال عضویت، یعنی احتمال بدون اطلاعات(مشاهده) یعنی احتمال پیشین و احتمال همراه با اطلاعات(مشاهده) یا احتمال پسین در نظر گرفته می شود
- بدیهی است مبنای تصمیم گیری، احتمال پسین خواهد بود ولی این احتمال مستقیما وجود ندارد و فرمول Bayes برای محاسبه احتمال پسین تدوین گردیده است و می گوید:

Posterior = (Likelihood×Prior) / Evidence

$$P(\omega_i \mid x) = \frac{p(x \mid \omega_i) \times P(\omega_i)}{P(x)}$$

تئوری طبقه بندی Bayes...

- در این فرمول چهار مفهوم وجود دارد که عبارتند از:

1. احتمال پیشین عضویت prior probability هر نمونه در یکی از کلاس هاکه مستقل از مشاهدات و اندازه گیری های ما بوده و جزو خصوصیات مسئله است. $P(\omega_j)$ مانند:
   1. احتمال آمدن شیر در یک آزمایش پرتاب سکه ۵۰٪ است.
   2. احتمال اینکه یک انسان مرد باشد ۵۰٪ است.
   3. احتمال اینکه یک انسان چپ دست باشد ۲۰٪ است.
2. تشابه likelihood(تکرار پذیری) که با توزیع اعضای یک کلاس توصیف می شود. $p(x \mid \omega_j)$
   - به عبارتی می گوید توزیع احتمال وقوع اعضای هر کلاس در نواحی مختلف فضای مشخصه ها چگونه است.

تئوری طبقه بندی Bayes...

3. شهود Evidence که به احتمال خود مدرک اشاره دارد
   - احتمال اینکه یک مشاهده (بدست آوردن شواهد و مدارک) صورت بگیرد (در بحث ما اغلب این احتمال ۱۰۰٪ است) $P(x)$
   - در یک مسئله C کلاسه این شهود از رابطه زیر قابل محاسبه است:

$$P(x) = \sum_{i=1}^{c} (p(x \mid \omega_j).P(\omega_j))$$

4. احتمال پسین posterior probability احتمال عضویت یک نمونه در یک کلاس را پس از مشاهده مدارک واستفاده از آنها به ما می دهد. $P(\omega_j \mid x)$ h

- اگر نمونه به کلاسی نسبت داده شود که در بین کلاس ها، دارای بیشترین احتمال پسین عضویت است، بدیهی است که احتمال وقوع خطای انتساب حداقل خواهد شد.

تئوری طبقه بندی Bayes...

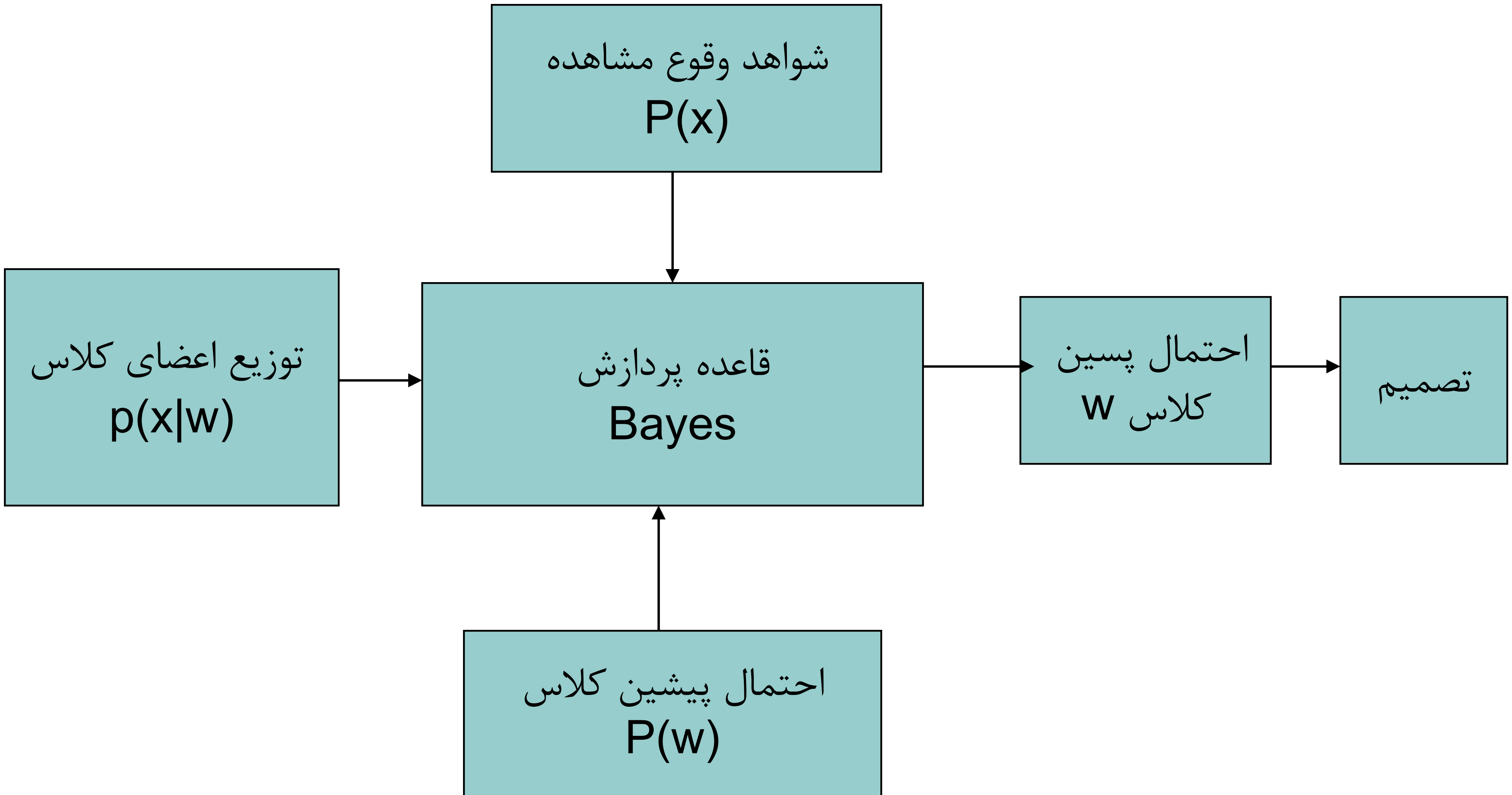
حال (در یک مسئله دو کلاسه) با یک مشاهده (pattern) مانند x پس از محاسبه احتمال پسین وقوع کلاس های ۱ و ۲ را محاسبه نموده و تصمیم گیری به شرح زیر صورت می پذیرد:

if $P(\omega_1 \mid x) > P(\omega_2 \mid x)$ ⇒ True state of nature = ω_1

if $P(\omega_1 \mid x) < P(\omega_2 \mid x)$ ⇒ True state of nature = ω_2

- یادآوری مهم: در روش آماری برای تصمیم گیری، همه اطلاعاتی که از اندازه گیری ها و استدلال ها و محاسبات حین پردازش بدست آمده اند بایستی تبدیل به کمیت های بیان کننده احتمال و یا چگالی توزیع آن شوند. و هر نوع اطلاعتی باعث افزایش و یا کاهش مقادیر توابع احتمال در کلاس ها خواهد شد.

ماشین Bayes


شواهد وقوع مشاهده
P(x)
توزیع اعضای کلاس
p(x|w)
قاعده پردازش
Bayes
احتمال پسین
کلاس w
تصمیم
احتمال پیشین کلاس
P(w)

استراتژی Bayes درطراحی کلاسیفایر بهینه

- حداقل نمودن خطای طبقه بندی:

- Decide ω_1 if $P(\omega_1 \mid x) > P(\omega_2 \mid x)$;
otherwise ω_2

- بنابراین:

$$P(error \mid x) = min\ [P(\omega_1 \mid x), P(\omega_2 \mid x)]$$

(*Bayes decision*)

نتیجه مهم:

قانون Bayes به حداقل کردن خطای انتساب منجر خواهد شد.

خلاصه روش Bayes :

- اگر بدانیم که کلاس ها، دارای چگونه چگالی توزیعی هستند با داشتن :
  - یک ورودی که موقعیت pattern در فضای مشخصه ها را نشان می دهد
  - احتمال پیشین هر کلاس
  - احتمال مشاهده شواهد(X)(که در بحث ما این احتمال یک خواهد بود)
- می توان احتمال پسین عضویت نمونه در هر یک از کلاس ها را محاسبه کرد.
- بر اساس احتمال پسین عضویت نمونه در کلاس ها، برای انتساب نمونه تصمیم گیری کرد.

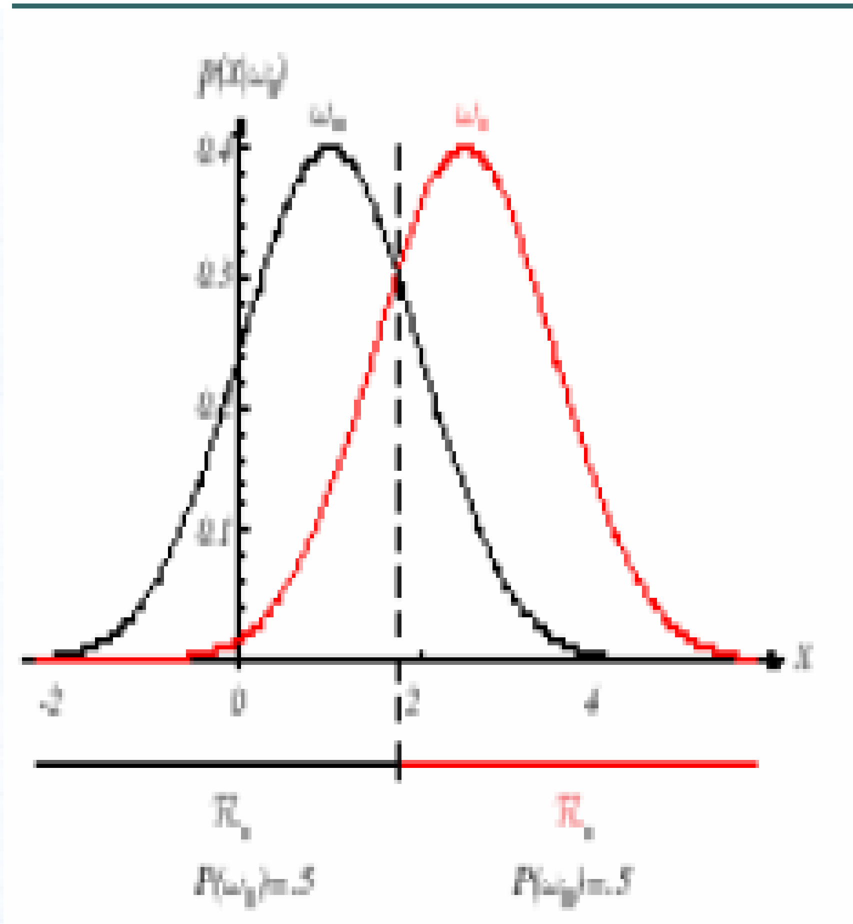
مثال: طراحی classifier بر اساس حد اقل کردن خطا یک بعد، دو کلاس، تصمیم گیری بدون مشاهدات

- فرض کنیم یک مسئله دو کلاسه داریم $c=2$
- ورودی ها یک بعدی هستند $d=1$
- احتمال پیشین کلاس ها: برای $w1$ ۳۰٪ و برای $w2$ ۷۰٪
- نمونه آمده ولی هیچ اندازه گیری روی آن صورت نگرفته است
- در این شرایط رتبه بندی کلاس ها، از حیث احتمال های پیشین و پسین یکسان شمرده می شوند
- هر ورودی را (به اصطلاح ندید) به کلاس ۲ نسبت می دهیم که احتمال پیشین بزرگتری دارد. به این ترتیب در ۷۰٪ موارد درست عمل کرده ایم.
  - خطا زمانی رخ خواهد داد که نمونه متعلق به کلاس ۱ باشد که احتمال آن ۳۰٪ می باشد.

مثال: طراحی classifier در شرایط وجود اندازه گیری

- منظور از طرح classifier یافتن قانون تصمیم یا مرز تصمیم است.
- می توان بجای استفاده از احتمال پسین، ازچگالی توزیع برای تصمیم گیری استفاده کرد. (تفاوت در یک ضریب است)
- با داشتن مقادیر اندازه گیری(X)و تابع توزیع، می توان مقدار چگالی احتمال $p(x|w)$ برای هرکلاس در نقطه اندازه گیری را بدست آورد
- در صورتی که احتمالات پیشین ، متفاوت باشد، منحنی های توزیع کلاس ها با یک فاکتور scaling تغییر می یابد.
- حال منحنی های حاصل را با هم مقایسه کرده و محل تلاقی منحنی ها با هم، مرز تصمیم را پیدا می کند.

مثال



- دو کلاس با توزیع نرمال دارای دو مقدار متوسط متفاوت ولی واریانس های یکسان (مطابق شکل)
- مرز تصمیم: موقعیتی که مقدار دو توزیع با هم مساوی است
- نواحی تصمیم(مثلا) سمت چپ نقطه تلاقی، ناحیه کلاس اول و سمت راست ناحیه کلاس دوم
- قانون تصمیم(طراحی کلاسیفایر): اگر ورودی X در سمت چپ مرز تصمیم است آن را به کلاس اول نسبت بده
- احتمال خطا: سطح زیر ناحیه همپوشانی

چند نکته

- بدیهی است اگر نوع توابع توزیع معلوم باشد، میزان ناحیه همپوشانی آنها، تعیین کننده احتمال وقوع خطا خواهد بود.
- در عین حال، می توان مسئله را به شکل معکوس نگریست. یعنی برای توابع توزیع احتمال معلوم می توان از روی حدود خطا، دریافت که همپوشانی چقدر بوده و به تبع آن می توان فاصله و مقدار پارامتر های توابع توزیع را پیدا کرد.
- این امر در بررسی کارآیی دیگر انواع کلاسیفایر ها می تواند کاربرد داشته باشد.

تعریف مسئله(هفته بعد)

- دو کلاس از داده های اتفاقی یک بعدی با توزیع نرمال با واریانس یکسان طرح شود که احتمال وقوع یکسان داشته باشند به گونه ای که خطای جدا سازی آنها به طور تئوریک 15 ٪ باشد.
  - (فاصله مقادیر متوسط دو تابع)
- دو کلاس از داده های اتفاقی یک بعدی با توزیع ۲ به ۱ با احتمال خطای Bayes 15%

Bayesian Decision Theory – Continuous Features

- تعمیم قانون Bayes :
  - استفاده از تعداد بیشتری از مشخصه ها (فضای چند بعدی)
  - وجود تعداد بیشتری از کلاس ها (مسئله چند کلاسه)
  - در نظر گرفتن پارامتر های دیگری پیش از تصمیم (مبادرت نکردن به تصمیم بلافاصله پس از محاسبه یا تخمین احتمال پسین....)
  - استفاده از معیار "هزینه" loss به جای صرف خطا برای طراحی تصمیم
- چگونگی محاسبه مقدار خطا؟
  - روش تحلیلی
  - روس تجربی

تغییر استراتژی:
اعمال فاکتور ریسک

- اگر فرض کنیم برای اقدامات ناشی از تصمیم های انتساب کلاس های مختلف $\{\omega_1, \omega_2, \ldots, \omega_c\}$، $\{\alpha_1, \alpha_2, \ldots, \alpha_a\}$ هزینه (ریسک) متفاوتی داشته باشیم $\lambda(\alpha_i \mid \omega_j)$، بدیهی است، برای تصمیم های خود در مورد مثال های مختلف می توان هزینه نهایی را به شرح زیر داشت:

$$R(\alpha_i \mid x) = \sum_{j=1}^{j=c} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid x)$$

- مرز تصمیم در این شرایط با مینیمم کردن تابع هزینه فوق به دست می آید.

مثال : کلاسیفایر دو کلاسه:

- فرض کنید برای یک مسئله داشته باشیم:

$\lambda_{ij} = \lambda(\alpha_i \mid \omega_j)$ α_2 : deciding ω_2 α_1 : deciding ω_1

λ_{ij} در این جا ضریب ریسک انتخاب کلاس i است اگر کلاس صحیح j باشد

در این حالت تابع ریسک عبارت خواهد بود از:

$$R(\alpha_1 \mid x) = \lambda_{11}P(\omega_1 \mid x) + \lambda_{12}P(\omega_2 \mid x)$$
$$R(\alpha_2 \mid x) = \lambda_{21}P(\omega_1 \mid x) + \lambda_{22}P(\omega_2 \mid x)$$

و مبنای تصمیم گیری ریسک یا هزینه کمتر خواهد بود:

decide ω_1 if:

$$(\lambda_{21}- \lambda_{11})\, P(x \mid \omega_1)\, P(\omega_1) > (\lambda_{12}- \lambda_{22})\, P(x \mid \omega_2)\, P(\omega_2)$$

and ω_2 otherwise

تاثیرتوام احتمال های پیشین و فاکتور های ریسک

با تشکیل نسبت شباهتLikelihood ratio و مقایسه آن با ترشهلدی که از روی نسبت احتمال های پسین و ضرایب ریسک بدست می آید، تصمیم گیری صورت می گیرد:

$$if \ \frac{P(x|\omega_1)}{P(x|\omega_2)} > \frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}} \cdot \frac{P(\omega_2)}{P(\omega_1)}$$

α_1 (decide ω_1)

Otherwise α_2 (decide ω_2)

Optimal decision property
اساس تصمیم گیری بهینه:

- دیدیم که مرز تصمیم بهینه توسط انتخاب یک ترشهلد روی نسبت شباهت ها بدست می آید.
- در موارد عملی که:
  - توابع توزیع احتمال و نسبت آنها بصورت فرمولی موجود نباشد
  - یا فرمول آنها خیلی پیچیده باشد

برای یافتن مرز تصمیم باید از جستجو روی عده ای از مثال ها یا نمونه ها استفاده کرد

- به این ترتیب: مرز تصمیم در یک فرآیند بهینه سازی روی داده های یادگیری حاصل می شود که به این فاز پروسه، آموزش یا یادگیری می گویند.
- محاسبه میزان خطا یا ریسک نیز ممکن است بطور تحلیلی یا با آزمایش کلاسیفایر روی عده دیگری از نمونه ها (نمونه های تست) صورت گیرد و بر اساس آن میزان قدرت تعمیم generalization محاسبه می شود.

مثال(تکلیف):
یافتن و رسم مرز تصمیم و میزان خطای کلاسیفایر به صورت تئوری

دو کلاس از بردار های اتفاق $\{\omega_1, \omega_2\}$ با توزیع نرمال با پارامتر های زیر موجود است:

$P(x \mid \omega_1)$ → N(2, 0.5) (Normal distribution)

$P(x \mid \omega_2)$ → N(1.5, 0.2)

نسبت تعداد داده ها (احتمال پیشین) و هزینه انتساب به صورت زیر است:

$P(\omega_1) = 2/3$

$P(\omega_2) = 1/3$

$$\lambda = \begin{bmatrix} 1 & 2 \\ 3 & 1.5 \end{bmatrix}$$

مرز تصمیم (با رسم شکل) و میزان خطای کلاسیفایر ناشی از آن را بدست آورید

شرایط معادل بودن استراتژی بهینه سازی خطا و ریسک

- در اکثر کاربرد ها، ضرایب هزینه های انتساب به شکل صفر و یک است:

$$\lambda(\alpha_i, \omega_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \qquad i, j = 1, \ldots, c$$

بنابر این هزینه شرطی تصمیم عبارت خواهد بود از:

$$R(\alpha_i \mid x) = \sum_{j=1}^{j=c} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid x)$$

$$= \sum_{j \neq 1} P(\omega_j \mid x) = 1 - P(\omega_i \mid x)$$

در این حالت دو استراتژی مینیمم کردن ریسک و خطا یکسان هستند.

نواحی تصمیم در حالت هزینه های باینری:

$$Let\ \frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}}.\frac{P(\omega_2)}{P(\omega_1)}=\theta_\lambda\ then\ decide\ \omega_1\ \ if:\ \frac{P(x|\omega_1)}{P(x|\omega_2)}>\theta_\lambda$$

- اگر ضرایب هزینه عبارت باشد از:

$$\lambda = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$then\quad \theta_\lambda = \frac{P(\omega_2)}{P(\omega_1)} = \theta_a$$

$$if\quad \lambda = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}\ then\quad \theta_\lambda = \frac{2P(\omega_2)}{P(\omega_1)} = \theta_b$$

مثال بررسی نقش تابع هزینه در نواحی تصمیم
مثال جدا سازی ماهی ها با استفاده از روشنی رنگ
سطح زیر منحنی برای هر کلاس ۱ است

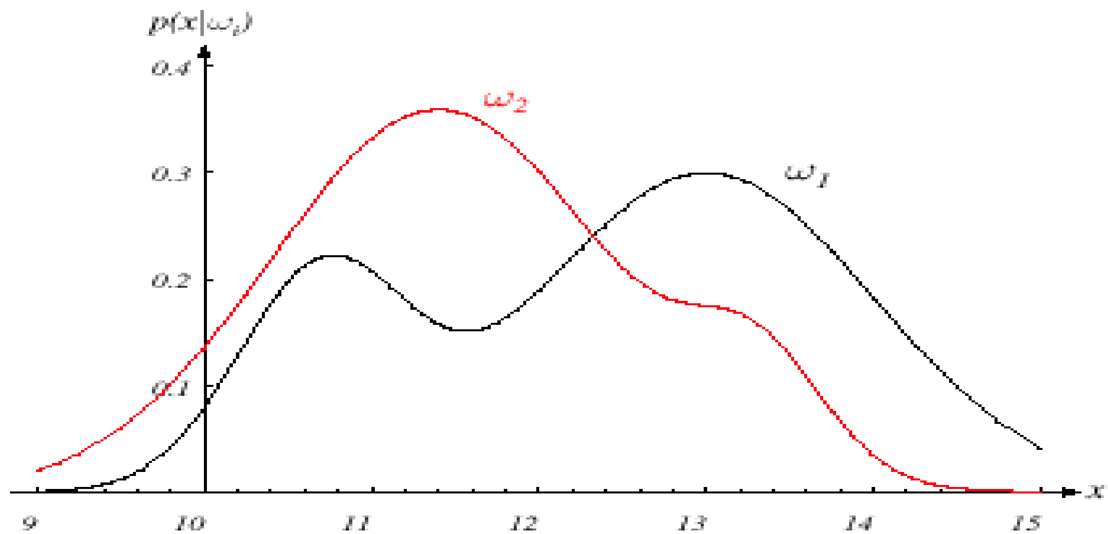


FIGURE 2.1. Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value x given the pattern is in category ω_i. If x represents the lightness of a fish, the two curves might describe the difference in lightness of populations of two types of fish. Density functions are normalized, and thus the area under each curve is 1.0. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

بررسی مسئله ماهی ها با استفاده از نسبت تشابه با دو ترشهلد مختلف

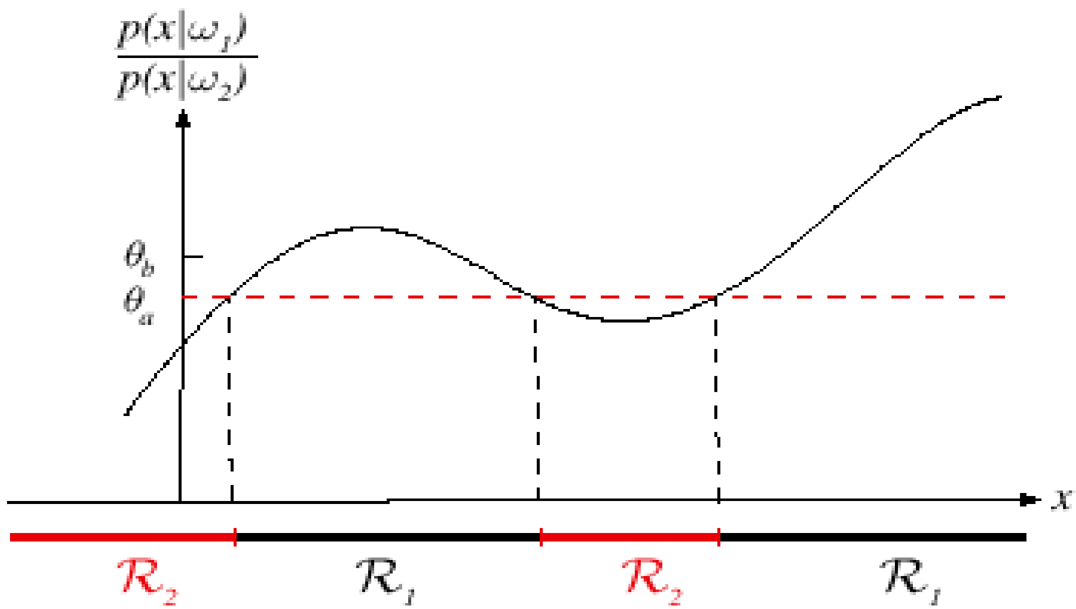


FIGURE 2.3. The likelihood ratio $p(x|\omega_1)/p(x|\omega_2)$ for the distributions shown in Fig. 2.1. If we employ a zero-one or classification loss, our decision boundaries are determined by the threshold θ_a. If our loss function penalizes miscategorizing ω_2 as ω_1 patterns more than the converse, we get the larger threshold θ_b, and hence $\mathcal{R}_1$ becomes smaller. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

حالت های مهم در توزیع گوسی چند بعدی (با فرض مساوی بودن ماتریس برای دو کلاس)

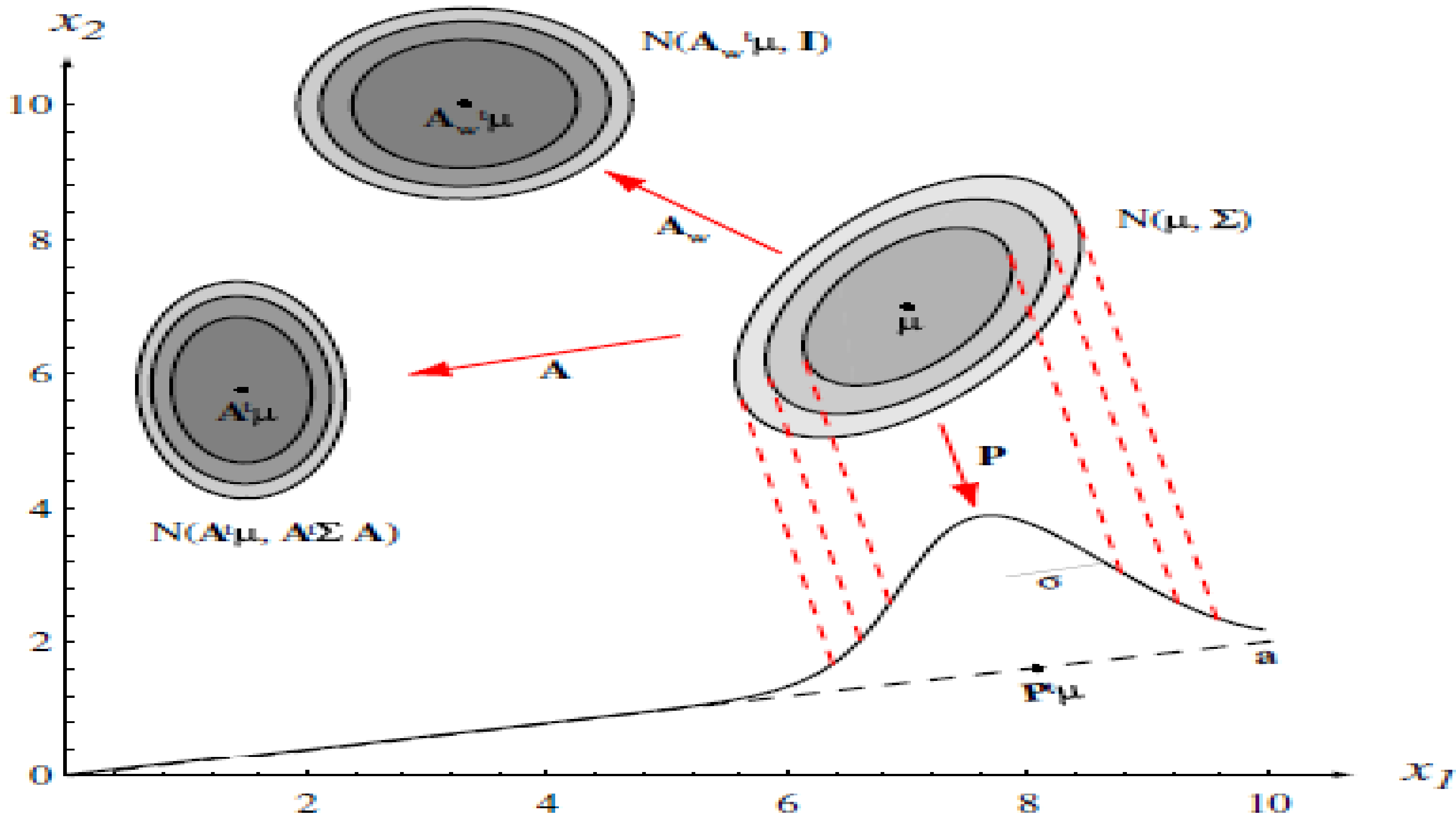
- حالت متقارن و مستقل $\Sigma_i = \sigma^2.I$
  - در این حالت اولا ارزش تفکیک کنندگی همه مشخصه ها یکسان است (پراکندگی همه مشخصه ها یکی است) و ثانیا مشخصه ها از هم کاملا مستقل هستند.
  - مبنای تفکیک می تواند فاصله اقلیدسی (یا جذر آن) از بدار های متوسط μ باشد
- حالت قطری بودن ماتریس کوواریانس
  - مشخصه ها تاثیر متفاوتی در تفکیک دارند(وزن متفاوت برای مشخصه ها) ولی از هم مستقل هستند
  - مبنای تفکیک می تواند فاصله ماهانالوبیس باشد. (فاصله وزن دار مشخصه ها از هم)

$$r^2 = (\mathrm{x} - \mu)^t \Sigma^{-1} (\mathrm{x} - \mu)$$

حالت های مهم در توزیع گوسی چند بعدی (با فرض مساوی بودن ماتریس برای دو کلاس)......

- اختیاری بودن ماتریس کوواریانس
  - در این حالت مشخصه ها از هم مستقل نیستندو ارزش یکسانی هم ندارند
  - تجزیه و تحلیل در این مورد چندان ساده نیست.
  - ساده ترین روش برای این کار آن است که یک تغییر (خطی) روی مشخصه ها ایجاد شود به گونه ای که در مشخصه های جدید ماتریس کوواریانس قطری باشد
  - یک تبدیل خطی (مثل A)روی مشخصه هائی که توزیع نرمال دارند، سبب ایجاد مشخصه های خطی جدیدی که آنها هم توزیع نرمال دارند می شود. البته پارامتر های توزیع جدید (Σو μ) تغییر میکند

تغيير پرامتر های توزيع نرمال با تبديل خطی مشخصه ها



Discreminent functions توابع جدا ساز

- در حالات چند کلاسه:
  - هدف طرح مجموعه توابعی است که هر یک منسوب به یک کلاس بوده و با توجه به مقدار آنها برای هر نمونه بتوان تصمیم گیری را برای انتساب آن نمونه انجام داد $g_i(\underline{x}),\ i = 1,\ldots,c$

- Assign **$\underline{x}$** to class ω_i

  if:

  $g_i(\underline{x}) > g_j(\underline{x})\ \ \forall j \neq i$

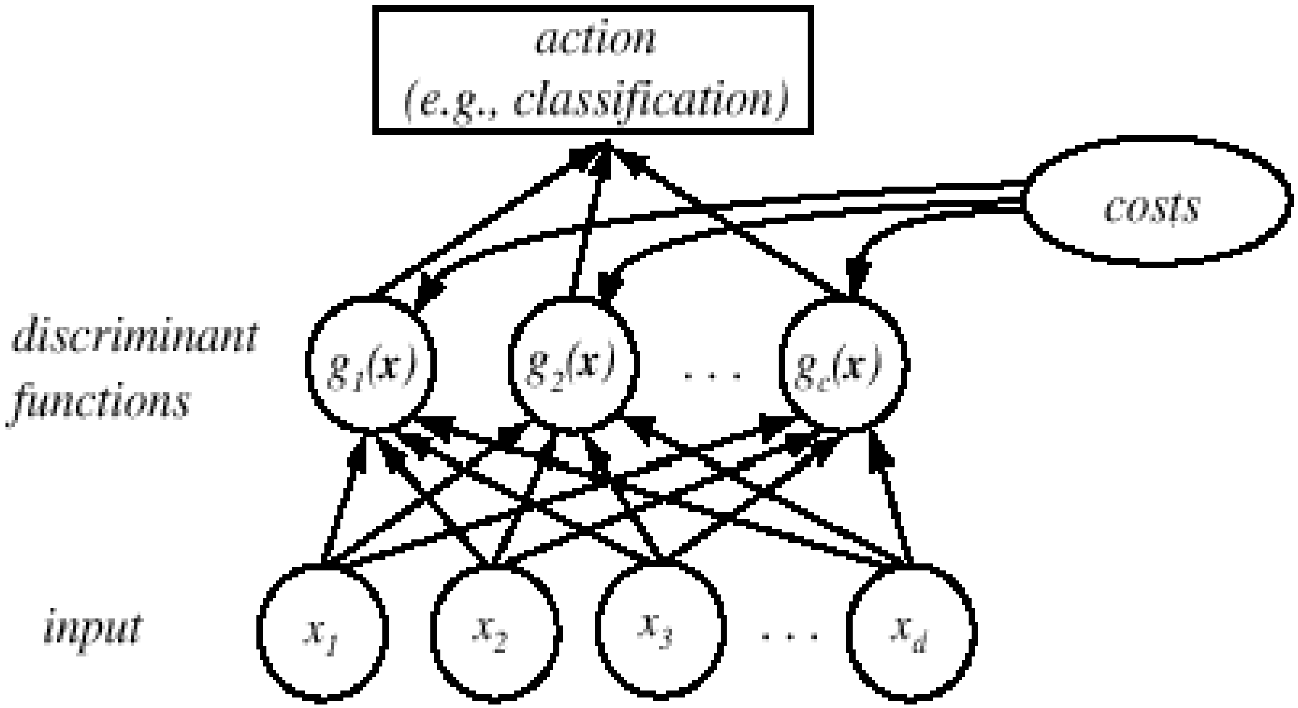


FIGURE 2.5. The functional structure of a general statistical pattern classifier which includes d inputs and c discriminant functions $g_i(\mathbf{x})$. A subsequent step determines which of the discriminant values is the maximum, and categorizes the input pattern accordingly. The arrows show the direction of the flow of information, though frequently the arrows are omitted when the direction of flow is self-evident. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

توابع جدا ساز ...

- توسط توابع جدا ساز می توان فضای مشخصه ها را به n ناحیه تقسیم کرد که در هر ناحیه، مقدار تابع جدا ساز مر بوط به یکی از کلاس ها، بیشتر(یا کمتر) از دیگران است یعنی:

$$\text{if } g_i(x) > g_j(x) \;\; \forall j \neq i \text{ then } x \text{ is in } \mathcal{R}_i$$

در حالت دو کلاسه با توجه به اینکه تنها دو تابع جدا ساز داریم می توان آن را باهم ترکیب نموده و به قانون ساده تری رسید:

$$g(x) \equiv g_1(x) - g_2(x)$$

$$\text{Decide} \quad \begin{cases} \omega_1 & \text{if } g(x) > 0\,; \\ \omega_2 & \text{Otherwise} \end{cases}$$

گزینه های معمول برای توابع جدا ساز:

- تابع ریسک $g_i(x) = - R(\alpha_i \mid x)$
- یا برای حالت کمترین خطا : $g_i(x) = P(\omega_i \mid x)$

در حالتی که $P(x)=1$ (احتمال مشاهده) باشد انتخاب تابع جدا ساز راحت تر می شود: $g_i(x) \equiv p(x \mid \omega_i)\, P(\omega_i)$

چون می توان از طرفین هر تابع یکنوایی گرفت ب اینکه دون نتیجه مقایسه توابع جدا ساز تغییر کند، بخصوص اگر توابع توزیع از نوع نمایی باشد، با گرفتن لگاریتم خواهیم داشت:

$$g_i(x) = \ln p(x \mid \omega_i) + \ln P(\omega_i)$$

محاسبه تابع جدا ساز از روی فرمول Bayes:

- به این ترتیب با توجه به موارد گفته شده:

$$g(x) = P(\omega_1 | x) - P(\omega_2 | x)$$

$$= \ln \frac{P(x | \omega_1)}{P(x | \omega_2)} + \ln \frac{P(\omega_1)}{P(\omega_2)}$$

- اگر توابع توزیع از نوع نمایی باشد، ترم اول تابع ساده ای (اکثرا خطی) از X و ترم دوم مقداری ثابت خواهد بود.
- بدیهی است تصمیم گیری از روی این تابع می تواند خیلی ساده تر از خود توابع توزیع باشد.

توابع جدا ساز برای چگالی توزیع نرمال

- اگر توزیع کلاس ها نرمال باشد:

$$p(x \mid \omega_i) = \frac{1}{(2\pi)^{d/2} . \Sigma^{1/2}} \exp[\frac{-1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)]$$

- تابع جدا ساز عبارت خواهد بود از:

$$g_i(x) = \ln P(x \mid \omega_i) + \ln P(\omega_i)$$

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^t \sum_i^{-1} (x-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln\left|\Sigma_i\right| + \ln P(\omega_i)$$

Case $\Sigma_i = \sigma^2.I$:حالت خاص

- در این شرایط می بینیم که تابع جدا ساز، تابعی خطی خواهد شد که ضرایب خط به مشخصات توزیع وابسته است. این حالت به فاصله اقلیدسی بین نمونه و متوسط هر کلاس باز می گردد.

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

$g_i(x) = w_i^t x + w_{i0}$ (linear discriminant function)

where:

$$w_i = \frac{\mu_i}{\sigma^2}; \quad w_{i0} = -\frac{1}{2\sigma^2}\mu_i^t \mu_i + \ln P(\omega_i)$$

(ω_{i0} is called the threshold for the ***i*** th category!)

ماشین خطی

- کلاسیفایر خطی که تابع جدا ساز خطی تولید می کنند " ماشین خطی " نامیده می شوند
- سطوح تصمیم درچنین مواردی، صفحات چند بعدی hyper planes هستند که برای یافتن آنها از رابطه زیر استفاده می شود:

$$g_i(x) = g_j(x)$$

- بسته به اینکه ارزش مشخصه های مختلف در تفکیک کلاس ها (توزیع آنهادر فضا) به چه شکل باشد، موقعیت صفحه نسبت به زیر فضا های متعلق به کلاس ها متفاوت خواهد بود.

مثال: کلاسیفایر برای داده هایی با پراکندگی مشابه

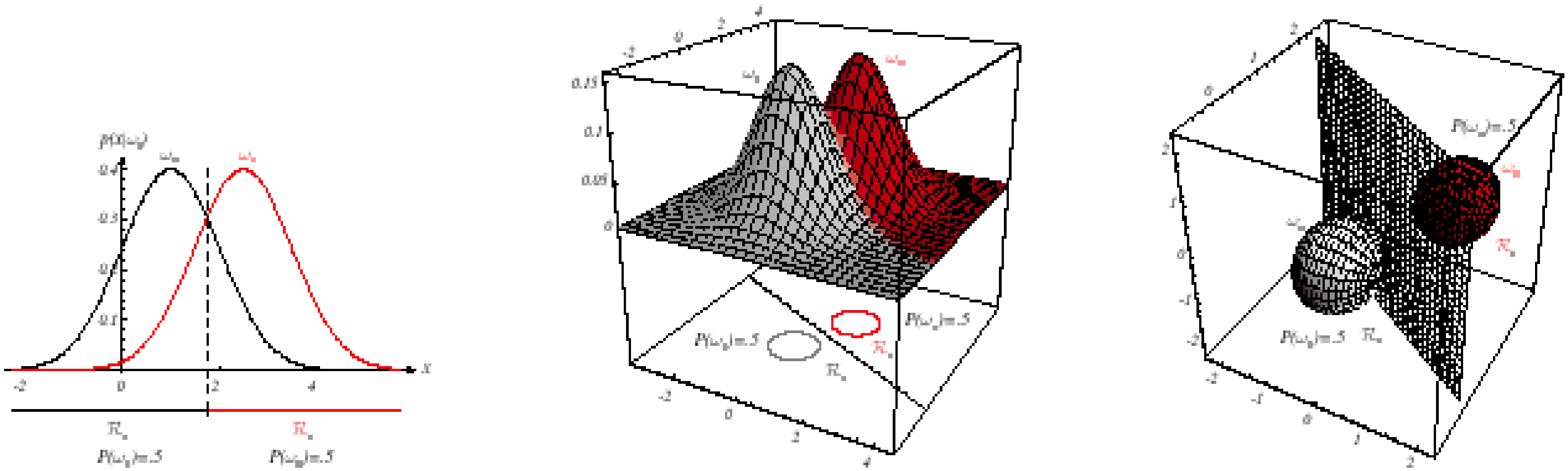


FIGURE 2.10. If the covariance matrices for two distributions are equal and proportional to the identity matrix, then the distributions are spherical in d dimensions, and the boundary is a generalized hyperplane of $d-1$ dimensions, perpendicular to the line separating the means. In these one-, two-, and three-dimensional examples, we indicate $p(\mathbf{x}|\omega_i)$ and the boundaries for the case $P(\omega_1) = P(\omega_2)$. In the three-dimensional case, the grid plane separates $\mathcal{R}_1$ from $\mathcal{R}_2$. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

The hyper plane separating $\mathcal{R}_i$ and $\mathcal{R}_j$

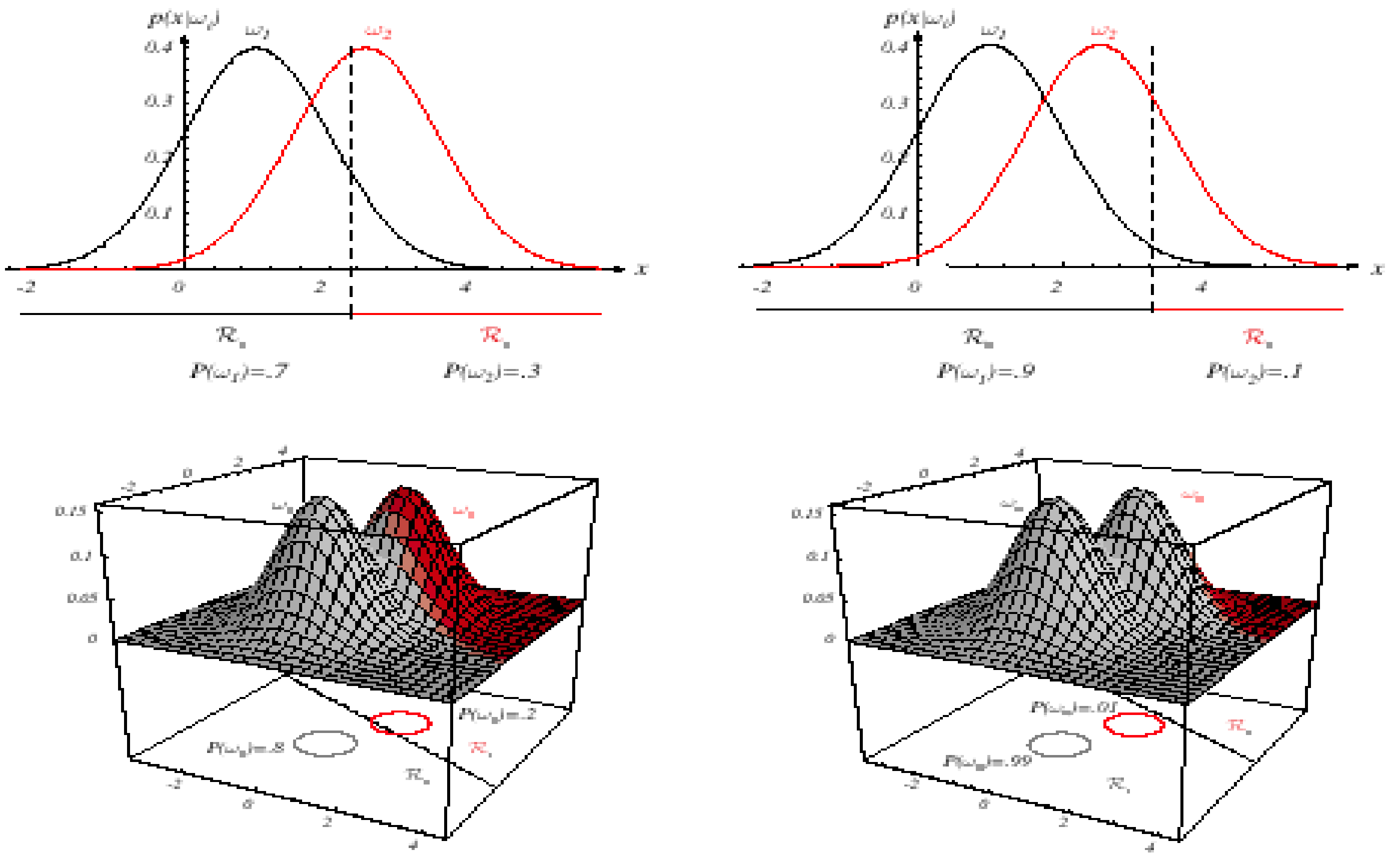
$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\sigma^2}{\|\mu_i - \mu_j\|^2} \ln \frac{P(\omega_i)}{P(\omega_j)} (\mu_i - \mu_j)$$

اگر ماتریس کوواریانس با یک عدد مشخص شود یا ماتریس یکه باشد (ارزش مشخصه ها در تفکیک با هم مساوی باشد) صفحات مزبور همواره بر خطی که مراکز توزیع را به هم مرتبط می کند عمود هستند.

عامل دیگر، احتمال پیشین وقوع کلاس هاست که در صورت متفاوت بودن، به صورت شیفت دهنده خط ظاهرمی شود.

$$if \quad P(\omega_i) = P(\omega_j) \quad then \quad x_0 = \frac{1}{2}(\mu_i + \mu_j)$$

بررسی پراکندگی مشابه در ابعاد مختلف



صفحه جدا ساز در حالت متفاوت بودن احتمالات پيشين

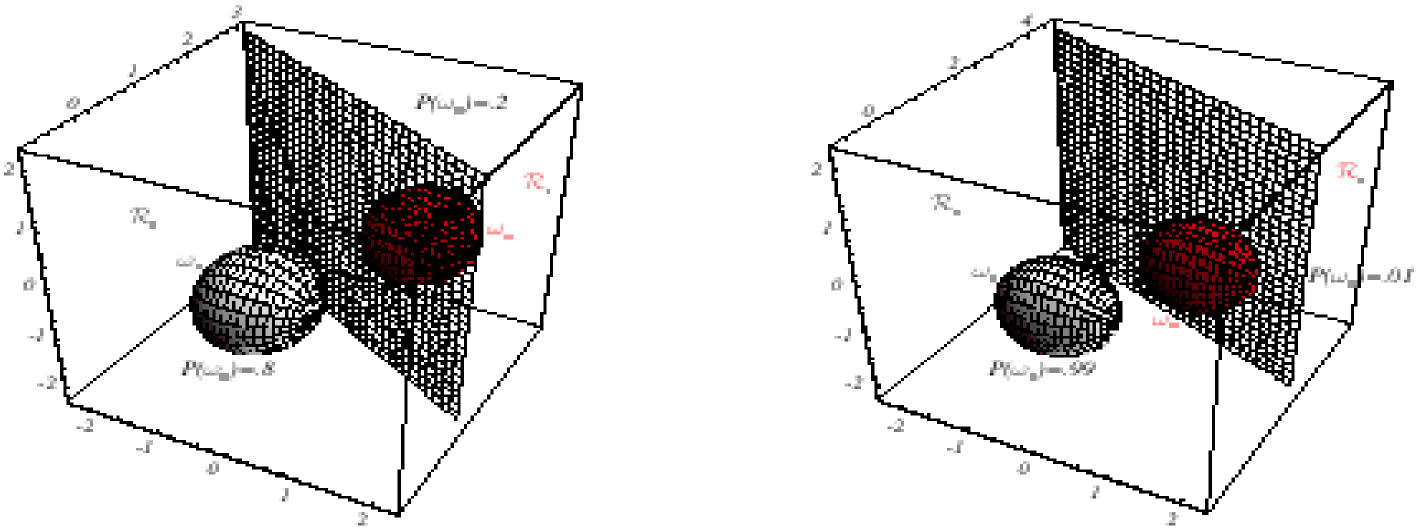
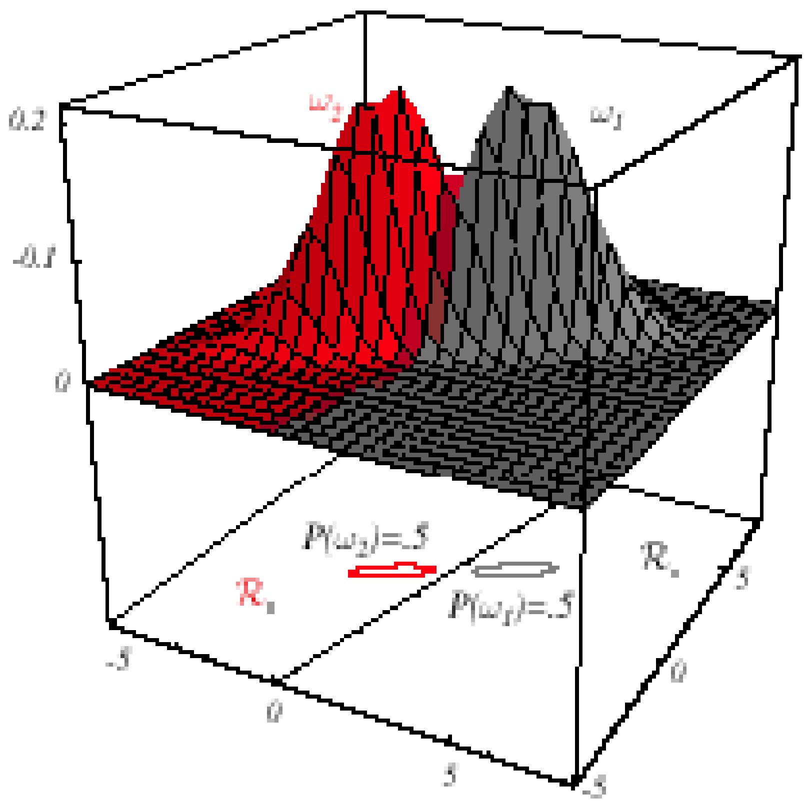


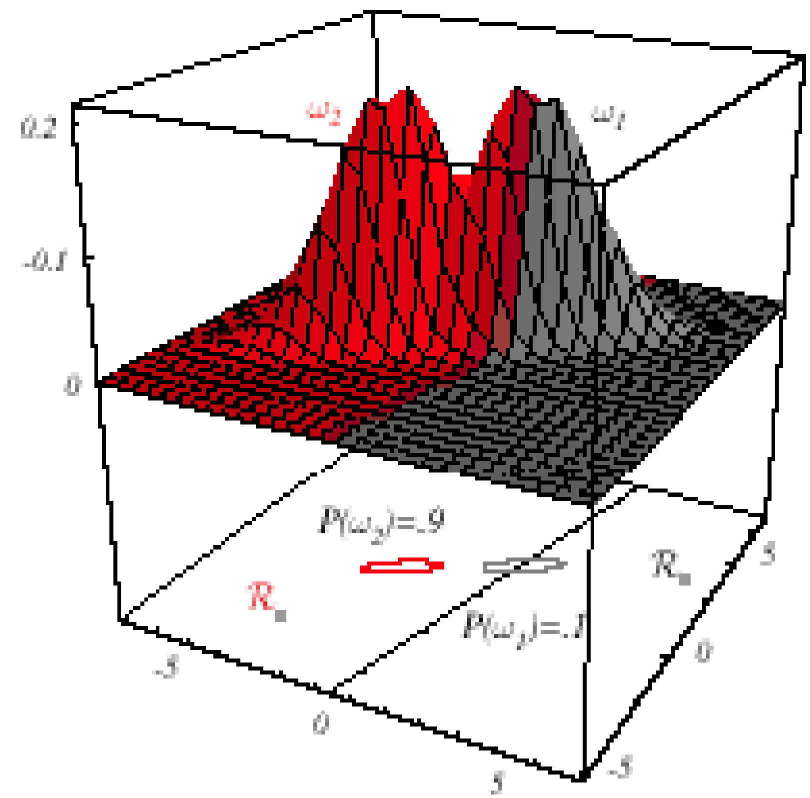
FIGURE 2.11. As the priors are changed, the decision boundary shifts; for sufficiently disparate priors the boundary will not lie between the means of these one-, two- and three-dimensional spherical Gaussian distributions. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

حالت غیر قطری بودن پراکندگی $\Sigma_i = \Sigma$

- اگر ارزش مشخصه ها(توزیع آنها) مساوی نباشد: صفحه جدا کننده بر خط واصل مراکز عمود نخواهد بود.

$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\ln\left[P(\omega_i)/P(\omega_j)\right]}{(\mu_i - \mu_j)^t \Sigma^{-1} (\mu_i - \mu_j)} \cdot (\mu_i - \mu_j)$$


0.2
-0.1
0
ω_2
ω_1
$P(\omega_2)=.5$
$P(\omega_1)=.5$
-5
0
5
-5
0
5


0.2
-0.1
0
ω_2
ω_1
$P(\omega_2)=.9$
$P(\omega_1)=.1$
-5
0
5
-5
0
5

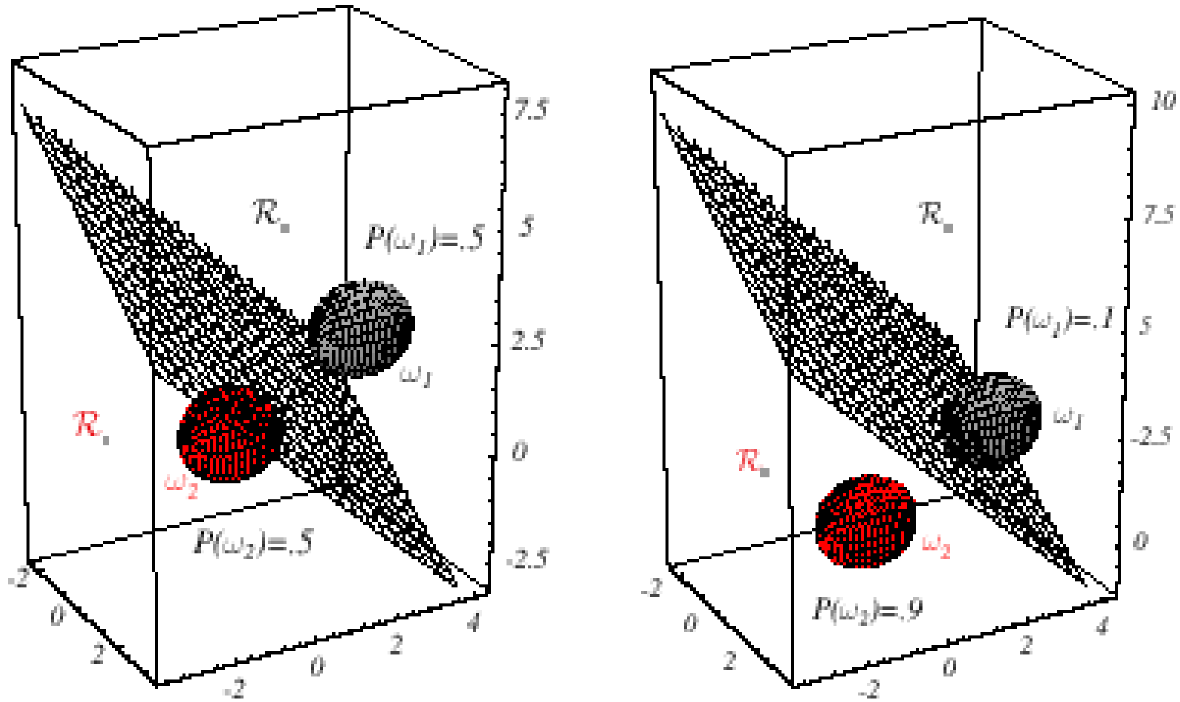


FIGURE 2.12. Probability densities (indicated by the surfaces in two dimensions and ellipsoidal surfaces in three dimensions) and decision regions for equal but asymmetric Gaussian distributions. The decision hyperplanes need not be perpendicular to the line connecting the means. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

Σ_i = دلخواه

- اگر ماتریس های کوواریانس برای دو کلاس متفاوت باشند:

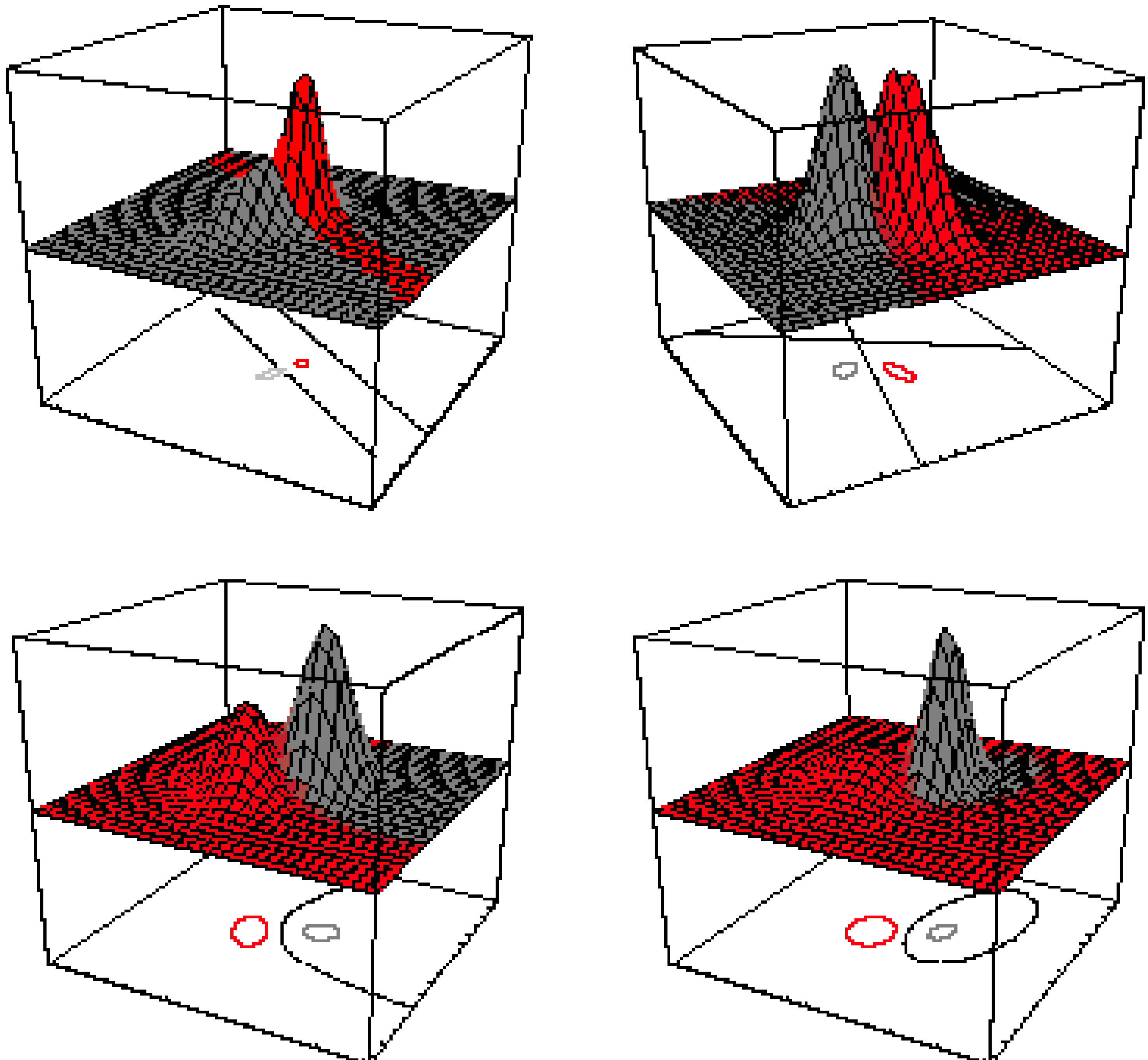
$$g_i(x) = x^t W_i x + w_i^t x = w_{i0}$$

where :

$$W_i = -\frac{1}{2}\Sigma_i^{-1}$$

$$w_i = \Sigma_i^{-1}\mu_i$$

$$w_{i0} = -\frac{1}{2}\mu_i^t \Sigma_i^{-1}\mu_i - \frac{1}{2}\ln\left|\Sigma_i\right| + \ln P(\omega_i)$$



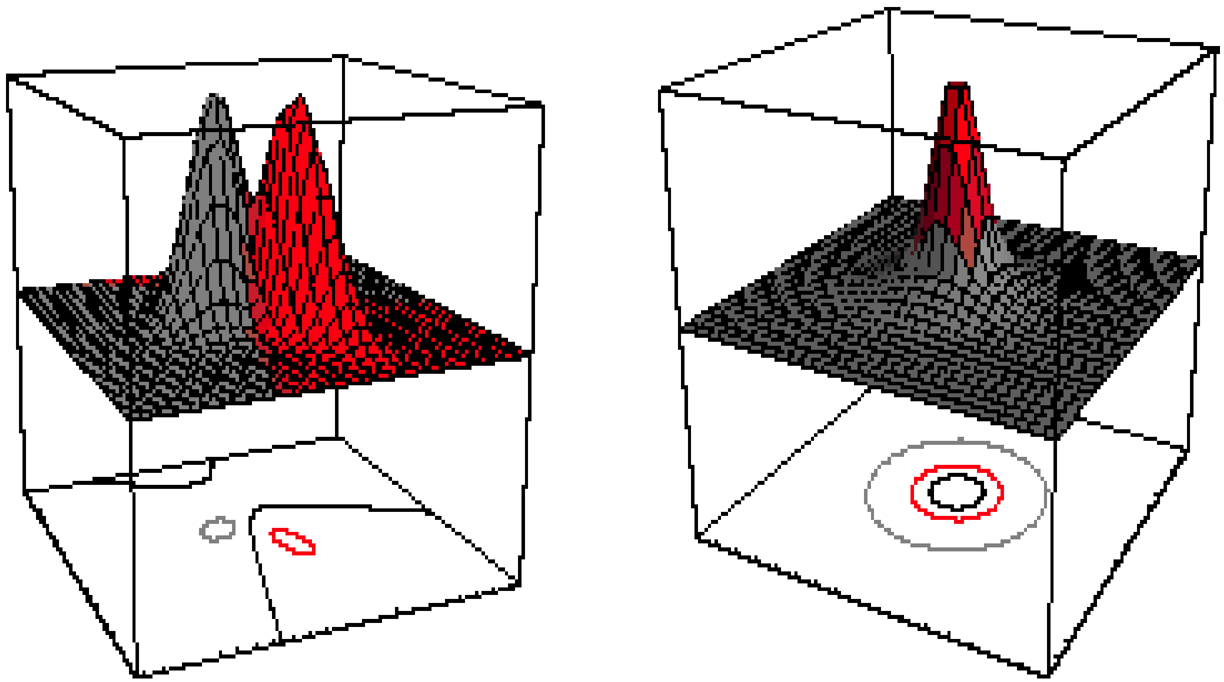


FIGURE 2.14. Arbitrary Gaussian distributions lead to Bayes decision boundaries that are general hyperquadrics. Conversely, given any hyperquadric, one can find two Gaussian distributions whose Bayes decision boundary is that hyperquadric. These variances are indicated by the contours of constant probability density. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

تئوری Bayes برای مشخصه های گسسته

- مبنای کار در این شرایط تغییری نمی کند الا اینکه توابع چگالی توزیع به توابع احتمال نقطه ای تبدیل شده و توابع تجمعی از شکل انتگرال به شکل جمع تبدیل می شوند.
- در نتیجه فرمول Bayes به شکل زیر در می آید:

$$\boxed{P(\omega_j|\mathrm{x}) = \frac{P(\mathrm{x}|\omega_j)P(\omega_j)}{P(\mathrm{x})},}$$

$$P(\mathrm{x}) = \sum_{j=1}^{c} P(\mathrm{x}|\omega_j)P(\omega_j).$$

مشخصه های باینری مستقل

- اگر مشخصه ها تنها مقادیر صفر و یک را گرفته و از هم مستقل هم باشند، و اگر احتمال یک بودن یکی از این مشخصه ها را در کلاس های دو گانه، به p و q نشان دهیم، با فرض وجود d مشخصه در x ، می توان احتمال عضویت x در هر کلاس را به شکل زیر به احتمال ۱ بودن مشخصه های مختلف به شکل زیر ارتباط داد:

$$P(\mathrm{x}|\omega_1) = \prod_{i=1}^{d} p_i^{x_i}(1-p_i)^{1-x_i}$$

$$P(\mathrm{x}|\omega_2) = \prod_{i=1}^{d} q_i^{x_i}(1-q_i)^{1-x_i}$$

مشخصه های باینری مستقل...

- بر این اساس می توان توابع شباهت (likelihood) و سپس توابع جدا ساز خطی را به شکل زیر بدست آورد:

$$\frac{P(\mathrm{x}|\omega_1)}{P(\mathrm{x}|\omega_2)} = \prod_{i=1}^{d} \left(\frac{p_i}{q_i}\right)^{x_i} \left(\frac{1-p_i}{1-q_i}\right)^{1-x_i}$$

- و در نتیجه:

$$g(\mathrm{x}) = \sum_{i=1}^{d} \left[x_i \ln \frac{p_i}{q_i} + (1-x_i) \ln \frac{1-p_i}{1-q_i} \right] + \ln \frac{P(\omega_1)}{P(\omega_2)}.$$

- با توجه به فرمول خط جدا ساز:

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

$$w_i = \ln \frac{p_i(1-q_i)}{q_i(1-p_i)} \qquad i = 1, ..., d$$

$$w_0 = \sum_{i=1}^{d} \ln \frac{1-p_i}{1-q_i} + \ln \frac{P(\omega_1)}{P(\omega_2)}$$

خلاصه مباحث

- چگونگی طراحی کلاسیفایر
- (یافتن مرز تصمیم):
  - تبدیل مسئله به شکل یک مسئله بهینه سازی
    - یافتن معیار مناسب (بر مبنای مینیمم کردن خطا)
    - روش جستجو ؟؟ (در تعداد کمی از مسائل به شکل فرمولی)
- تعریف توابع جدا ساز برای مشخصه های پیوسته
  - توابع جدا ساز بر مبنای احتمال
  - توابع جدا ساز خطی برای توزیع های نرمال

خلاصه مباحث

- برای مشخصه های گسسته و بخصوص باینری، مبانی کار همان است با این تفاوت که در این حالت می توان از احتمال یک یا صفر بودن هر مشخصه در اعضای یک کلاس، به احتمال عضویت X در یک کلاس رسید

پایان

سوال؟