4. a) For the following **cost matrix**:

		Predicted		
		a	b	c
Actual	a	0	10	5
	b	50	0	30
	c	20	10	0

If a classification model predicts a with probability 0.4, b with probability 0.3 and c with probability 0.3, what is the optimal (i.e. **lowest cost**) decision?

(7 marks)

b) Compare and contrast **K-means** and **agglomerative clustering**. Make a list of similarities and differences between the two approaches.

(8 marks)

c) Suppose we have market basket data consisting of 100 transactions and 20 items. If the support for item $\{a\}$ is 25%, the support for item $\{b\}$ is 90% and the support for itemset $\{a, b\}$ is 20%. What is the **confidence** of the association rule $\{a\} \rightarrow \{b\}$?

(3 marks)

d) A doctor can run a test for the horrible disease *Examophobia*. The test has two possible outcomes: positive and negative.

It is known that among all students, if Examophobia is present, the test comes out positive 80% of the time, and negative 20% of the time. If Examophobia is not present, the test comes out positive 1% of the time, negative 99%.

Among the general student population, Examophobia is known to occur in 35% of all students.

A student enters the clinic and tests positive for the disease. What is the probability they really have Examophobia?

(7 marks)