17. Derive the formula (Eq. 44) for the volume V of a hyperellipsoid of constant Mahalanobis distance r (Eq. 43) for a Gaussian distribution having covariance $\mathbf{\Sigma}$.

18. Consider two normal distributions in one dimension: $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. Imagine that we choose two random samples x_1 and x_2, one from each of the normal distributions and calculate their sum $x_3 = x_1 + x_2$. Suppose we do this repeatedly.

(a) Consider the resulting distribution of the values of x_3. Show from first principles that this is also a normal distribution.

(b) What is the mean, μ_3, of your new distribution?

(c) What is the variance, σ_3^2?

(d) Repeat the above with two distributions in a multi-dimensional space, i.e., $N(\boldsymbol{\mu}_1, \mathbf{\Sigma}_1)$ and $N(\boldsymbol{\mu}_2, \mathbf{\Sigma}_2)$.

19. Starting from the definition of entropy (Eq. 36), derive the general equation for the maximum-entropy distribution given constraints expressed in the general form

$$\int b_k(x) p(x)\, dx = a_k, \quad k = 1, 2, ..., q$$

as follows:

(a) Use Lagrange undetermined multipliers $\lambda_1, \lambda_2, ..., \lambda_q$ and derive the synthetic function:

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

State why we know $a_0 = 1$ and $b_0(x) = 1$ for all x.

(b) Take the derivative of H_s with respect to $p(x)$. Equate the integrand to zero, and thereby prove that the minimum-entropy distribution obeys

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

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

20. Use the final result from Problem 19 for the following.

(a) Suppose we know only that a distribution is non-zero in the range $x_l \leq x \leq x_u$. Prove that the maximum entropy distribution is uniform in that range, i.e.,

$$p(x) \sim U(x_l, x_u) = \begin{cases} 1/|x_u - x_l| & x_l \leq x \leq x_u \\ 0 & \text{otherwise.} \end{cases}$$

(b) Suppose we know only that a distribution is non-zero for $x \geq 0$ and that its mean is μ. Prove that the maximum entropy distribution is

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

(c) Now suppose we know solely that the distribution is normalized, has mean μ, and standard deviation σ^2, and thus from Problem 19 our maximum entropy distribution must be of the form

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

Write out the three constraints and solve for λ_0, λ_1, and λ_2 and thereby prove that the maximum entropy solution is a Gaussian, i.e.,

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

21. Three distributions — a Gaussian, a uniform distribution, and a triangle distribution (cf., Problem 4) — each have mean zero and standard deviation σ^2. Use Eq. 36 to calculate and compare their entropies.

22. Calculate the entropy of a multidimensional Gaussian $p(\mathbf{x}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

⊕*Section 2.6*

23. Consider the three-dimensional normal distribution $p(\mathbf{x}|\omega) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{pmatrix}$.

(a) Find the probability density at the point $\mathbf{x}_0 = (.5, 0, 1)^t$.

(b) Construct the whitening transformation $\mathbf{A}_w$. Show your $\boldsymbol{\Lambda}$ and $\boldsymbol{\Phi}$ matrices. Next, convert the distribution to one centered on the origin with covariance matrix equal to the identity matrix, $p(\mathbf{x}|\omega) \sim N(\mathbf{0}, \mathbf{I})$.

(c) Apply the same overall transformation to $\mathbf{x}_0$ to yield a transformed point $\mathbf{x}_w$.

(d) By explicit calculation, confirm that the Mahalanobis distance from $\mathbf{x}_0$ to the mean $\boldsymbol{\mu}$ in the original distribution is the same as for $\mathbf{x}_w$ to $\mathbf{0}$ in the transformed distribution.

(e) Does the probability density remain unchanged under a general linear transformation? In other words, is $p(\mathbf{x}_0|N(\boldsymbol{\mu}, \boldsymbol{\Sigma})) = p(\mathbf{T}^t\mathbf{x}_0|N(\mathbf{T}^t\boldsymbol{\mu}, \mathbf{T}^t\boldsymbol{\Sigma}\mathbf{T}))$ for some linear transform $\mathbf{T}$? Explain.

(f) Prove that a general whitening transform $\mathbf{A}_w = \boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2}$ when applied to a Gaussian distribution insures that the final distribution has covariance proportional to the identity matrix $\mathbf{I}$. Check whether normalization is preserved by the transformation.

24. Consider the multivariate normal density for which $\sigma_{ij} = 0$ and $\sigma_{ii} = \sigma_i^2$, i.e., $\mathbf{\Sigma} = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_d^2)$.

(a) Show that the evidence is

$$p(\mathbf{x}) = \frac{1}{\prod\limits_{i=1}^{d} \sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2} \sum_{i=1}^{d} \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2 \right].$$

(b) Plot and describe the contours of constant density.

(c) Write an expression for the Mahalanobis distance from $\mathbf{x}$ to $\boldsymbol{\mu}$.

25. Fill in the steps in the derivation from Eq. 57 to Eqs. 58–63.

26. Let $p(\mathbf{x}|\omega_i) \sim N(\boldsymbol{\mu}_i, \mathbf{\Sigma})$ for a two-category d-dimensional problem with the same covariances but arbitrary means and prior probabilities. Consider the squared Mahalanobis distance

$$r_i^2 = (\mathbf{x} - \boldsymbol{\mu}_i)^t \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_i).$$

(a) Show that the gradient of r_i^2 is given by

$$\nabla r_i^2 = 2\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i).$$

(b) Show that at any position on a given line through $\boldsymbol{\mu}_i$ the gradient ∇r_i^2 points in the same direction. Must this direction be parallel to that line?

(c) Show that ∇r_1^2 and ∇r_2^2 point in opposite directions along the line from $\boldsymbol{\mu}_1$ to $\boldsymbol{\mu}_2$.

(d) Show that the optimal separating hyperplane is tangent to the constant probability density hyperellipsoids at the point that the separating hyperplane cuts the line from $\boldsymbol{\mu}_1$ to $\boldsymbol{\mu}_2$.

(e) True of False: For a two-category problem involving normal densities with arbitrary means and covariances, and $P(\omega_1) = P(\omega_2) = 1/2$, the Bayes decision boundary consists of the set of points of equal Mahalanobis distance from the respective sample means. Explain.

27. Suppose we have two normal distributions with the same covariances but different means: $N(\boldsymbol{\mu}_1, \mathbf{\Sigma})$ and $N(\boldsymbol{\mu}_2, \mathbf{\Sigma})$. In terms of their prior probabilities $P(\omega_1)$ and $P(\omega_2)$, state the condition that the Bayes decision boundary *not* pass between the two means.

28. Two random variables $\mathbf{x}$ and $\mathbf{y}$ are called "statistically independent" if $p(\mathbf{x}, \mathbf{y}|\omega) = p(\mathbf{x}|\omega)p(\mathbf{y}|\omega)$.

(a) Prove that if $x_i - \mu_i$ and $x_j - \mu_j$ are statistically independent (for $i \neq j$) then σ_{ij} as defined in Eq. 42 is 0.

(b) Prove that the converse is true for the Gaussian case.

(c) Show by counterexample that this converse is *not* true in the general case.

24. Consider the multivariate normal density for which $\sigma_{ij} = 0$ and $\sigma_{ii} = \sigma_i^2$, i.e., $\mathbf{\Sigma} = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_d^2)$.

(a) Show that the evidence is

$$p(\mathbf{x}) = \frac{1}{\prod\limits_{i=1}^{d} \sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\sum_{i=1}^{d}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right].$$

(b) Plot and describe the contours of constant density.

(c) Write an expression for the Mahalanobis distance from $\mathbf{x}$ to $\boldsymbol{\mu}$.

25. Fill in the steps in the derivation from Eq. 57 to Eqs. 58–63.

26. Let $p(\mathbf{x}|\omega_i) \sim N(\boldsymbol{\mu}_i, \mathbf{\Sigma})$ for a two-category d-dimensional problem with the same covariances but arbitrary means and prior probabilities. Consider the squared Mahalanobis distance

$$r_i^2 = (\mathbf{x} - \boldsymbol{\mu}_i)^t \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_i).$$

(a) Show that the gradient of r_i^2 is given by

$$\nabla r_i^2 = 2\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i).$$

(b) Show that at any position on a given line through $\boldsymbol{\mu}_i$ the gradient ∇r_i^2 points in the same direction. Must this direction be parallel to that line?

(c) Show that ∇r_1^2 and ∇r_2^2 point in opposite directions along the line from $\boldsymbol{\mu}_1$ to $\boldsymbol{\mu}_2$.

(d) Show that the optimal separating hyperplane is tangent to the constant probability density hyperellipsoids at the point that the separating hyperplane cuts the line from $\boldsymbol{\mu}_1$ to $\boldsymbol{\mu}_2$.

(e) True of False: For a two-category problem involving normal densities with arbitrary means and covariances, and $P(\omega_1) = P(\omega_2) = 1/2$, the Bayes decision boundary consists of the set of points of equal Mahalanobis distance from the respective sample means. Explain.

27. Suppose we have two normal distributions with the same covariances but different means: $N(\boldsymbol{\mu}_1, \mathbf{\Sigma})$ and $N(\boldsymbol{\mu}_2, \mathbf{\Sigma})$. In terms of their prior probabilities $P(\omega_1)$ and $P(\omega_2)$, state the condition that the Bayes decision boundary *not* pass between the two means.

28. Two random variables $\mathbf{x}$ and $\mathbf{y}$ are called "statistically independent" if $p(\mathbf{x}, \mathbf{y}|\omega) = p(\mathbf{x}|\omega)p(\mathbf{y}|\omega)$.

(a) Prove that if $x_i - \mu_i$ and $x_j - \mu_j$ are statistically independent (for $i \neq j$) then σ_{ij} as defined in Eq. 42 is 0.

(b) Prove that the converse is true for the Gaussian case.

(c) Show by counterexample that this converse is *not* true in the general case.

⊕*Section 2.10*

47. Suppose we have three categories in two dimensions with the following underlying distributions:

- $p(\mathbf{x}|\omega_1) \sim N(\mathbf{0}, \mathbf{I})$
- $p(\mathbf{x}|\omega_2) \sim N\left(\binom{1}{1}, \mathbf{I}\right)$
- $p(\mathbf{x}|\omega_3) \sim \frac{1}{2}N\left(\binom{.5}{.5}, \mathbf{I}\right) + \frac{1}{2}N\left(\binom{-.5}{.5}, \mathbf{I}\right)$

with $P(\omega_i) = 1/3$, $i = 1, 2, 3$.

(a) By explicit calculation of posterior probabilities, classify the point $\mathbf{x} = \binom{.3}{.3}$ for minimum probability of error.

(b) Suppose that for a particular test point the first feature is missing. That is, classify $\mathbf{x} = \binom{*}{.3}$.

(c) Suppose that for a particular test point the second feature is missing. That is, classify $\mathbf{x} = \binom{.3}{*}$.

(d) Repeat all of the above for $\mathbf{x} = \binom{.2}{.6}$.

48. Show that Eq. 93 reduces to Bayes rule when the true feature is $\boldsymbol{\mu}_i$ and $p(\mathbf{x}_b|\mathbf{x}_t) \sim N(\mathbf{x}_t, \boldsymbol{\Sigma})$. Interpret this answer in words.

⊕*Section 2.11*

49. Suppose we have three categories with $P(\omega_1) = 1/2, P(\omega_2) = P(\omega_3) = 1/4$ and the following distributions

- $p(x|\omega_1) \sim N(0, 1)$
- $p(x|\omega_2) \sim N(.5, 1)$
- $p(x|\omega_3) \sim N(1, 1)$,

and that we sample the following four points: $x = 0.6, 0.1, 0.9, 1.1$.

(a) Calculate explicitly the probability that the sequence actually came from $\omega_1, \omega_3, \omega_3, \omega_2$. Be careful to consider normalization.

(b) Repeat for the sequence $\omega_1, \omega_2, \omega_2, \omega_3$.

(c) Find the sequence having the maximum probability.