

CSCI4390/6390 – Data Mining
Fall 2007, Quiz 9
Total Points: 20

Name: _____

1. Given the one dimensional data: $x_1 = 1, x_2 = 4, x_3 = 5, x_4 = 2$. Given that the data was generated from two univariate normal distributions with $\mu_1 = 2, \sigma_1^2 = \frac{1}{2}$ and $\mu_2 = 3, \sigma_2^2 = \frac{1}{2}$, and assuming $P(C_1) = P(C_2) = \frac{1}{2}$. Compute $P(C_1|x_1)$. You may use the following values of e : $e^{-1} = 0.37, e^{-2} = 0.14, e^{-3} = 0.05$ and $e^{-4} = 0.02$. You may ignore any smaller value of e , i.e. $e^{-5} = 0$, and so on.

2. Using the distance matrix below, use the average link method to generate hierarchical clusters. Show the merging distance thresholds.

	A	B	C	D	E
A	0	1	3	2	4
B		0	3	2	3
C			0	1	3
D				0	5
E					0