UNIVERSITY OF BRISTOL

September 2020 Examination Period

FACULTY OF ENGINEERING

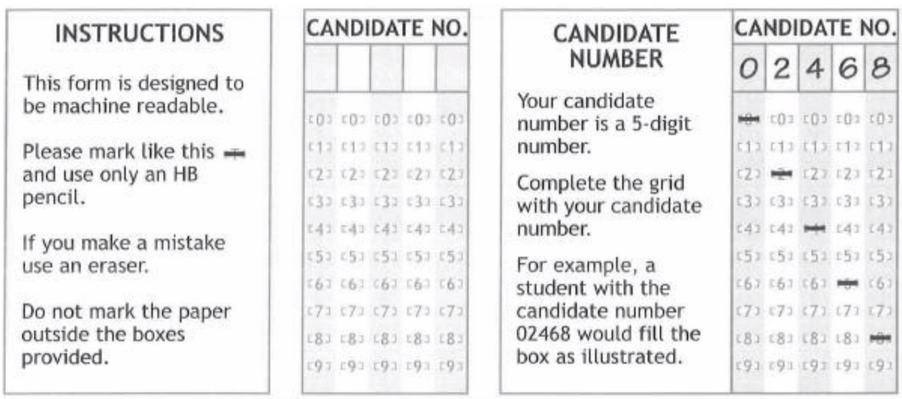
Second Year Examination for the Degrees of BSc, BEng and MEng

COMS21202
Symbols, Patterns and Signal

TIME ALLOWED:
2 hours

This paper contains 20 multiple choice questions worth a maximum of 60 marks
The number of marks allocated for each multiple choice question is indicated
For each multiple choice, choose only one answer
Each question has exactly one correct answer. You must select exactly one answer per question, anything else will not be counted. You must mark your answers with a pencil and only as indicated on the image below.

All questions will be used for assessment.
Calculators must have the Engineering Faculty seal of approval.


INSTRUCTIONS

This form is designed to be machine readable.

Please mark like this and use only an HB pencil.

If you make a mistake use an eraser.

Do not mark the paper outside the boxes provided.

CANDIDATE NO.

CANDIDATE NUMBER

Your candidate number is a 5-digit number.

Complete the grid with your candidate number.

For example, a student with the candidate number 02468 would fill the box as illustrated.

CANDIDATE NO.

0 2 4 6 8

TURN OVER ONLY WHEN TOLD TO START WRITING

Help Formulas:

One-dimensional Gaussian/Normal probability density function:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Multi-dimensional Gaussian/Normal probability density function:

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^M |\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

Least Squares Matrix Form:

$$\mathbf{a}_{LS} = (\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T\,\mathbf{y}$$

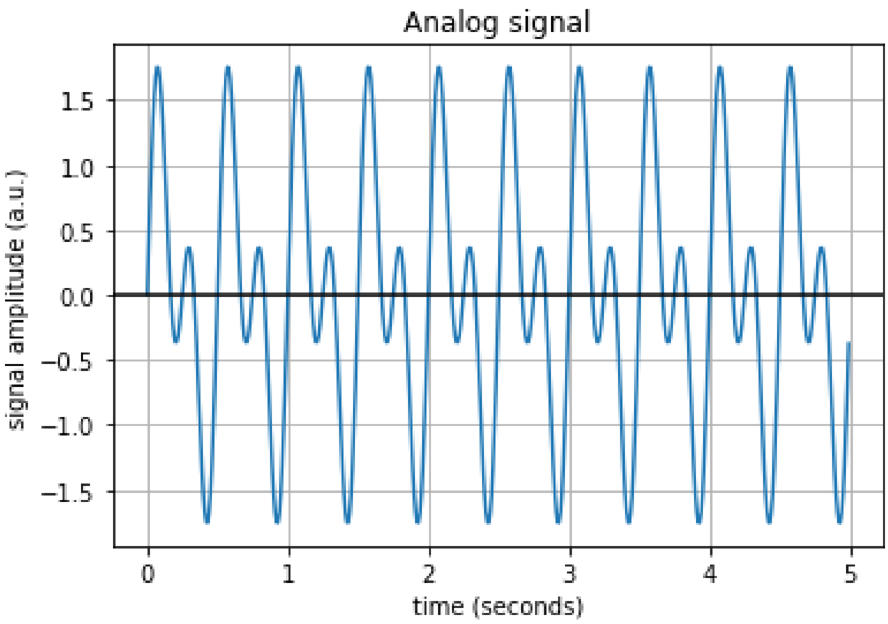
Matrix inversion:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Matrix Determinant:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

1. The image below shows an analogue one-dimensional signal. You are asked to digitise this signal. For that you have to decide which sampling frequency and quantisation to use. Select which of the following would provide a good enough sampling rate and split the signal into 8 levels.



(a) Sampling frequency = 12Hz, quantisation = 4 bits

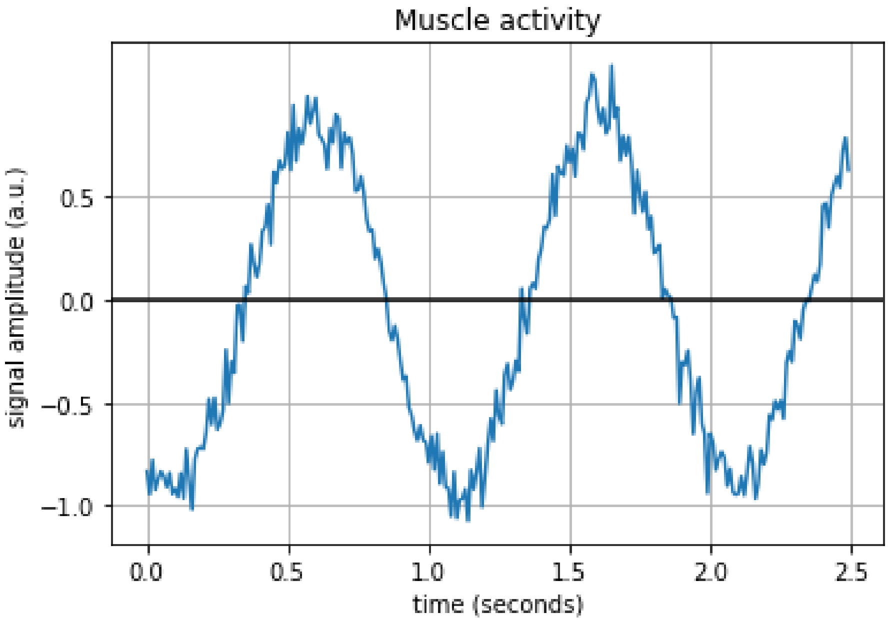
(b) Sampling frequency = 8Hz, quantisation = 3 bits

(c) Sampling frequency = 2Hz, quantisation = 4 bits

(d) Sampling frequency = 2Hz, quantisation = 3 bits

(e) None of the above *[2 marks]*

2. A collaborator records the muscle activity of an olympic athlete while running, and asks for your help to analyse it. You have already decided that the best sampling frequency is 2Hz and that you are going to use 2 bits to digitise the analog signal. What would be the correct binary representation of the signal plotted in the image below.



(a) 001100110110

(b) 001100110111

(c) 011101110110

(d) 011101110111

(e) 001101110111 *[3 marks]*

3. Compute the Hamming and Edit distance between these two strings: datascience and datedscience

(a) Hamming = 2, Edit distance = 3

(b) Hamming = 2, Edit distance = 2

(c) Hamming = 1, Edit distance = 3

(d) Hamming = 1, Edit distance = 2

(e) None of the above *[2 marks]*

4. You collected a four dimensional dataset of values $\mathbf{x} = (x_1, x_2, x_3, x_4)$ with means $(2, 3, 2.4, 1.6)$ and the following covariance matrix

$$\begin{bmatrix} 4 & -1.5 & -2 & -3 \\ -1.5 & 2 & -0.1 & 0 \\ -2 & -0.1 & 1 & 0.1 \\ -3 & 0 & 0.1 & 0.5 \end{bmatrix}$$

You are asked to select x_1 and one another variable to be processed by a machine learning algorithm. Which variable would you pick to best complement the information provided by x_1?

(a) x_1

(b) x_2

(c) x_3

(d) x_4

(e) x_5 *[3 marks]*

5. You are asked to find the most informative dimension of a 2D dataset with the following covariance matrix

$$\begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}$$

What is the most informative eigenvalue and eigenvector with unit vector length?

(a) $\lambda \sim 4.62$ and $v = [0.95 \;\; 0.54]$

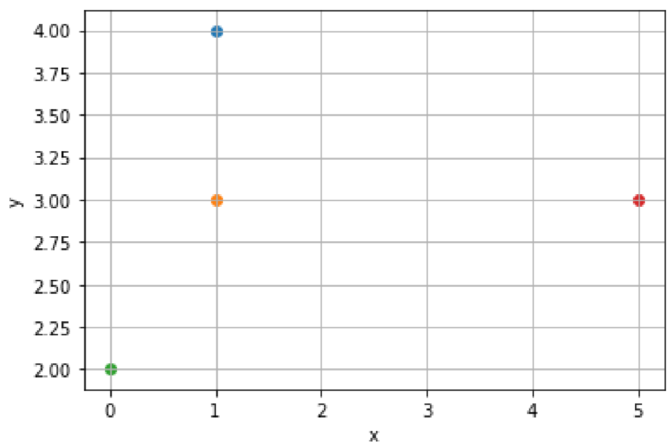
(b) $\lambda \sim 4.62$ and $v = [0.85 \;\; 0.53]$

(c) $\lambda \sim 4.62$ and $v = [0.75 \;\; 0.52]$

(d) $\lambda \sim 2.38$ and $v = [0.15 \;\; 0.54]$

(e) $\lambda \sim 2.38$ and $v = [0.45 \;\; 0.59]$ *[4 marks]*

6. You are given four 2D data points (x,y) represented in the figure below. Assuming a linear model of the form $y = a + bx$ use the general matrix form least squares method to tune the model's parameters. What is the result that you obtain:



(a) $y = 2.88 + 0.07x$

(b) $y = 2.78 + 0.08x$

(c) $y = 2.68 + 0.06x$

(d) $y = 2.61 + 0.05x$

(e) $y = 2.9 + 0.1x$ *[3 marks]*

7. Using the multivariate normal distribution, calculate the probability of the datapoint $x = (1, -1.9)$ using mean $(1, -2)$ and covariance matrix $\begin{bmatrix} 1 & 0.3 \\ 0.3 & 6 \end{bmatrix}$.

   (a) 0.055
   (b) 0.065
   (c) 0.032
   (d) 0.019
   (e) 0.075 *[3 marks]*

8. For the data displayed in the table, compute the least-squares parameter fit for a model of the form, $\hat{y} = w_0 + w_1 x$.

x	y
-1.0	-0.3
0.0	3.2
1.0	6.1

   (a) $\hat{y} = 3.00 + 3.20x$
   (b) $\hat{y} = 3.50 + 3.02x$
   (c) $\hat{y} = 3.57 + 3.32x$
   (d) $\hat{y} = 3.25 + 3.32x$
   (e) $\hat{y} = 3.22 + 3.36x$

[3 marks]

9. Given the test set in the table, which model has the lowest test-squared-error?

x	y
-0.5	-2.2
0.0	0.3
0.5	1.2

   (a) $\hat{y} = -0.5$
   (b) $\hat{y} = 2x - 0.5$
   (c) $\hat{y} = 3x - 1$
   (d) $\hat{y} = -0.3x^2 + 2x - 1$
   (e) $\hat{y} = 0.2x^3 + 2x - 1$

[3 marks]

10. Which of these is not a common cause of overfitting?

   (a) small data

(b) complex model

(c) inputs concentrated in small regions of input space

(d) no regularisation

(e) Bayesian inference

[2 marks]

11. For the following dataset, fit Gaussian distributions to each class using maximum-likelihood, then compute the corresponding posterior over the class-label for $x = 0.1$ (specifically $\mathrm{P}(y = 1|x_0 = 0.1)$).

x	y
-0.5	0
-1.3	0
0.3	0
0.4	1
1.0	1
0.6	1

(a) 0.16

(b) 0.21

(c) 0.25

(d) 0.31

(e) 0.46

12. Assuming K-means with $K = 2$ cluster centers which are initialized at -1.9 and 1.4, assign datapoints to the nearest clusters, and compute the loss function (i.e. the sum-square distance between the cluster centers and assigned data points) for the following data,

x
-2.3
-1.5
1.2
1.3
1.7

(a) 0.40

(b) 0.53

(c) 0.57

(d) 0.42

(e) 0.46

[3 marks]

13. Do a GMM M-step for a single cluster. In particular, x is the location of each datapoint, and p is the posterior probability that this datapoint belongs to the first cluster, compute the optimal mean for the first cluster (by taking the weighted mean of the datapoints).

x	p
-3.0	0.1
-2.3	0.3
-1.5	0.2
1.2	0.7
1.3	0.9
1.7	0.95

(a) 0.62

(b) 0.74

(c) 1.03

(d) 1.23

(e) 0.82

[3 marks]

14. The 2D Fourier transform can be performed as two 1D transforms. The same is true for some spatial filters. What is the equivalent 2D filter corresponding to applying the 1D filter $f = (-2\ \ 2\ -2)$ twice, once horizontally and once vertically?

(a) $\begin{pmatrix} -4 & 4 & -4 \end{pmatrix}$

(b) $\begin{pmatrix} -4 & 4 & -4 \\ 4 & -4 & 4 \\ -4 & 4 & -4 \end{pmatrix}$

(c) (12)

(d) $\begin{pmatrix} 4 & -4 & 4 \\ -4 & 4 & -4 \\ 4 & -4 & 4 \end{pmatrix}$

(e) $\begin{pmatrix} -4 & 8 & -4 \end{pmatrix}$

[3 marks]

15. Matrix K is a covariance matrix of some 3D data:

$$K = \begin{bmatrix} 1 & -1 & 4 \\ 3 & 2 & -1 \\ 2 & 1 & -1 \end{bmatrix}$$

Which of the following sets is the correct eigenvalues and eigenvectors of K?

(a) Set A $\longrightarrow$ $\lambda_A = 7$ and $\mathbf{e}_A = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$,

(b) Set B $\longrightarrow$ $\lambda_B = -5$ and $\mathbf{e}_B = \begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix}$.

(c) Set C $\longrightarrow$ $\lambda_C = -2$ and $\mathbf{e}_C = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$,

(d) Set D $\longrightarrow$ $\lambda_D = 1$ and $\mathbf{e}_D = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$.

(e) none of the above

[4 marks]

16. The eigenvalues of a dataset are: [25, 20, 11, 4, 1, 0.78, 0.22]. Approximately what variance in the dataset do the first 4 eigenvalues represent (when rounded)?

(a) 97%

(b) 94%

(c) 96%

(d) 95%

(e) 92%

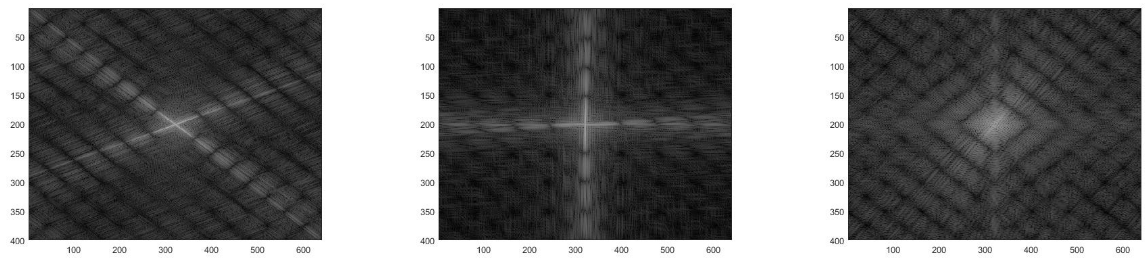
[4 marks]

17. Figure 1 shows handwritten graffiti type letters S, T, and V which are correspondingly labelled (S, T, V).

T S V

Figure 1: Handwritten images of the letters T, S, and V

Below in Figure 2, there are three results, labelled (X, Y, Z) that represent, *in an arbitrary order*, the FFT of the images in Figure 1. Select the choice that correctly states which FFT image corresponds to which graffiti image, using the image labels.



X Y Z

Figure 2: FFT results

(a) (X, Y, Z) corresponds to (V, S, T)

(b) (X, Y, Z) corresponds to (V, T, S)

(c) (X, Y, Z) corresponds to (T, S, V)

(d) (X, Y, Z) corresponds to (T, V, S)

(e) (X, Y, Z) corresponds to (S, V, T)

[5 marks]

18. How are the $(\boldsymbol{a_i}, \boldsymbol{b_i}, \mathbf{cos}, \mathbf{sin})$ terms commonly referred as in the Fourier Series equation below:

$$f(x) = \sum_{n=0}^{\infty} \left[a_n \,\mathbf{cos}(\frac{2\pi\, n\, x}{T}) + b_n \,\mathbf{sin}(\frac{2\pi\, n\, x}{T}) \right]$$

(a) The $\cos$ and $\sin$ terms are the Fourier coefficients, and the $\boldsymbol{a_n}$ and $\boldsymbol{b_n}$ terms are the basis functions.

(b) The $\cos$ and $\sin$ terms are the basis coefficients, and the $\boldsymbol{a_n}$ and $\boldsymbol{b_n}$ terms are the Fourier coefficients.

(c) The $\cos$ and $\sin$ terms are the Fourier coefficients, and the $\boldsymbol{a_n}$ and $\boldsymbol{b_n}$ terms are the basis coefficients.

(d) The $\cos$ and $\sin$ terms are the basis functions, and the $\boldsymbol{a_n}$ and $\boldsymbol{b_n}$ terms are the Fourier coefficients.

(e) none of the above

[2 marks]

19. A 5x5 spatial filter has all its elements set to -0.25, except for the central element which is set to 7. It must then have a:

(a) normalisation factor of 1/6

(b) normalisation factor of 1/1

(c) normalisation factor of 1/13

(d) normalisation factor of 1/7

(e) normalisation factor of 1/25

[2 marks]

END OF PAPER