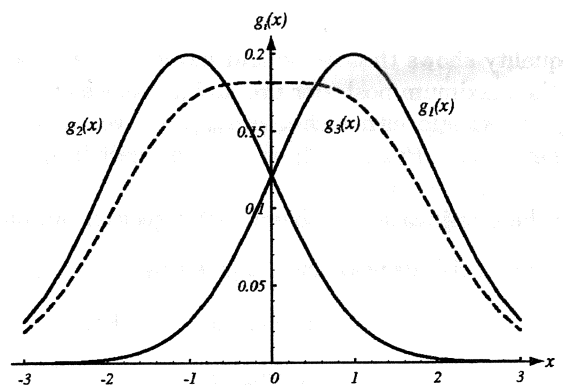
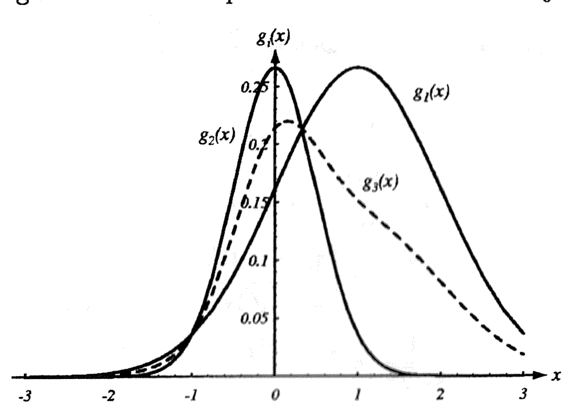


(c) If $\lambda_r/\lambda_s = 0$, there is no cost in rejecting as unrecognizable. Furthermore, $P(\omega_i|\mathbf{x}) \geq 1 - \lambda_r/\lambda_s$ is never satisfied if $\lambda_r/\lambda_s = 0$. In that case, the decision rule will always reject as unrecognizable. On the other hand, as $\lambda_r/\lambda_s \to 1$, $P(\omega_i|\mathbf{x}) \geq 1-\lambda_r/\lambda_s$ is always satisfied (there is a high cost of not recognizing) and hence the decision rule is the Bayes decision rule of choosing the class ω_i that maximizes the posterior probability $P(\omega_i|\mathbf{x})$.

(d) Consider the case $p(x|\omega_1) \sim N(1,1)$, $p(x|\omega_2) \sim N(0,1/4)$, $P(\omega_1) = 1/3$, $P(\omega_2) = 2/3$ and $\lambda_r/\lambda_s = 1/2$. In this case, the discriminant functions of part (a) give

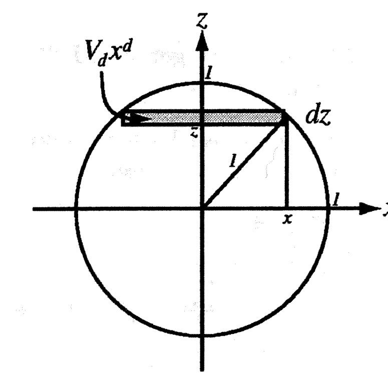
$$
\begin{aligned}
g_1(x) &= p(x|\omega_1)P(\omega_1) = \frac{2}{3}\frac{e^{-(x-1)^2/2}}{\sqrt{2\pi}} \\
g_2(x) &= p(x|\omega_2)P(\omega_2) = \frac{1}{3}\frac{2e^{-2x^2}}{\sqrt{2\pi}} \\
g_3(x) &= \left(1 - \frac{\lambda_r}{\lambda_s}\right)[p(x|\omega_1)P(\omega_1) + p(x|\omega_2)P(\omega_2)] \\
&= \frac{1}{2}\cdot\frac{2}{3}\left[\frac{e^{-(x-1)^2/2}}{\sqrt{2\pi}} + \frac{e^{-2x^2}}{\sqrt{2\pi}}\right] \\
&= \frac{1}{3\sqrt{2\pi}}\left[e^{-(x-1)^2/2} + e^{-2x^2}\right] = \frac{1}{2}[g_1(x) + g_2(x)].
\end{aligned}
$$

Note from the figure that for this problem we should never reject.



Section 2.5

15. We consider the volume of a d-dimensional hypersphere of radius 1.0, and more generally radius x, as shown in the figure.



(a) We use Eq. 47 in the text for $d =$ odd, that is, $V_d = 2^d\pi^{(d-1)/2}(\frac{d-1}{2})!/d!$. When applied to $d = 1$ (a line) we have $V_1 = 2^1\pi^0 1 = 2$. Indeed, a line segment $-1 \leq x \leq +1$ has generalized volume (length) of 2. More generally, a line of "radius" x has volume of $2x$.

(b) We use Eq. 47 in the text for $d =$ even, that is, $V_d = \pi^{(d/2)}/(d/2)!$. When applied to $d = 2$ (a disk), we have $V_2 = \pi^1/1! = \pi$. Indeed, a disk of radius 1 has generalized volume (area) of π. More generally, a disk of radius x has volume of πx^2.

(c) Given the volume of a line in $d = 1$, we can derive the volume of a disk by straightforward integration. As shown in the figure, we have

$$V_2 = 2\int_0^1 \sqrt{1-z^2}\, dz = \pi,$$

as we saw in part (a).

(d) As can be seen in the figure, to find the volume of a generalized hypersphere in $d+1$ dimensions, we merely integrate along the z (new) dimension the volume of a generalized hypersphere in the d-dimensional space, with proper factors and limits. Thus we have:

$$V_{d+1} = 2\int_0^1 V_d(1-z^2)^{d/2}dz = \frac{V_d\sqrt{\pi}\Gamma(d/2+1)}{\Gamma(d/2+3/2)},$$

where for integer k the gamma function obeys

$$\Gamma(k+1) = k! \text{ and } \Gamma(k+1/2) = 2^{-2k+1}\sqrt{\pi}(2k-1)!/(k-1)!.$$

(e) Using this formula for $d = 2k$ even, and V_d given for even dimensions, we get that for the next higher (odd) dimension d^*:

$$V_{d^*} = V_{d+1} = \frac{2\pi^{d/2}}{(d/2)!}\left[\frac{\sqrt{\pi}}{2}\frac{(d/2)!}{\Gamma(d/2+3/2)}\right]$$

$$
\begin{aligned}
&= \frac{\pi^{d/2}\ k!\ 2^{2k+1}}{(2k+1)!} \\
&= \frac{\pi^{(d^*-1)/2}(\frac{d^*-1}{2})!\ 2^{d^*}}{(d^*)!},
\end{aligned}
$$

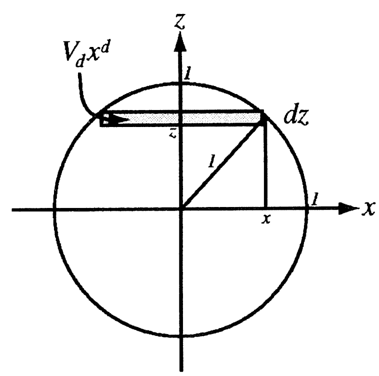
where we have used $2k = d$ for some integer k and $d^* = d + 1$. This confirms Eq. 47 for odd dimension given in the text.

(f) We repeat the above steps, but now use V_d for d odd, in order to derive the volume of the hypersphere in an even dimension:

$$
\begin{aligned}
V_{d+1} &= V_d \frac{\sqrt{\pi}}{2} \frac{\Gamma(\frac{d}{2}+1)}{\Gamma(\frac{d}{2}+\frac{3}{2})} \\
&= \frac{2^d\ \pi^{(d-1)/2}\ (\frac{d-1}{2})!}{d!}\ \frac{\sqrt{\pi}}{2}\ \frac{\Gamma((k+1)+\frac{1}{2})}{(k+1)!} \\
&= \frac{\pi^{d^*/2}}{(d^*/2)!},
\end{aligned}
$$

where we have used that for odd dimension $d = 2k + 1$ for some integer k, and $d^* = d+1$ is the (even) dimension of the higher space. This confirms Eq. 47 for even dimension given in the text.

16. We approach the problem analogously to problem 15, and use the same figure.



(a) The "volume" of a line from $-1 \leq x \leq 1$ is indeed $V_1 = 2$.

(b) Integrating once for the general case (according to the figure) gives

$$
V_{d+1} = 2 \int_0^1 V_d (1 - z^2)^{d/2} dz = \frac{V_d \sqrt{\pi}\Gamma(d/2+1)}{\Gamma(d/2+3/2)},
$$

where for integer k the gamma function obeys

$$
\Gamma(k+1) = k! \text{ and } \Gamma(k+1/2) = 2^{-2k+1}\sqrt{\pi}(2k-1)!/(k-1)!.
$$

Integrating again thus gives:

$$
V_{d+2} = \underbrace{V_d \left[\frac{\sqrt{\pi}\Gamma(d/2+1)}{\Gamma(d/2+3/2)}\right]}_{V_{d+1}} \left[\frac{\sqrt{\pi}\Gamma((d+1)/2+1)}{\Gamma((d+1)/2+3/2)}\right]
$$

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

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

To find the eigenvectors, we solve $\mathbf{\Sigma x} = \lambda_i \mathbf{x}$ for $(i = 1, 2, 3)$:

$$\mathbf{\Sigma x} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{pmatrix} \mathbf{x} = \begin{pmatrix} x_1 \\ 5x_2 + 2x_3 \\ 2x_2 + 5x_3 \end{pmatrix} = \lambda_i \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad i = 1, 2, 3.$$

The three eigenvectors are given by:

$$\begin{aligned}
\lambda_1 &= 1: \begin{pmatrix} x_1 \\ 5x_2 + 2x_3 \\ 2x_2 + 5x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \Rightarrow \boldsymbol{\phi}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \\
\lambda_2 &= 3: \begin{pmatrix} x_1 \\ 5x_2 + 2x_3 \\ 2x_2 + 5x_3 \end{pmatrix} = \begin{pmatrix} 3x_1 \\ 3x_2 \\ 3x_3 \end{pmatrix} \Rightarrow \boldsymbol{\phi}_2 = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}, \\
\lambda_3 &= 7: \begin{pmatrix} x_1 \\ 5x_2 + 2x_3 \\ 2x_2 + 5x_3 \end{pmatrix} = \begin{pmatrix} 7x_1 \\ 7x_2 \\ 7x_3 \end{pmatrix} \Rightarrow \boldsymbol{\phi}_3 = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}.
\end{aligned}$$

Thus our final $\mathbf{\Phi}$ and $\mathbf{A}_w$ matrices are:

$$\mathbf{\Phi} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

and

$$\begin{aligned}
\mathbf{A}_w = \mathbf{\Phi\Lambda}^{-1/2} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 7 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{6} & 1/\sqrt{14} \\ 0 & -1/\sqrt{6} & 1/\sqrt{14} \end{pmatrix}.
\end{aligned}$$

We have then, $\mathbf{Y} = \mathbf{A}_w^t(\mathbf{x} - \boldsymbol{\mu}) \sim N(\mathbf{0}, \mathbf{I})$.

(c) The transformed point is found by applying $\mathbf{A}$, that is,

$$\begin{aligned}
\mathbf{x}_w &= \mathbf{A}_w^t(\mathbf{x}_o - \boldsymbol{\mu}) \\
&= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{6} & 1/\sqrt{14} \\ 0 & -1/\sqrt{6} & 1/\sqrt{14} \end{pmatrix} \begin{pmatrix} -0.5 \\ -2 \\ -1 \end{pmatrix} = \begin{pmatrix} -0.5 \\ -1/\sqrt{6} \\ -3/\sqrt{14} \end{pmatrix}.
\end{aligned}$$

(d) From part (a), we have that the squared Mahalanobis distance from $\mathbf{x}_o$ to $\boldsymbol{\mu}$ in the original coordinates is $r^2 = (\mathbf{x}_o - \boldsymbol{\mu})^t\mathbf{\Sigma}^{-1}(\mathbf{x}_o - \boldsymbol{\mu}) = 1.06$. The Mahalanobis distance from $\mathbf{x}_w$ to $\mathbf{0}$ in the transformed coordinates is $\mathbf{x}_w^t\mathbf{x}_w = (0.5)^2 + 1/6 + 3/14 = 1.06$. The two distances are the same, as they must be under any linear transformation.

(e) A Gaussian distribution is written as

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

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

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

We note that $|\mathbf{\Sigma}'| = |\mathbf{T}^t\mathbf{\Sigma T}| = |\mathbf{\Sigma}|$ for transformations such as translation and rotation, and thus

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

The volume element is proportial to $|\mathbf{T}|$ and for transformations such as scaling, the transformed covariance is proportional to $|\mathbf{T}|^2$, so the transformed normalization contant contains $1/|\mathbf{T}|$, which exactly compensates for the change in volume.

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

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

and this implies that

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

the identity matrix.

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

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

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

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

To find the eigenvectors, we solve $\mathbf{\Sigma x} = \lambda_i \mathbf{x}$ for $(i = 1, 2, 3)$:

$$\mathbf{\Sigma x} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{pmatrix} \mathbf{x} = \begin{pmatrix} x_1 \\ 5x_2 + 2x_3 \\ 2x_2 + 5x_3 \end{pmatrix} = \lambda_i \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad i = 1, 2, 3.$$

The three eigenvectors are given by:

$$\begin{aligned}
\lambda_1 &= 1: & \begin{pmatrix} x_1 \\ 5x_2 + 2x_3 \\ 2x_2 + 5x_3 \end{pmatrix} &= \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \Rightarrow \boldsymbol{\phi}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \\
\lambda_2 &= 3: & \begin{pmatrix} x_1 \\ 5x_2 + 2x_3 \\ 2x_2 + 5x_3 \end{pmatrix} &= \begin{pmatrix} 3x_1 \\ 3x_2 \\ 3x_3 \end{pmatrix} \Rightarrow \boldsymbol{\phi}_2 = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}, \\
\lambda_3 &= 7: & \begin{pmatrix} x_1 \\ 5x_2 + 2x_3 \\ 2x_2 + 5x_3 \end{pmatrix} &= \begin{pmatrix} 7x_1 \\ 7x_2 \\ 7x_3 \end{pmatrix} \Rightarrow \boldsymbol{\phi}_3 = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}.
\end{aligned}$$

Thus our final $\mathbf{\Phi}$ and $\mathbf{A}_w$ matrices are:

$$\mathbf{\Phi} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

and

$$\begin{aligned}
\mathbf{A}_w = \mathbf{\Phi \Lambda}^{-1/2} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 7 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{6} & 1/\sqrt{14} \\ 0 & -1/\sqrt{6} & 1/\sqrt{14} \end{pmatrix}.
\end{aligned}$$

We have then, $\mathbf{Y} = \mathbf{A}_w^t(\mathbf{x} - \boldsymbol{\mu}) \sim N(\mathbf{0}, \mathbf{I})$.

(c) The transformed point is found by applying $\mathbf{A}$, that is,

$$\begin{aligned}
\mathbf{x}_w &= \mathbf{A}_w^t(\mathbf{x}_o - \boldsymbol{\mu}) \\
&= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{6} & 1/\sqrt{14} \\ 0 & -1/\sqrt{6} & 1/\sqrt{14} \end{pmatrix} \begin{pmatrix} -0.5 \\ -2 \\ -1 \end{pmatrix} = \begin{pmatrix} -0.5 \\ -1/\sqrt{6} \\ -3/\sqrt{14} \end{pmatrix}.
\end{aligned}$$

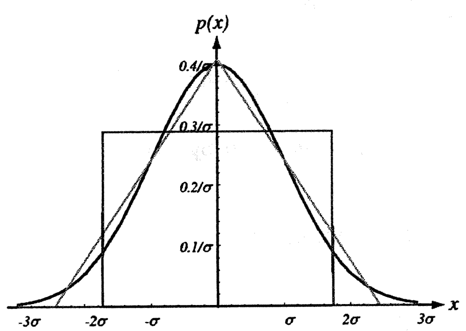
(d) From part (a), we have that the squared Mahalanobis distance from $\mathbf{x}_o$ to $\boldsymbol{\mu}$ in the original coordinates is $r^2 = (\mathbf{x}_o - \boldsymbol{\mu})^t \mathbf{\Sigma}^{-1}(\mathbf{x}_o - \boldsymbol{\mu}) = 1.06$. The Mahalanobis distance from $\mathbf{x}_w$ to $\mathbf{0}$ in the transformed coordinates is $\mathbf{x}_w^t \mathbf{x}_w = (0.5)^2 + 1/6 + 3/14 = 1.06$. The two distances are the same, as they must be under any linear transformation.

The entropy of the triangle distribution is then

$$\begin{aligned}
H(p(x)) &= -\int_{-w}^{w} \frac{w-|x|}{w^2}\ln\left[\frac{w-|x|}{w^2}\right]dx \\
&= \int_{0}^{w} \frac{w-x}{w^2}\ln\left[\frac{w-x}{w^2}\right]dx - \int_{-w}^{0} \frac{w+x}{w^2}\ln\left[\frac{w+x}{w^2}\right]dx \\
&= \ln w + 1/2 = \ln[\sqrt{6}\sigma] + 1/2 = \ln[\sqrt{6e}\sigma],
\end{aligned}$$

where we used the result $w = \sqrt{6}\sigma$ from the variance condition.

Thus, in order of decreasing entropy, these equal-variance distributions are Gaussian, uniform then triangle, as illustrated in the figure, where each has the same variance σ^2.



22. As usual, we denote our multidimensional Gaussian distribution by $p(\mathbf{x}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, or

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

According to Eq. 37 in the text, the entropy is

$$\begin{aligned}
H(p(\mathbf{x})) &= -\int p(\mathbf{x})\ln p(\mathbf{x})d\mathbf{x} \\
&= -\int p(\mathbf{x})\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) - \underbrace{\ln\left[(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}\right]}_{indep.\ of\ \mathbf{x}}\right]d\mathbf{x} \\
&= \frac{1}{2}\int\left[\sum_{i=1}^{d}\sum_{j=1}^{d}(x_i-\mu_i)[\boldsymbol{\Sigma}^{-1}]_{ij}(x_j-\mu_j)\right]d\mathbf{x} + \frac{1}{2}\ln[(2\pi)^d|\boldsymbol{\Sigma}|] \\
&= \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\int(x_j-\mu_j)(x_i-\mu_i)\underbrace{[\boldsymbol{\Sigma}^{-1}]_{ij}}_{indep.\ of\ \mathbf{x}}d\mathbf{x} + \frac{1}{2}\ln[(2\pi)^d|\boldsymbol{\Sigma}|] \\
&= \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}[\boldsymbol{\Sigma}]_{ji}[\boldsymbol{\Sigma}^{-1}]_{ij} + \frac{1}{2}\ln[(2\pi)^d|\boldsymbol{\Sigma}|] \\
&= \frac{1}{2}\sum_{j=1}^{d}[\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}]_{jj} + \frac{1}{2}\ln[(2\pi)^d|\boldsymbol{\Sigma}|]
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2}\underbrace{\sum_{j=1}^{d}[\mathbf{I}]_{jj}}_{d} + \frac{1}{2}\ln[(2\pi)^d|\boldsymbol{\Sigma}|] \\
&= \frac{d}{2} + \frac{1}{2}\ln[(2\pi)^d|\boldsymbol{\Sigma}|] \\
&= \frac{1}{2}\ln[(2\pi e)^d|\boldsymbol{\Sigma}|],
\end{aligned}$$

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

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

For this case we have

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

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:

$$p(\mathbf{x}_o|\omega) = \frac{1}{(2\pi)^{3/2}(21)^{1/2}}\exp\left[-\frac{1}{2}(1.06)\right] = 8.16\times10^{-3}.$$

(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

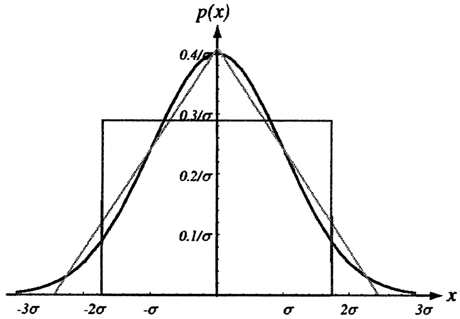
$$\begin{aligned}
\begin{vmatrix}1-\lambda&0&0\\0&5-\lambda&2\\0&2&5-\lambda\end{vmatrix} &= (1-\lambda)\left[(5-\lambda)^2-4\right] \\
&= (1-\lambda)(3-\lambda)(7-\lambda) = 0.
\end{aligned}$$

The entropy of the triangle distribution is then

$$\begin{aligned} H(p(x)) &= -\int_{-w}^{w} \frac{w-|x|}{w^2}\ln\left[\frac{w-|x|}{w^2}\right]dx \\ &= \int_{0}^{w} \frac{w-x}{w^2}\ln\left[\frac{w-x}{w^2}\right]dx - \int_{-w}^{0} \frac{w+x}{w^2}\ln\left[\frac{w+x}{w^2}\right]dx \\ &= \ln w + 1/2 = \ln[\sqrt{6}\sigma] + 1/2 = \ln[\sqrt{6e}\sigma], \end{aligned}$$

where we used the result $w = \sqrt{6}\sigma$ from the variance condition.

Thus, in order of decreasing entropy, these equal-variance distributions are Gaussian, uniform then triangle, as illustrated in the figure, where each has the same variance σ^2.



22. As usual, we denote our multidimensional Gaussian distribution by $p(\mathbf{x}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, or

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

According to Eq. 37 in the text, the entropy is

$$\begin{aligned} H(p(\mathbf{x})) &= -\int p(\mathbf{x})\ln p(\mathbf{x})d\mathbf{x} \\ &= -\int p(\mathbf{x})\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) - \underbrace{\ln\left[(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}\right]}_{indep.\ of\ \mathbf{x}}\right]d\mathbf{x} \\ &= \frac{1}{2}\int\left[\sum_{i=1}^{d}\sum_{j=1}^{d}(x_i-\mu_i)[\boldsymbol{\Sigma}^{-1}]_{ij}(x_j-\mu_j)\right]d\mathbf{x} + \frac{1}{2}\ln[(2\pi)^d|\boldsymbol{\Sigma}|] \\ &= \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\int(x_j-\mu_j)(x_i-\mu_i)\underbrace{[\boldsymbol{\Sigma}^{-1}]_{ij}}_{indep.\ of\ \mathbf{x}}d\mathbf{x} + \frac{1}{2}\ln[(2\pi)^d|\boldsymbol{\Sigma}|] \\ &= \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}[\boldsymbol{\Sigma}]_{ji}[\boldsymbol{\Sigma}^{-1}]_{ij} + \frac{1}{2}\ln[(2\pi)^d|\boldsymbol{\Sigma}|] \\ &= \frac{1}{2}\sum_{j=1}^{d}[\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}]_{jj} + \frac{1}{2}\ln[(2\pi)^d|\boldsymbol{\Sigma}|] \end{aligned}$$

Likewise, the mean constraint gives

$$\int_0^\infty e^{\lambda_1} e^{\lambda_0-1} x\,dx = e^{\lambda_0-1}\left(\frac{1}{\lambda_1^2}\right) = \mu.$$

Hence $\lambda_1 = -1/\mu$ and $\lambda_0 = 1 - \ln\mu$, and the density is

$$p(x) = \begin{cases} (1/\mu)e^{-x/\mu} & x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

(c) Here the density has three free parameters, and is of the general form

$$p(x) = \exp[\lambda_0 - 1 + \lambda_1 x + \lambda_2 x^2],$$

and the constraint equations are

$$\int_{-\infty}^{\infty} p(x)dx = 1 \qquad (*)$$

$$\int_{-\infty}^{\infty} xp(x)dx = \mu \qquad (**)$$

$$\int_{-\infty}^{\infty} x^2 p(x)dx = \sigma^2. \qquad (***)$$

We first substitute the general form of $p(x)$ into $(*)$ and find

$$\frac{1}{\sqrt{-\lambda_2}}\frac{\sqrt{\pi}}{2}e^{\lambda_0-1-\lambda_1^2/(4\lambda_2)}\text{erf}\left[\sqrt{-\lambda_2}x - \frac{\lambda_1}{2\sqrt{-\lambda_2}}\right]\Bigg|_{-\infty}^{\infty} = 1.$$

Since $\lambda_2 < 0$, $\text{erf}(\infty) = 1$ and $\text{erf}(-\infty) = -1$, we have

$$\frac{\sqrt{\pi}\exp[\lambda_0 - 1 - \lambda_1^2/(4\lambda_2)]}{\sqrt{-\lambda_2}} = 1.$$

Likewise, next substitute the general form of $p(x)$ into $(**)$ and find

$$-\frac{\lambda_1\sqrt{\pi}\exp[\lambda_0 - 1 - \lambda_1^2/(4\lambda_2)]}{4\lambda_2\sqrt{-\lambda_2}}\text{erf}\left[\sqrt{-\lambda_2}x - \lambda_2/(2\sqrt{-\lambda_2})\right]\Bigg|_{-\infty}^{\infty} = \mu,$$

which can be simplified to yield

$$\frac{\lambda_1\sqrt{\pi}}{2\lambda_2\sqrt{-\lambda_2}}\exp[\lambda_0 - 1 - \lambda_1^2/(4\lambda_2)] = -\mu.$$

Finally, we substitute the general form of $p(x)$ into $(***)$ and find

$$\frac{\sqrt{\pi}}{2\lambda_2\sqrt{-\lambda_2}}\exp[\lambda_0 - 1 - \lambda_1^2/(4\lambda_2)] = -\sigma^2.$$

We combine these three results to find the constants:

$$\begin{aligned} \lambda_0 &= 1 - \frac{\mu^2}{2\sigma_2} + \ln[1/(\sqrt{2\pi}\sigma)] \\ \lambda_1 &= \mu/\sigma^2 \\ \lambda_2 &= -1/(2\sigma^2). \end{aligned}$$

We substitute these values back into the general form of the density and find

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

that is, a Gaussian.

21. A Gaussian centered at $x = 0$ is of the form

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp[-x^2/(2\sigma^2)].$$

The entropy for this distribution is given by Eq. 37 in the text:

$$\begin{aligned} H(p(x)) &= -\int_{-\infty}^{\infty} p(x)\ln p(x)dx \\ &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp[-x^2/(2\sigma^2)]\ln\left[\frac{1}{\sqrt{2\pi}\sigma}\exp[-x^2/(2\sigma^2)]\right]dx \\ &= \ln[\sqrt{2\pi}\sigma] + 1/2 = \ln[\sqrt{2\pi e}\sigma]. \end{aligned}$$

For the uniform distribution, the entropy is

$$H(p(x)) = -\int_{x_l}^{x_u} \frac{1}{|x_u - x_l|}\ln\left[\frac{1}{|x_u - x_l|}\right]dx = -\ln\left[\frac{1}{|x_u - x_l|}\right] = \ln[x_u - x_l].$$

Since we are given that the mean of the distribution is 0, we know that $x_u = -x_l$. Further, we are told that the variance is σ^2, that is

$$\int_{x_l}^{x_h} x^2 p(x)dx = \sigma^2$$

which, after integration, implies

$$x_u^2 + x_u x_l + x_l^2 = 3\sigma^2.$$

We put these results together and find for the uniform distribution $H(p(x)) = \ln[2\sqrt{3}\sigma]$.

We are told that the variance of the triangle distribution centered on 0 having half-width w is σ^2, and this implies

$$\int_{-w}^{w} x^2\frac{w - |x|}{w^2}dx = \int_0^{w} x^2\frac{w - x}{w^2}dx + \int_{-w}^{0} x^2\frac{w + x}{w^2}dx = w^2/6 = \sigma^2.$$

(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 matrix are particularly simple:

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

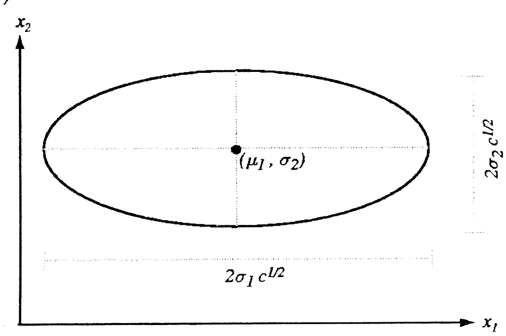
This leads to the density being expressed as:

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

(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:

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