Sample for COMP393 Quiz 4

Notes:

1. A "zero-one loss function" assigns a cost of 0 to any correct decision and 1 to any incorrect decision.
2. A "Bayes-optimal" classifier minimizes expected cost.
3. The first two notes above will not appear on the real quiz.

Question 1 (20 points)

Widgets are made of wood (W), metal (M), or fiberglass (F) in proportions 70%, 20% and 10% respectively. A single sensor measures the reflectance R of each widget, and assigns a value of high (H), low (L), or none (N). Reflectance likelihoods for this sensor are:

		P(R|T)		
		R=H	R=L	R=N
widget type T	T=W	0.8	0.1	0.1
	T=M	0.2	0.3	0.5
	T=F	0.3	0.6	0.1

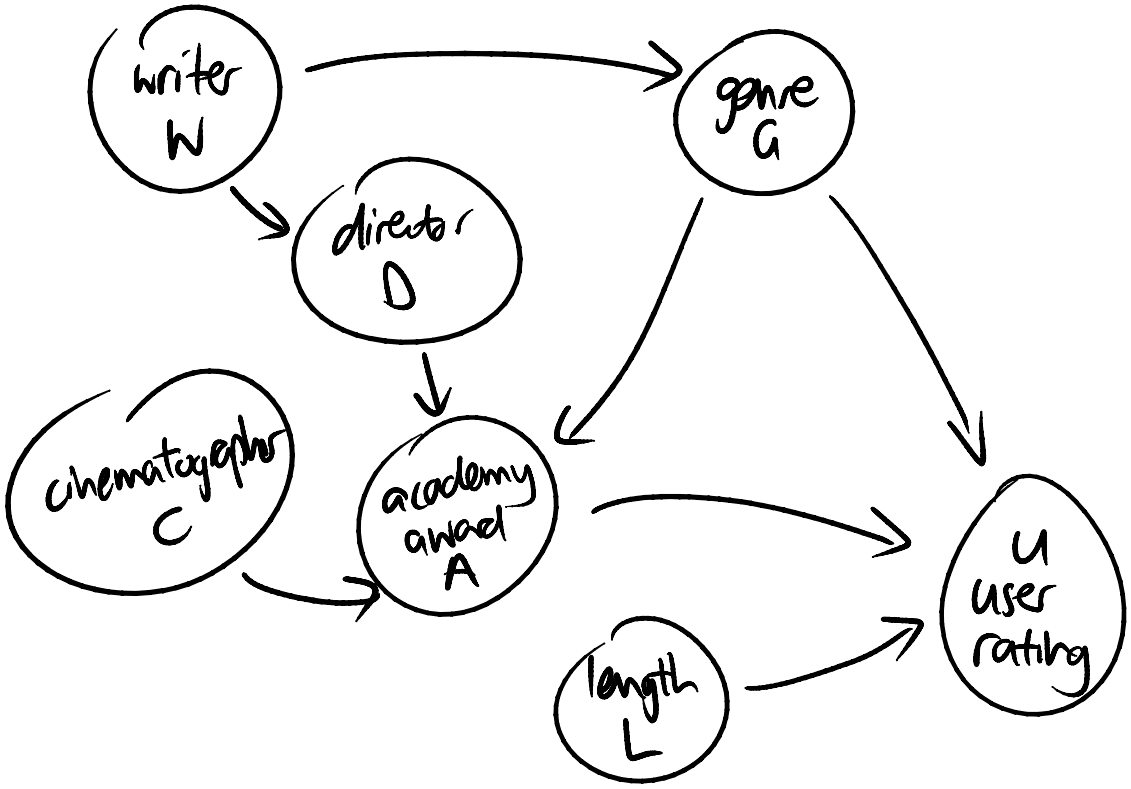
(a) Compute the Bayes-optimal classifier for a zero-one loss function.

(b) Compute the Bayes-optimal classifier for the following cost matrix:

		Estimated T		
		W	M	F
Actual T	W	0	1	1
	M	5	0	4
	F	3	1	0

Question 2 (10 points)

This question refers to the following Bayesian network.



Circle all statements that are true:

P(U|G,L,A) = P(U|G,L,A,C,D)

P(A|C,D,G) = P(A|C,D,G,W)

P(G|D,W) = P(G|D)

P(U|G,L,A) = P(U|G,L,A,D)

P(U|A,G,C,D,W) = P(U|A,G)

Question 3 (10 points)

The temperature values observed in a particular data set are:

42, 65, 94, 58, 37, 94, 35, 34.

What bins would result from discretizing this data into 4 bins using (a) equally-sized bins, and (b) frequency-equalized bins?

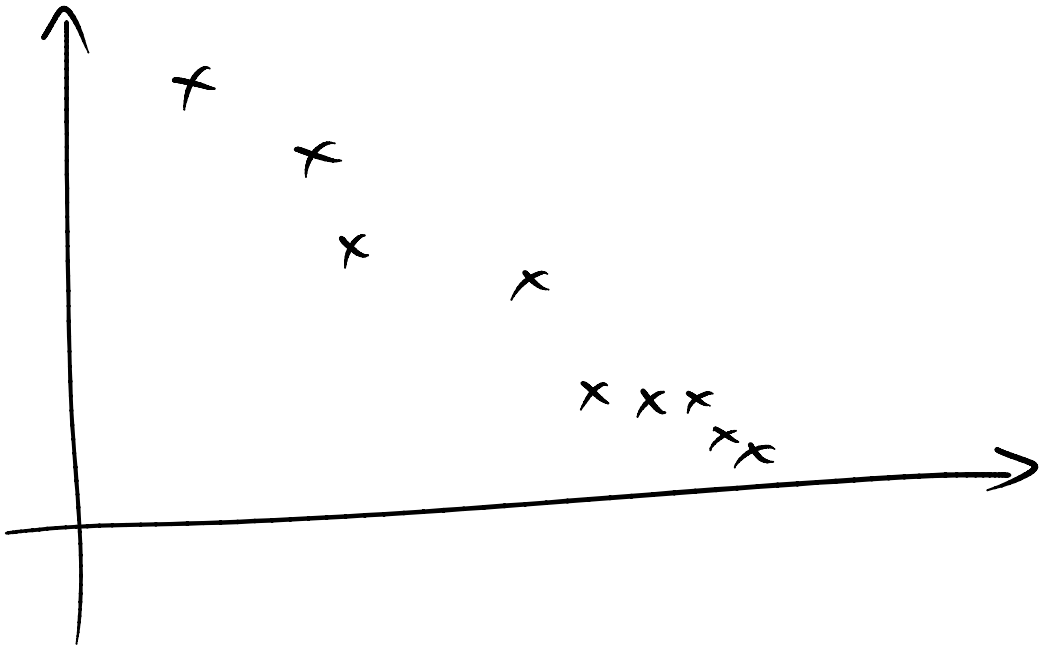
Question 4 (10 points)

Using the two-dimensional data set below, and without doing any precise mathematical calculations:

(a) estimate the first principal component of the data, and sketch the new data set that would result from extracting just the first principal component.

(b) estimate very approximately the proportion of variance captured by the first principal component, explaining your answer clearly.

Data set:



Question 5 (10 points)

A particular classification problem has attributes A, B, and C. Applying a particular classifier to a particular training set achieves the following error rates as estimated by 10-fold cross-validation:

Subset of features used	Error rate (%)
A	30
B	70
C	80
AB	98
AC	90
BC	94
ABC	93

What sequence of subsets is produced by (a) forward selection and (b) backward elimination?