UNIVERSITY OF BRISTOL

Summer 2024 Examination Period

FACULTY OF ENGINEERING

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

COMS20011
Data-Driven Computer Science

TIME ALLOWED:
2 Hours

Answers to COMS20011: Data-Driven Computer Science

Intended Learning Outcomes:

Help Formulas:

Minkowski distance:

$$D(x, y) = (\sum_{i=1}^{n} |x_i - y_i|^p)^{\frac{1}{p}}$$

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:

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

2D Convolution:

$$g(x, y) = \sum_{m=-1}^{1} \sum_{n=-1}^{1} h(m, n) f(x - m, y - n)$$

Cosine Similarity:

$$cos(\theta) = \frac{x.y}{||x||\ ||y||}$$

Least Squares Matrix Form:

$$\mathbf{w}^* = (\mathbf{X}^T \mathbf{X})^{-1}\ \mathbf{X}^T\ \mathbf{y}$$

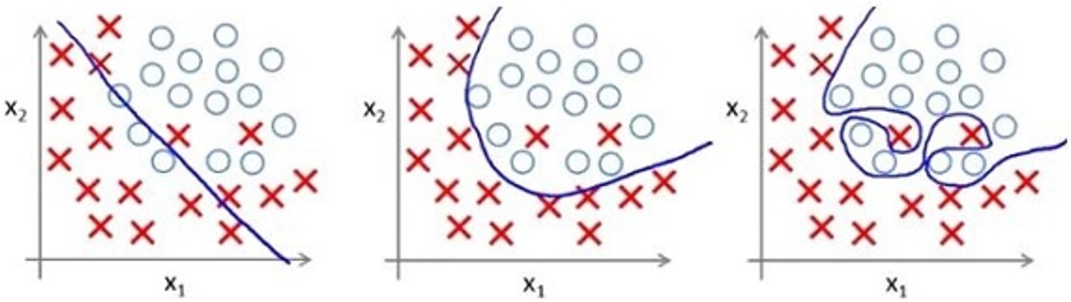
Matrix inversion:

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

Matrix Determinant:

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

Q1. Below are three scatter plots (W, Y, and Z, left to right) of some training data for measurements of two features (x_1, x_2) of different kinds of fish. Also shown, are hand-drawn decision boundaries for modelling regression on the data:



Plot W Plot Y Plot Z

Consider the following statements about these scatter plots:

1. The training error in plot Z is the least compared to plots W and Y.
2. The worst model for this regression problem is in plot Y because it has few training errors.
3. Plot Y is more robust than W and Z because it will perform best on unseen data.
4. All models will perform the same because we have not seen the testing data.
5. The shape of the decision boundary in plot W is least affected by the different number of sample points in the two clusters.
6. The model in plot Z is overfitting the training data compared to W and Y.

Which of the above statements are FALSE conclusions:

A. 1, 3, and 6

B. 2, 4, and 5

C. 1, 5, and 6

D. 3, 4, and 5

E. 2, 4, and 6

[5 marks]

Solution: B - 2, 4, and 5 are FALSE.

Q2. Here are a number of distances measured between two entities (points or vectors etc.) as stated:

(i) M = (5,0,4), N = (-1,5,2) - Distance MN using 1-norm (L_1) is 13

(ii) P = (5,6,2), Q = (1,1,1) - Distance PQ using 3-norm (L_3) is 5.7489

(iii) E = (-4,4,6), F = (2,-1,2) - Distance EF using ∞-norm (L_∞) is 6

(iv) W = 'Aeroplane', X = 'Airbusses' - Distance WX using Hamming Distance is 7

(v) Y = 'Emptied', Z = 'Unloaded' - Distance YZ using Edit Distance is 6

Which of the results in the above statements are CORRECT:

A. (i), (ii), (iii), and (iv)

B. (i), (iv) and (v)

C. (i), (ii), and (iv)

D. (i), (ii), (iv) and (v)

E. They are ALL correct.

[5 marks]

Solution: All distance are correct as shown.

Q3. Consider the following 2D matrix:

$$M = \begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix}$$

Choose the most valid option:

A. The only possible eigenvectors are $\begin{bmatrix} 2 \\ \frac{2}{3} \end{bmatrix}$, $\begin{bmatrix} \frac{1}{3} \\ -1 \end{bmatrix}$.

B. The only possible eigenvectors are $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$, $\begin{bmatrix} -1 \\ -3 \end{bmatrix}$.

C. The only possible eigenvectors are $\begin{bmatrix} 1 \\ \frac{1}{3} \end{bmatrix}$, $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

D. The only possible eigenvectors are $\begin{bmatrix} 1 \\ \frac{1}{3} \end{bmatrix}$, $\begin{bmatrix} 1 \\ -3 \end{bmatrix}$.

E. The eigenvectors in A and D are correct.

[5 marks]

Solution:

To solve for the eigenvalues and eigenvectors, we need to find the roots of the characteristic equation.

Using $det(C - \lambda I) = 0$, where I is the 2x2 identity matrix:

$\left|M - \lambda I\right| = \left|\begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right| = 0$

Expanding the determinant and solving for λ:

$$\begin{vmatrix} 7-\lambda & 3 \\ 3 & -1-\lambda \end{vmatrix} = 0 \quad \Rightarrow \quad \lambda^2 - 6\lambda - 16 = 0$$

Simplifying this to: $(\lambda - 8)(\lambda + 2) = 0$,

we get the eigenvalues as: $\lambda_1 = 8,\ \lambda_2 = -2$

Using $Cv = \lambda v$ to work out the eigenvectors. You should get the following corresponding eigenvectors:

$\begin{bmatrix} 1 \\ \frac{1}{3} \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -3 \end{bmatrix} \quad .$

Q4. Here is the Fourier Series equation,

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

How are the $(a_n, b_n, \cos, \sin)$ terms commonly referred as?

A. The cos and sin terms are the Fourier coefficients, and the a_n and b_n terms are the basis functions.

B. The cos and sin terms are the basis coefficients, and the a_n and b_n terms are the Fourier coefficients.

C. The cos and sin terms are the Fourier coefficients, and the a_n and b_n terms are the basis coefficients.

D. The cos and sin terms are the basis functions, and the a_n and b_n terms are the Fourier coefficients.

E. None of the above are correct.

[3 marks]

Solution: D

Q5. The eigenvalues of a 9-Dimensional dataset are: [76.6, 22.2, 14.4, 10.0, 5.5, 2.5, 1.0, 0.55, 0.15]. We wish to reduce the dimensionality of the dataset while retaining approximately 96.84% of the variance in the data. How many of the eigenvalues should be retained to facilitate this?

A. 3

B. 4

C. 5

D. 6

E. 7

[5 marks]

Q6. Given function $F(t) = \sum_{0}^{11} 5cos(2\pi nt)$, what is the minimum number of sample points needed to avoid the aliasing effect?

A. 12

B. 24

C. 48

D. 60

E. 240

[5 marks]

Solution: Since the number of data points should be "at least twice" the highest harmonic component present in the function, which is 12 (0 to 11), then the answer is 24.

Q7. The 4x4 matrix below was convolved with a filter f to produce the result shown (ignoring border effects):

$$\begin{bmatrix} 2 & 2 & 2 & 2 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \end{bmatrix} * f = \begin{bmatrix} 6 & 6 \\ 8 & 8 \end{bmatrix}$$

Which is the correct filter f that produced the above result?

A. $f = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

B. $f = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 6 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

C. $f = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 4 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

D. $f = \begin{bmatrix} 3 & 0 & 3 \\ 0 & 6 & 0 \\ 0.5 & 0 & 0.5 \end{bmatrix}$

E. $f = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 4 & 0 \\ 1 & 1 & 1 \end{bmatrix}$

[7 marks]

Solution: C - The correct answer can be found by attempting the convolution with the possible answers.

Q8. Here are six statements about PCA (Principal Component Analysis):

1. PCA is applied when multi-collinearity is suspected among features

(cont.)

2. After performing PCA in a 3D feature space, the relationship between the 1st principal component and the 3rd principal component is that they are orthogonal (uncorrelated) to each other.
3. The curse of dimensionality is when the number of features decreases, so does the number of samples, resulting in a complex model.
4. The purpose of eigenvectors in PCA is to determine the number of principal components to retain.
5. the typical range of values for eigenvalues in PCA when analyzing a dataset is any real value.
6. PCA is performed to increase the number and effectiveness of the features.

Select the option below that is fully correct:

A. 1 and 2 are TRUE, the rest are FALSE.

B. 4 and 5 and 6 are TRUE, the rest are FALSE.

C. 1 and 2 and 4 are TRUE, the rest are FALSE.

D. 1 and 4 are TRUE, the rest are FALSE.

E. 1 and 4 and 5 are TRUE, the rest are FALSE.

[5 marks]

Solution: 3, 4, 5, and 6 are all FALSE. 3 should say "increases" to be true. In 4, the purpose of the eigenvalues is to explain the variance in each PC. In 5, the range of values is any positive value. I6 is FALSE because PCA does not increase the number of features.

Q9. You are given a four-dimensional data set, where each sample is a four-dimensional vector $\mathbf{v} = (v_1, v_2, v_3, v_4)$, with the following covariance matrix:

$$\begin{bmatrix} 9.1 & 8.0 & -1.5 & 1.0 \\ 8.0 & 7.5 & 0.5 & 1.0 \\ -1.5 & 0.5 & 7.6 & 1.0 \\ 1.0 & 1.0 & 1.0 & 0.5 \end{bmatrix}$$

Which of the following conclusions can definitively be demonstrated by the covariance matrix?

A. v_2 has strong correlation to v_4

B. either v_2 or v_3 can be ignored as a redundant feature dimension as they have very similar variances

C. v_1 is weakly correlated with v_2

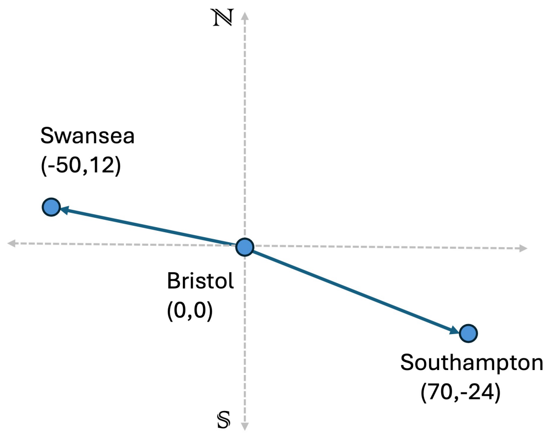
D. v_3 has a negative correlation with v_1

E. v_2 has the lowest variance

[5 marks]

Solution: D - is the only fact that is demonstrated by the covariance matrix.

Q10. The diagram below (approximately) shows relative coordinate positions to indicate distances between Bristol, Swansea, and Southampton, located at (0,0), (-50,12), and (70,-24) respectively.



What is the Cosine Distance between Swansea and Southampton?

A. 0.898

B. 0.102

C. 170.0

D. 200.0

E. None of the above

[5 marks]

Solution: First we apply the Cosine Similarity equation to the vector distances from Bristol: $x = (-50, 12), y = (70, -24)$

$x.y = (-3500) + (-288) = -3788$

$||x|| = \sqrt{2500 + 144} = 51.42$

$||y|| = \sqrt{4900 + 576} = 74.00$

$Distance = 1 - Similarity = 1 - \frac{-3788}{51.42 \ x \ 74.00} = 1 - 0.898 = 0.102$

Q11. Which of the following is NOT a reason that we use log likelihood over likelihood?

A. To avoid running out of floating point precision.

B. Because products are numerically stable.

C. Because it is easier to sum than to multiply.

D. Because log is a monotonically increasing function.

E. To simplify the calculation.

[4 marks]

Q12. Which statement about regularisation is false?

A. Regularisation smoothes overfitted hypersurfaces.

B. We can alter the strength of regularisation using cross-validation.

C. Regularisation can lead to underfitting.

D. Regularisation allows you to fit volatile functions more accurately.

E. Regularisation penalises large regression weights.

[4 marks]

Q13. Which of these is a feature of overfitting?

A. Better predictions in test performance.

B. Small least squares error against training data.

C. Small loss function result against a broad range of test data.

D. Poor performance on the "training" data used to fit the function.

E. Accurate predictions mean minimal need for regularisation.

[4 marks]

~~**Q14**. Which of these statements about the Kronecker delta is FALSE?~~

A. $\cancel{\sum_{i,j} \delta_{ij} x_i y_j = \mathbf{x} \cdot \mathbf{y}}$.

B. $\cancel{\sum_i a_i \delta_{ij} = a_j}$.

C. $\cancel{\delta_{ij} = \delta_{ji}}$.

D. If $j = l$ then $\delta_{jm}\delta_{kl} = \delta_{kl}$.

E. $\cancel{\delta_{ij}\delta_{jk}\delta_{kl} = 1}$ ~~if and only if~~ $\cancel{i = j = k = l}$.

~~*[4 marks]*~~

Q15. For the data in the table, fit a model of the form $\hat{y} = w_1 x + w_2 x^2$

x	y
-2	-6.2
-1	-2.6
0	0.5
1	2.5
3.1	5.7

A. $w_1 = 2.644$, $w_2 = 0.241$

B. $w_1 = -1.55$, $w_2 = 0.49$

C. $w_1 = 2.589$, $w_2 = -0.24$

D. $w_1 = -0.524$, $w_2 = 2.29$

E. $w_1 = 2.958$, $w_2 = -0.293$

[4 marks]

Solution: Seek $(X^T X)^{-1} X^T y$

$$X = \begin{bmatrix} -2 & 4 \\ -1 & 1 \\ 0 & 0 \\ 1 & 1 \\ 3.1 & 9.61 \end{bmatrix} \tag{1}$$

$$X^T X = \begin{bmatrix} 15.61 & 21.791 \\ 21.791 & 110.3521 \end{bmatrix} \tag{2}$$

$$(X^T X)^{-1} = \begin{bmatrix} 0.0088 & -0.0175 \\ -0.0175 & 0.0125 \end{bmatrix} \tag{3}$$

$$X^T y = \begin{bmatrix} 35.17 \\ 29.877 \end{bmatrix} \tag{4}$$

$$(X^T X)^{-1} X^T y = \begin{bmatrix} 2.5887 \\ -0.2404 \end{bmatrix} \tag{5}$$

$$\tag{6}$$

Q16. For the data in the table, fit a model of the form $\hat{y}_i = w_1 X_{i1} + w_2 X_{i2}^2$

X_{i1}	X_{i2}	y_i
-1	0	0.2
0	1	-0.5
1	0	1.2
1	-2	0.9

A. $w_1 = 0.569$, $w_2 = 0.049$

B. $w_1 = 0.545$, $w_2 = 0.046$

C. $w_1 = 0.523$, $w_2 = 0.048$

D. $w_1 = 0.672$, $w_2 = 0.051$

E. $w_1 = 0.401$, $w_2 = 0.044$

[4 marks]

Solution: Seek $(X^T X)^{-1} X^T y$

$$X = \begin{bmatrix} -1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 4 \end{bmatrix} \tag{7}$$

$$X^T X = \begin{bmatrix} 3 & 4 \\ 4 & 17 \end{bmatrix} \tag{8}$$

$$(X^T X)^{-1} = \begin{bmatrix} 0.4857 & -0.1143 \\ -0.1143 & 0.0857 \end{bmatrix} \tag{9}$$

$$X^T y = \begin{bmatrix} 1.9 \\ 3.1 \end{bmatrix} \tag{10}$$

$$(X^T X)^{-1} X^T y = \begin{bmatrix} 0.5686 \\ 0.0486 \end{bmatrix} \tag{11}$$

$$\tag{12}$$

Q17. Compute $\sum_i \log P(y_i|x_i)$ for binary classification, where

$$P(y_i = 1|x_i) = \sigma(-3 + x + 2x^2)$$

with data,

x	y
-1.1	0
-0.9	0
0.2	0
1.2	1
1.4	1

A. -0.69

B. -0.75

C. -0.62

D. -0.74

E. -0.72

[5 marks]

Solution: Seek $f(x) = \sigma(-3 + x + 2x^2)$ for x values where $\sigma(x) = \frac{1}{1+\exp(-x)}$. Then for $y = 1$ seek $\log(f(x))$ and for $y = 0$ seek $\log(1 - f(x))$

x	$f(x)$	$\log(f(x))$	$\log(1-f(x))$
-1.1	0.1571		-0.1709
-0.9	0.0928		-0.0974
0.2	0.0618		-0.0638
1.2	0.7465	-0.2924	
1.4	0.9105	-0.0937	

(13)

(14)

Finally sum the logged values to get -0.7182

Q18. Given the normal density,

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{-(x-\mu)^2}{2\sigma^2},$$

and an independent and identically distributed sample of N datapoints, $x_1, \ldots, x_N$. What is the maximum likelihood solution for the vector $\Theta = (\mu, \sigma)$ given,

$$P(x_i|\Theta) = f(x_i|\mu, \sigma).$$

A. $\mu = \frac{1}{N}\sum_i x_i$, $\sigma = \sqrt{\frac{2}{N}\sum_i (x_i - \mu)}$

B. $\mu = \frac{1}{N}\sum_i x_i$, $\sigma = \frac{1}{N}\sum_i (x_i - \mu)^2$

C. $\mu = \frac{1}{N}\sum_i x_i$**,** $\sigma = \sqrt{\frac{1}{N}\sum_i (x_i - \mu)^2}$

D. $\mu = \frac{1}{N}\sum_i x_i$, $\sigma = \frac{2}{N}\sum_i (x_i - \mu)$

E. $\mu = \frac{2}{N}\sum_i x_i$, $\sigma = \sqrt{\frac{2}{N}\sum_i (x_i - \mu)}$

[7 marks]

(cont.)

Solution:

$$\log P(x|\mu,\sigma) = \log \prod_i P(x_i|\mu,\sigma) \tag{15}$$

$$= \log \left[\prod_i \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{-(x_i-\mu)^2}{2\sigma^2} \right] \tag{16}$$

$$= \log \left[\frac{1}{(\sqrt{2\pi})^N \sigma^N} \exp \sum_i \frac{-(x_i-\mu)^2}{2\sigma^2} \right] \tag{17}$$

$$= -\frac{N}{2}\log(2\pi) - N\log(\sigma) - \frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2 \tag{18}$$

$$\frac{\partial \log P}{\partial \mu} = \frac{1}{\sigma^2}\sum_i (x_i-\mu) = 0 \tag{19}$$

$$\mu = \frac{1}{N}\sum_i x_i \tag{20}$$

$$\frac{\partial \log P}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_i (x_i-\mu)^2 = 0 \tag{21}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_i (x_i-\mu)^2} \tag{22}$$

Q19. Given observations on the scalar value x_i, $i = 1, \dots, N$, and given that each y_i is independently drawn according to the probability distribution function,

$$f(y_i|x_i, \theta) = (x_i\theta)^{-1} \exp \frac{-y_i}{x_i\theta}.$$

What is the maximum-likelihood solution for θ?

A. $\frac{1}{N}\sum_i \frac{x_i}{y_i}$

B. $\frac{1}{N}\sum_i \frac{y_i}{x_i}$

C. $\frac{1}{2N}\sum_i \frac{y_i}{x_i}$

D. $\sqrt{\frac{1}{N}\sum_i \frac{y_i}{x_i}}$

E. $\sqrt{\frac{2}{N}\sum_i \frac{y_i}{x_i}}$

[7 marks]

Solution:

$$\log P(y_i|x_i, \theta) = \sum_i (-\log x_i - \log\theta - \frac{y_i}{x_i\theta}) \tag{23}$$

$$0 = \frac{\partial}{\partial\theta} \log P(x_i|\lambda) = \sum_i (\frac{-1}{\theta} + \frac{y_i}{x_i\theta^2}) \tag{24}$$

$$0 = -\frac{N}{\theta} + \sum_i (\frac{y_i}{x_i}\frac{1}{\theta^2}) \tag{25}$$

$$0 = -\frac{N}{\theta} + \frac{1}{\theta^2}\sum_i (\frac{y_i}{x_i}) \tag{26}$$

$$\theta = \tfrac{1}{N}\sum_i \frac{y_i}{x_i} \tag{27}$$

Q20. An experiment consists of N Bernoulli trials with a success probability of θ,

$$P(x_i|\theta) = \prod_i \theta^{x_i}(1-\theta)^{1-x_i}.$$

What is the maximum-likelihood solution for θ?

A. $(\frac{1}{N}\sum_i x_i)^{\frac{1}{N}}$

B. $\frac{1}{2N}\sum_i x_i$

C. $\frac{1}{N}\sum_i (1 - x_i)$

D. $\frac{1}{N}\sum_i (x_i)^N$

(cont.)

E. $\frac{1}{N}\sum_i x_i$

[7 marks]

Solution:

$$\log P(x_i|\theta) = \log\theta\sum_i x_i + \log(1-\theta)\sum_i(1-x_i) \quad (28)$$

$$0 = \frac{\partial}{\partial\theta}\log P(x_i|\theta) \quad (29)$$

$$= \frac{1}{\theta}\sum_i x_i - \frac{1}{1-\theta}\sum_i(1-x_i) \quad (30)$$

$$= \frac{1}{\theta}\sum_i x_i - \frac{N}{1-\theta} + \frac{1}{1-\theta}\sum_i x_i \quad (31)$$

$$\frac{N}{1-\theta} = (\frac{1}{\theta} + \frac{1}{1-\theta})\sum_i x_i \quad (32)$$

$$\frac{N}{1-\theta} = \frac{1}{\theta(1-\theta)}\sum_i x_i \quad (33)$$

$$\frac{N\theta(1-\theta)}{1-\theta} = \sum_i x_i \quad (34)$$

$$\theta = \frac{1}{N}\sum_i x_i \quad (35)$$

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.