1. Exercise 3.4 in text.

**Exercise 3.4**  (a) Let $X_1 \sim \mathcal{N}(0, \sigma_1^2)$ and let $X_2 \sim \mathcal{N}(0, \sigma_2^2)$ be independent of $X_1$. Conolve the density of $X_1$ with that of $X_2$ to show that $X_1 + X_2$ is Gaussian, $\mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$.

(b) Let $W_1, W_2$ be IID normalized Gaussian r.v.s. Show that $a_1 W_1 + a_2 W_2$ is Gaussian, $\mathcal{N}(0, a_1^2 + a_2^2)$. Hint: You could repeat all the equations of (a), but the insightful approach is to let $X_i = a_i W_i$ for $i = 1, 2$ and then use (a) directly.

(c) Combine (b) with induction to show that all linear combinations of IID normalized Gaussian r.v.s are Gaussian.

2. Exercise 3.9 in text.

**Exercise 3.9**  Let $X$ and $Y$ be jointly Gaussian with means $m_X$, $m_Y$, variances $\sigma_X^2$, $\sigma_Y^2$, and normalized covariance $\rho$. Find the conditional density $f_{X|Y}(x | y)$.

3. Exercise 3.17 in text.

**Exercise 3.17**  Let $X$ and $Z$ be statistically independent Gaussian r.v.s of arbitrary dimension $n$ and $m$ respectively. Let $Y = [H]X + Z$, where $[H]$ is an arbitrary real $n \times m$ matrix.

(a) Explain why $X_1, \ldots, X_n, Z_1, \ldots, Z_m$ must be jointly Gaussian r.v.s. Then explain why $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ must be jointly Gaussian.

(b) Show that if $[K_X]$ and $[K_Z]$ are non-singular, then the combined covariance matrix $[K]$ for $(X_1, \ldots, X_n, Y_1, \ldots, Y_m)'$ must be non-singular.
4. Let $X$ and $Y$ be a $n$-dimensional random vector and a $m$-dimensional random vector respectively. Define $K_{XY} := E[XY^T]$ as the cross-covariance matrix between $X$ and $Y$.

(a) What is the dimension of $K_{XY}$? Is it square? Is it symmetric? Is $K_{XY} = K_{YX}$?

(b) Suppose $m = n$ and let $Z = X + Y$. Compute the covariance of $K_Z$ in terms of $K_X, K_Y$ and $K_{XY}$. When is $K_Z = K_X + K_Y$?

(c) For general $m, n$ and $k$, let $A$ be a $k$ by $n$ matrix and $B$ be a $k$ by $m$ matrix. Compute the covariance of $Z = AX + BY$.

5. Exercise 8.1 in text.

**Exercise 8.1** In this exercise, we evaluate $Pr\{e_\eta | X = a\}$ and $Pr\{e_\eta | X = b\}$ for binary detection from vector signals in Gaussian noise directly from (8.40) and (8.41).

(a) By using (8.40) for each sample value $y$ of $Y$, show that

$$E[LLR(Y) | X=a] = -\frac{(b-a)^T(b-a)}{2\sigma^2}.$$ 

(b) Defining $\gamma = \frac{\|b - a\|}{2\sigma}$, show that

$$E[LLR(Y) | X=a] = -2\gamma^2.$$ 

(c) Show that

$$VAR[LLR(Y) | X=a] = 4\gamma^2.$$ 

Hint: Note that, given $X = a$, $Y = a + Z$.

(d) Show that, conditional on $X = a$, $LLR(Y) \sim \mathcal{N}(-2\gamma^2, 4\gamma^2)$. Show that, conditional on $X = a$, $LLR(Y)/2\gamma \sim \mathcal{N}(-\gamma, 1)$.

(e) Show that the first half of (8.44) is valid, i.e., that

$$Pr\{e_\eta | X=a\} = Pr\{LLR(Y) \geq \ln \eta | X=a\} = Q\left(\frac{\ln \eta}{2\gamma} + \gamma\right).$$

(f) By essentially repeating (a)–(e), show that the second half of (8.44) is valid, i.e., that

$$Pr\{e_\eta | X=b\} = Q\left(\frac{-\ln \eta}{2\gamma} + \gamma\right).$$
6. Exercise 8.9 in text.

**Exercise 8.9** A disease has two strains, 0 and 1, which occur with a priori probabilities \( p_0 \) and \( p_1 = 1 - p_0 \) respectively.

(a) Initially, a rather noisy test was developed to find which strain is present for patients with the disease. The output of the test is the sample value \( Y_1 \) of a rv \( Y \). Given strain 0 (\( X = 0 \)), \( Y_1 = 5 + Z_1 \), and given strain 1 (\( X = 1 \)), \( Y_1 = 1 + Z_1 \). The measurement noise \( Z_1 \) is independent of \( X \) and is Gaussian, \( Z_1 \sim \mathcal{N}(0, \sigma^2) \). Give \( \text{Pr}[e | X=0] \) and \( \text{Pr}[e | X=1] \) in terms of the function \( Q(x) \).

(b) A budding medical researcher determines that the test is making too many errors. A new measurement procedure is devised with two observation rvs \( Y_1 \) and \( Y_2 \). \( Y_1 \) is the same as in (a). \( Y_2 \), under hypothesis 0, is given by \( Y_2 = 5 + Z_1 + Z_2 \), and, under hypothesis 1, is given by \( Y_2 = 1 + Z_1 + Z_2 \). Assume that \( Z_2 \) is independent of both \( Z_1 \) and \( X \), and that \( Z_2 \sim \mathcal{N}(0, \sigma^2) \). Find the MAP decision rule for \( \hat{x} \) in terms of the joint observation \((y_1, y_2)\), and find \( \text{Pr}[e | X=0] \) and \( \text{Pr}[e | X=1] \). Hint: Find \( f_{y_2|y_1}(y_2 | y_1, 0) \) and \( f_{y_2|y_1}(y_2 | y_1, 1) \).

(c) Explain in laymen’s terms why the medical researcher should learn more about probability.

(d) Now suppose that \( Z_2 \), in (b), is uniformly distributed between 0 and 1 rather than being Gaussian. We are still given that \( Z_2 \) is independent of both \( Z_1 \) and \( X \). Find the MAP decision rule for \( \hat{x} \) in terms of the joint observation \((y_1, y_2)\) and find \( \text{Pr}[e | X=0] \) and \( \text{Pr}[e | X=1] \).

(e) Finally, suppose that \( Z_2 \) is also uniformly distributed between 0 and 1. Again find the MAP decision rule and error probabilities.

7. This problem will explore optimal classification based on the features selected via principal component analysis (PCA). Download the MNIST handwritten digit dataset from the course website [http://web.stanford.edu/class/ee278/homeworks/hw3-data.zip](http://web.stanford.edu/class/ee278/homeworks/hw3-data.zip). The folders train0 and train2 contain the same set of images we used in Homework 2. We will use these images to train a classifier that can distinguish between the digits “0” and “2.” We also have added two test sets (each with 500 images) of the handwritten digit “0” and of the digit “2” in the folders test0 and test2.

   a) As in Homework 2, consider each image as vector \( \mathbf{X}_i \in \mathbb{R}^{784} \). Combine all the training images (in folders train0 and train2) and generate an estimator of the covariance matrix. Compute the first 20 eigenvectors \( \mathbf{U}_i \) (\( 1 \leq i \leq 20 \)) corresponding to the largest eigenvalues. Now project each training image \( \tilde{\mathbf{X}}_i \) onto the new set of basis vectors \( \mathbf{U}_i \). The result, denoted \( \tilde{\mathbf{X}}_i \in \mathbb{R}^{20} \), is a lower dimensional feature vector that we will use to represent the data.

   b) Estimate the mean and covariance matrix of the training vectors \( \tilde{\mathbf{X}}_i \) corresponding to the digit “0” and the mean and covariance matrix corresponding to the digit “2.” Suppose that each digit-0 (resp. digit-2) image \( \tilde{\mathbf{X}}_i \) is independently drawn
from a jointly Gaussian distribution \( P_0 \) (resp. \( P_2 \)). Propose a maximum likelihood detector that classifies a given image as a “0” or “2”, i.e. assuming an equal prior. (This is called Gaussian discriminative analysis in machine learning.)

c) Run your classifier on the test dataset in the folders test0 and test2. Report the empirical error rates.