1. Problem 2.13
2. Problem 2.18
3. Problem 3.17
Note that part (e) asks what $\hat{\theta}$ needs to be so that $p(x|\hat{\theta}) = p(\theta|D)$. (This would be the result if the posterior was concentrated around a single point, i.e. $p(\theta|D) = \delta(\theta - \hat{\theta})$, where $\delta(\cdot)$ is the Dirac delta function satisfying $\int f(x)\delta(x - x_0)dx = f(x_0)$.)
4. This problem demonstrates that if we make a poor choice in defining our model, the maximum-likelihood estimate for the model parameters no longer results in an optimal classifier. In fact, it is possible to choose parameter values in this case which are different from the ML estimate (i.e. suboptimal with respect to the ML criteria) and yet result in improved classification performance. This means that in the presence of model error, the ML classifier can be suboptimal even among the possible space of classifiers defined by our incorrect model.

After you have downloaded the zip archive from the class website, look at the file `ps1_4.m`, which defines the model for this problem. We assume that the Gaussian parameters for class 1 are perfectly known. We want to estimate the mean of class 2, and we are using an incorrect model of the covariance. (The true variances are large, but we have incorrectly chosen them to be 1.)

(a) Using the software, calculate the ML estimate for the mean and the associated classifier (assuming that the covariance matrices are diagonal). How can you explain the difference in the shape of the decision boundaries in the ML and Bayes cases? Repeat the estimation three times and record the error in the mean estimate and the testing error. How does the estimate of the mean affect the decision boundary? We are using a fairly large number of samples in this experiment, why does the result vary so much from run to run?
(b) Derive the analytic forms for the maximum likelihood and Bayes discriminant functions for this problem. Your results should be equations, and it is OK to approximate very small numbers with zero. Check your formulas against the decision boundaries drawn by the tool in the case where the ML estimate for the mean of class 2 is $[1, 1]$ (i.e. the parameter estimate with zero error). Does your derivation confirm the shape of the ML and Bayes boundaries displayed by the tool?
(c) Using your answer in (b), come up with a new value for the mean vector estimate for class 2 which improves the classification performance. Compare your new mean estimate to the previous case of $[1 \ 1]$. Implement this change using the software and report the improvement in the classification error. Why does your new estimate produce lower classification error?

(d) **Extra Credit:** See the instructions in `ps1_4.m`. 
5. The goal of this problem is to develop some intuition for the properties of distributions in high-dimensional vector spaces. It is adapted from Exercises 1.1 and 1.2 in Chris Bishop’s book, *Neural Networks for Pattern Recognition* (but having access to the book is no advantage in working the problem, as I have already provided more of the details).

(a) Preliminaries: Consider the following expression for the volume, $V_n(R)$, of a hypersphere of radius $R$ in $n$ dimensions:

$$V_n(R) = \int \cdots \int dX = C_{n-1} \int_0^R r^{n-1} dr,$$

with $C_{n-1} = \int \cdots \int d\Omega_{n-1}$, and $dX = dx_1 dx_2 \ldots dx_n$. The second integral is the result of changing variables to hyperspherical coordinates, analogous to changing from Cartesian to spherical coordinates in 3-D. In $n$ dimensions we have

$$dX = r^{n-1} dr d\Omega_{n-1},$$

where $d\Omega_{n-1}$ contains all of the angle-dependent factors, e.g. $d\Omega_1 = d\theta$ (polar), $d\Omega_2 = \sin \theta d\theta d\phi$ (spherical), etc. The constant $C_{n-1}$ is the result of integrating over all of the angles in $n - 1$ dimensions. Note that we can also use Eq. 1 to integrate some function $f(\cdot)$ over the hypersphere.

The first step in obtaining closed-form expressions for the volume and surface area is to calculate $C_{n-1}$. Since this constant does not depend upon $R$, we can use Eq. 1 with $R \to \infty$ and an appropriately-chosen $f(\cdot)$ to simplify the calculation. Consider the following identities:

1. $$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}$$
2. $$\Gamma(m) = \int_0^\infty r^{m-1}e^{-r} dr,$$

where $\Gamma(\cdot)$ is the Gamma function and we have $\Gamma(1) = 1$, $\Gamma(3/2) = \sqrt{\pi}/2$, and $m\Gamma(m) = \Gamma(m+1)$. Use Eqs. 1, 2 and 3 to compute the integral

$$\prod_{i=1}^n \int_{-\infty}^{\infty} e^{-x_i^2} dx_i$$

and solve for $C_{n-1}$, obtaining the expression

$$C_{n-1} = \frac{2\pi^{n/2}}{\Gamma(n/2)}.$$
(b) We can also express the volume as

\[ V_n(R) = \int_0^R S_n(r) \, dr, \]

where we define \( S_n(r) \) to be the surface area of a hypersphere of radius \( r \) in \( n \) dimensions. The expression follows from the fact that we can construct the volume \( V_n(R) \) by adding together infinitesimal spherical shells ranging in radius from 0 to \( R \). Using Eqs. 4 and 5, show that

\[ V_n(R) = \frac{1}{n} C_{n-1} R^n \quad \text{and} \quad S_n(R) = C_{n-1} R^{n-1}. \]

Verify that the formulas for \( V_n(R) \) and \( S_n(R) \) give the expected answers for \( n = 2 \) and \( n = 3 \). What is the geometric interpretation of \( C_{n-1} \)?
(c) Consider a hypersphere inscribed within a hypercube. If the hypersphere has radius $R$ then the hypercube has edge length $2R$. Show that the ratio of the volume of the hypersphere to the volume of the hypercube is given by

\[
\frac{\text{volume of hypersphere}}{\text{volume of hypercube}} = \frac{\pi^{n/2}}{n2^{n-1}\Gamma(n/2)}.
\]

Use Stirling’s approximation:

\[
\Gamma(m + 1) \approx (2\pi)^{1/2} e^{-m} m^{m+1/2},
\]

which is valid when $m$ is large, to show that the ratio in Eq. 6 goes to zero as $n \to \infty$. Use this result to argue that as $n$ increases, more and more of the volume of the hypercube is concentrated in its corners.

(d) Similarly, show that the ratio of the distance from the center of the hypercube to one of its corners, divided by the perpendicular distance to one of the edges, is $\sqrt{n}$, and therefore goes to infinity as $n \to \infty$. What does this imply about the shape of a hypercube in high dimensions?