Sharif University of Technology

پردازش سیگنال‌های حیاتی
مبحث نهم- طبقه‌بندی

محمدباقر شمس‌الهی

mbshams@sharif.edu

دانشکده برق دانشگاه صنعتی شریف

نیمسال دوم ۱۴۰۱-۱۴۰۲

مبحث نهم- طبقه‌بندی

- مقدمه
- طبقه‌بند آماری بیز
- طبقه‌بند آماری بیز با در نظر گرفتن ریسک متفاوت برای خطا
- طبقه‌بند آماری بیز با فرض توزیع گوسی ویژگی‌ها
- طبقه‌بند K نزدیک‌ترین همسایه
- کاهش بعد ویژگی
- ارزیابی عملکرد طبقه‌بندی

مقدمه

- در یادگیری ماشین که شاخه‌ای از هوش مصنوعی است، مبحثی به نام شناسایی الگو وجود دارد که هدفش تخصیص یک برچسب به یک ورودی است.

Artificial Intelligent/Machine Learning/Pattern Recognition

- هدف شناسایی الگو، دسته‌بندی (برچسب زدن) یک سری داده/ورودی/شیء/الگو به یکی از چند کلاس/گروه/دسته/خوشه است.

Date/Input/Object/Pattern Class/Group/Category/Cluster

- تقسیم‌بندی
  - شناسایی آماری الگو
  - شناسایی ساختاری الگو

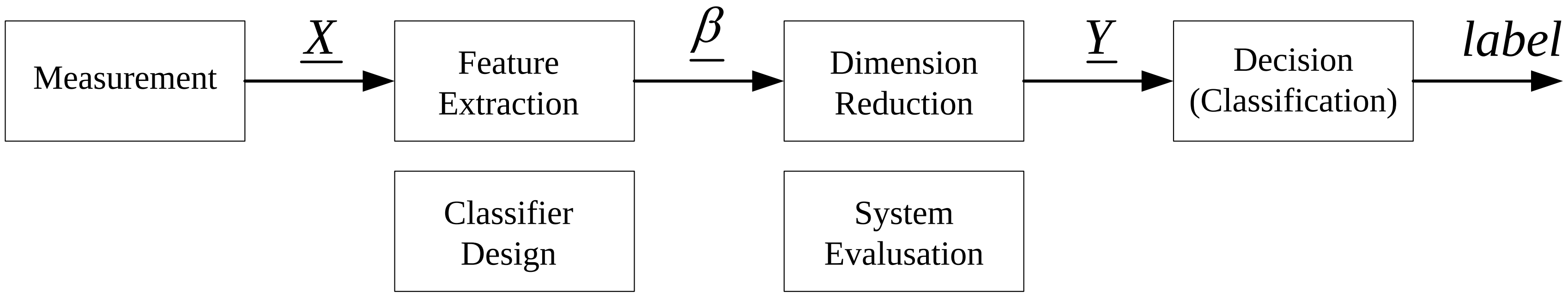
Statistical
Structural (Syntactic)

- تقسیم‌بندی
  - باسرپرست: طبقه‌بندی
  - بی‌سرپرست: خوشه‌بندی

Supervised: Classification
Unsupervised: Clustering

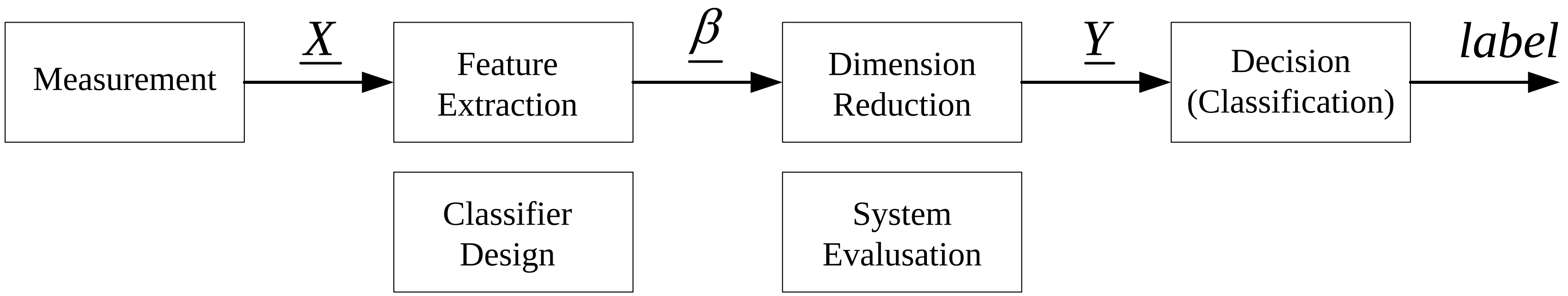
مقدمه

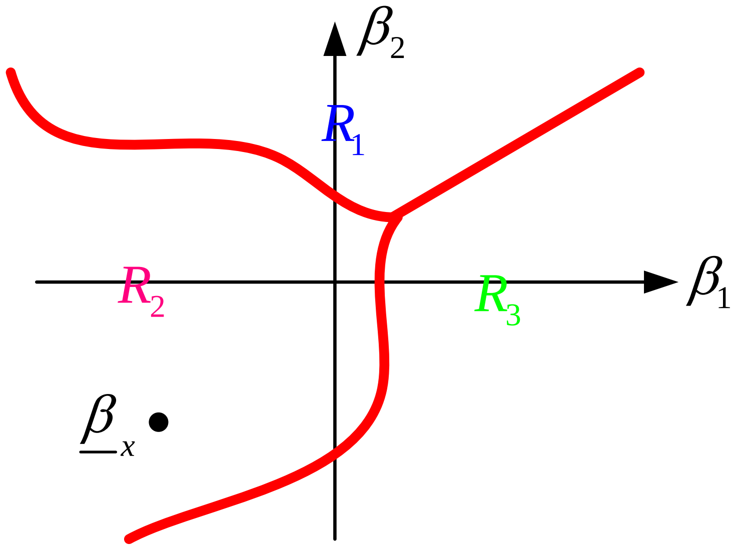
- هدف طبقه‌بندی تخصیص برچسب به یک داده بدون برچسب است.
  - تعداد کلاس‌ها معلوم
  - اطلاعات پیشین در مورد کلاس‌ها
    - اطلاعات آماری دقیق معلوم است.
    - تعدادی داده با برچسب موجود است: آموزش طبقه‌بند (تخمین اطلاعات آماری)
- کاربردها
  - تشخیص بیماری‌ها/تشخیص هویت/تشخیص متن/کاراکتر/...
- طبقه‌بندی دو کلاسی/آشکارسازی و ارتباط با تخمین/رگرسیون
- مراحل مختلف یک پروسه طبقه‌بندی



مقدمه

- مراحل مختلف یک پروسه طبقه‌بندی

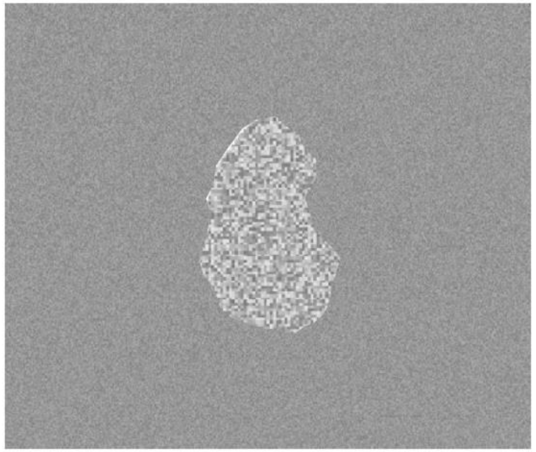


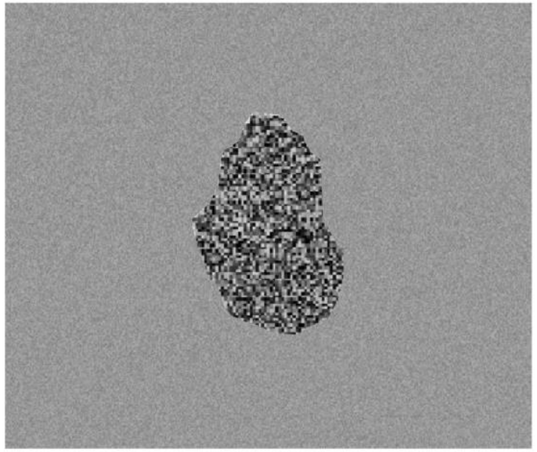


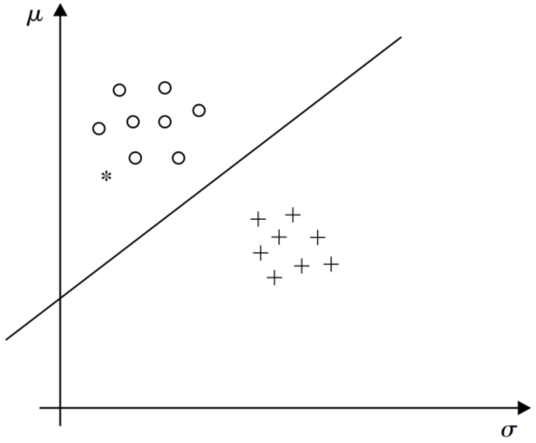
  – اندازه‌گیری اولیه و بدست آوردن داده خام اولیه
  – استخراج ویژگی
    - اطلاعات کمی توصیف کننده داده/متمایز برای کلاس‌های مختلف
  – کاهش بعد
  – تصمیم‌گیری و تعیین برچسب داده
- دو مرحله اولیه
  – طراحی طبقه‌بند (فاز آموزش) و تعیین مرز کلاس‌ها و نواحی مربوط به هر کلاس در فضای ویژگی
  – ارزیابی عملکرد سیستم طراحی شده (فاز آزمون/تست: دادن برچسب به داده‌های بدون برچسب)

مقدمه

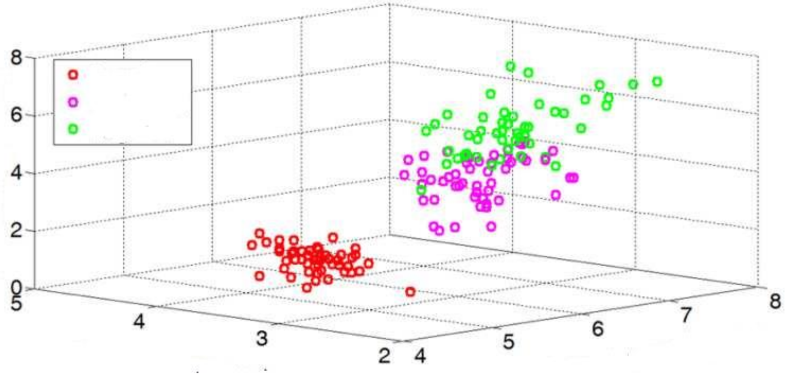
- طبقه‌بندی تصاویر مربوط به تومور خوش‌خیم و تومور بدخیم

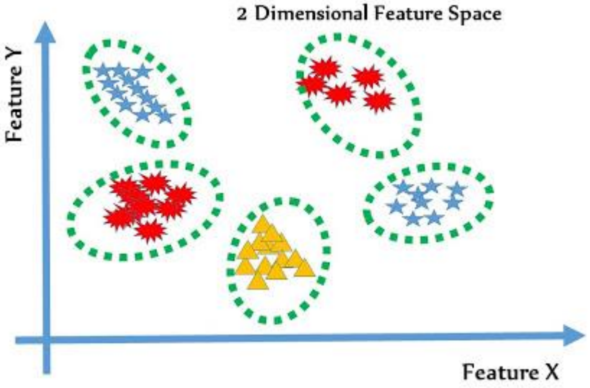






  – دو ویژگی ساده: متوسط و انحراف معیار مقدار پیکسل‌های ناحیه تومور
  – استخراج ویژگی از داده‌های برچسب‌دار (فاز آموزش)





- تعداد کلاس‌ها

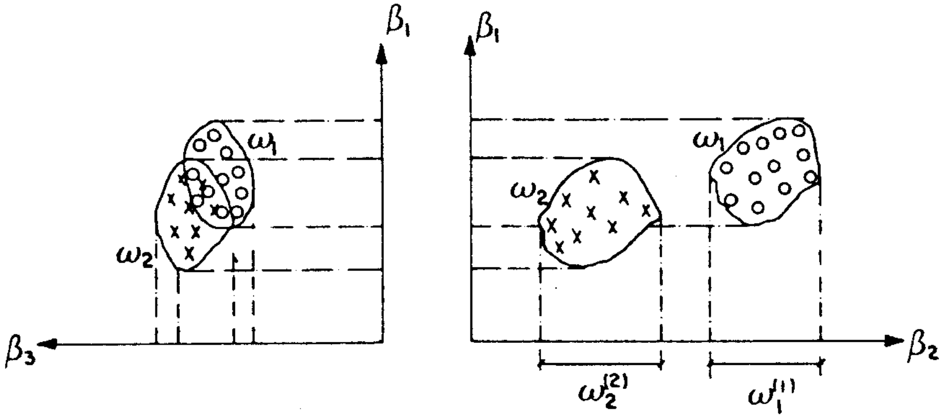
$$\underline{\beta} \in \omega_i, i = 1, 2, \cdots, M$$

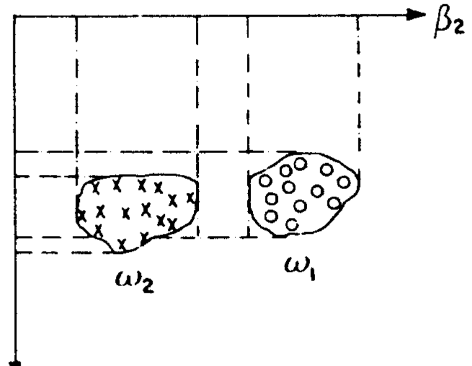
- بعد بردار ویژگی

$$\underline{\beta} \in \Re^N$$

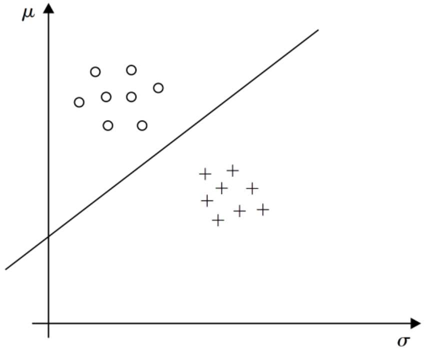
مقدمه

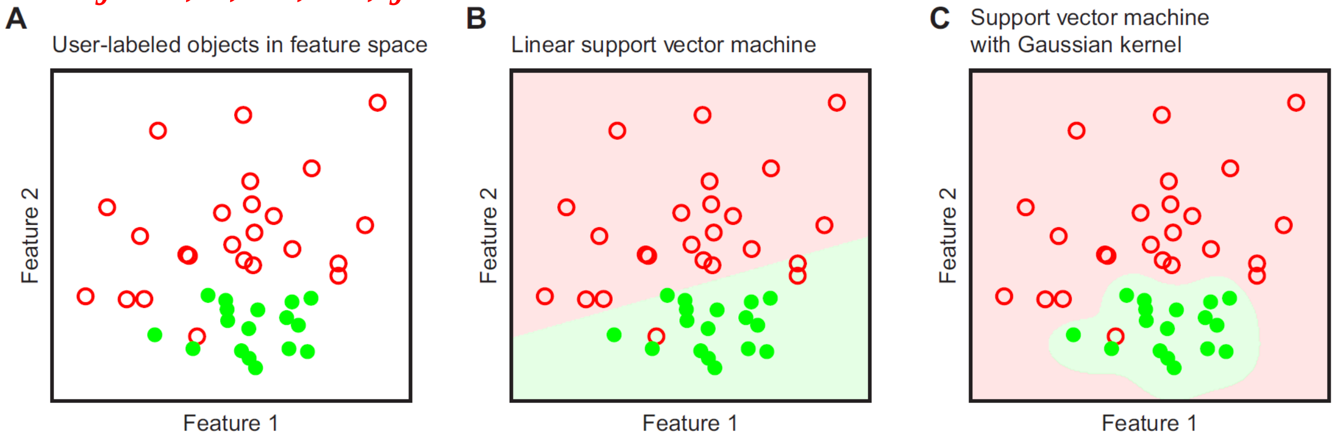
- جدایی‌پذیری کلاس‌ها در فضای ویژگی
  - جداپذیری/جدایی ناپذیری خطی
  - جدایی پذیری غیرخطی
- طبقه‌بندهای مختلف
  - طبقه‌بند آماری بیز/شبکه عصبی/SVM/kNN
- ایده کلی طبقه‌بندی
  - تابع تصمیم‌گیری/تابع تفکیک (هزینه/هدف)
    - می‌نیمم/ماکزیمم





$$d_i(\underline{\beta}) > d_j(\underline{\beta}) \Rightarrow \underline{\beta} \in \omega_i$$
$$\forall j = 1, 2, \cdots, M,\ j \neq i$$





طبقه‌بند آماری بیز Bayes Classifier

- توابع چگالی احتمال و احتمالات موجود در یک مسئله طبقه‌بندی از روی بردار ویژگی
  - تابع چگالی احتمال توام بردار ویژگی و هر کلاس $\underline{\beta} \in \omega_1, \omega_2, \cdots, \omega_M \quad \underline{\beta} \in \Re^N$
    $$p(\omega_i, \underline{\beta}) = P(\omega_i).p(\underline{\beta} \mid \omega_i) = p(\underline{\beta}).P(\omega_i \mid \underline{\beta})$$
  - احتمال پیشین هر کلاس $P(\omega_1), P(\omega_2), \cdots, P(\omega_M)$
  - احتمال پسین هر کلاس $P(\omega_1 \mid \underline{\beta}), P(\omega_2 \mid \underline{\beta}), \cdots, P(\omega_M \mid \underline{\beta})$
  - تابع چگالی درستنمایی هر کلاس $p(\underline{\beta} \mid \omega_1), p(\underline{\beta} \mid \omega_2), \cdots, p(\underline{\beta} \mid \omega_M)$
  - معلوم فرض کردن احتمال پیشین هر کلاس
  - معلوم فرض کردن تابع چگالی درستنمایی هر کلاس (اطلاعات آماری ویژگی در هر کلاس)
- معیار تصمیم‌گیری در طبقه‌بند بیز
  - احتمال پسین هر کلاس $d_i(\underline{\beta}) = P(\omega_i \mid \underline{\beta})$
  - ماکزیمم کردن $\forall j = 1, 2, \cdots, M, \; j \neq i \qquad P(\omega_i \mid \underline{\beta}) > P(\omega_j \mid \underline{\beta}) \Rightarrow \underline{\beta} \in \omega_i$
  - تقسیم فضای ویژگی به M ناحیه $R_1, R_2, \cdots, R_M$
  - رابطه برای دو کلاس
    $$P(\omega_1 \mid \underline{\beta}) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} P(\omega_2 \mid \underline{\beta})$$
  - عدم تصمیم در صورت تساوی

طبقه‌بند آماری بیز Bayes Classifier

- فرم دیگر معیار بیز با استفاده از قضیه بیز
  - پیچیدگی محاسبه تابع احتمال پسین هر کلاس

$$P(\omega_i \mid \underline{\beta}) = \frac{p(\omega_i, \underline{\beta})}{p(\underline{\beta})} = \frac{P(\omega_i).p(\underline{\beta} \mid \omega_i)}{p(\underline{\beta})}$$

$$\frac{P(\omega_i).p(\underline{\beta} \mid \omega_i)}{p(\underline{\beta})} > \frac{P(\omega_j).p(\underline{\beta} \mid \omega_j)}{p(\underline{\beta})} \Rightarrow \underline{\beta} \in \omega_i \qquad \forall j = 1, 2, \cdots, M, \; j \neq i$$

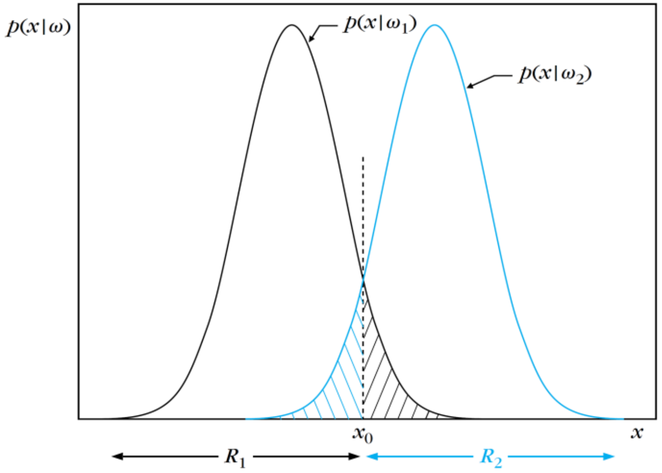
$$P(\omega_i).p(\underline{\beta} \mid \omega_i) > P(\omega_j).p(\underline{\beta} \mid \omega_j) \Rightarrow \underline{\beta} \in \omega_i \qquad \forall j = 1, 2, \cdots, M, \; j \neq i$$

$$d_i(\underline{\beta}) = P(\omega_i).p(\underline{\beta} \mid \omega_i)$$

- مزیت معیار جدید
  - وابستگی به توزیع ویژگی‌ها در هر کلاس
  - قابل تخمین از روی داده‌های برچسب‌دار آموزش

$$P(\omega_1).p(\underline{\beta} \mid \omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} P(\omega_2).p(\underline{\beta} \mid \omega_2)$$

- حالت خاص دو کلاسی
  - بررسی دو نوع خطا
  - یکسان بودن اهمیت دو نوع خطا



- معادل بودن معیار بیز با می‌نیمم کردن خطای زیر

$$P_e = \sum_{i=1}^{M} p(\underline{\beta} \in R_i, \omega \neq \omega_i)$$

طبقه‌بندی با انتخاب کلاس با ریسک می‌نیمم

- تمایز بین انواع خطا

$$\lambda_{ik} = \lambda(\underline{\beta} \in R_i \,|\, \omega = \omega_k),\;\; \lambda_{ii} = 0$$

  - تعریف تلفات (جریمه) برای انتخاب یک کلاس وقتی کلاس واقعی کلاس دیگری است
  - تعریف ریسک انتخاب کلاس

$$r_i(\underline{\beta}) = \sum_{k=1}^{M} \lambda_{ik} P(\omega_k \,|\, \underline{\beta})$$

  - معیار تصمیم‌گیری
    - می‌نیمم کردن ریسک انتخاب کلاس

$$r_i(\underline{\beta}) < r_j(\underline{\beta}) \Rightarrow \underline{\beta} \in \omega_i \qquad \forall j = 1, 2, \cdots, M,\; j \neq i$$

$$d_i(\underline{\beta}) = -r_i(\underline{\beta})$$

- حالت دو کلاسی

$$r_2(\underline{\beta}) \mathop{\gtrless}_{\omega_2}^{\omega_1} r_1(\underline{\beta}) \Rightarrow \lambda_{21} P(\omega_1 \,|\, \underline{\beta}) + \lambda_{22} P(\omega_2 \,|\, \underline{\beta}) \mathop{\gtrless}_{\omega_2}^{\omega_1} \lambda_{11} P(\omega_1 \,|\, \underline{\beta}) + \lambda_{12} P(\omega_2 \,|\, \underline{\beta})$$

$$\Rightarrow (\lambda_{21} - \lambda_{11}) P(\omega_1 \,|\, \underline{\beta}) \mathop{\gtrless}_{\omega_2}^{\omega_1} (\lambda_{12} - \lambda_{22}) P(\omega_2 \,|\, \underline{\beta}) \Rightarrow (\lambda_{21} - \lambda_{11}) P(\omega_1).p(\underline{\beta} \,|\, \omega_1) \mathop{\gtrless}_{\omega_2}^{\omega_1} (\lambda_{12} - \lambda_{22}) P(\omega_2).p(\underline{\beta} \,|\, \omega_2)$$

$$\Rightarrow \frac{p(\underline{\beta} \,|\, \omega_1)}{p(\underline{\beta} \,|\, \omega_2)} \mathop{\gtrless}_{\omega_2}^{\omega_1} \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(\omega_2)}{P(\omega_1)}$$

  - مقایسه نسبت درستنمایی با یک آستانه

- معادل بودن معیار با می‌نیمم کردن ریسک متوسط

$$r_{av} = \sum_{i=1}^{M} r_i(\underline{\beta}) P(\omega_i)$$

طبقه‌بندی با انتخاب کلاس با ریسک می‌نیمم

- مثال۱– تعیین مرز تصمیم‌گیری در طبقه‌بندی دو کلاسی با یک ویژگی با توزیع گوسی با متوسط صفر و یک و واریانس یکسان ۰.۵

$$p(\beta|\omega_1)=\frac{1}{\sqrt{2\pi(0.5)}}e^{-\frac{(\beta-0)^2}{2(0.5)}},\ p(\beta|\omega_2)=\frac{1}{\sqrt{2\pi(0.5)}}e^{-\frac{(\beta-1)^2}{2(0.5)}}$$

$$\frac{p(\underline{\beta}|\omega_1)}{p(\underline{\beta}|\omega_2)}\underset{\omega_2}{\overset{\omega_1}{\gtrless}}\frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}}\frac{P(\omega_2)}{P(\omega_1)}$$

$$a)P(\omega_1)=P(\omega_2),\lambda=\begin{pmatrix}0&1\\1&0\end{pmatrix}\Rightarrow\frac{p(\underline{\beta}|\omega_1)}{p(\underline{\beta}|\omega_2)}\underset{\omega_2}{\overset{\omega_1}{\gtrless}}1\Rightarrow\beta_0=\frac{1}{2}$$

قسمت اول همان طبقه‌بند بیز است.

$$b)P(\omega_1)=P(\omega_2),\lambda=\begin{pmatrix}0&0.5\\1&0\end{pmatrix}\Rightarrow\frac{p(\underline{\beta}|\omega_1)}{p(\underline{\beta}|\omega_2)}\underset{\omega_2}{\overset{\omega_1}{\gtrless}}\frac{1}{2}\Rightarrow\beta_0=\frac{1+\ln 2}{2}$$

قسمت سوم همان طبقه‌بند بیز است.

$$c)P(\omega_1)=0.5P(\omega_2),\lambda=\begin{pmatrix}0&1\\1&0\end{pmatrix}\Rightarrow\frac{p(\underline{\beta}|\omega_1)}{p(\underline{\beta}|\omega_2)}\underset{\omega_2}{\overset{\omega_1}{\gtrless}}2\Rightarrow\beta_0=\frac{1-\ln 2}{2}$$

$$d)P(\omega_1)=2P(\omega_2),\lambda=\begin{pmatrix}0&0.5\\1&0\end{pmatrix}\Rightarrow\frac{p(\underline{\beta}|\omega_1)}{p(\underline{\beta}|\omega_2)}\underset{\omega_2}{\overset{\omega_1}{\gtrless}}\frac{1}{4}\Rightarrow\beta_0=\frac{1+\ln 4}{2}$$

- تمرین: در قسمت اول تابع احتمال پسین هر دو کلاس را بدست آورده و رسم کنید.

طبقه‌بند آماری بیز با توزیع گوسی ویژگی‌ها

- تابع تفکیک طبقه‌بند بیز

$$d_i(\underline{\beta}) = P(\omega_i \mid \underline{\beta}) = P(\omega_i).p(\underline{\beta} \mid \omega_i)$$

- توزیع گوسی ویژگی‌ها

$$\omega_i : \begin{cases} \underline{\mu}_i = E\left\{\underline{\beta} \mid \omega_i\right\} \\ \Sigma_i = E\left\{(\underline{\beta}-\underline{\mu}_i)(\underline{\beta}-\underline{\mu}_i)^T \mid \omega_i\right\} \end{cases} \Rightarrow p(\underline{\beta} \mid \omega_i) = \frac{1}{\sqrt{(2\pi)^N \det \Sigma_i}} e^{-\frac{1}{2}(\underline{\beta}-\underline{\mu}_i)^T \Sigma_i^{-1} (\underline{\beta}-\underline{\mu}_i)}$$

$$d_i(\underline{\beta}) = \ln P(\omega_i \mid \underline{\beta}) = \ln\left(P(\omega_i).p(\underline{\beta} \mid \omega_i)\right) = \ln P(\omega_i) + \ln p(\underline{\beta} \mid \omega_i)$$

$$= \ln P(\omega_i) + \ln\left(\frac{1}{\sqrt{(2\pi)^N \det \Sigma_i}} e^{-\frac{1}{2}(\underline{\beta}-\underline{\mu}_i)^T \Sigma_i^{-1} (\underline{\beta}-\underline{\mu}_i)}\right)$$

- تابع تفکیک درجه دو از بردار ویژگی

$$= \ln P(\omega_i) - \frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln(\det \Sigma_i) - \frac{1}{2}(\underline{\beta}-\underline{\mu}_i)^T \Sigma_i^{-1} (\underline{\beta}-\underline{\mu}_i)$$

- ساده‌سازی تابع تفکیک
  - حذف جملات مستقل از کلاس

$$d_i(\underline{\beta}) = \ln P(\omega_i) - \frac{1}{2}\ln(\det \Sigma_i) - \frac{1}{2}(\underline{\beta}-\underline{\mu}_i)^T \Sigma_i^{-1} (\underline{\beta}-\underline{\mu}_i)$$

$$= \underline{\beta}^T A_i \underline{\beta} + \underline{b}_i^T \underline{\beta} + c_i \qquad A_i = -\frac{1}{2}\Sigma_i^{-1}, \quad \underline{b}_i = \Sigma_i^{-1}\underline{\mu}_i, \quad c_i = \ln P(\omega_i) - \frac{1}{2}\ln(\det \Sigma_i) - \frac{1}{2}\underline{\mu}_i^T \Sigma_i^{-1} \underline{\mu}_i$$

طبقه‌بند آماری بیز با توزیع گوسی ویژگی‌ها

- حالت خاص ۱: یکسان بودن ماتریس کوواریانس کلاس‌ها
  - پراکندگی یکسان هر یک از ویژگی‌ها در همه کلاس‌ها
  - تابع تفکیک خطی

$$\Sigma_i = \Sigma \Rightarrow A_i = A$$

$$d_i(\underline{\beta}) = \underline{\beta}^T A_i \underline{\beta} + \underline{b}_i^T \underline{\beta} + c_i \Rightarrow d_i(\underline{\beta}) = \underline{b}_i^T \underline{\beta} + c_i \qquad \underline{b}_i = \Sigma^{-1}\underline{\mu}_i, \quad c_i = \ln P(\omega_i) - \frac{1}{2}\underline{\mu}_i^T \Sigma^{-1} \underline{\mu}_i$$

  - معادل فاصله ماهالانوبیس Mahalanobis

$$d_i(\underline{\beta}) = \ln P(\omega_i) - \frac{1}{2}(\underline{\beta} - \underline{\mu}_i)^T \Sigma^{-1} (\underline{\beta} - \underline{\mu}_i)$$

- حالت خاص ۲: ماتریس واحد ماتریس کوواریانس کلاس‌ها
  - ناهمبسته بودن ویژگی‌ها و پراکندگی یکسان همه ویژگی‌ها در همه کلاس
  - تابع تفکیک خطی (عمود منصف در صورت هم‌احتمال بودن کلاس‌ها)

$$\Sigma_i = \Sigma = \sigma^2 I \Rightarrow A_i = A$$

$$d_i(\underline{\beta}) = \underline{\beta}^T A_i \underline{\beta} + \underline{b}_i^T \underline{\beta} + c_i \Rightarrow d_i(\underline{\beta}) = \underline{b}_i^T \underline{\beta} + c_i \qquad \underline{b}_i = \frac{1}{\sigma^2}\underline{\mu}_i, \quad c_i = \ln P(\omega_i) - \frac{1}{2\sigma^2}\underline{\mu}_i^T \underline{\mu}_i$$

  - معادل فاصله اقلیدسی

$$d_i(\underline{\beta}) = \ln P(\omega_i) - \frac{1}{2\sigma^2}(\underline{\beta} - \underline{\mu}_i)^T (\underline{\beta} - \underline{\mu}_i) = \ln P(\omega_i) - \frac{1}{2\sigma^2}\left\| \underline{\beta} - \underline{\mu}_i \right\|^2$$

طبقه‌بند آماری بیز با توزیع گوسی ویژگی‌ها

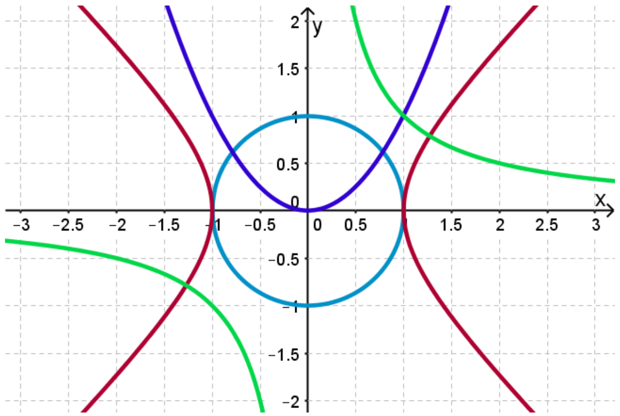
- مرز تصمیم‌گیری در طبقه بندی بیز دو کلاسی با توزیع گوسی دو ویژگی
  - مقاطع مخروطی
  - دایره/بیضی
  - سهمی/هذلولی
  - خط

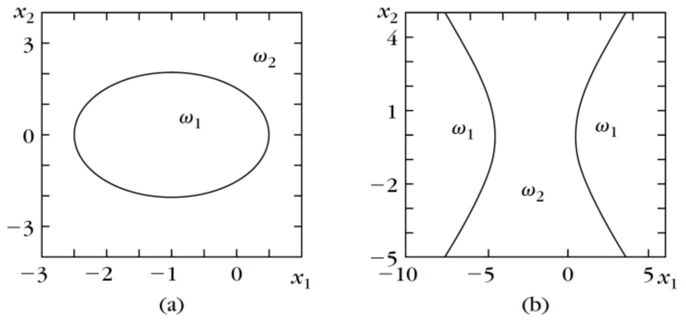
$$d_i(\underline{\beta}) = \ln P(\omega_i) - \frac{1}{2}\ln(\det \Sigma_i) - \frac{1}{2}(\underline{\beta} - \underline{\mu}_i)^T \Sigma_i^{-1} (\underline{\beta} - \underline{\mu}_i)$$

$$d_1(\underline{\beta}) = d_2(\underline{\beta})$$

$$\ln P(\omega_1) - \frac{1}{2}\ln(\det \Sigma_1) - \frac{1}{2}\begin{pmatrix}\beta_1 - \mu_{11} & \beta_2 - \mu_{21}\end{pmatrix}\begin{pmatrix}\sigma_{11}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22}^2\end{pmatrix}^{-1}\begin{pmatrix}\beta_1 - \mu_{11} \\ \beta_2 - \mu_{21}\end{pmatrix}$$

$$= \ln P(\omega_2) - \frac{1}{2}\ln(\det \Sigma_2) - \frac{1}{2}\begin{pmatrix}\beta_1 - \mu_{12} & \beta_2 - \mu_{22}\end{pmatrix}\begin{pmatrix}\sigma_{11}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22}^2\end{pmatrix}^{-1}\begin{pmatrix}\beta_1 - \mu_{12} \\ \beta_2 - \mu_{22}\end{pmatrix}$$





طبقه‌بند آماری بیز با توزیع گوسی ویژگی‌ها

- مثال ۲- صفحه ۴۴ جلد دوم کتاب مثال ۳-۱ (یک اشتباه تایپی در کتاب)

Example 3.1

Assume two classes w_1, w_2 given in a two-dimensional feature space (β_1,β_2) as described in Figure 4. The training set consists of five signals from each class. Assume the features are normally distributed with the same covariance matrix. The discriminant given by Equations 3.19A and 3.19B can be used. The training set and four unknown signals (x) are given below.

Since no information is given on the probabilities of the classes, we shall use $P(w_1) = P(w_2) = 0.5$.

w_1	w_2	x
$\beta^T_{1,1} = [-0.5,1]$	$\beta^T_{1,2} = [-2,0]$	$\beta^T_{1,x} = [-1,0]$
$\beta^T_{2,1} = [0,2]$	$\beta^T_{2,2} = [-0.5,-0.5]$	$\beta^T_{2,x} = [0,0]$
$\beta^T_{3,1} = [1,1]$	$\beta^T_{3,2} = [-1,-1.5]$	$\beta^T_{3,x} = [2,-2]$
$\beta^T_{4,1} = [0.5,0.5]$	$\beta^T_{4,2} = [0.5,-1]$	$\beta^T_{4,x} = [2,0]$
$\beta^T_{5,1} = [1,-0.5]$	$\beta^T_{5,2} = [0,-2]$	

The expectation of the two classes is estimated by:

$$\hat{\underline{\mu}} = \frac{1}{5}\sum_{j=1}^{5} \underline{\beta}_{j,i}; \quad i = 1,2$$

which yields $\hat{\underline{\mu}}_1^T = [0.4,0.8]$ and $\hat{\underline{\mu}}_2^T = [-0.6,-1]$. Since classes are assumed to possess identical covariance matrix, it is estimated by:

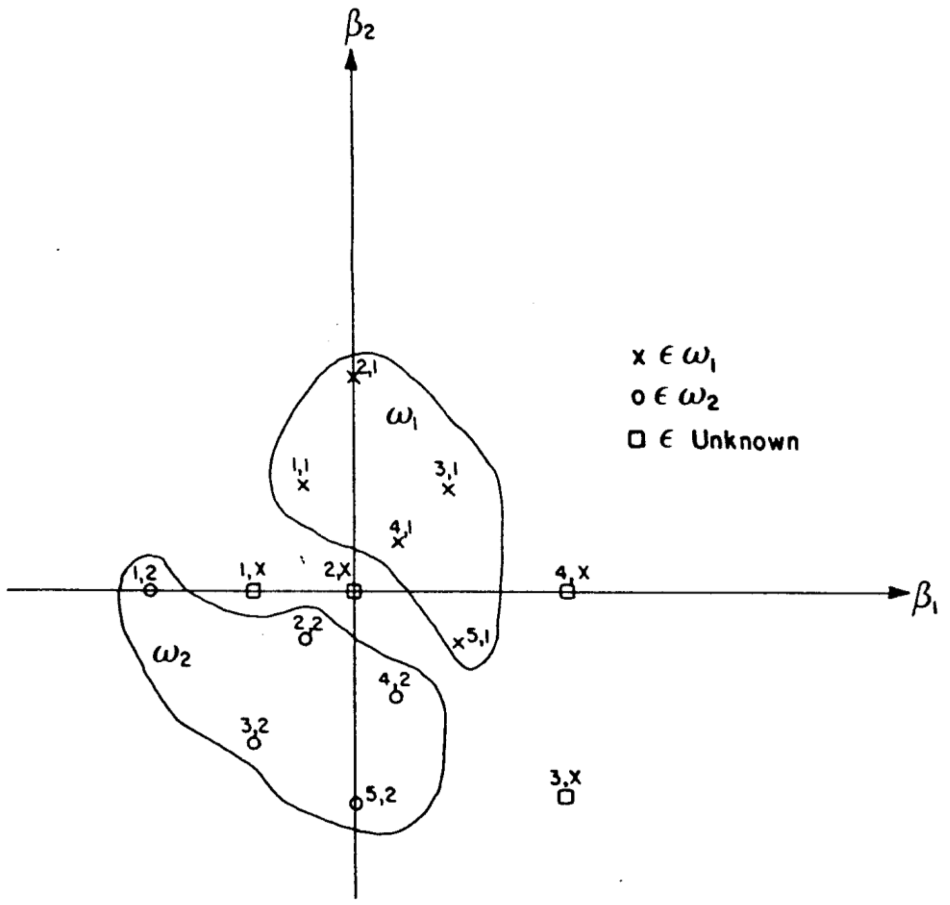


FIGURE 4. Data for Example 1.

طبقه‌بند آماری بیز با توزیع گوسی ویژگی‌ها

$$\hat{\Sigma} = \frac{1}{10}\left[\sum_{j=1}^{5}\left[(\underline{\beta}_{j,1} - \hat{\underline{\mu}}_1)(\underline{\beta}_{j,1} - \hat{\underline{\mu}}_1)^T + (\underline{\beta}_{j,2} - \hat{\underline{\mu}}_2)(\underline{\beta}_{j,2} - \hat{\underline{\mu}}_2)^T\right]\right]$$

which for the data given is

$$\hat{\Sigma} = \begin{bmatrix} 0.54 & -0.35 \\ -0.35 & 0.68 \end{bmatrix} \quad \hat{\Sigma}^{-1} = \begin{bmatrix} 2.779 & 1.430 \\ 1.430 & 2.207 \end{bmatrix}$$

The discriminant functions $d_1(\underline{\beta})$ and $d_2(\underline{\beta})$ are calculated from Equation 3.19A to give:

$$d_1(\underline{\beta}) = (2.256,\ 2.337)\underline{\beta} - 2.079$$

$$d_2(\underline{\beta}) = (-3.098,\ -3.065)\underline{\beta} - 3.155$$

$$5.354\beta_1 + 5.402\beta_2 + 1.076 = 0 \Rightarrow \beta_2 \cong -\beta_1 - 0.2$$

The discriminant functions for the data set are given below:

$$d_1(\underline{\beta}_{1,1}) = -0.869; \quad d_2(\underline{\beta}_{1,1}) = -4.671$$

$$d_1(\underline{\beta}_{2,1}) = 2.596; \quad d_2(\underline{\beta}_{2,1}) = -9.285$$

$$d_1(\underline{\beta}_{3,1}) = 2.514; \quad d_2(\underline{\beta}_{3,1}) = -9.318$$

طبقه‌بند آماری بیز با توزیع گوسی ویژگی‌ها

$$5.354\beta_1 + 5.402\beta_2 + 1.076 = 0 \Rightarrow \beta_2 \cong -\beta_1 - 0.2$$

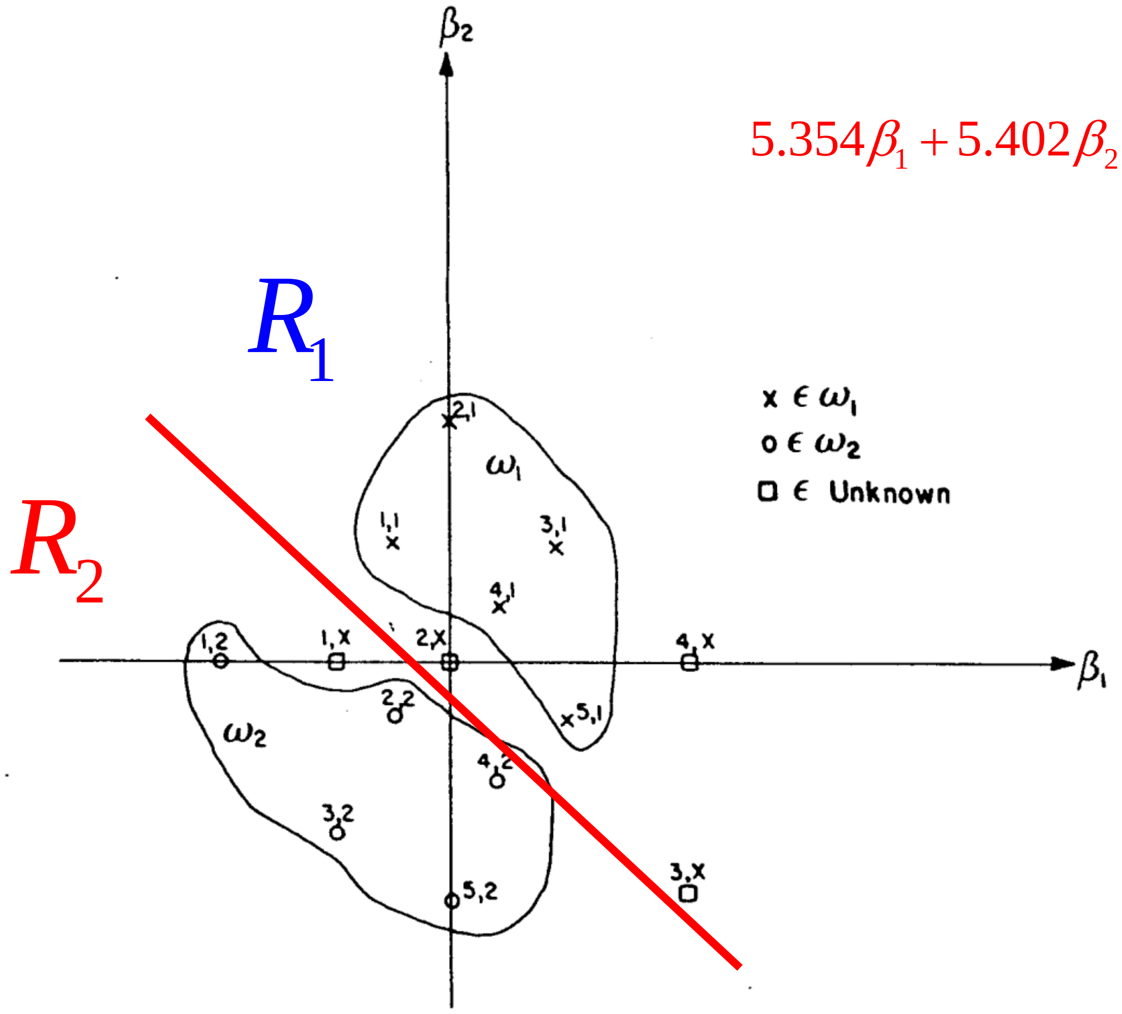


FIGURE 4. Data for Example 1.

طبقه‌بند آماری بیز با توزیع گوسی ویژگی‌ها

$$d_1(\underline{\beta}_{4,1}) = 0.218; \quad d_2(\underline{\beta}_{4,1}) = -6.236$$

$$d_1(\underline{\beta}_{5,1}) = 0.992; \quad d_2(\underline{\beta}_{5,1}) = -4.720$$

$$d_1(\underline{\beta}_{1,2}) = -6.590; \quad d_2(\underline{\beta}_{1,2}) = 3.040$$

$$d_1(\underline{\beta}_{2,2}) = -4.376; \quad d_2(\underline{\beta}_{2,2}) = -0.074$$

$$d_1(\underline{\beta}_{3,2}) = -7.841; \quad d_2(\underline{\beta}_{3,2}) = 4.540$$

$$d_1(\underline{\beta}_{4,2}) = 3.219; \quad d_2(\underline{\beta}_{4,2}) = -1.639$$

$$d_1(\underline{\beta}_{5,2}) = -6.754; \quad d_2(\underline{\beta}_{5,2}) = 2.975$$

The discriminant functions have correctly classified all training data, since $d_1(\underline{\beta}_{j,1}) > d_2(\underline{\beta}_{j,1})$ and $d_1(\underline{\beta}_{j,2}) < d_2(\underline{\beta}_{j,2})$ for all j's. Consider now the four unknown signals $\underline{\beta}_{j,x}$:

$$d_1(\underline{\beta}_{1,x}) = -4.335 \qquad d_2(\underline{\beta}_{1,x}) = -0.057 \rightarrow \underline{\beta}_{1,x} \in w_2$$

$$d_1(\underline{\beta}_{2,x}) = -2.079 \qquad d_2(\underline{\beta}_{2,x}) = -3.155 \rightarrow \underline{\beta}_{2,x} \in w_2 \ \omega_1$$

$$d_1(\underline{\beta}_{3,x}) = -2.243 \qquad d_2(\underline{\beta}_{3,x}) = -3.220 \rightarrow \underline{\beta}_{3,x} \in w_1$$

$$d_1(\underline{\beta}_{4,x}) = 2.432 \qquad d_2(\underline{\beta}_{4,x}) = -9.285 \rightarrow \underline{\beta}_{4,x} \in w_1$$

طبقه‌بند آماری بیز با توزیع گوسی ویژگی‌ها

- مثال ۳– صفحه ۴۶ جلد دوم کتاب (مثال ۳–۲) (دو کلاس با ۲ آستانه و سه ناحیه)
  - سیگنال صوتی گریه نوزاد مرتبط با درد و گرسنگی
  - ضرایب مدل AR به عنوان ویژگی

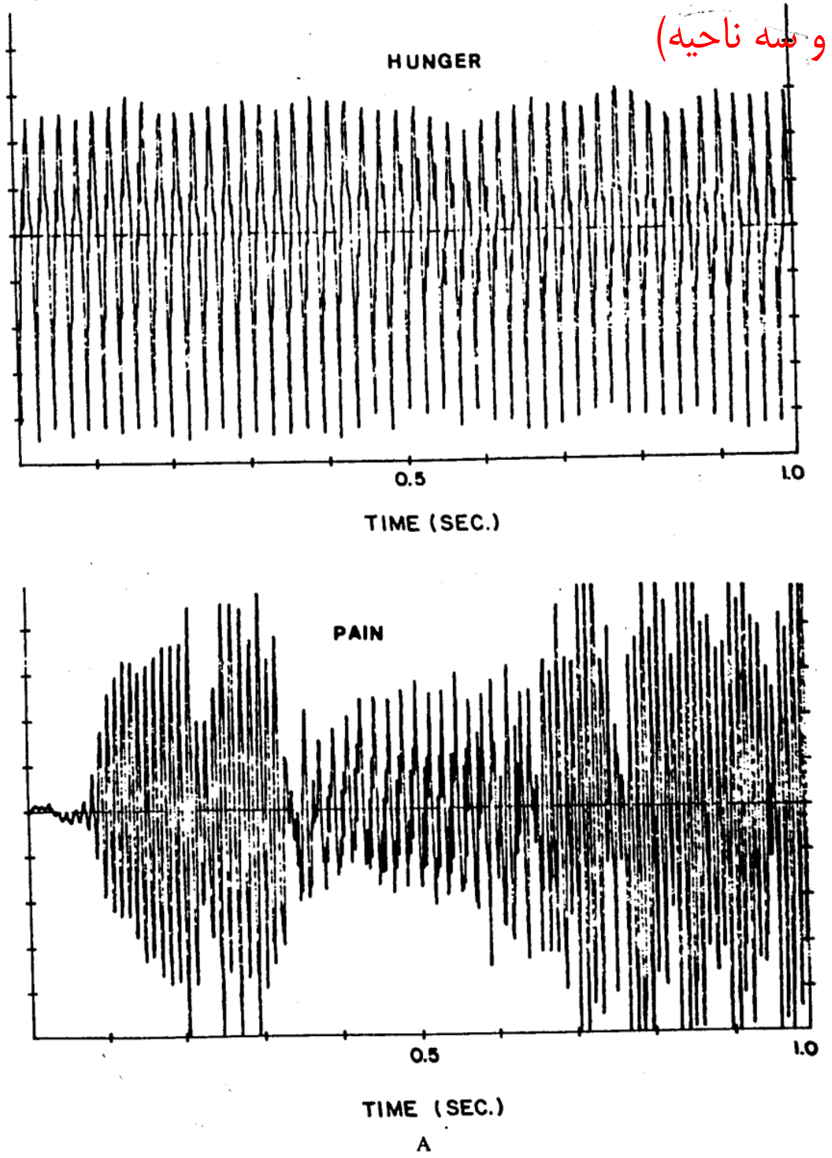
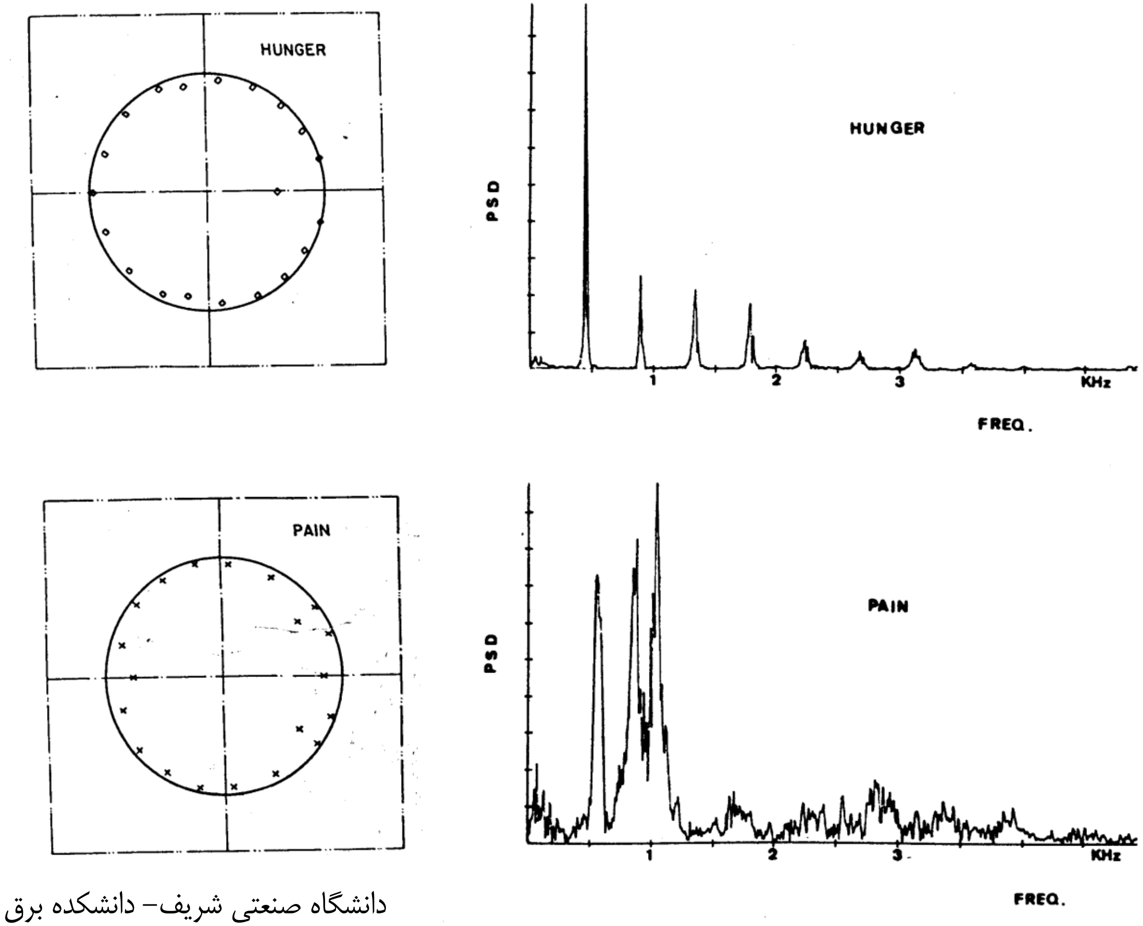


FIGURE 5. Typical hunger and pain cry. (A) Time Records; (B) power spectral density estimated by FFT.



طبقه‌بند آماری بیز با توزیع گوسی ویژگی‌ها

- مثال ۳– صفحه ۴۶ جلد دوم کتاب (مثال ۳–۲) (دو کلاس با ۲ آستانه و سه ناحیه)
  - طبقه بندی دو کلاسی بیز
    - تعیین یک آستانه
  - طبقه بندی سه کلاسی
    - تعیین دو آستانه
    - کلاس سوم: عدم تصمیم‌گیری

$$\hat{\Sigma}_i = \frac{1}{N_i}\sum_{j=1}^{N_i}(\underline{\beta}_i^j - \hat{\underline{\mu}}_i)(\underline{\beta}_i^j - \hat{\underline{\mu}}_i)^T \qquad i = H,P \tag{3.24}$$

The Bayes rule (Equation (3.5)) in its log form with the probability of Equation 3.16 for this case becomes:

$$\frac{1}{2}(\underline{\beta} - \hat{\underline{\mu}}_H)^T \hat{\Sigma}_H^{-1}(\underline{\beta} - \hat{\underline{\mu}}_H) - \frac{1}{2}(\underline{\beta} - \underline{\mu}_P)^T \Sigma_P^{-1}(\underline{\beta} - \hat{\underline{\mu}}_P) + \frac{1}{2}\ln\frac{|\hat{\Sigma}_H|}{|\hat{\Sigma}_P|} \lessgtr \ln\left(\frac{P(w_H)}{P(w_P)}\right) \rightarrow \underline{\beta} \in \begin{cases} w_H \\ w_P \end{cases} \tag{3.25}$$

where w_H and w_P denote the hunger and pain classes, respectively. Define the quadratic distance, $D_{j.\underline{\beta}}$ (Mahalanobis distance):

$$D_{j.\underline{\beta}} = (\underline{\beta} - \hat{\underline{\mu}}_j)^T \Sigma_j^{-1}(\underline{\beta} - \hat{\underline{\mu}}_j) \qquad j = H,P \tag{3.26}$$

and

$$D = D_{H.\underline{\beta}} - D_{P.\underline{\beta}} \tag{3.27}$$

and

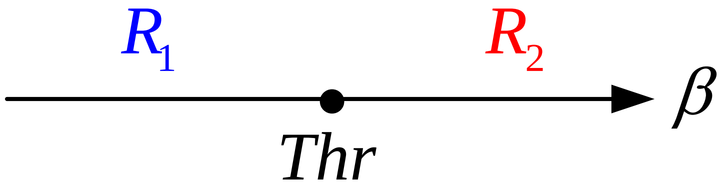
$$THR = 2\ln\left(\frac{p(w_H)}{p(w_P)}\right) - \ln\left(\frac{|\Sigma_H|}{|\Sigma_P|}\right) \tag{3.28}$$

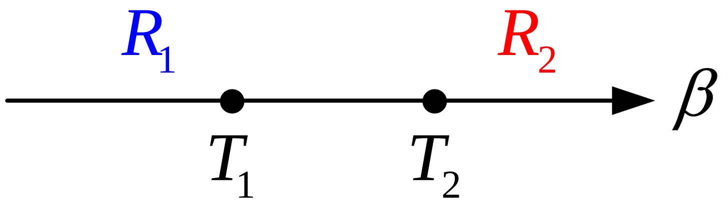
Then the decision rule (Equation 3.25) can be rewritten:

$$D \lessgtr THR \rightarrow \underline{\beta} \in \begin{cases} w_H \\ w_P \end{cases} \tag{3.29}$$

Equation 3.29 is the quadratic Bayes test for minimum error. This classifier does not allow rejects. Each data record is forced to be classified even if it is a ''bad'' record in the sense that it includes artifacts or does not belong to each one of the two classes. Consider now the case where we introduce two threshold levels, R_1 and R_2, such that:

$$D = D_{H.\underline{\beta}} - D_{P.\underline{\beta}} \begin{cases} \leqslant R_1 & \rightarrow \underline{\beta} \in w_H \\ \geqslant R_2 & \rightarrow \underline{\beta} \in w_P \\ \text{otherwise} & \rightarrow \underline{\beta} \in \text{Reject} \end{cases} \tag{3.30}$$





طبقه‌بند آماری بیز با توزیع گوسی ویژگی‌ها

- مثال ۴– دوکلاس هم‌احتمال با دو ویژگی گوسی با ماتریس کوواریانس یکسان

$$\underline{\mu}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \underline{\mu}_2 = \begin{pmatrix} 3 \\ 3 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 1.1 & 0.3 \\ 0.3 & 1.9 \end{pmatrix} \quad \underline{\beta}_x = \begin{pmatrix} 1 \\ 2.2 \end{pmatrix}$$

الف) بدون پیدا کردن مرز، به ساده‌ترین روش، برچسب بردار داده شده را بدست آورید.

– فاصله ماهالانوبیس

$$\begin{cases} d_{M1}(\underline{\beta}_x) = (\underline{\beta}_x - \underline{\mu}_1)^T \Sigma^{-1} (\underline{\beta}_x - \underline{\mu}_1) = 2.952 \\ d_{M2}(\underline{\beta}_x) = (\underline{\beta}_x - \underline{\mu}_2)^T \Sigma^{-1} (\underline{\beta}_x - \underline{\mu}_2) = 3.672 \end{cases} \Rightarrow \underline{\beta}_x \in \omega_1$$

ب) با فاصله اقلیدسی برچسب بردار داده شده را بدست آورید.

$$\begin{cases} d_{O1}(\underline{\beta}_x) = (\underline{\beta}_x - \underline{\mu}_1)^T (\underline{\beta}_x - \underline{\mu}_1) = 5.84 \\ d_{O2}(\underline{\beta}_x) = (\underline{\beta}_x - \underline{\mu}_2)^T (\underline{\beta}_x - \underline{\mu}_2) = 4.62 \end{cases} \Rightarrow \underline{\beta}_x \in \omega_2$$

پ) مرز طبقه‌بند را بدست آورید.

$$6\beta_1 + 8\beta_2 - 21 = 0$$

طبقه‌بند آماری بیز با توزیع گوسی ویژگی‌ها

- مثال ۵– تعیین مرزها و برچسب $\underline{\mu}_1=\begin{pmatrix}-4\\0\end{pmatrix},\underline{\mu}_2=\begin{pmatrix}4\\0\end{pmatrix},\underline{\mu}_3=\begin{pmatrix}0\\4\end{pmatrix},\Sigma=\begin{pmatrix}2&b\\b&2\end{pmatrix},P(\omega_i)=\frac{1}{3},\underline{\beta}_x=\begin{pmatrix}1\\2\end{pmatrix}$

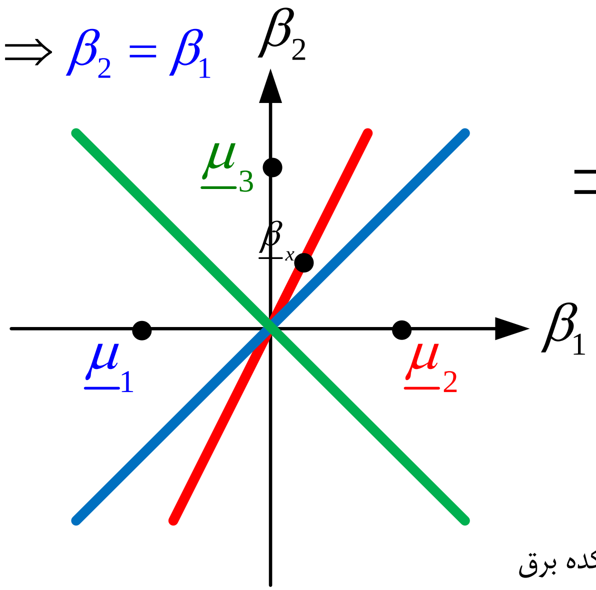
به داده مجهول: الف) اگر b=1

$$\Sigma^{-1}=\frac{1}{3}\begin{pmatrix}2&-1\\-1&2\end{pmatrix}\qquad d_i(\underline{\beta})=\underline{b}_i^T\underline{\beta}+c_i\qquad \underline{b}_i=\Sigma^{-1}\underline{\mu}_i,\quad c_i=\ln P(\omega_i)-\frac{1}{2}\underline{\mu}_i^T\Sigma^{-1}\underline{\mu}_i$$

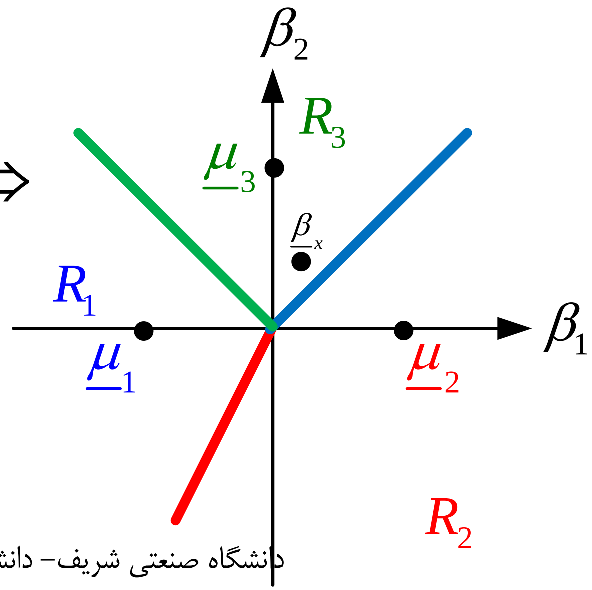
$$d_1(\underline{\beta})=\begin{pmatrix}-\frac{8}{3}&\frac{4}{3}\end{pmatrix}\begin{pmatrix}\beta_1\\\beta_2\end{pmatrix}-\frac{16}{3},\ d_2(\underline{\beta})=\begin{pmatrix}\frac{8}{3}&-\frac{4}{3}\end{pmatrix}\begin{pmatrix}\beta_1\\\beta_2\end{pmatrix}-\frac{16}{3},\ d_3(\underline{\beta})=\begin{pmatrix}-\frac{4}{3}&\frac{8}{3}\end{pmatrix}\begin{pmatrix}\beta_1\\\beta_2\end{pmatrix}-\frac{16}{3}$$

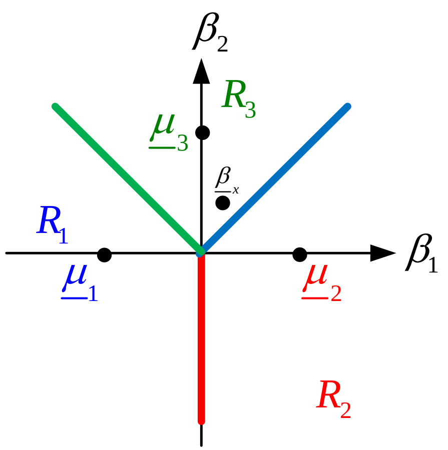
$$\begin{cases}d_1(\underline{\beta})=d_2(\underline{\beta})\Rightarrow \beta_2=2\beta_1\\ d_1(\underline{\beta})=d_3(\underline{\beta})\Rightarrow \beta_2=-\beta_1\\ d_2(\underline{\beta})=d_3(\underline{\beta})\Rightarrow \beta_2=\beta_1\end{cases}$$

ب) اگر b=0

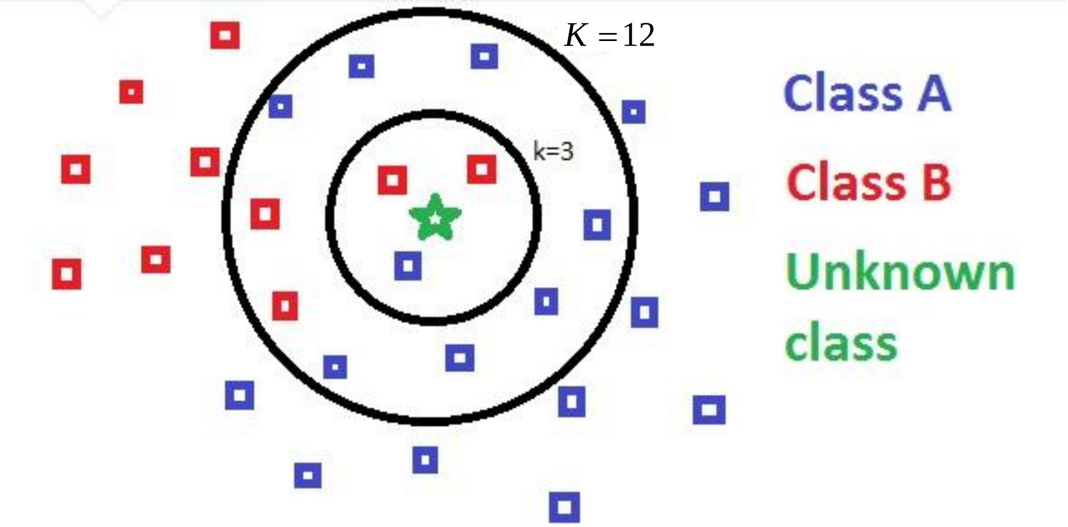


$\Rightarrow$

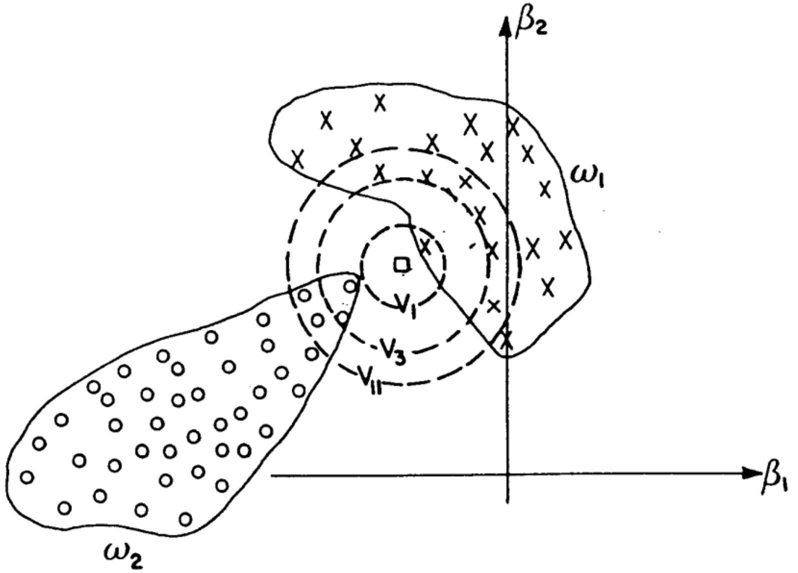




طبقه‌بند k نزدیک‌ترین همسایه k Nearest Neighbors



- طبقه‌بند k نزدیک‌ترین همسایه
  - روش ساده بدون مرحله آموزش
  - نیاز به داده‌های برچسب‌دار
  - استفاده از معیاری برای فاصله دو نقطه
  - مقدار فرد برای k در حالت دو کلاسی



کاهش بعد ویژگی‌ها

- کاهش بعد Dimension Reduction
  - انتخاب تعداد زیادی ویژگی
  - رابطه تقریبی بین ابعاد ویژگی و تعداد داده‌های آموزش (برچسب دار)
- دو روش کلی کاهش بعد
  - انتخاب ویژگی Feature Selection
  - کاهش ویژگی Feature Reduction
    - بی‌سرپرست: معیار آنتروپی/PCA
    - با سرپرست: معیار فیشر
- کاهش ویژگی: انتخاب چند ترکیب از ویژگی‌ها
  - ترکیب خطی/ترکیب غیرخطی
  - باسرپرست/بی‌سرپرست

$$\underline{\beta} \in \Re^N \Rightarrow \begin{cases} \underline{y} = A^T \underline{\beta} \\ \underline{y} \in \Re^d, A_{N\times d}, d < N \end{cases}$$

کاهش ویژگی (ترکیب خطی)

- روش بی‌سرپرست مبتنی بر ماتریس کواریانس کل داده‌ها
  - تجزیه به مقادیر ویژه و بردارهای ویژه

$$\Sigma = U\Lambda U^T$$

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_N \end{pmatrix}, U = (\underline{u}_1 \quad \underline{u}_2 \quad \cdots \quad \underline{u}_N), UU^T = U^T U = I$$

  - آنالیز مولفه‌های اساسی Principle Component Analysis (PCA)
    - تبدیل یک بردار به برداری با همان بعد با مولفه‌های ناهمبسته و با واریانس نزولی

$$\underline{y} = U^T \underline{\beta}, \quad U = (\underline{u}_1 \quad \underline{u}_2 \quad \cdots \quad \underline{u}_{N-1} \quad \underline{u}_N), 0 \le \lambda_1 < \lambda_2 < \cdots < \lambda_{N-1} < \lambda_N$$

  - روش کاهش ویژگی مبتنی بر آنالیز مولفه‌های اساسی
    - انتخاب d بردار ویژه متناظر با d بزرگترین مقدار ویژه $A = (\underline{u}_{N-d+1} \quad \underline{u}_{N-d+2} \quad \cdots \quad \underline{u}_N), A_{N\times d}$
    - مناسب برای فشرده‌سازی/عدم تضمین حفظ و بهبود تمایزپذیری
  - روش کاهش ویژگی مبتنی بر آنتروپی $A = (\underline{u}_1 \quad \underline{u}_2 \quad \cdots \quad \underline{u}_d), A_{N\times d}$
    - انتخاب d بردار ویژه متناظر با d کوچکترین مقدار ویژه
    - افزایش تمایزپذیری در صورت کوچک نبودن مقادیر ویژه انتخاب شده

کاهش ویژگی (ترکیب خطی)

- روش باسرپرست مبتنی بر ماتریس کوواریانس کلاس‌ها
  - روش Fisher’s Linear Discriminant (FLD)
  - کاهش بعد به *d=M-1*
  - ارائه روش برای دو کلاس
  - تعریف دو ماتریس در فضای اولیه ویژگی
    - ماتریس کواریانس درون کلاسی
    - ماتریس کواریانس بین کلاسی
  - تعریف دو ماتریس در فضای جدید
  - تعریف معیار فیشر
    - کم کردن پراکندگی درون کلاسی
    - زیاد کردن پراکندگی بین کلاسی
    - ماکزیمم کردن معیار فیشر
  - حل مسئله با تجزیه به مقادیر ویژه تعمیم‌یافته

$$\underline{\beta} \in \Re^N \Rightarrow \begin{cases} \underline{y} = A^T \underline{\beta} \\ \underline{y} \in \Re^d,\ A_{N\times d},\ d = M-1 \end{cases}$$

$$\underline{\beta} \in \Re^N \Rightarrow \begin{cases} y = \underline{\rho}^T \underline{\beta} \\ y \in \Re,\ \underline{\rho}_{N\times 1},\ d = 1 \end{cases}$$

$$\underline{\beta}: \begin{cases} \underline{\mu}_1, \underline{\mu}_2 \\ \Sigma_1, \Sigma_2 \end{cases} \Rightarrow \begin{cases} \Sigma_\beta = P(\omega_1)\Sigma_1 + P(\omega_2)\Sigma_2 \\ B_\beta = (\underline{\mu}_1 - \underline{\mu}_2)(\underline{\mu}_1 - \underline{\mu}_2)^T \end{cases}$$

$$y = \underline{\rho}^T \underline{\beta}: \begin{cases} \tilde{\mu}_i = \underline{\rho}^T \underline{\mu}_i \\ \tilde{\Sigma}_i = \tilde{\sigma}_i^2 = \underline{\rho}^T \Sigma_i \underline{\rho} \end{cases} \Rightarrow \begin{cases} \tilde{\mu}_1, \tilde{\mu}_2 \\ \tilde{\sigma}_1^2, \tilde{\sigma}_2^2 \end{cases}$$

$$\Rightarrow \begin{cases} \sigma_y^2 = P(\omega_1)\tilde{\sigma}_1^2 + P(\omega_2)\tilde{\sigma}_1^2 = \underline{\rho}^T \Sigma_\beta \underline{\rho} \\ b_y = (\tilde{\mu}_1 - \tilde{\mu}_2)(\tilde{\mu}_1 - \tilde{\mu}_2)^T = \underline{\rho}^T B_\beta \underline{\rho} \end{cases}$$

$$J(\underline{\rho}) = \frac{b_y}{\sigma_y^2} = \frac{\underline{\rho}^T B_\beta \underline{\rho}}{\underline{\rho}^T \Sigma_\beta \underline{\rho}}$$

$$\underline{\rho}_{opt} = \arg \underset{\underline{\rho}}{Max} J(\underline{\rho}) = \alpha\, \Sigma_\beta^{-1}(\underline{\mu}_1 - \underline{\mu}_2)$$

کاهش ویژگی (ترکیب خطی)

$$\underline{\beta} \in \Re^N \Rightarrow \begin{cases} \underline{y} = T^T \underline{\beta} \\ \underline{y} \in \Re^d, T_{N\times d}, d = M-1 \end{cases}$$

- روش باسرپرست مبتنی بر ماتریس کوواریانس کلاس‌ها برای حالت چند کلاسی
  - روش FLD
  - کاهش بعد به *d=M-1*
  - تعریف دو ماتریس در فضای اولیه ویژگی
    - ماتریس کواریانس درون کلاسی
    - ماتریس کواریانس بین کلاسی
  - تعریف دو ماتریس در فضای جدید
  - تعریف معیار فیشر
    - کم کردن پراکندگی درون کلاسی
    - زیاد کردن پراکندگی بین کلاسی
    - ماکزیمم کردن معیار فیشر
  - حل مسئله با تجزیه به مقادیر ویژه تعمیم‌یافته
    - انتخاب *d=M-1* بردار ویژه تعمیم یافته
    - متناظر با *d=M-1* بزرگترین مقدار ویژه تعمیم یافته ماتریس $\Sigma_\beta^{-1} B_\beta$

$$\underline{\beta}: \begin{cases} \underline{\mu}_i, i = 1,2,\dots,M \\ \Sigma_i, i = 1,2,\dots,M \end{cases}$$

$$\Rightarrow \begin{cases} \Sigma_\beta = \sum_{i=1}^{M} P(\omega_i)\Sigma_i \\ B_\beta = \sum_{i=1}^{M} P(\omega_i)(\underline{\mu}_i - \underline{\mu}_0)(\underline{\mu}_i - \underline{\mu}_0)^T, \quad \underline{\mu}_0 = \sum_{i=1}^{M} P(\omega_i)\underline{\mu}_i \end{cases}$$

$$\underline{y} = T^T \underline{\beta}: \begin{cases} \tilde{\underline{\mu}}_i = T^T \underline{\mu}_i, i = 1,2,\dots,M \\ \tilde{\Sigma}_i = T^T \Sigma_i T, i = 1,2,\dots,M \end{cases}$$

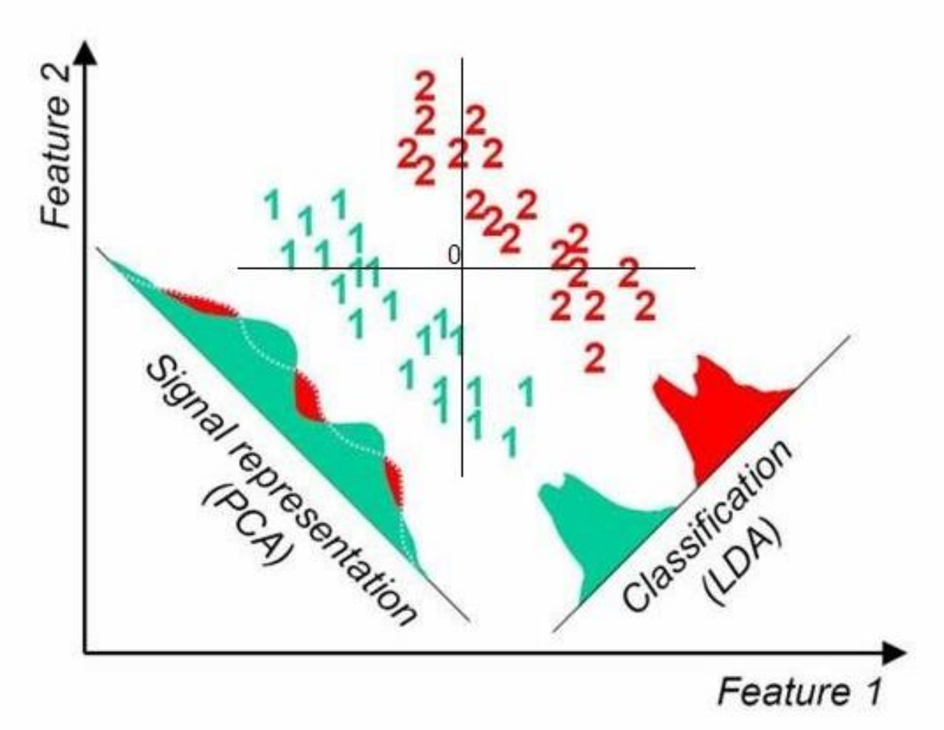
$$\Rightarrow \begin{cases} \Sigma_y = \sum_{i=1}^{M} P(\omega_i)\tilde{\Sigma}_i = T^T \Sigma_\beta T \\ B_y = \sum_{i=1}^{M} P(\omega_i)(\tilde{\underline{\mu}}_i - \tilde{\underline{\mu}}_0)(\tilde{\underline{\mu}}_i - \tilde{\underline{\mu}}_0)^T = T^T B_\beta T, \quad \tilde{\underline{\mu}}_0 = \sum_{i=1}^{M} P(\omega_i)\tilde{\underline{\mu}}_i \end{cases}$$

$$J(\underline{\rho}) = \frac{\det(B_y)}{\det(\Sigma_y)} = \frac{\det\left(T^T B_\beta T\right)}{\det\left(T^T \Sigma_\beta T\right)}$$

$$T_{opt} = \arg \underset{T}{Max} J(T)$$

کاهش ویژگی (ترکیب خطی)

- مقایسه PCA و FLD



- ارتباط LDA و FLD
  - یک روش کاهش بعد: Fisher's Linear Discriminant
  - یک طبقه بند: Linear Discriminant Analysis
    - کاهش بعد و طبقه‌بندی همزمان

انتخاب ویژگی

- انتخاب d بهترین ویژگی از بین N ویژگی
  - معیار انتخاب
  - روش جستجو

$$\underline{\beta} \in \Re^N \Rightarrow \{\underline{y} \in \Re^d, y_i = \beta_j$$

- معیار انتخاب
  - انتخاب یک ویژگی و یا یک دسته ویژگی
  - تقسیم بندی از نظر وابستگی به طبقه‌بندی کننده
    - وابسته به طبقه‌بندی کننده
    - مستقل از طبقه‌بندی کننده
      - معیارهای بی‌سرپرست
      - معیارهای باسرپرست
        - » مبتنی بر تست آماری
        - » مبتنی بر فاصله
        - » مبتنی بر ماتریس‌های پراکندگی

انتخاب ویژگی

- معیار مبتنی بر ماتریس‌های پراکندگی
  - هدف: انتخاب d ویژگی از بین N ویژگی با بیشترین پراکندگی بین کلاسی و کمترین پراکندگی درون کلاسی

$$\underline{\beta} \in \Re^N : \begin{cases} \underline{\mu}_i, i = 1, 2, \ldots, M \\ \Sigma_i, i = 1, 2, \ldots, M \end{cases} \Rightarrow \underline{y} \in \Re^d : \begin{cases} \underline{\tilde{\mu}}_i, i = 1, 2, \ldots, M \\ \tilde{\Sigma}_i, i = 1, 2, \ldots, M \end{cases}$$

$$\Rightarrow \begin{cases} S_w = \sum_{i=1}^{M} P(\omega_i) \tilde{\Sigma}_i \\ S_b = \sum_{i=1}^{M} P(\omega_i)(\underline{\tilde{\mu}}_i - \underline{\tilde{\mu}}_0)(\underline{\tilde{\mu}}_i - \underline{\tilde{\mu}}_0)^T, \quad \underline{\tilde{\mu}}_0 = \sum_{i=1}^{M} P(\omega_i) \underline{\tilde{\mu}}_i \end{cases}$$

$$J = \frac{\det(S_b)}{\det(S_w)}, \quad \frac{trace(S_b)}{trace(S_w)}, \quad \det(S_w^{-1} S_b), \quad trace(S_w^{-1} S_b)$$

- حالت خاص: انتخاب یک ویژگی
  - معیار Fischer Discriminant Ratio (FDR)
  - یکی شدن هر چهار معیار
  - برای دو کلاس هم‌احتمال
  - برای M کلاس هم‌احتمال

$$FDR = J = \frac{(\tilde{\mu}_1 - \tilde{\mu}_2)^2}{\tilde{\sigma}_1^2 + \tilde{\sigma}_2^2}$$

$$FDR = J = \sum_{i=1}^{M} \sum_{j=1}^{M} \frac{(\tilde{\mu}_i - \tilde{\mu}_j)^2}{\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2}$$

انتخاب ویژگی

- انتخاب d بهترین ویژگی از بین N ویژگی
  - معیار انتخاب
  - روش جستجو
- روش جستجو
  - روش اول: مرتب کردن تک تک ویژگی‌ها بر اساس معیار انتخاب تک ویژگی و انتخاب d تای اول
  - روش دوم: جستجوی سرتاسری
    - مرتب کردن همه دسته‌های d تایی بر اساس معیار انتخاب d ویژگی
  - روش جستجوی سرتاسری بهینه است ولی بسیار وقت‌گیر
    - اگر دو ویژگی به طور تکی دو بهترین ویژگی باشند، ترکیب انها الزاما بهترین ترکیب دوتایی نیست.
  - روش‌های زیر بهینه ولی سریع وجود دارد
    - جستجوی پسرو/پیشرو/شناور
    - جستجو به روش الگوریتم‌های تکاملی

کاهش بعد ویژگی‌ها

- مثال ۶– دوکلاس با دو ویژگی گوسی

$$\underline{\mu}_1=\begin{pmatrix}0\\0\end{pmatrix},\underline{\mu}_2=\begin{pmatrix}3\\-3\end{pmatrix},\Sigma=\begin{pmatrix}2&1\\1&2\end{pmatrix},P(\omega_1)=3P(\omega_2)$$

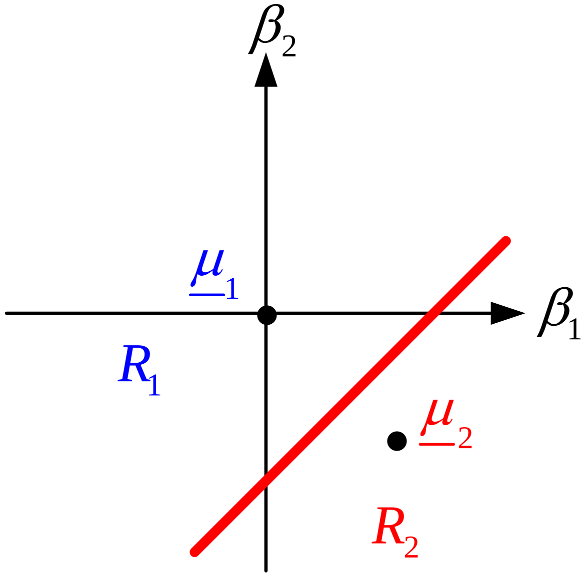
الف) مرز تصمیم‌گیری

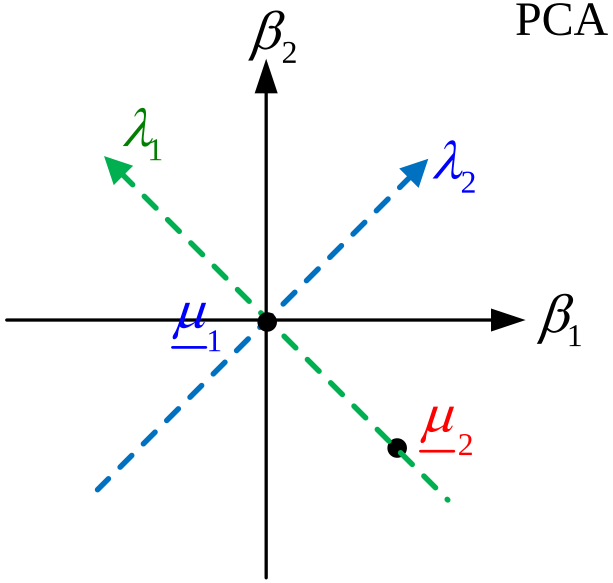
$$\Sigma^{-1}=\frac{1}{3}\begin{pmatrix}2&-1\\-1&2\end{pmatrix}\qquad d_i(\underline{\beta})=\underline{b}_i^T\underline{\beta}+c_i\qquad \underline{b}_i=\Sigma^{-1}\underline{\mu}_i,\quad c_i=\ln P(\omega_i)-\frac{1}{2}\underline{\mu}_i^T\Sigma^{-1}\underline{\mu}_i$$

$$d_1(\underline{\beta})=\begin{pmatrix}0&0\end{pmatrix}\begin{pmatrix}\beta_1\\\beta_2\end{pmatrix}+\ln\frac{3}{4},\ d_2(\underline{\beta})=\begin{pmatrix}3&-3\end{pmatrix}\begin{pmatrix}\beta_1\\\beta_2\end{pmatrix}+\ln\frac{1}{4}-9$$

$$d_1(\underline{\beta})=d_2(\underline{\beta})\Rightarrow\beta_2=\beta_1-3-\frac{1}{3}\ln 3=\beta_1-3.37$$

ب) کاهش بعد با روش آنتروپی و PCA





$$\Sigma=\begin{pmatrix}2&1\\1&2\end{pmatrix}\Rightarrow\begin{cases}\lambda_1=1\Rightarrow\underline{u}_1=\begin{pmatrix}-1\\1\end{pmatrix}\\\lambda_2=3\Rightarrow\underline{u}_2=\begin{pmatrix}1\\1\end{pmatrix}\end{cases}$$

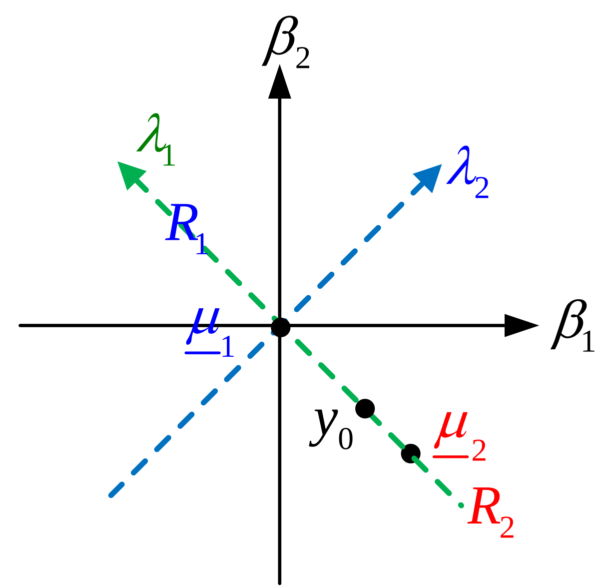
کاهش بعد ویژگی‌ها

- مثال ۶– دوکلاس با دو ویژگی گوسی

ب) کاهش بعد با روش آنتروپی و PCA

$$y=\underline{u}_1^T\underline{\beta}\Rightarrow\begin{cases}\tilde{\mu}_1=0\\ \tilde{\mu}_2=-6\\ \Sigma_y=\underline{u}_1^T\Sigma\underline{u}_1=2\end{cases}\Rightarrow d_1(y)=d_2(y)\Rightarrow y_0=-3-\frac{1}{3}\ln 3$$

$$y=\underline{u}_2^T\underline{\beta}\Rightarrow\begin{cases}\tilde{\mu}_1=0\\ \tilde{\mu}_2=0\end{cases}$$



پ) کاهش بعد با روش فیشر

$$\begin{cases}y=\underline{\rho}^T\underline{\beta}\\ \underline{\rho}=\alpha\,\Sigma_\beta^{-1}(\underline{\mu}_1-\underline{\mu}_2)=\alpha\begin{pmatrix}-3\\3\end{pmatrix}\end{cases}\Rightarrow\alpha=\frac{1}{3}\Rightarrow\begin{cases}\tilde{\mu}_1=0\\ \tilde{\mu}_2=-6\\ \Sigma_y=\underline{\rho}^T\Sigma\underline{\rho}=2\end{cases}\Rightarrow d_1(y)=d_2(y)\Rightarrow y_0=-3-\frac{1}{3}\ln 3$$

کاهش بعد ویژگی‌ها

$$\underline{\mu}_1=\begin{pmatrix}0\\0\end{pmatrix},\underline{\mu}_2=\begin{pmatrix}3\\-3\end{pmatrix},\Sigma=\begin{pmatrix}2&1\\1&2\end{pmatrix},P(\omega_1)=3P(\omega_2)$$

- مثال ۶- دوکلاس با دو ویژگی گوسی

ت) انتخاب یک ویژگی با معیار FDR

$$feature1\begin{cases}S_w=\sum_{i=1}^{M}P(\omega_i)\sigma^2=0.75*2+0.25*2=2\\ \tilde{\mu}_0=0.75*0+0.25*3=0.75\\ S_b=0.75*(0-0.75)^2+0.25*(3-0.75)^2=1.6875\end{cases}$$

$$\Rightarrow J_1=\frac{1.6875}{2}=0.8438$$

$$\Rightarrow J_2>J_1$$

$$feature2\begin{cases}S_w=\sum_{i=1}^{M}P(\omega_i)\sigma^2=0.75*2+0.25*2=2\\ \tilde{\mu}_0=0.75*0+0.25*(-3)=-0.75\\ S_b=0.75*(0+0.75)^2+0.25*(3+0.75)^2=3.9375\end{cases}$$

$$\Rightarrow J_2=\frac{3.9375}{2}=1.9688$$

ت) با فرض هم‌احتمال بودن کلاس‌ها انتخاب یک ویژگی با معیار FDR

$$J_1=\frac{(0-3)^2}{2+2}=\frac{9}{4}$$

$$J_2=\frac{(0-(-3))^2}{2+2}=\frac{9}{4}$$

$$\Rightarrow J_2=J_1$$

کاهش بعد ویژگی‌ها

$$\Sigma = \begin{pmatrix} 0.54 & -0.35 \\ -0.35 & 0.68 \end{pmatrix}$$

$$\Rightarrow \begin{cases} \lambda_1 = 0.967 \Rightarrow \underline{u}_1 = \begin{pmatrix} 0.634 \\ -0.773 \end{pmatrix} \\ \lambda_2 = 0.253 \Rightarrow \underline{u}_2 = \begin{pmatrix} 0.773 \\ 0.634 \end{pmatrix} \end{cases}$$

- مثال ۷– ادامه مثال ۲– صفحه ۴۴ جلد دوم کتاب مثال ۳–۱

الف) مرز تصمیم‌گیری $5.354\beta_1 + 5.402\beta_2 + 1.076 = 0 \Rightarrow \beta_2 \cong -\beta_1 - 0.2$

ب) کاهش بعد با روش PCA

– نیمی از داده‌های آموزش غلط طبقه‌بندی می‌شوند

– تمایزپذیری بدتر شده است

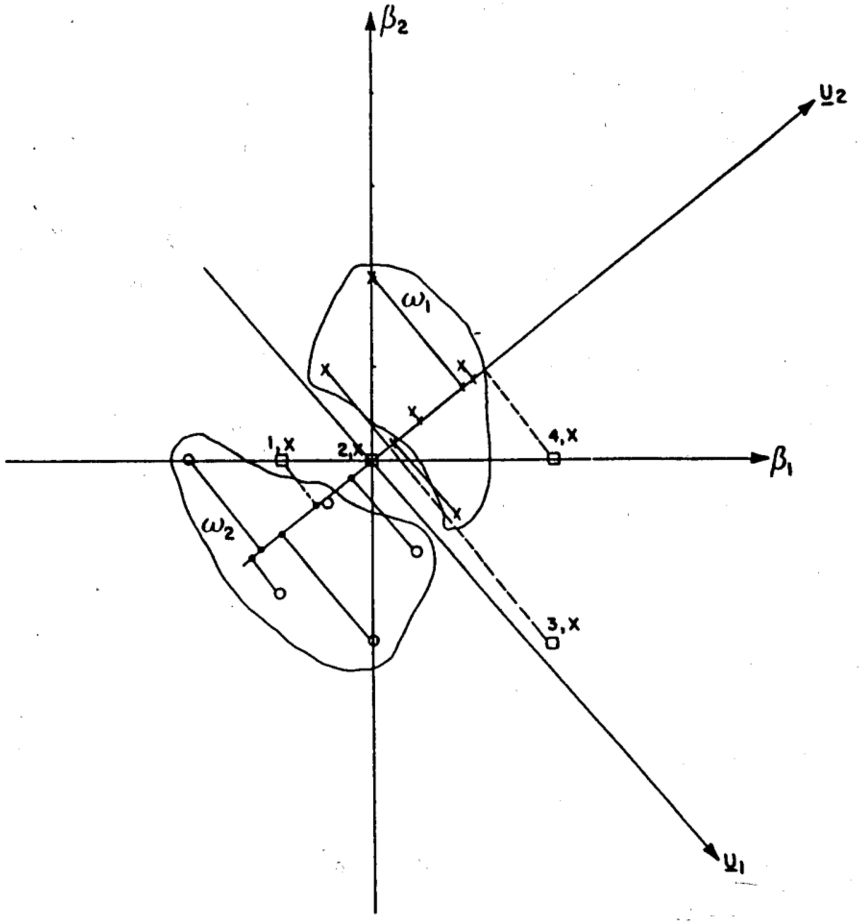
$$y = \underline{u}_1^T \underline{\beta} \Rightarrow \begin{cases} \tilde{\mu}_1 = -0.3648 \\ \tilde{\mu}_2 = 0.3926 \end{cases} \Rightarrow d_1(y) = d_2(y) \Rightarrow y_0 = 0.014$$

پ) کاهش بعد با روش آنتروپی

– همه داده‌های آموزش درست طبقه‌بندی می‌شوند

– تمایزپذیری حفظ شده است

$$y = \underline{u}_2^T \underline{\beta} \Rightarrow \begin{cases} \tilde{\mu}_1 = 0.8 \\ \tilde{\mu}_2 = -1.074 \end{cases} \Rightarrow d_1(y) = d_2(y) \Rightarrow y_0 = -0.141$$

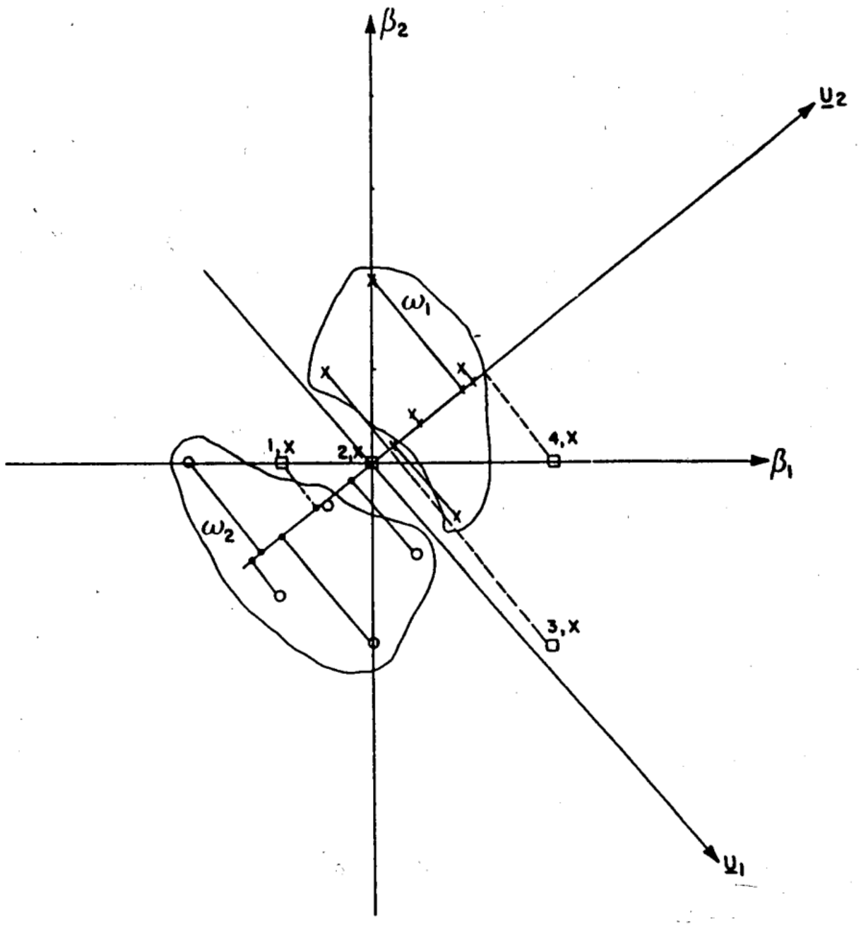


کاهش بعد ویژگی‌ها

- مثال ۷– ادامه مثال ۲– صفحه ۴۴ جلد دوم کتاب مثال ۳–۱

ت) کاهش بعد با روش فیشر

– شباهت با روش آنتروپی



$$\begin{cases} y = \underline{\rho}^T \underline{\beta} \\ \underline{\rho} = \alpha\, \Sigma_\beta^{-1}(\underline{\mu}_1 - \underline{\mu}_2) = \alpha \begin{pmatrix} 5.353 \\ 5.402 \end{pmatrix} \sim \begin{pmatrix} 0.704 \\ 0.710 \end{pmatrix} \end{cases}$$

$$\Rightarrow \begin{cases} \tilde{\mu}_1 = 6.464 \\ \tilde{\mu}_2 = -8.615 \end{cases} \Rightarrow d_1(y) = d_2(y) \Rightarrow y_0 = -1.076$$

ث) انتخاب یک ویژگی با معیار FDR

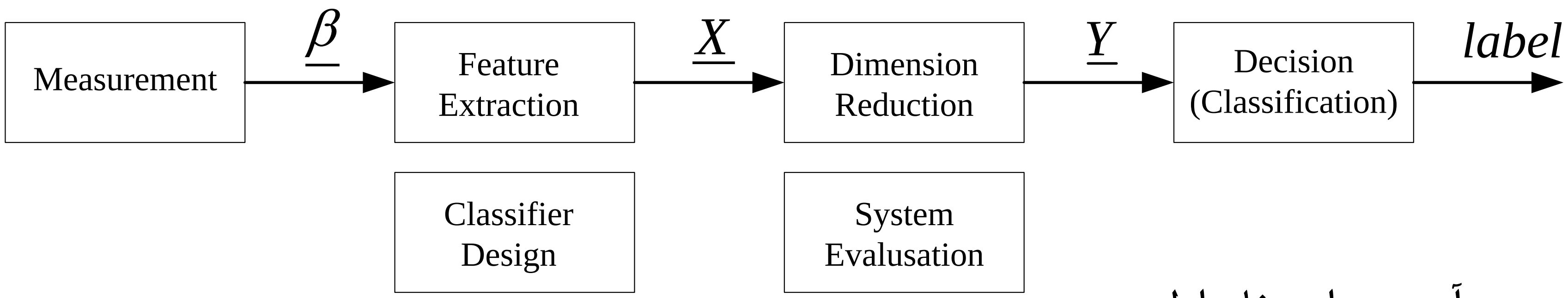
$$J_1 = \frac{(0.4-(-0.6))^2}{0.54+0.54} = 0.93$$

$$J_2 = \frac{(0.8-(-1))^2}{0.68+0.68} = 2.38$$

$$\Rightarrow J_2 > J_1$$

یادآوری مراحل مختلف پروسه طبقه‌بندی

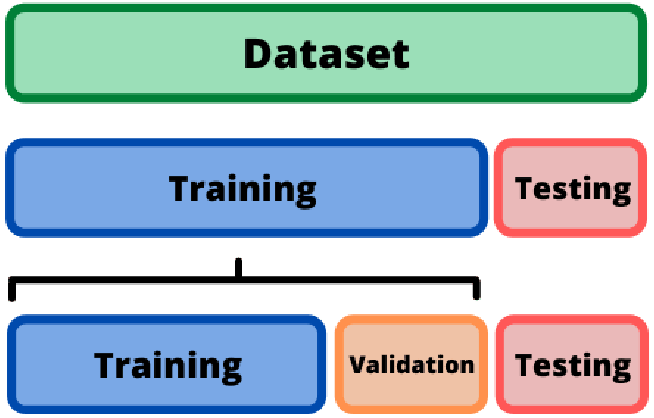
- مراحل مختلف یک پروسه طبقه‌بندی



  - اندازه‌گیری اولیه و بدست آوردن داده خام اولیه
  - استخراج ویژگی
    - اطلاعات کمی توصیف کننده داده/متمایز برای کلاس‌های مختلف
  - کاهش بعد
  - تصمیم‌گیری و تعیین برچسب داده
- دو مرحله اولیه
  - طراحی طبقه‌بند (فاز آموزش/یادگیری) و تعیین مرز کلاس‌ها و نواحی مربوط به آنها در فضای ویژگی
  - ارزیابی عمکرد سیستم طراحی شده (فاز آزمون/تست: دادن برچسب به داده‌های بدون برچسب

ارزیابی عملکرد طبقه‌بند

- تقسیم داده‌های با برچسب
  - داده آموزش/یادگیری/تعلیم Training/Learning Set
    - برای طراحی طبقه‌بند
  - داده ارزیابی/اعتبارسنجی Validation Set
    - برای انتخاب پارامترهای ثابت طبقه‌بند (مثلا C در C-SVM، k در KNN)
  - داده آزمون/آزمایش/تست Test Set
    - برای ارزیابی عملکرد طبقه‌بند
- چگونه داده‌ها را تقسیم کنیم؟
  - روش تصادفی
  - تعداد پارامترهای مدل/تعداد داده‌های بابرچسب موجود/بعد ویژگی
- عدم اطمینان به نتایج فاز آزمون
  - ضرورت تکرار
- انواع داده در مدل مسابقه

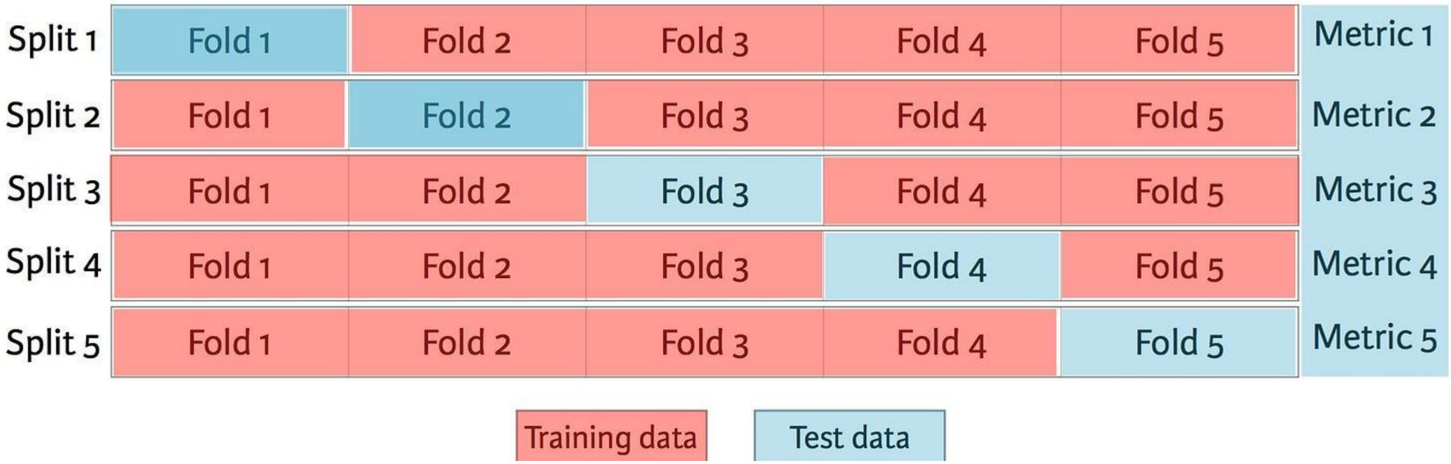


ارزیابی عمکرد طبقه‌بند

- ارزیابی متقاطع Cross Validation
  - تکرار تقسیم داده‌های برچسب‌دار به دو قسمت آموزش و آزمون
  - ارزیابی داده‌های آزمون و متوسط‌گیری روی دفعات مختلف و اعلام متوسط و انحراف معیار
  - روش‌های مختلف
    - روش Hold out
    - روش k-Fold
    - روش Leave One Out (LOO)
- داده آزمون واقعی در مسابقات
- عدم دخالت داده آزمون در هیچ مرحله از آموزش
  - هیچ بخشی از داده آزمون در هیچ بخشی از مراحل آموزش و یا تعیین پارامترهای طبقه‌بند نباید دخالت داشته باشد.

ارزیابی عمکرد طبقه‌بند

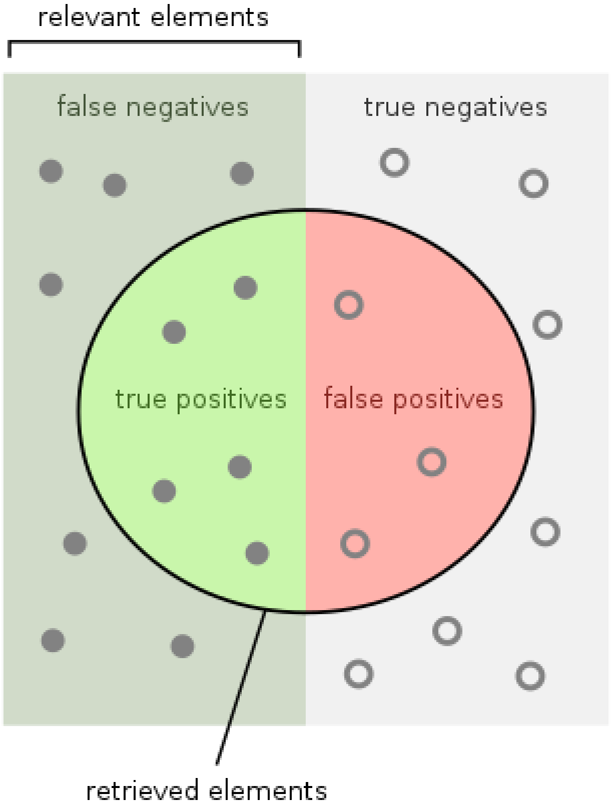
- ارزیابی متقاطع Cross Validation
  - روش k-Fold
    - بر زدن اولیه داده‌ها
    - تقسیم جداگانه برای هر کلاس



  - روش Hold out
  - روش Leave One Out (LOO)
    - مناسب برای داده‌های آموزشی کم

ارزیابی عملکرد طبقه‌بند

- ارزیابی داده‌های آزمون بعد از آموزش طبقه‌بند دوکلاسی
  - یک کلاس اصل در نظر گرفته می‌شود (مثبت)
  - تشکیل ماتریس Confusion
    - تشخیص صحیح–مثبت
      True-Positive (TP)
    - تشخیص نادرست–مثبت
      False-Positive (FP)
    - تشخیص صحیح–منفی
      True-Negative (TN)
    - تشخیص نادرست–منفی
      False-Negative (FN)



		True label		
		Class 1 (Presence)	Class 2 (Absence)	
Predicted Label	Class 1 (Presence)	TP	FP	
	Class 2 (Absence)	FN	TN	
		N1	N2	N

ارزیابی عملکرد طبقه‌بند

- معیارهای ارزیابی در طبقه‌بندی دو کلاسی
  – دقت طبقه‌بندی (صحت)
  – حساسیت
  – اختصاصیت
  – پیشگویی مثبت

$$Accuracy = C.C.R = \frac{TP+TN}{TP+TN+FP+FN} = \frac{TP+TN}{N_1+N_2} = \frac{TP+TN}{N}$$

$$Sensitivity = TP.R = \frac{TP}{TP+FN} = Accuracy\ of\ \ Presence = \color{red}{Recall}$$

$$Specifity = TN.R = \frac{TN}{TN+FP} = Accuracy\ of\ \ Absence$$

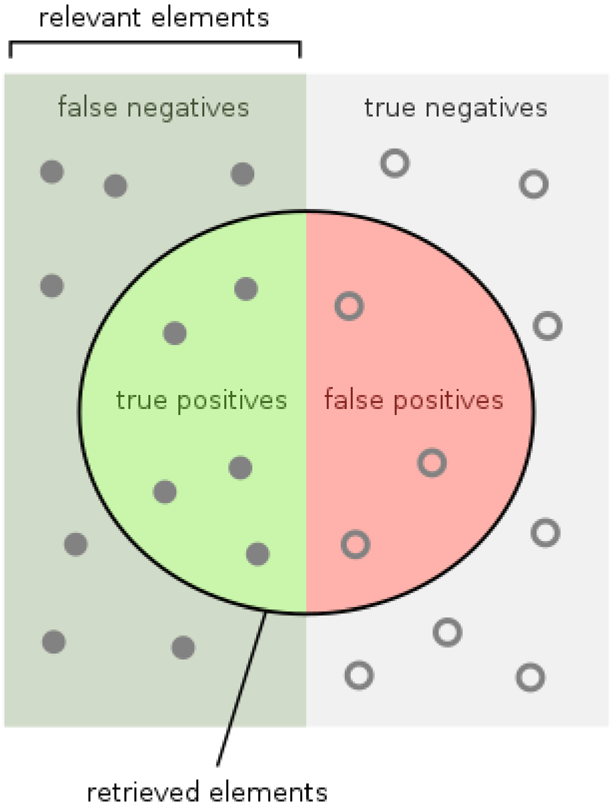
$$Positive\ Predictive\ Value = +P = \frac{TP}{TP+FP} = \color{red}{Precision}$$

$$Missed\ \ Rate = FN.R = \frac{FN}{TP+FN} = 1 - Sensitivity$$

$$FP.R = \frac{FP}{TN+FP} = 1 - Specifity$$

$$False\ \ Alarm\ Rate = \frac{FP}{TP+FP}$$

$$= 1 - Positive\ Predictive\ Value$$

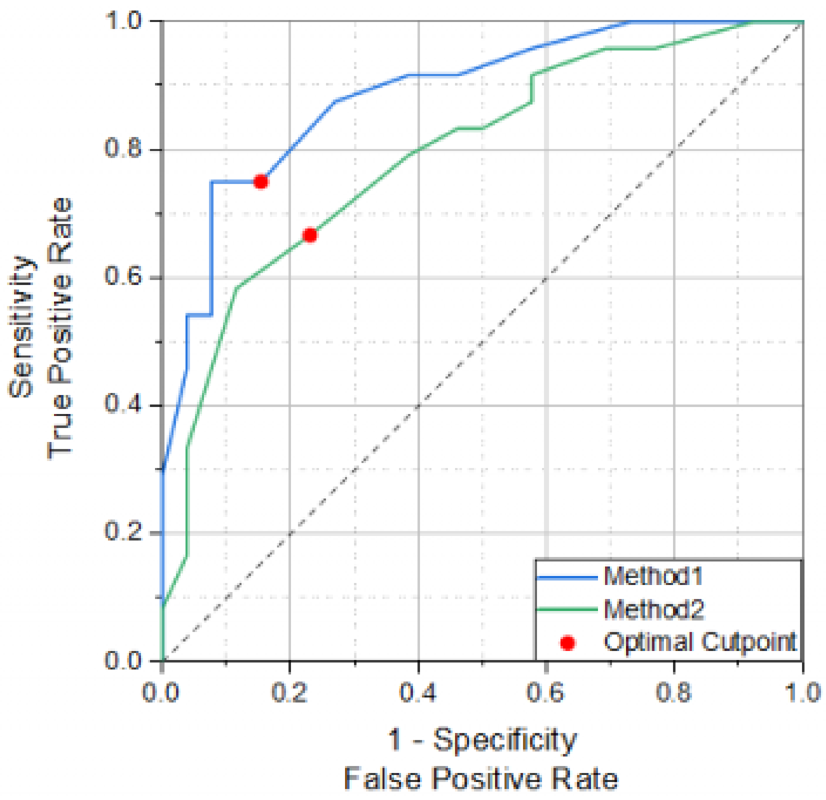


		True label	
		Class 1 (Presence)	Class 2 (Absence)
Predicted Label	Class 1 (Presence)	TP	FP
	Class 2 (Absence)	FN	TN
		N1	N2

N

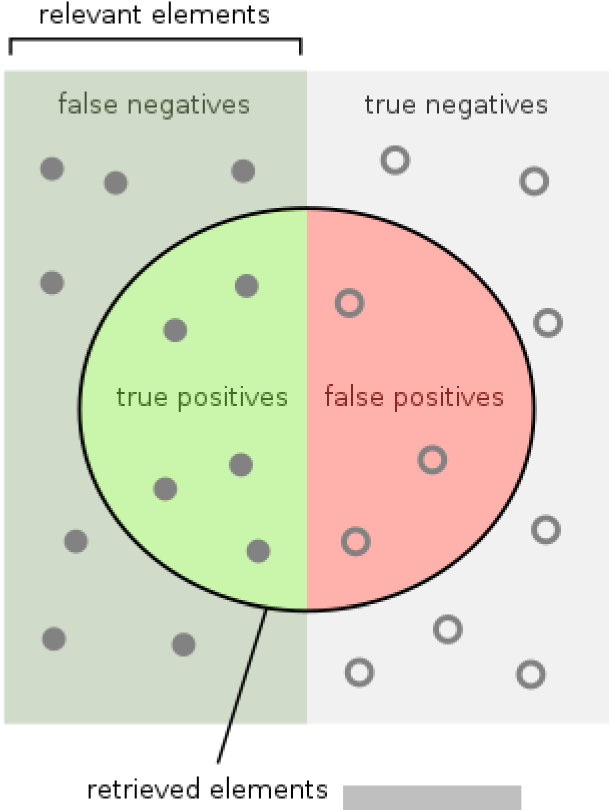
ارزیابی عمکرد طبقه‌بند

- منحنی ROC برای طبقه‌بند دو کلاسی
  - Receiver Operating Characteristic
  - خروجی احتمالاتی به جای خروجی ۰ و ۱
  - تغییر آستانه در خروجی و دادن برچسب
  - خط نیمساز: برای حالت تصادفی
  - نقطه بهینه
  - سطح زیر منحنی AUC
  - نامناسب برای داده‌های نامتعادل



		True label	
		Class 1 (Presence)	Class 2 (Absence)
Predicted Label	Class 1 (Presence)	TP	FP
	Class 2 (Absence)	FN	TN
		N1	N2	N

ارزیابی عمکرد طبقه‌بند



True label

		Class 1 (Presence)	Class 2 (Absence)	
Predicted Label	Class 1 (Presence)	TP	FP	
	Class 2 (Absence)	FN	TN	
		N1	N2	N

- معیارهای ارزیابی در طبقه‌بندی دو کلاسی
  - معیارهای خاص برای حالت داده‌های نامتعادل unbalanced data
    - معیار F1 $F-measure / balanced\ F-score / F_1\ score$

$$harmonic\ mean\ of\ \ precision\ and\ recall$$

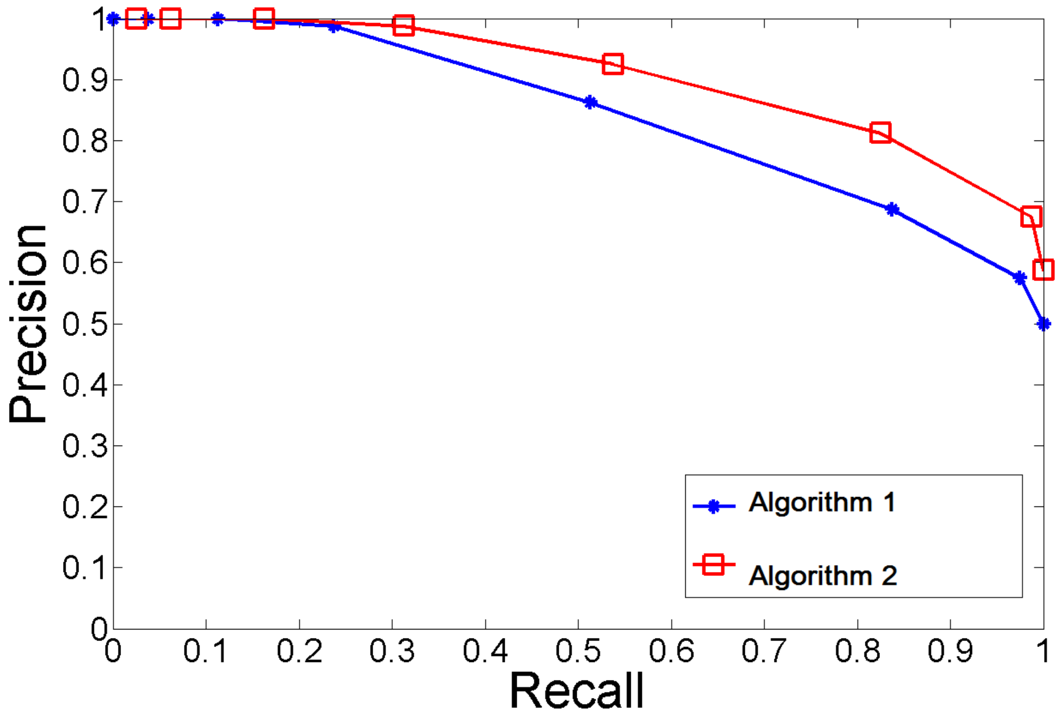
$$\frac{2}{F_1} = \frac{1}{Precision} + \frac{1}{Recall}$$

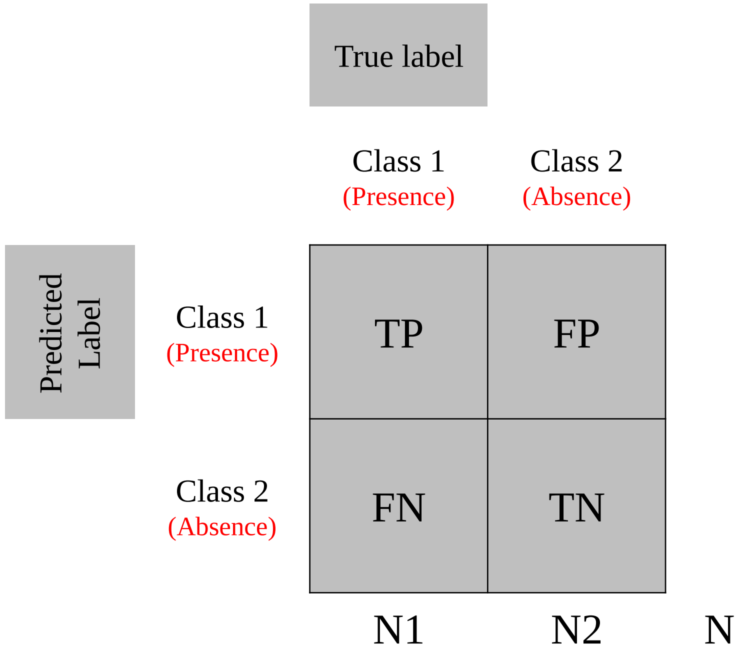
$$\Rightarrow F_1 = \frac{2}{\dfrac{1}{Precision} + \dfrac{1}{Recall}} = 2\frac{Precision.Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN}$$

  - حالت خاص داده‌های نامتعادل
    - داده‌های کلاس غیراصلی خیلی بیشتر است
    - نیازی به داده‌های TN نیست
    - معادل آشکارسازی آریتمی (آنومالی)
    - تفاوت آشکارسازی و طبقه‌بندی دو کلاسی

ارزیابی عملکرد طبقه‌بند

- منحنی PRC برای طبقه‌بند دو کلاسی با داده‌های نامتعادل
  - Precision Recall Curve
  - نقطه بهینه
  - سطح زیر منحنی AUPRC





ارزیابی عمکرد طبقه‌بند

		Predicted condition		Sources: [4][5][6][7][8][9][10][11][12] view · talk · edit		
	Total population $= P + N$	**Positive (PP)**	**Negative (PN)**	Informedness, bookmaker informedness (BM) $= TPR + TNR - 1$	Prevalence threshold (PT) $= \frac{\sqrt{TPR \times FPR} - FPR}{TPR - FPR}$
Actual condition	**Positive (P)**	**True positive (TP)**, hit	**False negative (FN)**, type II error, miss, underestimation	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power $= \frac{TP}{P} = 1 - FNR$	False negative rate (FNR), miss rate $= \frac{FN}{P} = 1 - TPR$
	Negative (N)	**False positive (FP)**, type I error, false alarm, overestimation	**True negative (TN)**, correct rejection	False positive rate (FPR), probability of false alarm, fall-out $= \frac{FP}{N} = 1 - TNR$	True negative rate (TNR), specificity (SPC), selectivity $= \frac{TN}{N} = 1 - FPR$
	Prevalence $= \frac{P}{P + N}$	Positive predictive value (PPV), precision $= \frac{TP}{PP} = 1 - FDR$	False omission rate (FOR) $= \frac{FN}{PN} = 1 - NPV$	Positive likelihood ratio (LR+) $= \frac{TPR}{FPR}$	Negative likelihood ratio (LR−) $= \frac{FNR}{TNR}$
	Accuracy (ACC) $= \frac{TP + TN}{P + N}$	False discovery rate (FDR) $= \frac{FP}{PP} = 1 - PPV$	Negative predictive value (NPV) $= \frac{TN}{PN}$ $= 1 - FOR$	Markedness (MK), deltaP (Δp) $= PPV + NPV - 1$	Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$
	Balanced accuracy (BA) $= \frac{TPR + TNR}{2}$	F_1 score $= \frac{2 PPV \times TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$	Fowlkes–Mallows index (FM) $= \sqrt{PPV \times TPR}$	Matthews correlation coefficient (MCC) $= \sqrt{TPR \times TNR \times PPV \times NPV} - \sqrt{FNR \times FPR \times FOR \times FDR}$	Threat score (TS), critical success index (CSI), Jaccard index $= \frac{TP}{TP + FN + FP}$

ارزیابی عمکرد طبقه‌بند

- مثال – در یک مسئله طبقه‌بندی دو کلاسی، در مرحله تست داریم:

$$TP = 37,\ TN = 52,\ FN = 3,\ FP = 8$$

الف) صحت، حساسیت، اختصاصیت، پیشگویی مثبت و خطای طبقه‌بندی را بدست آورید.

$$Accuracy = \frac{37+52}{37+52+3+8} = 0.89 \Rightarrow error = 1-0.89 = 0.11$$

$$Sensitivity = \frac{37}{37+3} = 0.925$$

$$Specifity = \frac{52}{52+8} = 0.867$$

$$Positive\ Predictive\ Value = \frac{37}{37+8} = 0.823$$

ب) احتمال پیشین کلاس‌ها را تخمین بزنید.

پ) آیا در یک مسئله با کلاس‌های غیرهم‌احتمال، داده‌های آزمون هم حتما باید نامساوی باشند؟

$$N1 = 37+3 = 40$$

$$N2 = 52+8 = 60$$

$$N = N1 + N2 = 100$$

$$P(\omega_1) = \frac{N1}{N} = 0.4,\ P(\omega_2) = \frac{N2}{N} = 0.6$$

ارزیابی عمکرد طبقه‌بند

- ارزیابی داده‌های آزمون بعد از آموزش طبقه‌بند چندکلاسی
  - تشکیل ماتریس Confusion

True label

Predicted Label	Class 1	Class 2	Class 3	
Class 1	N11	N21	N31	
Class 2	N12	N22	N32	
Class 3	N13	N23	N33	
	N1	N2	N3	N

- معیار اصلی ارزیابی در طبقه‌بندی چندکلاسی
  - دقت طبقه‌بندی (صحت)

$$Accuracy = C.C.R = \frac{\sum_{k=1}^{M} N_{kk}}{\sum_{i=1}^{M}\sum_{j=1}^{M} N_{ij}} = \frac{\sum_{k=1}^{M} N_{kk}}{\sum_{i=1}^{M} N_{i}} = \frac{\sum_{k=1}^{M} N_{kk}}{N}$$

- تعریف حساسیت در طبقه‌بندی چندکلاسی
  - تعریف برای هر کلاس
  - دقت طبقه‌بندی هر کلاس

$$Sensitivity_k = \frac{N_{kk}}{\sum_{i=1}^{M} N_{ik}} = \frac{N_{kk}}{N_k}$$

مثال‌هایی از طبقه‌بندی سیگنال‌های حیاتی

- طبقه‌بندی سیگنال قلبی
  - طبقه‌بندی ضربانی (سالم/بیمار یا انواع آریتمی)
  - طبقه‌بندی سیگنالی
  - ویژگی
    - سیگنالی
    - مورفولوژیکی
- طبقه‌بندی سیگنال مغزی
  - طبقه‌بندی سالم/بیمار
  - رابط مغز-رایانه
  - بازه‌های مختلف سیگنال صرعی/سطوح مختلف بیهوشی/سطوح مختلف درد
  - ویژگی
    - ویژگی‌های زمانی/فرکانسی/حوزه تبدیل/غیرخطی و آشوبی
    - ضرایب مدل‌های پارامتری
    - انرژی باندهای فرکانسی