(e) For non-pathological distributions, $P_d \propto e^{-k(1/2)}$ goes to zero as $d \to \infty$. This is because $k(1/2) \to \infty$ for $d \to \infty$.

(f) No. First note that

$$\begin{aligned} P_d(error) \leq P_d &= \sqrt{P(\omega_1)P(\omega_2)}e^{-k(1/2)} \\ P_{d+1}(error) \leq P_{d+1} &= \sqrt{P(\omega_1)P(\omega_2)}e^{-\tilde{k}(1/2)}. \end{aligned}$$

But, there is no clear relation between $P_{d+1}(error)$ and $P_d = \sqrt{P(\omega_1)(P(\omega_2)}e^{-k(1/2)}$. So, even if $\tilde{k}(1/2) > k(1/2)$, it is not guaranteed that $P_{d+1}(error) < P_d(error)$.

36. First note the definition of $k(\beta)$ given by Eq. 75 in the text:

$$\begin{aligned} k(\beta) &= \frac{\beta(1-\beta)}{2}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t(\beta\boldsymbol{\Sigma}_2 + (1-\beta)\boldsymbol{\Sigma}_1)^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) \\ &\quad + \frac{1}{2}\ln\left[\frac{|\beta\boldsymbol{\Sigma}_2 + (1-\beta)\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_2|^{\beta}|\boldsymbol{\Sigma}_1|^{1-\beta}}\right]. \end{aligned}$$

(a) Recall from Eq. 74 we have

$$\begin{aligned} e^{-k(\beta)} &= \int p^{\beta}(\mathbf{x}|\omega_1)p^{1-\beta}(\mathbf{x}|\omega_2)\, d\mathbf{x} \\ &= \int \left[\frac{\exp\left[-\frac{\beta}{2}\boldsymbol{\mu}^t\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 - \frac{(1-\beta)}{2}\boldsymbol{\mu}^t\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2\right]}{|\boldsymbol{\Sigma}_1|^{\beta/2}|\boldsymbol{\Sigma}_2|^{\beta/2}} \right. \\ &\quad \left. \times \frac{\exp\left[-\frac{1}{2}\{\mathbf{x}^t(\beta\boldsymbol{\Sigma}_1^{-1} + (1-\beta)\boldsymbol{\Sigma}_2^{-1})\mathbf{x} - 2\mathbf{x}^t(\beta\boldsymbol{\Sigma}_1^{-1} + (1-\beta)\boldsymbol{\Sigma}_2^{-1})\}\right]}{(2\pi)^{d/2}}\right] d\mathbf{x} \end{aligned}$$

(b) Again from Eq. 74 we have

$$\begin{aligned} e^{-k(\beta)} &= \frac{\exp\left[-\beta/2\; \boldsymbol{\mu}_1^t\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 - (1-\beta)/2\; \boldsymbol{\mu}_2^t\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2\right]}{|\boldsymbol{\Sigma}_1|^{\beta/2}|\boldsymbol{\Sigma}_2|^{(1-\beta)/2}} \\ &\quad \times \int \frac{\exp\left[-\frac{1}{2}\{\mathbf{x}^t\mathbf{A}^{-1}\mathbf{x} - 2\mathbf{x}\mathbf{A}^{-1}\boldsymbol{\theta}\}\right]}{(2\pi)^{d/2}} d\mathbf{x} \end{aligned}$$

where

$$\mathbf{A} = (\beta\boldsymbol{\Sigma}_1^{-1} + (1-\beta)\boldsymbol{\Sigma}_2^{-1})^{-1}$$

and

$$\mathbf{A}^{-1}\boldsymbol{\theta} = \beta\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + (1-\beta)\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2.$$

Thus we conclude that the vector $\boldsymbol{\theta}$ is

$$\boldsymbol{\theta} = \mathbf{A}(\beta\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + (1-\beta)\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2).$$

(c) For the conditions given, we have

$$\begin{aligned} \int \exp\left[\frac{1}{2}(\mathbf{x}^t\mathbf{A}^{-1}\mathbf{x} - 2\mathbf{x}^t\mathbf{A}^{-1}\boldsymbol{\theta})\right] d\mathbf{x} &= e^{\frac{1}{2}\boldsymbol{\theta}^t\mathbf{A}^{-1}\boldsymbol{\theta}} \int \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\theta})^t\mathbf{A}^{-1}(\mathbf{x}-\boldsymbol{\theta})\right] d\mathbf{x} \\ &= (2\pi)^{d/2}e^{1/2\boldsymbol{\theta}^t\mathbf{A}^{-1}\boldsymbol{\theta}}|\mathbf{A}|^{1/2} \end{aligned}$$

since

$$g(\mathbf{x}) = \frac{e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\theta})^t\mathbf{A}^{-1}(\mathbf{x}-\boldsymbol{\theta})}}{(2\pi)^{d/2}|\mathbf{A}|^{1/2}},$$

where $g(\mathbf{x})$ has the form of a d-dimensional Gaussian density. So it follows that

$$\begin{aligned} e^{-k(\beta)} &= \exp\left[-\frac{1}{2}\{-\boldsymbol{\theta}\mathbf{A}^{-1}\boldsymbol{\theta} + \beta\boldsymbol{\mu}_1^t\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu} + (1-\beta)\boldsymbol{\mu}_2^t\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2\}\right] \times \\ &\quad \frac{|\mathbf{A}|^{1/2}}{|\boldsymbol{\Sigma}_1|^{\beta/2}|\boldsymbol{\Sigma}_2|^{(1-\beta)/2}} \cdot \square \end{aligned}$$

PROBLEM NOT YET SOLVED

37. We are given that $P(\omega_1) = P(\omega_2) = 0.5$ and

$$\begin{aligned} p(\mathbf{x}|\omega_1) &\sim N(\mathbf{0}, \mathbf{I}) \\ p(\mathbf{x}|\omega_2) &\sim N(\mathbf{1}, \mathbf{I}) \end{aligned}$$

where $\mathbf{1}$ is a two-component vector of 1s.

(a) The inverse matrices are simple in this case:

$$\boldsymbol{\Sigma}_1^{-1} = \boldsymbol{\Sigma}_2^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We substitute these into Eqs. 53–55 in the text and find

$$\begin{aligned} g_1(\mathbf{x}) &= \mathbf{w}_1^t\mathbf{x} + w_{10} \\ &= \mathbf{0}^t\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + 0 + \ln(1/2) \\ &= \ln(1/2) \end{aligned}$$

and

$$\begin{aligned} g_2(\mathbf{x}) &= \mathbf{w}_2^t\mathbf{x} + w_{20} \\ &= (1,1)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \frac{1}{2}(1,1)\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \ln(1/2) \\ &= x_1 + x_2 - 1 + \ln(1/2). \end{aligned}$$

We set $g_1(\mathbf{x}) = g_2(\mathbf{x})$ and find the decision boundardy is $x_1 + x_2 = 1$, which passes through the midpoint of the two means, that is, at

$$(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}.$$

This result makes sense because these two categories have the same prior and conditional distributions except for their means.

(b) We use Eqs. 76 in the text and substitute the values given to find

$$\begin{aligned} k(1/2) &= \frac{1}{8}\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right)^t \left[\frac{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}{2}\right]^{-1} \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) + \frac{1}{2}\ln\frac{\left|\frac{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}{2}\right|}{\sqrt{\left|\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right|\left|\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right|}} \\ &= \frac{1}{8}\begin{pmatrix} 1 \\ 1 \end{pmatrix}^t\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{1}{2}\ln\frac{1}{1} \\ &= 1/4. \end{aligned}$$

Equation 77 in the text gives the Bhatacharyya bound as

$$P(error) \leq \sqrt{P(\omega_1)P(\omega_2)}e^{-k(1/2)} = \sqrt{0.5 \cdot 0.5}e^{-1/4} = 0.3894.$$

(c) Here we have $P(\omega_1) = P(\omega_2) = 0.5$ and

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad , \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix}$$

$$\boldsymbol{\mu}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad , \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}.$$

The inverse matrices are

$$\begin{aligned} \boldsymbol{\Sigma}_1^{-1} &= \begin{pmatrix} 8/5 & -2/15 \\ -2/15 & 8/15 \end{pmatrix} \\ \boldsymbol{\Sigma}_2^{-1} &= \begin{pmatrix} 5/9 & -4/9 \\ -4/9 & 5/9 \end{pmatrix}. \end{aligned}$$

We use Eqs. 66–69 and find

$$\begin{aligned} g_1(\mathbf{x}) &= -\frac{1}{2}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^t \begin{pmatrix} 8/5 & -2/15 \\ -2/15 & 8/15 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \left(\begin{pmatrix} 8/5 & -2/15 \\ -2/15 & 8/15 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right)^t \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &\quad -\frac{1}{2}\begin{pmatrix} 0 \\ 0 \end{pmatrix}^t \begin{pmatrix} 8/5 & -2/15 \\ -2/15 & 8/15 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \end{pmatrix} - \frac{1}{2}\ln\left|\begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix}\right| + \ln\frac{1}{2} \\ &= -\frac{4}{15}x_1^2 + \frac{2}{15}x_1x_2 - \frac{4}{15}x_2^2 - 0.66 + \ln\frac{1}{2}, \end{aligned}$$

and

$$\begin{aligned} g_2(\mathbf{x}) &= -\frac{1}{2}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^t \begin{pmatrix} 5/9 & -4/9 \\ -4/9 & 5/9 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \left(\begin{pmatrix} 5/9 & -4/9 \\ -4/9 & 5/9 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)^t \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &\quad -\frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}^t \begin{pmatrix} 5/9 & -4/9 \\ -4/9 & 5/9 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{2}\ln\left|\begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}\right| + \ln\frac{1}{2} \\ &= -\frac{5}{18}x_1^2 + \frac{8}{18}x_1x_2 - \frac{5}{18}x_2^2 + \frac{1}{9}x_1 + \frac{1}{9}x_2 - \frac{1}{9} - 1.1 + \ln\frac{1}{2}. \end{aligned}$$

The Bayes decision boundary is the solution to $g_1(\mathbf{x}) = g_2(\mathbf{x})$ or

$$x_1^2 + x_2^2 - 28x_1x_2 - 10x_1 - 10x_2 + 50 = 0,$$

which consists of two hyperbolas, as shown in the figure.

We use Eqs. 76 and 77 in the text and find

$$\begin{aligned} k(1/2) &= \frac{1}{8}\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right)^t \left[\frac{\begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix} + \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}}{2}\right]^{-1} \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) + \ln\frac{\left|\frac{\begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix} + \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}}{2}\right|}{\sqrt{\left|\begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix}\right|\left|\begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}\right|}} \\ &= \frac{1}{8}\begin{pmatrix} 1 \\ 1 \end{pmatrix}^t \begin{pmatrix} 3.5 & 2.25 \\ 2.25 & 3.5 \end{pmatrix}^{-1}\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{1}{2}\ln\frac{7.1875}{5.8095} \\ &= 0.1499. \end{aligned}$$

The simplest way to prove that $b_L(p)$ is a lower bound for p in the range $0 \leq p \leq 1/2$ is to note that $b_L(0) = 0$, and the derivative is always less than the derivative of $\min[p, 1-p]$. Indeed, at $p = 0$ we have

$$b_L(0) = \frac{1}{\beta}\ln\left[\frac{1+e^{-\beta}}{1+e^{-\beta}}\right] = \frac{1}{\beta}\ln[1] = 0.$$

Moreover the derivative is

$$\begin{aligned}\frac{\partial}{\partial p}b_L(p) &= \frac{e^{\beta} - e^{2\beta p}}{e^{\beta} + e^{2\beta p}} \\ &< 1 = \frac{\partial}{\partial p}\min[p, 1-p]\end{aligned}$$

in the range $0 \leq p \leq 1/2$ for $\beta < \infty$.

(b) To show that $b_L(p)$ is an arbitrarily tight bound, we need show only that in the limit $\beta \to \infty$, the derivative, $\partial b_L(p)/\partial p$ approaches 1, the same as

$$\frac{\partial}{\partial p}\min[p, 1-p]$$

in this range. Using the results from part (a) we find

$$\lim_{\beta\to\infty}\frac{\partial}{\partial p}b_L(p) = \lim_{\beta\to\infty}\frac{e^{\beta} - e^{2\beta p}}{e^{\beta} + e^{2\beta p}} = 1$$

in the range $0 \leq p < 1/2$.

(c) Our candidate upper bound is specified by

$$b_U(p) = b_L(p) + [1 - 2b_L(0.5)]b_G(p),$$

where $g_U(p)$ obeys several simple conditions, restated in part (d) below. We let $b_L(p) = p - \theta(p)$, where from part (a) we know that $\theta(p)$ is non-negative and in fact is at least linear in p. By the conditions given, we can write $b_G(p) = p + \phi(p)$, where $\phi(p)$ is non-negative and $\phi(0) = \phi(1/2) = 0$. Then our candidate upper limit obeys

$$\begin{aligned}b_U(p) &= p - \theta(p) + [1 - 2(1/2 - \theta(1/2))](p + \phi(p)) \\ &= p - \theta(p) + \theta(1/2)(p + \phi(p)).\end{aligned}$$

We show that this is an upper bound by calculating the difference between this bound and the Bayes limit (which is $\min[p, 1-p] = p$ in the range $0 \leq p \leq 1/2$). Thus we have

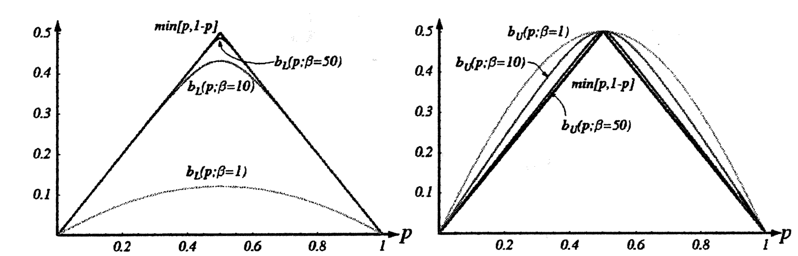
$$\begin{aligned}b_U(p) - p &= -\theta(p) + p\theta(1/2) + \theta(1/2)\phi(p) \\ &> 0.\end{aligned}$$

(d) We seek to confirm that $b_G(p) = 1/2\sin[\pi p]$ has the following four properties:

- $b_G(p) \geq \min[p, 1-p]$: Indeed, $1/2\sin[\pi p] \geq p$ for $0 \leq p \leq 1/2$, with equality holding at the extremes of the interval (that is, at $p = 0$ and $p = 1/2$). By symmetry (see below), the relation holds for the interval $1/2 \leq p \leq 1$.

- $b_G(p) = b_G(1-p)$: Indeed, the sine function is symmetric about the point $\pi/2$, that is, $1/2\ \sin[\pi/2 + \theta] = 1/2\ \sin[\pi/2 - \theta]$. Hence by a simple substitution we see that $1/2\ \sin[\pi p] = 1/2\ \sin[\pi(1-p)]$.
- $b_G(0) = b_G(1) = 0$: Indeed, $1/2\ \sin[\pi \cdot 0] = 1/2\ \sin[\pi \cdot 1] = 0$ — a special case of the fact $b_G(p) = b_G(1-p)$, as shown immediately above.
- $b_G(0.5) = 0.5$: Indeed, $1/2\ \sin[\pi 0.5] = 1/2 \cdot 1 = 0.5$.

(e) See figure.



Section 2.9

43. Here the components of the vector $\mathbf{x} = (x_1, \ldots, x_d)^t$ are binary-valued (0 or 1), and

$$p_{ij} = \Pr[x_i = 1|\omega_j] \qquad \begin{matrix} i = 1, \ldots, d \\ j = 1, \ldots, c. \end{matrix}$$

(a) Thus p_{ij} is simply the probability we get a 1 in feature x_i given that the category is ω_j. This is the kind of probability structure we find when each category has a set of independent binary features (or even real-valued features, thresholded in the form "$y_i > y_{i0}$?").

(b) The discriminant functions are then

$$g_j(\mathbf{x}) = \ln p(\mathbf{x}|\omega_j) + \ln P(\omega_j).$$

The components of $\mathbf{x}$ are statistically independent for all $\mathbf{x}$ in ω_j, then we can write the density as a product:

$$\begin{aligned}p(\mathbf{x}|\omega_j) &= p((x_1, \ldots, x_d)^t|\omega_j) \\ &= \prod_{i=1}^{d} p(x_i|\omega_j) = \prod_{i=1}^{d} p_{ij}^{x_i}(1 - p_{ij})^{1-x_i}.\end{aligned}$$

Thus, we have the discriminant function

$$\begin{aligned}g_j(\mathbf{x}) &= \sum_{i=1}^{d}[x_i \ln p_{ij} + (1 - x_i)\ln(1 - p_{ij})] + \ln P(\omega_j) \\ &= \sum_{i=1}^{d} x_i \ln\frac{p_{ij}}{1 - p_{ij}} + \sum_{i=1}^{d}\ln(1 - p_{ij}) + \ln P(\omega_j).\end{aligned}$$

We can express the probability of a Gaussian distribution in terms of the error function as:

$$P(x > x^*) = 1/2 - \text{erf}\left[\frac{x^* - \mu}{\sigma}\right]$$

and thus

$$\frac{x^* - \mu}{\sigma} = \text{erf}^{-1}\left[1/2 - P(x > x^*)\right].$$

We let $P_{hit} = P(x > x^*|x \in \omega_2)$ and $P_{false} = P(x > x^*|x \in \omega_1)$. The discriminability can be written as

$$d' = \frac{\mu_2 - \mu_1}{\sigma} = \frac{x^* - \mu_1}{\sigma} - \frac{x^* - \mu_2}{\sigma} = \text{erf}^{-1}[1/2 - P_{false}] - \text{erf}^{-1}[1/2 - P_{hit}].$$

We substitute the values for this problem and find

$$\begin{aligned} d_1' &= \text{erf}^{-1}[0.2] - \text{erf}^{-1}[-0.3] = 0.52 + 0.84 = 1.36 \\ d_2' &= \text{erf}^{-1}[0.1] - \text{erf}^{-1}[-0.2] = 0.26 + 0.52 = 0.78. \end{aligned}$$

(c) According to Eq. 70 in the text, we have

$$\text{Case 1}: \quad P(error) = \frac{1}{2}[0.3 + (1 - 0.8)] = 0.25$$

$$\text{Case 2}: \quad P(error) = \frac{1}{2}[0.4 + (1 - 0.7)] = 0.35.$$

(d) Because of the symmetry property of the ROC curve, the point (P_{hit}, P_{false}) and the point $(1 - P_{hit}, 1 - P_{false})$ will go through the same curve corresponding to some fixed d'. For case B, (0.1, 0.3) is also a point on ROC curve that (0.9, 0.7) lies. We can compare this point with case A, going through (0.8, 0.3) and the help of Fig. 2.20 in the text, we can see that case A has a higher discriminability d'.

40. We are to assume that the two Gaussians underlying the ROC curve have different variances.

(a) From the hit rate $P_{hit} = P(x > x^*|x \in \omega_2)$ we can calculate $(x^* - \mu_2)/\sigma_2$. From the false alarm rate $P_{false} = P(x > x^*|x \in \omega_1)$ we can calculate $(x^* - \mu_1)/\sigma_1$. Let us denote the ratio of the standard deviations as $\sigma_1/\sigma_2 = K$. Then we can write the discriminability in this case as

$$d_a' = \left|\frac{\mu_2 - \mu_1}{\sqrt{\sigma_1\sigma_2}}\right| = \left|\frac{\mu_2 - x^*}{\sqrt{\sigma_1\sigma_2}} - \frac{x^* - \mu_1}{\sqrt{\sigma_1\sigma_2}}\right| = \left|\frac{x^* - \mu_2}{\sigma_2/K} - \frac{x^* - \mu_1}{K\sigma_1}\right|.$$

Because we cannot determine K from $(\mu_2 - x^*)/\sigma_2$ and $(x^* - \mu_1)/\sigma_1$, we cannot determine d' uniquely with only $P_{hit} = P(x > x^*|x \in \omega_2)$ and $P_{false} = P(x > x^*|x \in \omega_1)$.

(b) Suppose we are given the following four experimental rates:

$$P_{hit1} = P(x > x_1^*|\omega_2) \qquad P_{false1} = P(x > x_1^*|\omega_1)$$
$$P_{hit2} = P(x > x_2^*|\omega_2) \qquad P_{false2} = P(x > x_2^*|\omega_1).$$

Then we can calculate the four quantities

$$a_1 = \frac{x_1^* - \mu_2}{\sigma_2} = \text{erf}^{-1}[1/2 - P_{hit1}] \qquad b_1 = \frac{x_1^* - \mu_1}{\sigma_1} = \text{erf}^{-1}[1/2 - P_{false1}] \text{ for } x_1^*$$

$$a_2 = \frac{x_2^* - \mu_2}{\sigma_2} = \text{erf}^{-1}[1/2 - P_{hit2}] \qquad b_2 = \frac{x_2^* - \mu_1}{\sigma_1} = \text{erf}^{-1}[1/2 - P_{false2}] \text{ for } x_2^*.$$

Then we have the following relations:

$$\begin{aligned} a_1 - a_2 &= \frac{x_1^* - x_2^*}{\sigma_2} \\ b_1 - b_2 &= \frac{x_1^* - x_2^*}{\sigma_1} \\ K &= \frac{a_1 - a_2}{b_1 - b_2} = \frac{\sigma_1}{\sigma_2}. \end{aligned}$$

Thus, with K we can calculate d_a' as

$$\begin{aligned} d_a' &= \left|\frac{\mu_2 - \mu_1}{\sqrt{\sigma_1\sigma_2}}\right| = \left|\frac{\mu_2 - x_1^*}{\sigma_2/K} - \frac{x_1^* - \mu_1}{K\sigma_1}\right| \\ &= \left|-\frac{(a_1 - a_2)a_1}{b_1 - b_2} - \frac{(b_1 - b_2)b_1}{a_1 - a_2}\right|. \end{aligned}$$

(c) For all those x_1^* and x_2^* that satisfy

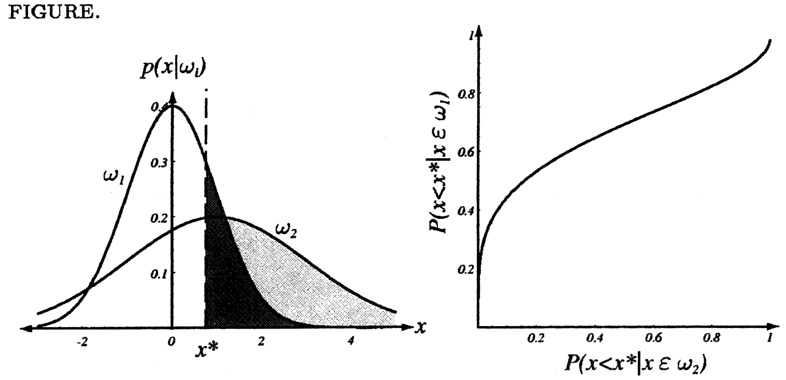
$$\frac{x_1^* - \mu_2}{\sigma_2} = -\frac{x_2^* - \mu_2}{\sigma_2}$$

or

$$\frac{x_1^* - \mu_1}{\sigma_1} = -\frac{x_2^* - \mu_1}{\sigma_1}.$$

That is, the two different thresholds do not provide any additional information and conveys the same information as only one observation. As explained in part (a), this kind of result would not allow us to determine d_a'.

(d) SEE FIGURE.



41. We use the notation shown in the figure.

We can express the probability of a Gaussian distribution in terms of the error function as:

$$P(x > x^*) = 1/2 - \text{erf}\left[\frac{x^* - \mu}{\sigma}\right]$$

and thus

$$\frac{x^* - \mu}{\sigma} = \text{erf}^{-1}\left[1/2 - P(x > x^*)\right].$$

We let $P_{hit} = P(x > x^*|x \in \omega_2)$ and $P_{false} = P(x > x^*|x \in \omega_1)$. The discriminability can be written as

$$d' = \frac{\mu_2 - \mu_1}{\sigma} = \frac{x^* - \mu_1}{\sigma} - \frac{x^* - \mu_2}{\sigma} = \text{erf}^{-1}[1/2 - P_{false}] - \text{erf}^{-1}[1/2 - P_{hit}].$$

We substitute the values for this problem and find

$$\begin{aligned} d'_1 &= \text{erf}^{-1}[0.2] - \text{erf}^{-1}[-0.3] = 0.52 + 0.84 = 1.36 \\ d'_2 &= \text{erf}^{-1}[0.1] - \text{erf}^{-1}[-0.2] = 0.26 + 0.52 = 0.78. \end{aligned}$$

(c) According to Eq. 70 in the text, we have

$$\begin{aligned} \text{Case 1}: \quad P(error) &= \frac{1}{2}[0.3 + (1 - 0.8)] = 0.25 \\ \text{Case 2}: \quad P(error) &= \frac{1}{2}[0.4 + (1 - 0.7)] = 0.35. \end{aligned}$$

(d) Because of the symmetry property of the ROC curve, the point (P_{hit}, P_{false}) and the point $(1 - P_{hit}, 1 - P_{false})$ will go through the same curve corresponding to some fixed d'. For case B, (0.1, 0.3) is also a point on ROC curve that (0.9, 0.7) lies. We can compare this point with case A, going through (0.8, 0.3) and the help of Fig. 2.20 in the text, we can see that case A has a higher discriminability d'.

40. We are to assume that the two Gaussians underlying the ROC curve have different variances.

(a) From the hit rate $P_{hit} = P(x > x^*|x \in \omega_2)$ we can calculate $(x^* - \mu_2)/\sigma_2$. From the false alarm rate $P_{false} = P(x > x^*|x \in \omega_1)$ we can calculate $(x^* - \mu_1)/\sigma_1$. Let us denote the ratio of the standard deviations as $\sigma_1/\sigma_2 = K$. Then we can write the discriminability in this case as

$$d'_a = \left|\frac{\mu_2 - \mu_1}{\sqrt{\sigma_1\sigma_2}}\right| = \left|\frac{\mu_2 - x^*}{\sqrt{\sigma_1\sigma_2}} - \frac{x^* - \mu_1}{\sqrt{\sigma_1\sigma_2}}\right| = \left|\frac{x^* - \mu_2}{\sigma_2/K} - \frac{x^* - \mu_1}{K\sigma_1}\right|.$$

Because we cannot determine K from $(\mu_2 - x^*)/\sigma_2$ and $(x^* - \mu_1)/\sigma_1$, we cannot determine d' uniquely with only $P_{hit} = P(x > x^*|x \in \omega_2)$ and $P_{false} = P(x > x^*|x \in \omega_1)$.

(b) Suppose we are given the following four experimental rates:

$$\begin{aligned} P_{hit1} &= P(x > x_1^*|\omega_2) & P_{false1} &= P(x > x_1^*|\omega_1) \\ P_{hit2} &= P(x > x_2^*|\omega_2) & P_{false2} &= P(x > x_2^*|\omega_1). \end{aligned}$$

Equation 77 in the text gives the Bhatacharyya bound as

$$P(error) \le \sqrt{P(\omega_1)P(\omega_2)}e^{-k(1/2)} = \sqrt{0.5 \cdot 0.5}e^{-1/4} = 0.3894.$$

(c) Here we have $P(\omega_1) = P(\omega_2) = 0.5$ and

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix}$$
$$\boldsymbol{\mu}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}.$$

The inverse matrices are

$$\begin{aligned} \boldsymbol{\Sigma}_1^{-1} &= \begin{pmatrix} 8/5 & -2/15 \\ -2/15 & 8/15 \end{pmatrix} \\ \boldsymbol{\Sigma}_2^{-1} &= \begin{pmatrix} 5/9 & -4/9 \\ -4/9 & 5/9 \end{pmatrix}. \end{aligned}$$

We use Eqs. 66–69 and find

$$\begin{aligned} g_1(\mathbf{x}) &= -\frac{1}{2}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^t \begin{pmatrix} 8/5 & -2/15 \\ -2/15 & 8/15 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \left(\begin{pmatrix} 8/5 & -2/15 \\ -2/15 & 8/15 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right)^t \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &\quad -\frac{1}{2}\begin{pmatrix} 0 \\ 0 \end{pmatrix}^t \begin{pmatrix} 8/5 & -2/15 \\ -2/15 & 8/15 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \end{pmatrix} - \frac{1}{2}\ln\left|\begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix}\right| + \ln\frac{1}{2} \\ &= -\frac{4}{15}x_1^2 + \frac{2}{15}x_1x_2 - \frac{4}{15}x_2^2 - 0.66 + \ln\frac{1}{2}, \end{aligned}$$

and

$$\begin{aligned} g_2(\mathbf{x}) &= -\frac{1}{2}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^t \begin{pmatrix} 5/9 & -4/9 \\ -4/9 & 5/9 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \left(\begin{pmatrix} 5/9 & -4/9 \\ -4/9 & 5/9 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)^t \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &\quad -\frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}^t \begin{pmatrix} 5/9 & -4/9 \\ -4/9 & 5/9 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{2}\ln\left|\begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}\right| + \ln\frac{1}{2} \\ &= -\frac{5}{18}x_1^2 + \frac{8}{18}x_1x_2 - \frac{5}{18}x_2^2 + \frac{1}{9}x_1 + \frac{1}{9}x_2 - \frac{1}{9} - 1.1 + \ln\frac{1}{2}. \end{aligned}$$

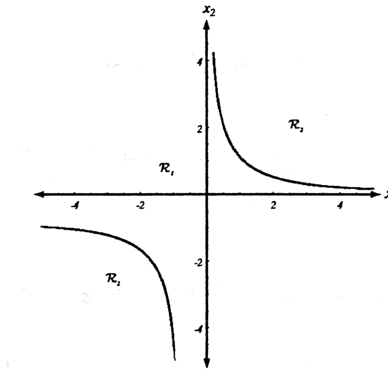
The Bayes decision boundary is the solution to $g_1(\mathbf{x}) = g_2(\mathbf{x})$ or

$$x_1^2 + x_2^2 - 28x_1x_2 - 10x_1 - 10x_2 + 50 = 0,$$

which consists of two hyperbolas, as shown in the figure.

We use Eqs. 76 and 77 in the text and find

$$\begin{aligned} k(1/2) &= \frac{1}{8}\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right)^t \left[\frac{\begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix} + \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}}{2}\right]^{-1} \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) + \ln\frac{\left|\frac{\begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix} + \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}}{2}\right|}{\sqrt{\left|\begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix}\right| \left|\begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}\right|}} \\ &= \frac{1}{8}\begin{pmatrix} 1 \\ 1 \end{pmatrix}^t \begin{pmatrix} 3.5 & 2.25 \\ 2.25 & 3.5 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{1}{2}\ln\frac{7.1875}{5.8095} \\ &= 0.1499. \end{aligned}$$



Equation 77 in the text gives the Bhatacharyya bound as

$$P(error) \le \sqrt{P(\omega_1)P(\omega_2)}e^{-k(1/2)} = \sqrt{0.5 \cdot 0.5}e^{-1.5439} = 0.4304.$$

38. We derive the Bhattacharyya error bound without first examining the Chernoff bound as follows.

(a) We wish to show that $\min[a, b] \le \sqrt{ab}$. We suppose without loss of generality that $a \le b$, or equivalently $b \le a + \delta$ for $\delta > 0$. Thus $\sqrt{ab} = \sqrt{a(a+\delta)} \ge \sqrt{a^2} = a = \min[a, b]$.

(b) Using the above result and the formula for the probability of error given by Eq. 7 in the text, we have:

$$\begin{aligned} P(error) &= \int \min\left[P(\omega_1)p(\mathbf{x}|\omega_1), P(\omega_2)p(\mathbf{x}|\omega_2)\right]\, d\mathbf{x} \\ &\le \underbrace{\sqrt{P(\omega_1)P(\omega_2)}}_{\le 1/2} \underbrace{\int \sqrt{p(\mathbf{x}|\omega_1)p(\mathbf{x}|\omega_2)}\, d\mathbf{x}}_{=\rho} \\ &\le \rho/2, \end{aligned}$$

where for the last step we have used the fact that $\min[P(\omega_1), P(\omega_2)] \le 1/2$, which follows from the normalization condition $P(\omega_1) + P(\omega_2) = 1$.

39. We assume the underlying distributions are Gaussian.

(a) Based on the Gaussian assumption, we can calculate $(x^* - \mu_2)/\sigma_2$ from the hit rate $P_{hit} = P(x > x^*|x \in \omega_2)$. We can also calculate $(x^* - \mu_1)/\sigma_1$ from the false alarm rate $P_{false} = P(x > x^*|x \in \omega_1)$. Since $\sigma_1 = \sigma_2 = \sigma$, the discriminability is simply

$$d' = \left|\frac{\mu_2 - \mu_1}{\sigma}\right| = \left|\frac{x^* - \mu_1}{\sigma_1} - \frac{x^* - \mu_2}{\sigma_2}\right|.$$

(b) Recall the error function from Eq. 96 in the Appendix of the text:

$$\mathrm{erf}[u] = \frac{2}{\sqrt{\pi}}\int_0^u e^{-x^2} dx.$$

44. The minimum probability of error is achieved by the following decision rule:

$$\text{Choose } \omega_k \text{ if } g_k(\mathbf{x}) \geq g_j(\mathbf{x}) \text{ for all } j \neq k,$$

where here we will use the discriminant function

$$g_j(\mathbf{x}) = \ln p(\mathbf{x}|\omega_j) + \ln P(\omega_j).$$

The components of $\mathbf{x}$ are statistically independent for all $\mathbf{x}$ in ω_j, and therefore,

$$p(\mathbf{x}|\omega_j) = p((x_1, \ldots, x_d)^t|\omega_j) = \prod_{i=1}^{d} p(x_i|\omega_j),$$

where

$$\begin{aligned} p_{ij} &= \Pr[x_i = \ \ 1 \ |\omega_j], \\ q_{ij} &= \Pr[x_i = \ \ 0 \ |\omega_j], \\ r_{ij} &= \Pr[x_i = -1 \ |\omega_j]. \end{aligned}$$

As in Sect. 2.9.1 in the text, we use exponents to "select" the proper probability, that is, exponents that have value 1.0 when x_i has the value corresponding to the particular probability and value 0.0 for the other values of x_i. For instance, for the p_{ij} term, we seek an exponent that has value 1.0 when $x_i = +1$ but is 0.0 when $x_i = 0$ and when $x_i = -1$. The simplest such exponent is $\frac{1}{2}x_i + \frac{1}{2}x_i^2$. For the q_{ij} term, the simplest exponent is $1 - x_i^2$, and so on. Thus we write the class-conditional probability for a single component x_i as:

$$p(x_i|\omega_j) = p_{ij}^{\frac{1}{2}x_i + \frac{1}{2}x_i^2} q_{ij}^{1-x_i^2} r_{ij}^{-\frac{1}{2}x_i + \frac{1}{2}x_i^2} \qquad \begin{array}{l} i = 1, \ldots, d \\ j = 1, \ldots, c \end{array}$$

and thus for the full vector $\mathbf{x}$ the conditional probability is

$$p(\mathbf{x}|\omega_j) = \prod_{i=1}^{d} p_{ij}^{\frac{1}{2}x_i + \frac{1}{2}x_i^2} q_{ij}^{1-x_i^2} r_{ij}^{-\frac{1}{2}x_i + \frac{1}{2}x_i^2}.$$

Thus the discriminant functions can be written as

$$\begin{aligned} g_j(\mathbf{x}) &= \ln p(\mathbf{x}|\omega_j) + \ln P(\omega_j) \\ &= \sum_{i=1}^{d} \left[\left(\frac{1}{2}x_i + \frac{1}{2}x_i^2 \right) \ln p_{ij} + (1 - x_i^2) \ln q_{ij} + \left(-\frac{1}{2}x_i + \frac{1}{2}x_i^2 \ln r_{ij} \right) \right] + \ln P(\omega_j) \\ &= \sum_{i=1}^{d} x_i^2 \ln \frac{\sqrt{p_{ij} r_{ij}}}{q_{ij}} + \frac{1}{2} \sum_{i=1}^{d} x_i \ln \frac{p_{ij}}{r_{ij}} + \sum_{i+1}^{d} \ln q_{ij} + \ln P(\omega_j), \end{aligned}$$

which are quadratic functions of the components x_i.

45. We are given that $P(\omega_1) = P(\omega_2) = 1/2$ and

$$\begin{aligned} p_{i1} &= p > 1/2 \\ p_{i2} &= 1 - p \qquad i = 1, \ldots, d, \end{aligned}$$

where d is the dimension, or number of features.