**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:

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

Least Squares Matrix Form:

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

COMS21202 - Symbols, Patterns and Signals

Name: ______________________________

CS Username: ______________________________

Test 1,

## Section 1. Choose One Answer [30 marks]

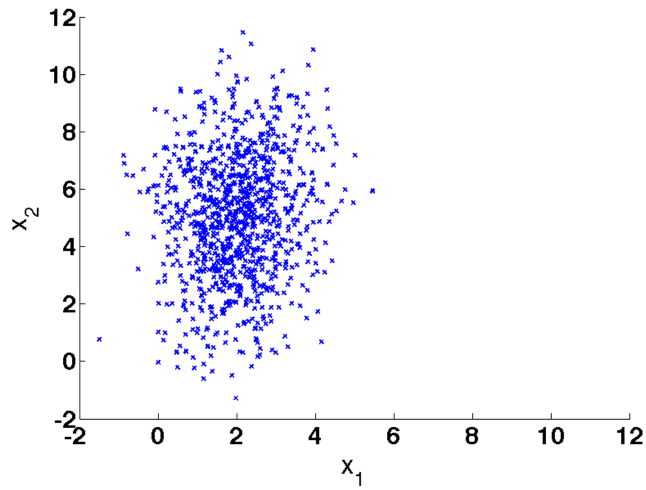
**[a] at the start of the question indicates the number of marks allocated to it**

1. **[2]** For $x = (10, 2)$ and $y = (6, 5)$ which of the following is a correct Minkowski distance?
   (a) For $p = 1$, D(x,y) = 1
   (b) For $p = 2$, D(x,y) = 4
   (c) For $p = 3$, D(x,y) = 9.5
   (d) For $p = \inf$, D(x,y) = 4
   (e) None of the above

2. **[2]** Which of these files has the largest size if stored, raw/uncompressed?
   (a) A one minute phone call with your friend. Recall that speech is sampled at 8Khz and quantised at 8bps.
   (b) 10 seconds of an Audio CD. Recall that Audio CD contains stereo data sampled at 44KHz and quantised at 16bps.
   (c) A colour photo on a 16Mega Pixels camera. Recall that each colour channel is quantised at 8bps.
   (d) A 0.5 second colour video recorded without audio using 1Mega Pixels camera. Note that videos are recorded at 30 frames per second. Recall that each colour channel is quantised at 8bps.
   (e) A whatsapp message

3. **[2]** When calculating the Hamming distance $D_H$ and the Edit distance $D_E$ given two words 'bridge' and 'burger', which of the following is correct?
   (a) $D_H$('bridge', 'burger') = 5, $D_E$('bridge', 'burger') = 4
   (b) $D_H$('bridge', 'burger') = 5, $D_E$('bridge', 'burger') = 5
   (c) $D_H$('bridge', 'burger') = 4, $D_E$('bridge', 'burger') = 4
   (d) $D_H$('bridge', 'burger') = 4 but $D_E$ cannot be calculated over words of the same length.
   (e) None of the above

4. **[2]** For the data sample of 1000 points shown here




which of the following is a reasonable estimate of the model parameters

(a) $\mu = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$, $\Sigma = \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix}$

(b) $\mu = \begin{bmatrix} 2 \\ 8 \end{bmatrix}$, $\Sigma = \begin{bmatrix} 3 & 3 \\ 3 & 1 \end{bmatrix}$

(c) $\mu = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, $\Sigma = \begin{bmatrix} 1 & -3 \\ -3 & 4 \end{bmatrix}$

(d) $\mu = \begin{bmatrix} 2.5 \\ 5 \end{bmatrix}$, $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

(e) None of the above

5. **[2]** When discussing the concepts of generalisation versus overfitting, which of the following statements are incorrect:

(a) An overfitted model achieves better results when tested on the 'training' data

(b) A general model achieves better results on 'future' data

(c) A general model is more complex than an overfitted model

(d) An overfitted model has a higher number of parameters to optimise when compared to a general model

(e) None of the above - all statements are correct

6. **[4]** For a sample of size $N$ of three dimensional points $(x_i, y_i, z_i)$, and considering the model:

$$y = a_0 + a_1 x + a_2 xz + a_3 x.^3$$

The size of the matrices $\mathbf{y}$, $\mathbf{X}$, $\mathbf{a}$ used in the matrix form of the least squares method would be

(a) $\mathbf{y}_{N\times1}$, $\mathbf{X}_{N\times4}$, $\mathbf{a}_{1\times4}$

(b) $\mathbf{y}_{N\times4}$, $\mathbf{X}_{N\times4}$, $\mathbf{a}_{4\times1}$

(c) $\mathbf{y}_{N\times1}$, $\mathbf{X}_{N\times4}$, $\mathbf{a}_{4\times1}$

(d) $\mathbf{y}_{N\times1}$, $\mathbf{X}_{N\times3}$, $\mathbf{a}_{4\times1}$

(e) Least squares cannot be used to solve for this polynomial due to the presence of the term $a_2xz$

7. **[2]** Which of the following pairs of a model and its parameters are incorrect

   (a) A normal distribution has a single parameter $\mu$

   (b) A uniform distribution has two parameters representing the range $[a, b]$

   (c) A linear function $y = mx + c$ has two parameters representing the slope and the y-intercept

   (d) A binomial distribution has one parameter representing the probability of a success $\alpha$

   (e) None of the above - all pairs are correct

8. **[3]** For a one-dimensional numeric data $D$, given a representation of $p(D|\theta)$ for a probabilistic model, **MLE** estimates the model parameter $\hat{\theta} = \arg\max_\theta p(D|\theta)$. Which of the following answers are incorrect:

   (a) $\hat{\theta} = \arg\max_\theta lnp(D|\theta)$ where $ln$ is the natural logarithm function

   (b) $\hat{\theta} = \arg\max_\theta \big(p(D|\theta) + c\big)$ where $c$ is a constant

   (c) $\hat{\theta} = \arg\min_\theta bp(D|\theta)$ where $b < 0$ is a constant

   (d) $\hat{\theta} = \arg\max_\theta p(D + c|\theta)$ where $c > 0$ is a constant

   (e) None of the above - all answers are correct

9. **[2]** The assumption that a sample is **i.i.d** implies that

   (a) The data has been sampled by an expert who has studied the full population.

   (b) The observations are believed to be independent.

   (c) The sample is large enough to estimate the model parameters.

   (d) The sample is multi-dimensional.

   (e) All of the above.

10. **[2]** For a one dimensional numeric data, given a probabilistic model with a single parameter $b$, $var(b_{ML})$ was calculated to be $var(b_{ML}) = \sigma^2 \sum_i x_i$. Based on this finding you advise the data collection team to:

    (a) Collect samples with large values of $x_i$ if possible.

    (b) Collect samples with small values of $x_i$ if possible.

    (c) Collect samples that achieve a uniform distribution of $x_i$ over its range.

    (d) Collect samples around the mean of the distribution.

    (e) Model parameter estimation does not depend on the sample collected, so no change in data collection is needed.

11. **[3]** In a certain COMS module, students were given three assessments ($G_1$, $G_2$, $G_3$). The grades for the three assessments for a sample of students is given below.

| | George | Amy | John | Judith | Adam |
|---|---|---|---|---|---|
| $G_1$ | 5 | 6 | 7 | 8 | 3 |
| $G_2$ | 4 | 6 | 6 | 6 | 2 |
| $G_3$ | 5 | 6 | 7 | 7 | 3 |

Assuming a 3-D Normal distribution, which of the following is the most likely mark for a sixth student 'Alexis' in the same cohort?

(a) $(G_1, G_2, G_3) = (5, 5, 5)$
(b) $(G_1, G_2, G_3) = (5, 5, 6)$
(c) $(G_1, G_2, G_3) = (5, 4, 5)$
(d) $(G_1, G_2, G_3) = (7, 6, 7)$
(e) $(G_1, G_2, G_3) = (8, 6, 4)$

12. **[4]** For the same COMS marks in Q11, a marker, keen on decreasing his workload, believes that the mark for the third assessment ($G_3$) could be estimated from the two assessments ($G_1, G_2$). You decided to select a linear relationship (polynomial of degree 1) to estimate $G_3$ from $G_1$ and $G_2$. Using the matrix form, the best fit prediction would be:

(a) $G_3 = 2G_1 - 1.5G_2$
(b) $G_3 = G_1$
(c) $G_3 = 0.65G_1 + 0.37G_2$
(d) $G_3 = 0.55G_1 + 0.48G_2$
(e) None of the above