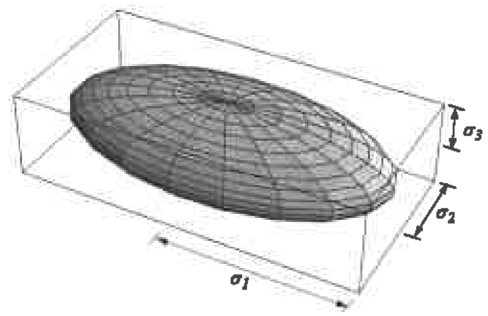


$$= \mathbf{x}^t \begin{pmatrix} 1/\sigma_1^2 & 0 & \dots & 0 \\ 0 & 1/\sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & 1/\sigma_d^2 \end{pmatrix} \mathbf{x}.$$

Thus, along each of the principal axes, the distance obeys $x_i^2 = \sigma_i^2 r^2$. Because the distance across the rectangular volume is twice that amount, the volume of the rectangular bounding box is

$$V_{rect} = (2x_1)(2x_2)\cdots(2x_d) = 2^d r^d \prod_{i=1}^{d} \sigma_i = 2^d r^d |\tilde{\mathbf{\Sigma}}|^{1/2}.$$

We let V be the (unknown) volume of the hyperellipsoid, V_d the volume of the unit hypersphere in d dimension, and V_{cube} be the volume of the d-dimensional cube having length 2 on each side. Then we have the following relation:

$$\frac{V}{V_{rect}} = \frac{V_d}{V_{cube}}.$$

We note that the volume of the hypercube is $V_{cube} = 2^d$, and substitute the above to find that

$$V = \frac{V_{rect} V_d}{V_{cube}} = r^d |\tilde{\mathbf{\Sigma}}|^{1/2} V_d,$$

where V_d is given by Eq. 47 in the text. Recall that the determinant of a matrix is unchanged by rotation of axes ($|\tilde{\mathbf{\Sigma}}|^{1/2} = |\mathbf{\Sigma}|^{1/2}$), and thus the value can be written as

$$V = r^d |\mathbf{\Sigma}|^{1/2} V_d.$$

18. Let $X_1, \dots, X_n$ be a random sample of size n from $N(\mu_1, \sigma_1^2)$ and let $Y_1, \dots, Y_m$ be a random sample of size m from $N(\mu_2, \sigma_2^2)$.

(a) Let $Z = (X_1 + \cdots + X_n) + (Y_1 + \cdots + Y_m)$. Our goal is to show that Z is also normally distributed. From the discussion in the text, if $\mathbf{X}_{d\times 1} \sim N(\boldsymbol{\mu}_{d\times 1}, \mathbf{\Sigma}_{d\times d})$

and $\mathbf{A}$ is a $k \times d$ matrix, then $\mathbf{A}^t\mathbf{X} \sim N(\mathbf{A}^t\boldsymbol{\mu}, \mathbf{A}^t\boldsymbol{\Sigma}\mathbf{A})$. Here, take

$$\mathbf{X}_{(n+m)\times 1} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \\ Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix};$$

Then, clearly $\mathbf{X}$ is normally distributed in $(n+m) \times 1$ dimensions. We can write Z as a particular matrix $\mathbf{A}^t$ operating on $\mathbf{X}$:

$$Z = X_1 + \cdots + X_n + Y_1 + \cdots + Y_m = \mathbf{1}^t\mathbf{X},$$

where $\mathbf{1}$ denotes a vector of 1's. By the above fact, it follows that Z has a univariate normal distribution.

(b) We let μ_3 be the mean of the new distribution. Then, we have

$$\begin{aligned} \mu_3 &= \mathcal{E}(Z) \\ &= \mathcal{E}[(X_1 + \cdots + X_n) + (Y_1 + \cdots + Y_m)] \\ &= \mathcal{E}(X_1) + \cdots + \mathcal{E}(X_n) + \mathcal{E}(Y_1) + \cdots + \mathcal{E}(Y_m) \\ &\quad \text{(since } X_1, \ldots, X_n, Y_1, \ldots, Y_m \text{ are independent)} \\ &= n\mu_1 + m\mu_2. \end{aligned}$$

(c) We let σ_3 be the variance of the new distribution. Then, we have

$$\begin{aligned} \sigma_3^2 &= \mathrm{Var}(Z) \\ &= \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) + \mathrm{Var}(Y_1) + \cdots + \mathrm{Var}(Y_m) \\ &\quad \text{(since } X_1, \ldots, X_n, Y_1, \ldots, Y_m \text{ are independent)} \\ &= n\sigma_1^2 + m\sigma_2^2 \end{aligned}$$

(d) Define a column vector of the samples, as:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \\ \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_m \end{pmatrix}.$$

Then, clearly $\mathbf{X}$ is $[(nd+md) \times 1]$-dimensional random variable that is normally distributed. Consider the linear projection operator $\mathbf{A}$ defined by

$$\mathbf{A}^t = \underbrace{(\mathbf{I}_{d\times d}\ \mathbf{I}_{d\times d} \cdots \mathbf{I}_{d\times d})}_{(n+m)\ times}.$$

Then we have

$$\mathbf{Z} = \mathbf{A}^t\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_n + \mathbf{Y}_1 + \cdots + \mathbf{Y}_m,$$

which must therefore be normally distributed. Furthermore, the mean and variance of the distribution are

$$\begin{aligned}
\boldsymbol{\mu}_3 = \mathcal{E}(\mathbf{Z}) &= \mathcal{E}(\mathbf{X}_1) + \cdots + \mathcal{E}(\mathbf{X}_n) + \mathcal{E}(\mathbf{Y}_1) + \cdots + \mathcal{E}(\mathbf{Y}_m) \\
&= n\boldsymbol{\mu}_1 + m\boldsymbol{\mu}_2. \\
\boldsymbol{\Sigma}_3 = \mathrm{Var}(\mathbf{Z}) &= \mathrm{Var}(\mathbf{X}_1) + \cdots + \mathrm{Var}(\mathbf{X}_n) + \mathrm{Var}(\mathbf{Y}_1) + \cdots + \mathrm{Var}(\mathbf{Y}_m) \\
&= n\boldsymbol{\Sigma}_1 + m\boldsymbol{\Sigma}_2.
\end{aligned}$$

19. The entropy is given by Eq. 37 in the text:

$$H(p(x)) = -\int p(x)\ \ln p(x) dx$$

with constraints

$$\int b_k(x)p(x)dx = a_k \qquad \text{for } k = 1, \ldots, q.$$

(a) We use Lagrange factors and find

$$\begin{aligned}
H_s &= \int p(x)\ln p(x)dx + \sum_{k=1}^{q}\left[\int b_k(x)p(x)dx - a_k\right] \\
&= -\int p(x)\left[\ln p(x) - \sum_{k=0}^{q}\lambda_k b_k(x)\right] - \sum_{k=0}^{q} a_k\lambda_k.
\end{aligned}$$

From the normalization condition $\int p(x)dx = 1$, we know that $a_0 = b_0 = 1$ for all x.

(b) In order to find the maximum or minimum value for H (having constraints), we take the derivative of H_s (having no constraints) with respect to $p(x)$ and set it to zero:

$$\frac{\partial H_s}{\partial p(x)} = -\int\left[\ln p(x) - \sum_{k=0}^{q}\lambda_k b_k(x) + 1\right] dx = 0.$$

The argument of the integral must vanish, and thus

$$\ln p(x) = \sum_{k=0}^{q}\lambda_k b_k(x) - 1.$$

We exponentiate both sides and find

$$p(x) = \exp\left[\sum_{k=0}^{q}\lambda_k b_k(x) - 1\right],$$

where the $q+1$ parameters are determined by the constraint equations.

where we used our common notation of $\mathbf{I}$ for the d-by-d identity matrix.

23. We have $p(\mathbf{x}|\omega) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{pmatrix}.$$

(a) The density at a test point $\mathbf{x}_o$ is

$$p(\mathbf{x}_o|\omega) = \frac{1}{(2\pi)^{3/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-1/2(\mathbf{x}_o - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\mathbf{x}_o - \boldsymbol{\mu})\right].$$

For this case we have

$$\begin{aligned}
|\boldsymbol{\Sigma}| &= \begin{vmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{vmatrix} = 1\begin{vmatrix} 5 & 2 \\ 2 & 5 \end{vmatrix} = 21, \\
\boldsymbol{\Sigma}^{-1} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \multicolumn{2}{c}{\multirow{2}{*}{$\begin{pmatrix} 5 & 2 \\ 2 & 5 \end{pmatrix}^{-1}$}} \\ 0 & & \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5/21 & -2/21 \\ 0 & -2/21 & 5/21 \end{pmatrix},
\end{aligned}$$

and the squared Mahalanobis distance from the mean to $\mathbf{x}_o = (.5, 0, 1)^t$ is

$$\begin{aligned}
&(\mathbf{x}_o - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\mathbf{x}_o - \boldsymbol{\mu}) \\
&= \left[\begin{pmatrix} .5 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}\right]^t \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5/21 & -2/21 \\ 0 & -2/21 & 5/21 \end{pmatrix}^{-1} \left[\begin{pmatrix} .5 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}\right] \\
&= \begin{bmatrix} -0.5 \\ -8/21 \\ -1/21 \end{bmatrix}^t \begin{bmatrix} -0.5 \\ -2 \\ -1 \end{bmatrix} = 0.25 + \frac{16}{21} + \frac{1}{21} = 1.06.
\end{aligned}$$

We substitute these values to find that the density at $\mathbf{x}_o$ is:

$$p(\mathbf{x}_o|\omega) = \frac{1}{(2\pi)^{3/2}(21)^{1/2}} \exp\left[-\frac{1}{2}(1.06)\right] = 8.16 \times 10^{-3}.$$

(b) Recall from Eq. 44 in the text that $\mathbf{A}_w = \boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2}$, where $\boldsymbol{\Phi}$ contains the normalized eigenvectors of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues. The characteristic equation, $|\boldsymbol{\Sigma} - \lambda\mathbf{I}| = 0$, in this case is

$$\begin{aligned}
\begin{vmatrix} 1-\lambda & 0 & 0 \\ 0 & 5-\lambda & 2 \\ 0 & 2 & 5-\lambda \end{vmatrix} &= (1-\lambda)\left[(5-\lambda)^2 - 4\right] \\
&= (1-\lambda)(3-\lambda)(7-\lambda) = 0.
\end{aligned}$$

The three eigenvalues are then $\lambda = 1, 3, 7$ can be read immediately from the factors. The (diagonal) $\boldsymbol{\Lambda}$ matrix of eigenvalues is thus

$$\boldsymbol{\Lambda} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 7 \end{pmatrix}.$$

Under a general linear transformation $\mathbf{T}$, we have that $\mathbf{x}' = \mathbf{T}^t\mathbf{x}$. The transformed mean is

$$\boldsymbol{\mu}' = \sum_{k=1}^{n} \mathbf{x}'_k = \sum_{k=1}^{n} \mathbf{T}^t\mathbf{x}_k = \mathbf{T}^t \sum_{k=1}^{n} \mathbf{x}_k = \mathbf{T}^t\boldsymbol{\mu}.$$

Likewise, the transformed covariance matrix is

$$\begin{aligned}
\boldsymbol{\Sigma}' &= \sum_{k=1}^{n} (\mathbf{x}'_k - \boldsymbol{\mu}')(\mathbf{x}'_k - \boldsymbol{\mu}')^t \\
&= \mathbf{T}^t \left[\sum_{k=1}^{n} (\mathbf{x}_k - \boldsymbol{\mu})(\mathbf{x}_k - \boldsymbol{\mu}) \right] \mathbf{T} \\
&= \mathbf{T}^t \boldsymbol{\Sigma} \mathbf{T}.
\end{aligned}$$

We note that $|\boldsymbol{\Sigma}'| = |\mathbf{T}^t\boldsymbol{\Sigma}\mathbf{T}| = |\boldsymbol{\Sigma}|$, and thus

$$p(\mathbf{x}_o|N(\boldsymbol{\mu}, \boldsymbol{\Sigma})) = p(\mathbf{T}^t\mathbf{x}_o|N(\mathbf{T}^t\boldsymbol{\mu}, \mathbf{T}^t\boldsymbol{\Sigma}\mathbf{T})).$$

(f) Recall the definition of a whitening transformation given by Eq. 44 in the text: $\mathbf{A}_w = \boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2}$. In this case we have

$$\mathbf{y} = \mathbf{A}_w^t\mathbf{x} \sim N(\mathbf{A}_w^t\boldsymbol{\mu}, \mathbf{A}_w^t\boldsymbol{\Sigma}\mathbf{A}_w),$$

and this implies that

$$\begin{aligned}
\mathrm{Var}(\mathbf{y}) &= \mathbf{A}_w^t(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^t\mathbf{A}_w \\
&= \mathbf{A}_w^t\boldsymbol{\Sigma}\mathbf{A} \\
&= (\boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2})^t\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^t(\boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2}) \\
&= \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Phi}^t\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^t\boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2} \\
&= \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1/2} \\
&= \mathbf{I},
\end{aligned}$$

the dentity matrix.

24. Recall that the general multivariate normal density in d-dimensions is:

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

(a) Thus we have if $\sigma_{ij} = 0$ and $\sigma_{ii} = \sigma_i^2$, then

$$\begin{aligned}
\boldsymbol{\Sigma} &= \mathrm{diag}(\sigma_1^2, \ldots, \sigma_d^2) \\
&= \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_d^2 \end{pmatrix}
\end{aligned}$$

Thus the determinant and inverse are particularly simple:

$$\begin{aligned}
|\boldsymbol{\Sigma}| &= \prod_{i=1}^{d} \sigma_i^2, \\
\boldsymbol{\Sigma}^{-1} &= \mathrm{diag}(1/\sigma_1^2, \ldots, 1/\sigma_d^2).
\end{aligned}$$

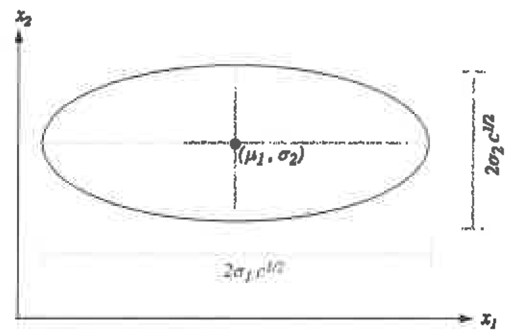
This leads to the density being expressed as:

$$
\begin{aligned}
p(\mathbf{x}) &= \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t \mathrm{diag}(1/\sigma_1^2, \ldots, 1/\sigma_d^2)(\mathbf{x}-\boldsymbol{\mu})\right] \\
&= \frac{1}{\prod\limits_{i=1}^{d} \sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\sum_{i=1}^{d}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right].
\end{aligned}
$$

(b) The contours of constant density are concentric ellipses in d dimensions whose centers are at $(\mu_1, \ldots, \mu_d)^t = \boldsymbol{\mu}$, and whose axes in the ith direction are of length $2\sigma_i\sqrt{c}$ for the density $p(\mathbf{x})$ held constant at

$$
\frac{e^{-c/2}}{\prod\limits_{i=1}^{d} \sqrt{2\pi}\sigma_i}.
$$

The axes of the ellipses are parallel to the coordinate axes. The plot in 2 dimensions ($d = 2$) is shown:



(c) The squared Mahalanobis distance from $\mathbf{x}$ to $\boldsymbol{\mu}$ is:

$$
\begin{aligned}
(\mathbf{x}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) &= (\mathbf{x}-\boldsymbol{\mu})^t \begin{pmatrix} 1/\sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_d^2 \end{pmatrix} (\mathbf{x}-\boldsymbol{\mu}) \\
&= \sum_{i=1}^{d}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2.
\end{aligned}
$$

Section 2.6

25. A useful discriminant function for Gaussians is given by Eq. 52 in the text,

$$
g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) + \ln P(\omega_i).
$$

We expand to get

$$
\begin{aligned}
g_i(\mathbf{x}) &= -\frac{1}{2}\left[\mathbf{x}^t\boldsymbol{\Sigma}^{-1}\mathbf{x} - \boldsymbol{\mu}_i^t\boldsymbol{\Sigma}^{-1}\mathbf{x} - \mathbf{x}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i + \boldsymbol{\mu}_i^t\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i\right] + \ln P(\omega_i) \\
&= -\frac{1}{2}\Big[\underbrace{\mathbf{x}^t\boldsymbol{\Sigma}^{-1}\mathbf{x}}_{\text{indep. of } i} - 2\boldsymbol{\mu}_i^t\boldsymbol{\Sigma}^{-1}\mathbf{x} + \boldsymbol{\mu}_i^t\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i\Big] + \ln P(\omega_i).
\end{aligned}
$$

and therefore we have:

$$\mathbf{w}^t(\boldsymbol{\mu}_1 - \mathbf{x}_o) > 0 \text{ and } \mathbf{w}^t(\boldsymbol{\mu}_2 - \mathbf{x}_o) > 0.$$

This last equation implies

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) > 2 \ln \left[\frac{P(\omega_1)}{P(\omega_2)}\right]$$

and

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) < -2 \ln \left[\frac{P(\omega_1)}{P(\omega_2)}\right].$$

Likewise, the conditions can be written as:

$$\mathbf{w}^t(\boldsymbol{\mu}_1 - \mathbf{x}_o) < 0 \text{ and } \mathbf{w}^t(\boldsymbol{\mu}_2 - \mathbf{x}_o) < 0$$

or

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) < 2 \ln \left[\frac{P(\omega_1)}{P(\omega_2)}\right] \text{ and}$$
$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) > -2 \ln \left[\frac{P(\omega_1)}{P(\omega_2)}\right].$$

In sum, the condition that the Bayes decision boundary does not pass between the two means can be stated as follows:

Case 1 : $P(\omega_1) \le P(\omega_2)$. Condition: $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) < 2 \ln \left[\frac{P(\omega_1)}{P(\omega_2)}\right]$ and this ensures $\mathbf{w}^t(\boldsymbol{\mu}_1 - \mathbf{x}_o) > 0$ and $\mathbf{w}^t(\boldsymbol{\mu}_2 - \mathbf{x}_o) > 0$.

Case 2 : $P(\omega_1) > P(\omega_2)$. Condition: $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) < 2 \ln \left[\frac{P(\omega_1)}{P(\omega_2)}\right]$ and this ensures $\mathbf{w}^t(\boldsymbol{\mu}_1 - \mathbf{x}_o) < 0$ and $\mathbf{w}^t(\boldsymbol{\mu}_2 - \mathbf{x}_o) < 0$.

28. We use Eqs. 42 and 43 in the text for the mean and covariance.

(a) The covariance obeys:

$$\begin{aligned}
\sigma_{ij}^2 &= \mathcal{E}\left[(x_i - \mu_i)(x_j - \mu_j)\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \underbrace{p(x_i, x_j)}_{\substack{=p(x_i)p(x_j) \\ \text{by indep.}}} (x_i - \mu_i)(x_j - \mu_j)dx_i dx_j \\
&= \int_{-\infty}^{\infty} (x_i - \mu_i)p(x_i)dx_i \int_{-\infty}^{\infty} (x_j - \mu_j)p(x_j)dx_j \\
&= 0,
\end{aligned}$$

where we have used the fact that

$$\int_{-\infty}^{\infty} x_i p(x_i)dx_i = \mu_i \qquad \text{and} \qquad \int_{-\infty}^{\infty} p(x_i)dx_i = 1.$$

(b) Suppose we had a two-dimensional Gaussian distribution, i.e.,

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \right),$$

where $\sigma_{12} = \mathcal{E}[(x_1 - \mu_1)(x_2 - \mu_2)]$. Furthermore, we have that the joint density is Gaussian, that is,

$$p(x_1, x_2) = \frac{1}{2\pi|\mathbf{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^t \mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right].$$

If $\sigma_{12} = 0$, then $|\mathbf{\Sigma}| = |\sigma_1^2 \sigma_2^2|$ and the inverse covariance matrix is diagonal, that is,

$$\mathbf{\Sigma}^{-1} = \begin{pmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{pmatrix}.$$

In this case, we can write

$$\begin{aligned} p(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{1}{2}\left\{ \left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right\} \right] \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{1}{2}\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 \right] \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{1}{2}\left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right] \\ &= p(x_1)p(x_2). \end{aligned}$$

Although we have derived this for the special case of two dimensions and $\sigma_{12} = 0$, the same method applies to the fully general case in d dimensions and two arbitrary coordinates i and j.

(c) Consider the following discrete distribution:

$$x_1 = \begin{cases} +1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2, \end{cases}$$

and a random variable x_2 conditioned on x_1 by

$$\begin{aligned} &\text{If } x_1 = +1, \quad x_2 = \begin{cases} +1/2 & \text{with probability } 1/2 \\ -1/2 & \text{with probability } 1/2. \end{cases} \\ &\text{If } x_1 = -1, \quad x_2 = 0 \text{ with probability } 1. \end{aligned}$$

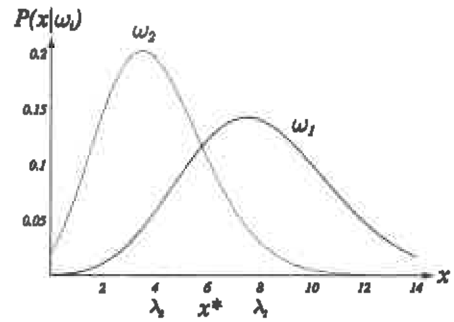
It is simple to verify that $\mu_1 = \mathcal{E}(x_1) = 0$; we use that fact in the following calculation:

$$\begin{aligned} \mathrm{Cov}(x_1, x_2) &= \mathcal{E}\left[(x_1 - \mu_1)(x_2 - \mu_2) \right] \\ &= \mathcal{E}[x_1 x_2] - \mu_2 \mathcal{E}[x_1] - \mu_1 \mathcal{E}[x_2] - \mathcal{E}[\mu_1 \mu_2] \\ &= \mathcal{E}[x_1 x_2] - \mu_1 \mu_2 \\ &= \frac{1}{2} P(x_1 = +1, x_2 = +1/2) + \left(-\frac{1}{2} \right) P(x_1 = +1, x_2 = -1/2) \\ &\quad +0 \cdot P(x_1 = -1) \\ &= 0. \end{aligned}$$

Thus the Bayes decision rule is

$$\begin{array}{rll} \text{Choose } \omega_2 & \text{if} & e^{\lambda_2-\lambda_1}\left(\dfrac{\lambda_1}{\lambda_1}\right)^x > 1, \\ \text{or equivalently} & \text{if} & x < \dfrac{(\lambda_2-\lambda_1)}{\ln[\lambda_1]-\ln[\lambda_2]}. \\ \text{Choose } \omega_1 & & \text{otherwise,} \end{array}$$

as illustrated in the figure (where the x values are discrete).



(e) The conditional Bayes error rate is

$$P(error|x) = \min\left[e^{-\lambda_1}\frac{\lambda_1^x}{x!}, e^{-\lambda_2}\frac{\lambda_2^x}{x!}\right].$$

The Bayes error, given the decision rule in part (d) is

$$P_B(error) = \sum_{x=0}^{x^*} e^{\lambda_2}\frac{\lambda_2^x}{x!} + \sum_{x=x^*}^{\infty} e^{-\lambda_1}\frac{\lambda_1^x}{x!},$$

where $x^* = \lfloor(\lambda_2-\lambda_1)/(\ln[\lambda_1]-\ln[\lambda_2])\rfloor$.

Section 2.10

40 ~~48~~. In two dimensions, the Gaussian distribution is

$$p(\mathbf{x}|\omega_i) = \frac{1}{2\pi|\boldsymbol{\Sigma}_i|^{1/2}}\exp\left[-1/2(\mathbf{x}-\boldsymbol{\mu}_i)^t\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right].$$

(a) By direct calculation using the densities stated in the problem, we find that for $\mathbf{x} = \binom{.3}{.3}$ that $p(\mathbf{x}|\omega_1)P(\omega_1) = 0.04849$, $p(\mathbf{x}|\omega_2)P(\omega_2) = 0.03250$ and $p(\mathbf{x}|\omega_3)P(\omega_3) = 0.04437$, and thus the pattern should be classified as category ω_1.

(b) To classify $\binom{*}{.3}$, i.e., a vector whose first component is missing and its second component is 0.3, we need to marginalize over the unknown feature. Thus we compute numerically

$$P(\omega_i)p\left(\left(\begin{matrix}*\\ .3\end{matrix}\right)\bigg|\omega_i\right) = P(\omega_i)\int_{-\infty}^{\infty} p\left(\left(\begin{matrix}x\\ .3\end{matrix}\right)\bigg|\omega_i\right)\,dx$$

and find that $P(\omega_1)p((*,.3)^t|\omega_1) = 0.12713$, $P(\omega_1)p((*,.3)^t|\omega_2) = 0.10409$, and $P(\omega_1)p((*,.3)^t|\omega_3) = 0.13035$. Thus the pattern should be categorized as ω_3.

(c) As in part (a), we calculate numerically

$$P(\omega_i)\tilde{p}\left(\begin{pmatrix}.3\\ *\end{pmatrix}\Big|\omega_i\right) = P(\omega_i)\int_{-\infty}^{\infty} p\left(\begin{pmatrix}.3\\ y\end{pmatrix}\Big|\omega_i\right)\,dy$$

and find that $P(\omega_1)p((.3,*)^t|\omega_1) = 0.12713$, $P(\omega_1)p((.3,*)^t|\omega_2) = 0.10409$, and $P(\omega_1)p((.3,*)^t|\omega_3) = 0.11346$. Thus the pattern should be categorized as ω_1.

(d) We follow the above procedure:

$\mathbf{x} = (.2, .6)^t$

- $P(\omega_1)p(\mathbf{x}|\omega_1) = 0.04344.$
- $P(\omega_2)p(\mathbf{x}|\omega_2) = 0.03556.$
- $P(\omega_3)p(\mathbf{x}|\omega_3) = 0.04589.$

Thus $\mathbf{x} = (.2, .6)^t$ should be categorized as ω_3.

$\mathbf{x} = (*, .6)^t$

- $P(\omega_1)p(\mathbf{x}|\omega_1) = 0.11108.$
- $P(\omega_2)p(\mathbf{x}|\omega_2) = 0.12276.$
- $P(\omega_3)p(\mathbf{x}|\omega_3) = 0.13232.$

Thus $\mathbf{x} = (*, .6)^t$ should be categorized as ω_3.

$\mathbf{x} = (.2, *)^t$

- $P(\omega_1)p(\mathbf{x}|\omega_1) = 0.11108.$
- $P(\omega_2)p(\mathbf{x}|\omega_2) = 0.12276.$
- $P(\omega_3)p(\mathbf{x}|\omega_3) = 0.10247.$

Thus $\mathbf{x} = (*, .6)^t$ should be categorized as ω_2.

49. PROBLEM NOT YET SOLVED

Section 2.11

50. We use the values from Example 4 in the text.

(a) For this case, the probabilities are:

$$\begin{aligned}
P(a_1) &= P(a_4) = 0.5\\
P(a_2) &= P(a_3) = 0\\
P(b_1) &= 1\\
P(b_2) &= 0\\
P(d_1) &= 0\\
P(d_2) &= 1.
\end{aligned}$$

Then using Eq. 99 in the text we have

$$\begin{aligned}
P_{\mathcal{P}}(x_1) &\sim P(x_1|a_1,b_1)P(a_1)P(b_1) + 0 + 0 + 0 + 0 + 0 + P(x_1|a_4,b_1)P(a_4)P(b_1) + 0\\
&= \frac{0.9\cdot 0.65}{0.9\cdot 0.65 + 0.1\cdot 0.35}\cdot 0.5\cdot 1 + \frac{0.8\cdot 0.65}{0.8\cdot 0.65 + 0.2\cdot 0.35}0.5\cdot 1\\
&= 0.472 + 0.441\\
&= 0.913.
\end{aligned}$$