(c) Suppose we know a fish is thin and medium lightness and that it was caught in the north Atlantic. What season is it, most likely? What is the probability of being correct?

51. Consider a Bayesian belief net with several nodes having unspecified values. Suppose that one such node is selected at random, with the probabilities of its nodes computed by the formulas described in the text. Next, another such node is chosen at random (possibly even a node already visited), and the probabilities are similarly updated. Prove that this procedure will converge to the desired probabilities throughout the full network.

Section 2.12

52. 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.

COMPUTER EXERCISES

Several of the computer exercises will rely on the following data.

	ω_1			ω_2			ω_3		
sample	x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
1	−5.01	−8.12	−3.68	−0.91	−0.18	−0.05	5.35	2.26	8.13
2	−5.43	−3.48	−3.54	1.30	−2.06	−3.53	5.12	3.22	−2.66
3	1.08	−5.52	1.66	−7.75	−4.54	−0.95	−1.34	−5.31	−9.87
4	0.86	−3.78	−4.11	−5.47	0.50	3.92	4.48	3.42	5.19
5	−2.67	0.63	7.39	6.14	5.72	−4.85	7.11	2.39	9.21
6	4.94	3.29	2.08	3.60	1.26	4.36	7.17	4.33	−0.98
7	−2.51	2.09	−2.59	5.37	−4.63	−3.65	5.75	3.97	6.65
8	−2.25	−2.13	−6.94	7.18	1.46	−6.66	0.77	0.27	2.41
9	5.56	2.86	−2.26	−7.39	1.17	6.30	0.90	−0.43	−8.71
10	1.03	−3.33	4.33	−7.50	−6.32	−0.31	3.52	−0.36	6.43

Section 2.5

1. You may need the following procedures for several exercises below.

(a) Write a procedure to generate random samples according to a normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in d dimensions.

(b) Write a procedure to calculate the discriminant function (of the form given in Eq. 49) for a given normal distribution and prior probability $P(\omega_i)$.

(c) Write a procedure to calculate the Euclidean distance between two arbitrary points.

(d) Write a procedure to calculate the Mahalanobis distance between the mean $\boldsymbol{\mu}$ and an arbitrary point $\mathbf{x}$, given the covariance matrix $\boldsymbol{\Sigma}$.

2. Refer to Computer exercise 1 (b) and consider the problem of classifying 10 samples from the table above. Assume that the underlying distributions are normal.

(a) Assume that the prior probabilities for the first two categories are equal ($P(\omega_1) = P(\omega_2) = 1/2$ and $P(\omega_3) = 0$) and design a dichotomizer for those two categories using only the x_1 feature value.

(b) Determine the empirical training error on your samples, that is, the percentage of points misclassified.

(c) Use the Bhattacharyya bound to bound the error you will get on novel patterns drawn from the distributions.

(d) Repeat all of the above, but now use *two* feature values, x_1 and x_2.

(e) Repeat, but use all *three* feature values.

(f) Discuss your results. In particular, is it ever possible for a finite set of data that the empirical error might be *larger* for more data dimensions?

3. Repeat Computer exercise 2 but for categories ω_1 and ω_3.

4. Consider the three categories in Computer exercise 2, and assume $P(\omega_i) = 1/3$.

(a) What is the Mahalanobis distance between each of the following test points and each of the category means in Computer exercise 2: $(1, 2, 1)^t$, $(5, 3, 2)^t$, $(0, 0, 0)^t$, $(1, 0, 0)^t$.

(b) Classify those points.

(c) Assume instead that $P(\omega_1) = 0.8$, and $P(\omega_2) = P(\omega_3) = 0.1$ and classify the test points again.

5. Illustrate the fact that the average of a large number of independent random variables will approximate a Gaussian by the following:

(a) Write a program to generate n random integers from a uniform distribution $U(x_l, x_u)$. (Some computer systems include this as a single, compiled function call.)

(b) Now write a routine to choose x_l and x_u randomly, in the range $-100 \leq x_l < x_u \leq +100$, and n (the number of samples) randomly in the range $0 < n \leq 1000$.

(c) Generate and plot a histogram of the accumulation of 10^4 points sampled as just described.

(d) Calculate the mean and standard deviation of your histogram, and plot it

(e) Repeat the above for 10^5 and for 10^6. Discuss your results.

Section 2.8

6. Explore how the empirical error does or does not approach the Bhattacharyya bound as follows: