September 28th, 2005

EE E6887 (Statistical Pattern Recognition)
Solutions for homework 1

P.1a Suppose two equally probable 1-dimensional densities are of the form

$$p(x|\omega_i) \propto \exp\left(-|x-a_i|/b_i\right), \text{ for } i = 1, 2, \text{ and } b_i > 0$$

(a) write an analytic expression for each density function. Namely, you have to normalize a_i and b_i for each function.

(b) Write the likelihood ration $p(x|\omega_1)/p(x|\omega_2)$

(c) Sketch a graph of the likelihood ratio for the case $a_1 = 1$, $b_1 = 2$, $a_2 = 0$, $b_2 = 1$.

Answer:

(a) Since $\int_{-\infty}^{\infty} p(x|\omega_i)dx = 1$, so we have

$$\int_{-\infty}^{a_i} A\exp\left((x-a_i)/b_i\right)dx + \int_{a_i}^{\infty} A\exp\left(-(x-a_i)/b_i\right)dx = 1$$

So $A = 1/2b_i$. Then we get the analytic expression:

$$p(x|\omega_i) = \frac{1}{2b_i}\exp\left(-|x-a_i|/b_i\right), \text{ for } i = 1, 2, \text{ and } b_i > 0$$

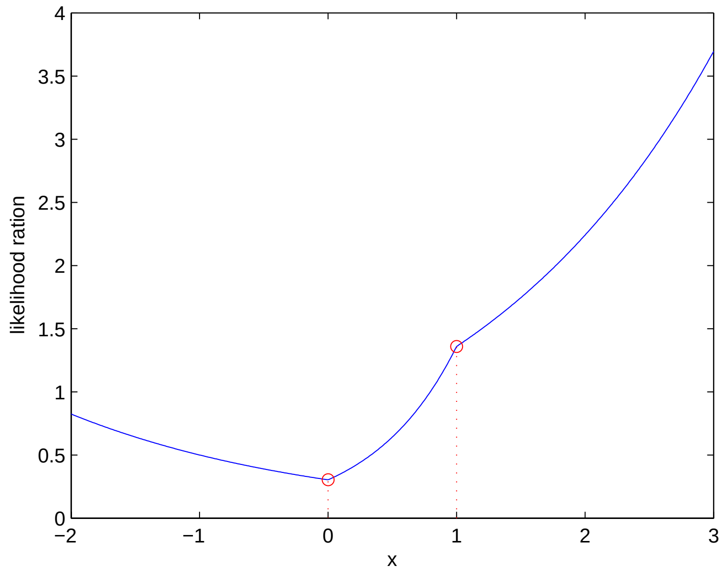
(b)

$p(x|\omega_1)/p(x|\omega_2) = \frac{b_2}{b_1}\exp\left(-\frac{|x-a_1|}{b_1} + \frac{|x-a_2|}{b_2}\right)$

(c) When $a_1 = 1$, $b_1 = 2$, $a_2 = 0$, $b_2 = 1$, we have the following

$$\begin{aligned} p(x|\omega_1)/p(x|\omega_2) &= \frac{1}{2}\exp\left(-\frac{|x-1|}{2} + |x|\right) \\ &= \begin{cases} \frac{1}{2}\exp\left(-\frac{1+x}{2}\right) & \text{if } x < 0 \\ \frac{1}{2}\exp\left(\frac{3x-1}{2}\right) & \text{if } 0 \le x < 1 \\ \frac{1}{2}\exp\left(\frac{x+1}{2}\right) & \text{if } x \ge 0 \end{cases} \end{aligned}$$

see figure below:



P.1b Under the Bayesian decision rule, the classification error is given by

$$P(error) = \int P(error|x)p(x)dx = \int \min[P(\omega_1|x), P(\omega_2|x)]p(x)dx$$

Show that for arbitrary density functions, and upper bound of the classification error can be found by replacing $\min[P(\omega_1|x), P(\omega_2|x)]$ with $2P(\omega_1|x)P(\omega_2|x)$. And a lower bound can be found by replacing $\min[P(\omega_1|x), P(\omega_2|x)]$ with $P(\omega_1|x)P(\omega_2|x)$.

Answer: Without loss of generality, we can assume that $P(\omega_1|x) \geq P(\omega_2|x)$, then we have $P(error) = \int P(\omega_2|x)p(x)dx$.

Obviously $P(\omega_2|x) \geq P(\omega_1|x)P(\omega_2|x)$ (since $0 \leq P(\omega_1|x) \leq 1$). So a lower bound of $P(error)$ is $\int P(\omega_1|x)P(\omega_2|x)p(x)dx$.

On the other hand, since $P(\omega_1|x) = 1 - P(\omega_2|x)$, so $P(\omega_1|x) \geq \frac{1}{2} \Rightarrow 2P(\omega_1|x) \geq 1$. Thus we have $P(\omega_2|x) \leq 2P(\omega_1|x)P(\omega_2|x)$. So a upper bound of $P(error)$ is $\int 2P(\omega_1|x)P(\omega_2|x)p(x)dx$.

P.3 Matlab Exercise

(a) Write a function to calculate the discriminant function of the following form

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^t\Sigma_i^{-1}(x-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

for a given mean vector and a covariance matrix

(b) Write a function to calculate the Mahalanobis distance between an arbitrary point x and the mean, μ, of a Gaussian distribution with covariance matrix Σ.

(c) Use the data set shown in the table on page 80 of the textbook, Assume each category has Gaussian distribution. Compute the mean and covariance matrix for each category. Assume the prior probabilities are $P(\omega_1) = 0.8$, $P(\omega_2) = P(\omega_3) = 0.1$. Then use your procedures developed in the previous parts (a) and (b) to classify the following test data points: (1,2,1), (5,3,1), (0,0,0) and (1,0,0).

Answer:

(a) The function is as follows:

```
function y = gaussiandiscriminant(x,mu,sigma,prior)
dim = size(x,1);
y = -0.5*((x-mu)')*(inv(sigma))*(x-mu)
    -(0.5*dim*log(2*pi))-(0.5*log(det(sigma)))+log(prior);
```

(b) The function is as follows:

```
function y = mahalanobis(x,mu,sigma)
y = sqrt((x-mu)'*(inv(sigma))*(x-mu));
```

(c)

```
function d = test()
x1 = [[-5.01,-5.43,1.08,0.86,-2.67,4.94,-2.51,-2.25,5.56,1.03]',[-8.12,-3.48,-
5.52,-3.78,0.63,3.29,2.09,-2.13,2.86,-3.33]',[-3.68,-3.54,1.66,-4.11,7.39,2.08,-
2.59,-6.94,-2.26,4.33]'];
```

```
x2 = [[-0.91,1.30,-7.75,-5.47,6.14,3.60,5.37,7.18,-7.39,-7.50]',[-0.18,-2.06,-4.54,0.50,5.72,1.26,-4.63,1.46,1.17,-6.32]',[-0.05,-3.53,-0.95,3.92,-4.85,4.36,-3.65,-6.66,6.30,-0.31]'];

x3 = [[5.35,5.12,-1.34,4.48,7.11,7.17,5.75,0.77,0.90,3.52]',[2.26,3.22,-5.31,3.42,2.39,4.33,3.97,0.27,-0.43,-0.36]',[8.13,-2.66,-9.87,5.19,9.21,-0.98,6.65,2.41,-8.71,6.43]'];

p1 = [1,2,1]'; p2 = [5,3,1]'; p3 = [0,0,0]'; p4 = [1,0,0]';

prior1 = 0.8; prior2 = 0.1; prior3 = 0.1;

mu1 = (mean(x1))'; mu2 = (mean(x2))'; mu3 = (mean(x3))';

sigma1 = cov(x1); sigma2 = cov(x2); sigma3 = cov(x3);

g11 = gaussiandiscriminant(p1,mu1,sigma1,prior1);

g12 = gaussiandiscriminant(p1,mu2,sigma2,prior2);

g13 = gaussiandiscriminant(p1,mu3,sigma3,prior3);

[g1,d1]=max([g11,g12,g13]);

g21 = gaussiandiscriminant(p2,mu1,sigma1,prior1);

g22 = gaussiandiscriminant(p2,mu2,sigma2,prior2);

g23 = gaussiandiscriminant(p2,mu3,sigma3,prior3);

[g2,d2]=max([g21,g22,g23]);

g31 = gaussiandiscriminant(p3,mu1,sigma1,prior1);

g32 = gaussiandiscriminant(p3,mu2,sigma2,prior2);

g33 = gaussiandiscriminant(p3,mu3,sigma3,prior3);

[g3,d3]=max([g31,g32,g33]);

g41 = gaussiandiscriminant(p4,mu1,sigma1,prior1);

g42 = gaussiandiscriminant(p4,mu2,sigma2,prior2);

g43 = gaussiandiscriminant(p4,mu3,sigma3,prior3);

[g4,d4]=max([g41,g42,g43]);

d = [d1,d2,d3,d4]';
```

The discriminant function for the four data points are

$$
\begin{aligned}
g_1(x_1) &= -7.4565;\ g_2(x_1) = -9.4979;\ g_3(x_1) = -11.6499 \\
g_1(x_2) &= -8.1313;\ g_2(x_2) = -10.2826;\ g_3(x_2) = -8.3497 \\
g_1(x_3) &= -7.0614;\ g_2(x_3) = -9.1658;\ g_3(x_3) = -10.5849 \\
g_1(x_4) &= -7.0601;\ g_2(x_4) = -9.2319;\ g_3(x_4) = -9.1420
\end{aligned}
$$

and by the rule of maximum discriminant function, the classification results are d1=1, d2=1, d3=1, and d4=1, that is, all the four test data points are classified to fall in the first category.

The Mahalanobis distances between the three points and mean vector of each Gaussian distribution are:

$$
\begin{aligned}
d(x_1, \mu_1) &= 1.0150;\ d(x_1, \mu_2) = 0.8581;\ d(x_1, \mu_3) = 2.6748 \\
d(x_2, \mu_1) &= 1.5427;\ d(x_2, \mu_2) = 1.5184;\ d(x_2, \mu_3) = 0.7443 \\
d(x_3, \mu_1) &= 0.4900;\ d(x_3, \mu_2) = 0.2684;\ d(x_3, \mu_3) = 2.2415 \\
d(x_4, \mu_1) &= 0.4872;\ d(x_4, \mu_2) = 0.4518;\ d(x_4, \mu_3) = 1.4623
\end{aligned}
$$

and by the rule of minimum Mahalanobis distance, the classification results are d1=2, d2=3, d3=2, and d4=2.