QUEEN'S UNIVERSITY FINAL EXAMINATION

FACULTY OF ENGINEERING AND APPLIED SCIENCE

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

ELEC 409 – Professor Michael Korenberg

April 13, 2015

INSTRUCTIONS TO STUDENTS:

This examination is THREE HOURS in length.

There are 3 questions in this examination, and the marks for each part of a question are indicated in square brackets.

Please answer all questions in the answer booklets.

The following aids are allowed:

Casio FX-991 or red sticker calculator

Datasheets, Math Tables, Notebooks, Textbooks, Photocopies

Put your student number on all pages of all answer booklets, including the front.

GOOD LUCK!

PLEASE NOTE:

"Proctors are unable to respond to queries about the interpretation of exam questions.

Do your best to answer exam questions as written.

1. Consider the matrix

$$A = \begin{bmatrix} 1 & 1 \\ 3 & 5 \\ 4 & 0 \\ 0 & 0 \end{bmatrix}$$

[2] a) What is the singular value decomposition (SVD) equation?

[18] b) Carry out the singular value decomposition of matrix A. If you multiply the matrices you obtain to see if you recover A you may find that the 2 columns are switched, but this does not matter for the purposes of the problem. Hint: one can do the problem without using a calculator by retaining square roots in your work and in the final answer, e.g. $\sqrt{2}$ rather than 1.414…

A determinant can be computed by Laplacian expansion along the first row, e.g. for the 4 x 4 case:

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{vmatrix} = 1 \cdot \begin{vmatrix} 6 & 7 & 8 \\ 10 & 11 & 12 \\ 14 & 15 & 16 \end{vmatrix} - 2 \cdot \begin{vmatrix} 5 & 7 & 8 \\ 9 & 11 & 12 \\ 13 & 15 & 16 \end{vmatrix} + 3 \cdot \begin{vmatrix} 5 & 6 & 8 \\ 9 & 10 & 12 \\ 13 & 14 & 16 \end{vmatrix} - 4 \cdot \begin{vmatrix} 5 & 6 & 7 \\ 9 & 10 & 11 \\ 13 & 14 & 15 \end{vmatrix}$$

2. A predictor was used to predict whether each of a group of 42 patients would survive or die. The number of patients predicted to die was 20 less than the number predicted to survive. The number of patients predicted to die plus the number predicted to survive equals 42. Overall, the predictor made 13 errors: 3 patients who actually survived were predicted to die, and 10 patients who actually died were predicted to survive.

[6] a) How many of the 42 patients actually died and how many actually survived?

[6] b) Enter the numbers for the actual and predicted survived and died into a 2 x 2 contingency table, and include the marginal totals. This table will be referred to as the observed array.

[6] c) Calculate Matthews' correlation coefficient ϕ_1 for the observed array.

[6] d) For the same marginal totals (in the same positions) as in the observed array, show the cell frequencies in each 2 x 2 contingency table having a Matthews' correlation coefficient larger than ϕ_1.

[6] e) For the same marginal totals (in the same positions) as in the observed array, show the cell frequencies in each 2 x 2 contingency table having a Matthews' correlation coefficient less than or equal to $-\phi_1$.

3. The Table below shows predictor scores based on gene expression data for 20 patients with a cardiac disease. Because of the way the predictor was built, it is expected that the more positive the score, the more likely is the outcome to be survival (S); and the more negative the score, the more likely is the outcome to be death (D). For example a score of -2.5 indicates that the patient is expected to be at greater risk of death than a patient with a score of -2.4. The scores are intrinsically continuous, they are presented here rounded to the 1st decimal.

Patient Identifier	Predictor Score	Actual Outcome
P1	17.2	S
P2	6.1	S
P3	-2.2	S
P4	5.1	S
P5	4.7	S
P6	-3.2	S
P7	-2.8	S
P8	-4.1	S
P9	4.1	S
P10	-3.5	S
P11	2.2	S
P12	2.0	S
P13	2.3	D
P14	-1.0	D
P15	3.2	D
P16	-3.8	D
P17	2.1	D
P18	-4.9	D
P19	-5.5	D
P20	-20.2	D

We want to test whether the predictor results are statistically significant using the Mann-Whitney test, which we assume can be validly applied here. Let Group A comprise the 12 survivors, and Group B the 8 patients who died.

[6] (a) Prepare a ranking of the patient outcomes according to their corresponding predictor scores, starting with the most strongly negative score (-20.2) at the top down to the most strongly positive score (17.2) at the bottom. For example, in rank 1 at the top

will be D corresponding to score -20.2, in rank 2 will be D corresponding to score -5.5, etc. At the bottom (rank 20) will be S corresponding to score 17.2.

[6] (b) Calculate the value of T_A, the observed sum of the ranks for Group A, the set of 12 patients with outcome S.

[6] (c) Calculate $T_{A(max)}$, the maximum possible value of T_A when there are 12 patients in Group A, and 8 patients in Group B.

[6] (d) Calculate U_A and U_B and explain how to check that they are consistent.

[6] (e) The observed value of U_A is significant at the 0.05 level if it is equal to or less than 26 on a directional test. The observed value of U_A is significant at the 0.05 level if it is equal to or less than 22 on a non-directional test. Explain whether the results are significant in this problem and whether it is proper to use a directional test.

[12] (f) Using the ranking from (a), sketch the Receiver Operating Characteristic (ROC) curve for plot of true positive fraction (sensitivity) against false positive fraction (equal to 1 - specificity), for each possible cut-off value. The curve should be plotted for predicting death: sensitivity refers to the fraction of death (D) cases that would be correctly classified for each possible cut-off value, and 1 - specificity is the fraction of survivor (S) cases that would be misclassified using the same cut-off. Clearly label the axes, and provide values on each axis showing where the changes in the curve occur.

[8] (g) Calculate the area under the ROC curve (AUC). Your graph and calculation should show enough detail to make it clear how the value of the AUC was obtained.