UNIVERSITY OF BRISTOL

Summer 2025 Examination Period

FACULTY OF ENGINEERING

Second Year Examination for the Degree of Bachelor of Science and Master of Engineering

COMS20017
Algorithms & DATA

TIME ALLOWED:
2 Hours

Answers to COMS20017: Algorithms & DATA

Intended Learning Outcomes:

Q1. What does the term 'Outlier' mean?

A. A value that is left out of the analysis because of missing data

B. When two data points are the same and one can be discarded

C. An extreme value at either end of a distribution

D. A type of variable that is not easy to quantify

E. None of the above

[2 marks]

Solution: C - an outlier is a point with a value significantly different from that of the other points

Q2. Given function $F(t) = \sum_{n=0}^{8} 12cos(2\pi nt)$, what is the least number of sample points needed to satisfy the Nyquist criterion?

A. 16

B. 24

C. 18

D. 96

E. 108

[4 marks]

Solution: A - Since the number of data points should be "at least twice" the highest harmonic component present in the function, which is 8, then the answer is 16.

Q3. Let the Fourier Series expansion of a function $f(x)$ to be:
$f(x) = \Sigma_{n=0}^{\infty} a_n cos(\frac{2\pi nx}{T}) + b_n sin(\frac{2\pi nx}{T})$.

Here are two different functions:

$$f_1(x) = x^2, \quad -\pi < x < \pi \qquad \text{and} \qquad f_2(x) = 1 + |x|, \quad -5 < x < 5$$

If functions $f_1(x)$ and $f_2(x)$ were expanded by their Fourier Series, then which is the correct statement below:

A. All the a_n coefficients in both functions are zero.

B. All the b_n coefficients in both functions are zero.

C. The a_n coefficients in $f_1(x)$ and the b_n coefficients in $f_2(x)$ are zero.

D. The b_n coefficients in $f_1(x)$ and the a_n coefficients in $f_2(x)$ are zero.

E. For even n values only, the b_n coefficients in both functions are zero.

[5 marks]

Solution: B - $f_1(x)$ and $f_2(x)$ are both EVEN functions, which means their b_n terms for all n are zero.

Q4. Here are five statements about PCA (Principal Component Analysis):

1. PCA can be effective when multi-collinearity is suspected among features
2. After performing PCA in a 6D feature space and retaining the first three principal components, this means that the discarded components were not orthogonal to all other components.
3. The curse of dimensionality is when the number of features increases, so does the number of samples, resulting in a more elaborate model.
4. The purpose of eigenvectors in PCA is to determine the number of principal components to retain.
5. PCA is performed to reduce the number and increase the effectiveness of the features used.

Select the option below that is fully correct:

A. 1, 2 and 3 are TRUE.

B. 2, 3 and 4 are TRUE.

C. Only 3 and 5 are TRUE.

D. 3, 4 and 5 are TRUE.

E. 1, 3 and 5 are TRUE.

[5 marks]

Solution: E - 2 and 4 are both FALSE. In 2, all feature dimensions/axes after PCA are orthogonal to each other. In 4, it's the purpose of the eigenvalues to explain the variance in each PC and allow us to determine how many dimensions to keep.

Q5. In the design of your autonomous driving system, you need to decide (as a binary classification problem) whether the path in front of the car shown by the front camera is 'road' or not. For this, you need to assess the performance of 3 different and independent features (f_a, f_b, f_c) by evaluating their t_{score} for 100 images taken by the front camera.

The mean and standard deviations of the resulting two classes for the 100 images and the size of each class (M and N) for each feature are:

(cont.)

$f_a : (\mu_1 = 3.5, \sigma_1 = 0.5), (\mu_2 = 1.5, \sigma_2 = 0.2), M = 75, N = 25$
$f_b : (\mu_1 = 12.2, \sigma_1 = 0.8), (\mu_2 = 5.7, \sigma_2 = 0.9), M = 67, N = 33$
$f_c : (\mu_1 = 1.5, \sigma_1 = 1.5), (\mu_2 = 4.4, \sigma_2 = 0.2), M = 78, N = 22$

Which of the following statements best explains the t_{score} values?

A. f_a is the best feature, and f_b is the second best.

B. f_b is the best feature, and f_c is the second best.

C. f_a is the best feature, and f_c is the worst.

D. f_b is the best feature, and f_a is the second best.

E. f_c is the best feature, and f_b is the worst.

[8 marks]

Solution: D - Putting the values given for each feature into the t_{score} significance test formulae, ranks the features in the following order f_b, f_a, f_c.

Q6. The eigenvalues of the covariance of a dataset are: [26.0, 20.2, 14.3, 10.0, 7.8, 5.3, 2.0, 1.1, 0.77, 0.15]. How many eigenvalues represent 95.41% of the dataset?

A. 5

B. 6

C. 7

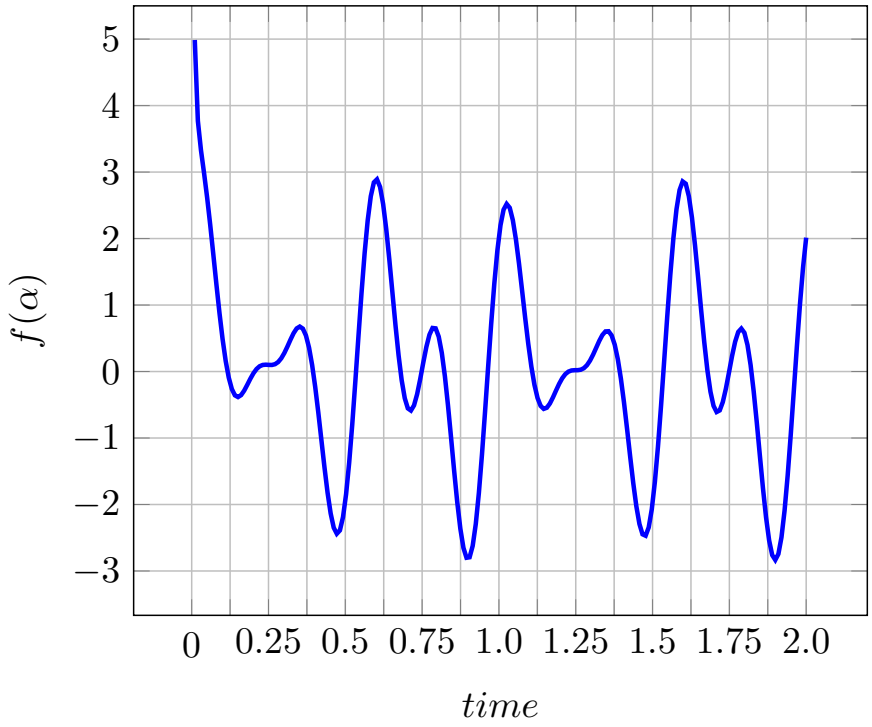
D. 8

E. 9

[4 marks]

Solution: B - Sum of the first 6 eigenvalues divided by the sum of all the eigenvalues, multiplied by 100 and rounded to 2 decimal points.

Q7. You are given a signal that needs to be sampled and quantised into binary values. You have already decided that the best sampling frequency is 4Hz and that you are going to use 3 bits to digitise the analog signal. What would be the correct binary representation of the signal plotted in the image below?



A. 00001100001010001100011100 1

B. 11101000000110001100000 1100

C. 1110110000101000110000 11100

D. 00001100001010011110111100 1

E. 11101100101011100010011110 0

[6 marks]

Solution: C - We encode/quantize the y-axis to values 0-7, represented by binary numbers, from 000 to 111. These are then assigned to values at each time point at 4Hz, i.e. 1/4 in time, starting from the beginning.

Q8. Consider you are in Shanghai (at the origin of a coordinate system of distances to other cities). The coordinates of the city of Seoul relative to you is (80,115), while the coordinates of the city of Manila relative to you is (10,-330). What is the Cosine Distance between Seoul and Manila?

A. 1.803

B. -0.803

C. -0.197

D. 0.197

(cont.)

E. 1.197

[5 marks]

Solution: We apply the Cosine Similarity equation to the vector distances from the origin at Shanghai.

$x = (80, 115), y = (10, -330)$

$x.y = (800) + (-37950) = -37150$

$||x|| = \sqrt{6400 + 13225} = 140.09$

$||y|| = \sqrt{100 + 108900} = 330.15$

$Distance = 1 - Similarity = 1 - \frac{-37150}{140.09 \ x \ 330.15} = 1 - -0.803 = 1.803$

Q9. Here are a number of distances measured between two entities as stated:

(i) M = (-8,7), N = (-2,-2) - Distance MN using ∞-norm (L_∞) is 10
(ii) P = (5,6,12), Q = (2,1,2) - Distance PQ using 3-norm (L_3) is 10.483
(iii) E = (-8,6,-5), F = (-6,-4,9) - Distance EF using ∞-norm (L_∞) is 4
(iv) W = 'Blasted', X = 'Plastered' - Distance WX using Hamming Distance is 3
(v) Y = 'Divides', Z = 'Conquer' - Distance YZ using Edit Distance is 6

Which of the results in the above statements are CORRECT:

A. (ii), (iv) and (v).
B. (i), (ii) and (iii).
C. (iv) and (v)
D. (i), (ii) and (iv).
E. (ii) and (v).

[6 marks]

Solution: E - For Hamming Distance the two strings must be the same length. MN and EF distances are incorrect as the Chebyshev Distance has not been applied properly.

Q10. Your friend was recently on an 11-hour, long-haul flight. Each hour during the flight, he or she made a note of the "outside air temperature" reading as displayed on the aircraft information screen. The readings are shown as signal R below:

$$R = \begin{pmatrix} 0 & 25 & -44 & -70 & -70 & 50 & -40 & 0 \end{pmatrix}.$$

You decide to apply an averaging filter h to smooth this R signal, which then results in signal S (noting edge effects when performing convolution), i.e. $S = R * h$:

$$S = \frac{1}{6}\begin{pmatrix} 0 & 50 & -38 & -178 & -368 & -180 & -120 & 20 & -80 & 0 \end{pmatrix}.$$

Given $\frac{1}{6}$ as the normalisation factor, which of these filters below is the correct filter h?

A. $h = \frac{1}{6}(-1 \quad 4 \quad -1)$

B. $h = \frac{1}{6}(2 \quad -2 \quad 2)$

C. $h = \frac{1}{6}(-2 \quad 2 \quad -2)$

D. $h = \frac{1}{6}(2 \quad 2 \quad 2)$

E. $h = \frac{1}{6}(2 \quad 3 \quad 2)$

[5 marks]

Solution: D - The only filter that gives the correct result is (2 2 2).

Q11. Which of these statements about the natural logarithm and its use in data science is FALSE:

A. The logarithm converts products into sums, i.e. $\ln x \cdot y = \ln x + \ln y$.

B. The logarithm converts powers into products, i.e. $\ln x^a = a \ln x$

C. The derivative of the natural logarithm is $\frac{\partial \ln x}{\partial x} = 2x^{-1}$

D. When computing the Maximum Likelihood Estimate (MLE), the parameters with the highest log-likelihood are the same as the parameters with the highest "raw" likelihood.

E. Using log-probabilities rather than "raw" probabilities helps us avoid numerical under/overflow.

[2 marks]

Solution: C - The derivative of the natural logarithm is in fact $\frac{\partial \ln x}{\partial x} = \frac{1}{x}$

Q12. Which of the following statements about the MLE is FALSE?

A. The MLE is an optimality criterion.

B. The MLE is asymptotically unbiased.

C. The MLE achieves the Cramer-Rao Lower Bound.

D. The MLE is a consistent estimator.

E. The MLE cannot always be obtained in closed form.

[3 marks]

Solution:

A - The ML estimation is a procedure, not an optimality criterion.

Q13. Let x denote a Rayleigh distributed random variable with probability density function given by

$$p(x|\theta) = \frac{x}{\theta^2} \exp\left(-\frac{x^2}{2\theta^2}\right)$$

What is the MLE of θ?

A. $\hat{\theta}_{MLE} = \frac{1}{2}\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$

B. $\hat{\theta}_{MLE} = \frac{1}{2N}\sqrt{\sum_{i=1}^{N} x_i^2}$

C. $\hat{\theta}_{MLE} = \frac{1}{2N}\sum_{i=1}^{N} x_i^2$

D. $\hat{\theta}_{MLE} = \sqrt{\frac{1}{2N}\sum_{i=1}^{N} x_i^2}$

E. $\hat{\theta}_{MLE} = \frac{1}{N}\sum_{i=1}^{N} x_i^2$

[5 marks]

Solution: D - The likelihood function can be written as

$$\begin{aligned} L(x|\theta) &= \prod_{i=1}^{N} p(x_i|\theta) \\ &= \frac{\prod_{i=1}^{N} x_i}{\theta^{2N}} \exp\left[-\frac{1}{2\theta^2}\sum_{i=1}^{N} x_i^2\right] \end{aligned}$$

Log-likelihood function is

$$l(\theta|x) = \sum_{i=1}^{N} \log x_i - 2N \log\theta - \frac{1}{2\theta^2}\sum_{i=1}^{N} x_i^2$$

Taking the derivative yields

$$\frac{\partial l(\theta|x)}{\partial\theta} = -\frac{2N}{\theta} + \frac{2}{2\theta^3}\sum_{i=1}^{N} x_i^2$$

The ML estimate of θ is obtained by setting to zero the derivative of l w.r.t. θ

$$\frac{\partial l(\theta|x)}{\partial\theta} = 0 \quad \rightarrow \quad \theta^2 = \frac{1}{2N}\sum_{i=1}^{N} x_i^2$$

which leads to the ML estimate of θ

$$\hat{\theta}_{MLE} = \sqrt{\frac{1}{2N}\sum_{i=1}^{N} x_i^2}$$

Q14. What is the estimator that minimizes the Bayes risk when the cost function used for its derivation is the absolute error between the estimate and its true value, i.e. $C\left[\theta, \hat{\theta}(x)\right] = C(\epsilon) = |\epsilon|$?

A. The mode of the posterior distribution.

B. The mode of the prior distribution.

C. The median of the posterior distribution.

D. The mean of the likelihood function.

E. The median of the likelihood function.

[4 marks]

Solution: C - The Bayesian estimator that minimizes the absolute error cost function is the median of the posterior distribution.

See lecture AA06 for a proof.

Q15. For a Maximum-a-Posteriori (MAP) estimator assume that the likelihood PDF is Gaussian with variance σ_n^2 defined as:

$$p(x|\theta) = \frac{1}{\sigma_n\sqrt{2\pi}}\exp\left[-\frac{(x-\theta)^2}{2\sigma_n^2}\right],$$

and that the prior pdf is a zero mean Gaussian with variance σ^2 defined as

$$p(\theta) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{\theta^2}{2\sigma^2}\right).$$

What is the expression for the MAP estimate of θ?

A. $\hat{\theta} = \frac{\sigma_n^2}{\sigma^2} \cdot x$

B. $\hat{\theta} = \frac{\sigma^2}{\sigma_n^2} \cdot x$

C. $\hat{\theta} = \frac{\sigma^2}{\sigma_n^2+\sigma^2} \cdot x$

D. $\hat{\theta} = \frac{\sigma_n^2}{\sigma_n^2+\sigma^2} \cdot x$

E. A MAP estimate cannot be obtained in closed form.

[7 marks]

Solution: C - The MAP estimator is the mode of the posterior pdf:

$$\hat{\theta}(x) = \arg\max_{\hat{\theta}} p(\theta|x)$$

Using Bayes theorem together with the above equation, one gets

$$\begin{aligned}\hat{\theta} &= \arg\max_{\hat{\theta}} p(x|\theta)p(\theta) \\ &= \arg\max_{\hat{\theta}} [\ln p(x|\theta) + \ln p(\theta)]\end{aligned}$$

Plugging the given likelihood and prior PDF's in the last equation above

$$\hat{\theta}(x) = \arg\max_{\hat{\theta}} \left[-\frac{(x-\theta)^2}{2\sigma_n^2} - \frac{\theta^2}{2\sigma^2}\right]$$

Taking the derivative w.r.t θ

$$\frac{d}{d\theta}\left[-\frac{(x-\theta)^2}{2\sigma_n^2} - \frac{\theta^2}{2\sigma^2}\right] = 0$$

Setting equal to 0 and solving for θ:

$$\begin{aligned}\frac{(x-\hat{\theta})}{\sigma_n^2} - \frac{\hat{\theta}}{\sigma^2} &= 0 \\ \hat{\theta} &= \frac{\sigma^2}{\sigma_n^2 + \sigma^2} \cdot x\end{aligned}$$

Q16. For the following data,

x	y
-2.1	0
-0.9	0
0.2	0
1.2	1
2.4	1

Fit a Gaussian distribution to each class, and compute the posterior probability that $x = 1.3$ is in class 1, given a prior of $P(y = 1) = 0.4$.

A. 0.92

B. 0.95

C. 0.97
D. 0.99
E. 1.00

[8 marks]

Solution: To fit a Gaussian distribution, estimate μ and σ^2 for each of the two classes.

Class $y = 0$:

$$\mu_0 = \frac{1}{3}\sum_{i=1}^{3} x_i = \frac{-2.1 - 0.9 + 0.2}{3} = -0.93$$

$$\sigma_0^2 = \frac{1}{3}\sum_{i=1}^{3}(x_i - \mu_0)^2 =$$

$$= \frac{1}{3}[(-2.1 + 0.93)^2 + (-0.9 + 0.93)^2 + (0.2 + 0.93)^2] = 0.8822$$

Class $y = 1$:

$$\mu_1 = \frac{1.2 + 2.4}{2} = 1.8$$

$$\sigma_1^2 = \frac{[(1.2 - 1.8)^2 + (2.4 - 1.8)^2]}{2} = 0.36$$

The conditional (posterior) pdf of $y = 1$ given $x = 1.3$ is:

$$p(y = 1|x = 1.3) = \frac{p(x = 1.3|y = 1)p(y = 1)}{p(x = 1.3)}$$

The Gaussian likelihood is given by:

$$p(x|y) = \frac{1}{\sigma_y\sqrt{2\pi}} \exp\left[-\frac{(x - \mu)^2}{2\sigma_y^2}\right]$$

For $x = 1.3|y = 1$ we have:

$$p(x = 1.3|y = 1) = \frac{1}{0.36\sqrt{2\pi}} \exp\left[-\frac{(1.3 - 1.8)^2}{2 \times 0.36}\right]$$

The denominator of the conditional pdf can be calculated as:

$$p(x = 1.3) = p(x = 1.3|y = 0)p(y = 0) + p(x = 1.3|y = 1)p(y = 1) = 0.203$$

Similarly as for $x = 1.3|y = 1$ above, one can get:

$$p(x = 1.3|y = 0) = 0.0251$$

and finally $p(y = 0) = 1 - p(y = 1) = 0.6$

Substituting into the expression for the conditional pdf above, we get:

$$p(y = 1|x = 1.3) = \frac{0.469 \times 0.4}{0.203} = 0.924$$

Q17. In a two-class classification problem with a single feature, x, the pdfs are Gaussians with variances $\sigma^2 = 1/2$ for both classes and mean values 0 and 1, respectively:

$$P(x|\omega_1) = \frac{1}{\sqrt{\pi}} exp(-x^2)$$

and

$$P(x|\omega_2) = \frac{1}{\sqrt{\pi}} exp\left[-(x-1)^2\right]$$

If $P(\omega_1) = P(\omega_2)$ and the loss matrix is $L = \begin{pmatrix} 0 & 0.5 \\ 1 & 0 \end{pmatrix}$, what is the threshold value, x_0, for minimum risk classification?

A. $1/2$

B. $1 + \ln 2/2$

C. $(1 + \ln 2)/2$

D. $(1 - \ln 2)/2$

E. $1 - \ln 2/2$

[7 marks]

Solution: D - The likelihood ratio is given by

$$l_{12} = \frac{P(x|\omega_1)}{P(x|\omega_2)}$$

The threshold is the point x_0 where

$$l_{12} = \frac{P(x|\omega_1)}{P(x|\omega_2)} = \frac{P(\omega_2)\lambda_{21} - \lambda_{22}}{P(\omega_1)\lambda_{12} - \lambda_{11}}$$

Substituting the given pdf's and loss matrix elements:

$$x_0 : exp(-x^2) = 2exp\left[-(x-1)^2\right]$$

Taking the natural logarithm:

$$x_0 : -x^2 = \ln 2 - (x-1)^2$$

Finally:

$$x_0 = (1 - \ln 2)/2$$

Q18. For the data displayed in the following table, compute the least-squares parameter fit for a model of the form $\hat{y} = w_1 + w_2 x$

x	y
0.5	3.2
1.0	2.4
1.5	2.1
2.0	1.1

A. $\hat{y} = 4.89 + 0.78x$

B. $\hat{y} = 3.67 - 0.92x$

C. $\hat{y} = 4.33 + 1.24x$

D. $\hat{y} = 3.85 - 1.32x$

E. $\hat{y} = 3.85 - 1.36x$

[7 marks]

Solution: D -

$$w_2 = \frac{\sum_{i=1}^{N} x_i y_i - N\overline{xy}}{\sum_{i=1}^{N} x_i^2 - N\overline{x}^2} =$$

$$= \frac{0.5 \times 3.2 + 1 \times 2.4 + 1.5 \times 2.1 + 2.0 \times 1.1 - 4 \times 1.25 \times 2.2}{0.5^2 + 1 + 1.5^2 + 2.0^2} =$$

$$= \frac{-1.65}{1.25} = -1.32$$

$$w_1 = \overline{y} - w_2 \overline{x} = 2.2 + 1.32 \times 1.25 = 3.85$$

Q19. When is overfitting least likely to be a serious issue?

A. When fitting a straight line (i.e. $y = w_0 x + w_1$) with lots of training data.

B. When little training data is available.

C. When fitting a complex nonlinear function.

D. when input data, **x**, is a high-dimensional vector.

(cont.)

E. When fitting a high-order polynomial.

[4 marks]

Solution: A - There is little risk of overfitting when fitting a straight line with lots of data.

Q20. Which of the following statements is FALSE?

A. Regularisation mitigates overfitting.

B. Regularisation always improves prediction performance.

C. Regularisation penalises very large values of the weights.

D. The regularization parameter controls the strength of the penalty applied to the model's loss function.

E. Regularization is particularly useful when fitting a complex function.

[3 marks]

Solution:

B - Regularization does not always lead to improved performance. If the penalty term is weighted too high by the regularization parameter, it can be detrimental to prediction performance.

END OF PAPER

The following pages are left blank for your rough workings. They will not be collected or marked. You must enter your answers on the provided answer sheet only.