Bayesian Classifier

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

(a) Write a procedure to generate random samples according to a normal distribution $N(\mu,\Sigma)$ in d dimensions, and try to visualize an example in 3D space (take 3 dimension as example) .

(b) Write a procedure to calculate the discriminant function as shown below for a given normal distribution and prior probability $P(\omega_i)$.

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

(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 μ and an arbitrary point x, given the covariance matrix Σ.

You may find some MATLAB functions to facilitate your programming.

2. Refer to exercise 1(b) and consider the problem of classifying 10 samples from the following table (also in the attached MATLAB file T1.mat). Assume that the underlying distributions are normal.

	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

(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 (两类分类器) for those two categories using only the x_1 feature value.

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

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

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

(e) 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 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 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.

(*adopted from Duda et al. Pattern Classification*)

5. Read the following Nature Views "*The currency of guessing*" **and write a paragraph in Chinese to explain the possible mechanisms/circuits of brain's Bayesian decision. For more information on the original Nature Article or want to enjoy the videos, please visit:**

http://www.nature.com/nature/journal/v447/n7148/abs/nature05852.html