In The Name of God, The Compassionate, The Merciful

Name: .....

Student ID#: .....

Statistical Pattern Recognition (CE-725)
Department of Computer Engineering
Quiz #2 solution (Feature Selection)
Spring 2011

1. (10 points) Suppose we have dataset A={x1,x2,y} in which x1 and x2 are our features and y is the label. Also suppose that the covariance matrix between these three random variables are as follows:

	x1	**x2**	**y**
x1	a	d	e
x2	d	b	f
y	e	f	c

a. For which covariance values, we may infer that we can ignore x1 in the process of classification without loss of accuracy?

If e=0, label y is independent of x1 (Since they are binary-valued). So we can ignore x1 in classification, regardless of other values.

b. For which covariance values, we may infer that we can ignore either one of features in the process of classification without loss of accuracy?

If |d|>>0, there is high correlation between x1 and x2 and hence information redundancy. So we may ignore one of them without any problem.

c. For which covariance values, we may infer that we **can not** ignore neither x1 nor x2 in the process of classification?

If d=0 and none of e or f are zero, then the features are independent and both are influential on label. So we may not ignore any of them.