# of Questions: 22
Total Exam Points: 62.00

Question #: 1

Consider the matrix

$$\mathbf{A} = \begin{bmatrix} X_{1,1} & X_{1,2} \\ X_{2,1} & X_{2,2} \end{bmatrix}$$

Suppose each entry $X_{i,j}$ is drawn uniformly from the set $\{0, 1\}$, i.e.,

$$\Pr(X_{i,j} = 0) = \Pr(X_{i,j} = 1) = 1/2$$

What is the probability that $\mathbf{A}$ has full rank?

A. 1/8
B. 3/16
✓C. 3/8
D. 1/4
E. None of the others.

Item ID: 247996 / 1
Item Description: MCQ_5
Item Weight: 3.0

Question #: 2

Randomized Controlled Trial (RCT) is a way to conduct causal data analysis. Since we only need to assigns participants into an experimental group or a control group, RCT is always feasible to conduct.

A. True
✓B. False

Item ID: 247883 / 1
Item Description: Copy of Copy of Part1_TF5
Item Weight: 2.0

Question #: 3

The elements in a set are ordered, while the elements in matrix are not ordered.

A. True
✓B. False

Item ID: 247882 / 1
Item Description: Copy of Copy of Part1_TF4
Item Weight: 2.0

Question #: 4

It is a common practice to replace missing data (or feature) in a dataset with NA. When we compute the mean and standard deviation for a specific feature, we may practically treat the NA values as -1 and include them in the mean/standard deviation calculation.

A. True
✓B. False

Item ID: 247881 / 1
Item Description: Copy of Copy of Part1_TF3
Item Weight: 2.0

Question #: 5

We are given the average daily temperature of Singapore in the past 20 years, and we are asked to train a machine learning model that predicts the average temperature next Friday. The model, once trained, conducts deductive reasoning.

A. True
✓B. False

Item ID: 247880 / 1
Item Description: Copy of Copy of Part1_TF2
Item Weight: 2.0

Question #: 6

Artificial intelligence is a subset of machine learning which adopts deductive reasoning to make inference.

A. True
✓B. False

Item ID: 247879 / 1
Item Description: Copy of Copy of Part1_TF1
Item Weight: 2.0

Question #: 7

We have an unbiased die with six faces (i.e., 1,2,3,4,5,6), where each face has equal probability to show up in a rolling game.

Suppose we roll the die twice, and then count the sum of the faces showing up. For example, if the first time 1 shows up and the second time 3 shows up, the sum would be 1+3=4.

What is the probability of obtaining an sum being an **odd number greater than 8**?

A. 1/12

B. 1/9
C. 1/36
D. 1/8
✓E. 1/6
F. None of others is correct.

Item ID: 247878 / 1
Item Description: Copy of Copy of Part1_MCQ4
Item Weight: 3.0

Question #: 8

Please select the correct option.

A. Binary Coding is a common practice to convert categories into binary form, and is treated as a method for data cleaning.
B. Z-score standardization always outputs a value between 0 and 1.
C. Clipping outlier is as a method for data formatting.
D. Imputation is as a method for data formatting.
E. (B) and (C)
F. (A), (B), and (C)
G. (A), and (B)
✓H. None of others is correct.

Item ID: 247877 / 1
Item Description: Copy of Copy of Part1_MCQ3
Item Weight: 3.0

Question #: 9

Please select the correct option.

A. Classification works with labels that are only continuous.
B. Classification works with labels that only belong to ordinal data.
C. Classification works with features that are only continu ous.
D. Classification works with features that are only discrete.
E. Regression works with features that are only discrete.
F. b) and d)
G. b) and c)
✓H. None of others is correct.

Item ID: 247876 / 1
Item Description: Copy of Copy of Part1_MCQ2
Item Weight: 3.0

Question #: 10

We have a collection of 1,000 images from three class labels: cat, bird, and dog. With these images, we would like to train an image classifier that categorizes an input image into one of the three classes.

To ensure the 1,000 images are of good quality for training the classifier, we ask a well-trained human inspector to go through all the images, to label the images and remove noisy ones. Eventually, we revmoved 200 images of low quality suggested by the inspector, and use the remaining 800 images to train the classifier; the 800 images comprise 200 cat images, 300 bird images, and 300 dog images.

For each image, we then use the following features:

1) The number of pixels with black color; this will give us a non-negative integer.

2) The number of pixels with white color; this will also give us a non-negative integer.

We then apply nearest neighbor classifier based on the two features.

Please select the correct option.

A. The human inspection process can be considered as a feature classification step.

B. The used features (number of pixels with black and white colors) belong to nominal data.

C. The image labels (cat, bird, and dog) belong to ordinal data.

D. The nearest neighbor classifier is an unsupervised learning method.

E. (a) and (b)

F. (a), (b), and (d)

✓G. None of others is correct.

Item ID: 247875 / 2
Item Description: Copy of Copy of Part1_MCQ1
Item Weight: 3.0

Question #: 11

There are 300 students in a training camp, who are divided into three training groups, 150 students in Group A, 50 students in Group B, and 100 students in Group C. The students take part in a fitness test.

Let G be the variable that stands for the group that a student belongs to, which can be 'a' (Group A), or 'b' (Group B), or 'c' (Group C). Hence, we know, for example, Pr(G=a) =150/300 =0.5.

Let T be the variable that stands for the result of the fitness test, which can be either 'p' (passing the test) or 'f' (failing the test).

In group A, 10% students fail the test, i.e., P(T=f | G=a)=0.1.

In group B, 20% students fail the test, i.e., P(T=f | G=b)=0.2.

In group C, 20% students fail the test, i.e., P(T=f | G=c)=0.2.

What is the probability that a student in this training camp **passes** the fitness test, i.e., P(T=p)?

Your answner here : __1__ (Please write your answer to 3 decimal places, between 0 and 1).

If a student in this camp fails a test, it is most likely that he or she comes from __2__. Please select a group here. Hint: the probablity that a student who fails the test coming from Group A is denoted as P(G=a | T=f).

1. Range - Min:0.84 Max:0.86
2. Choice of: Group A | Group B | Group C - Correct Answer:Group C

Item ID: 247874 / 1
Item Description: Copy of Copy of Part1_FIB2
Item Weight: 4.0

Question #: 12

This question is related to the understanding of linear systems and partial derivatives. Which of the following statements below is/are correct?

I. The function may have multiple global minimums.

II. In over-determined linear systems, the number of unknown equations is less than the number of parameters.

III. The system

$$\begin{bmatrix} 5 & -6 \\ 2 & 0 \\ 4 & 7 \\ 11 & -8 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} 3 \\ -5.5 \\ 9 \\ 1 \end{bmatrix}$$

has no *exact* solution but an approximated solution is available using the left inverse.

IV. If **f**(**x**) is a vector-valued function of *h outputs a*nd **x** is a 4-dimensional vector, then differentiation of **f**(**x**) with respect to **x** is an 4 *x h* matrix.

V. Consider the linear system

$$\mathbf{Xw} = \mathbf{y}, \mathbf{X} \in R^{m \times d}$$

is the input data matrix,

$$\mathbf{w} \in R^{d \times 1}$$

is the parameter vector, and

$$\mathbf{y} \in R^{m \times 1}$$

is the target vector. If d <m, the system has less parameters than equations.

I & III
II & III
II, III & IV
III & V
None of the others.

Item ID: 247872 / 1
Item Description: Copy of S2-Midterm_Part2_MCQ5
Item Weight: 3.0

Question #: 13

A vector function has the mapping of

$$\mathbf{f}(\mathbf{x}) : \mathbf{R}^4 \rightarrow \mathbf{R}^5$$

. The gradient of this function with respect to **x** is a matrix of partial derivatives, which is a 4×5 matrix.

A. True
✓B. False

Item ID: 247871 / 1
Item Description: Copy of S2-Midterm-Part2M-TF5
Item Weight: 2.0

Question #: 14

This question is related to understanding of modelling assumptions.

$$f(\mathbf{x}) = 3x_1 + 5x_2 - 2x_3 + 8$$

is an affine function, and the input is a 3-dimensional vector.

✓A. True
B. False

Item ID: 247870 / 1
Item Description: Copy of S2-Midterm-Part2M-TF3
Item Weight: 2.0

Question #: 15

The values of feature **x** and their corresponding values of target y are shown in the table below, where

$$\mathbf{X}\mathbf{w} = \mathbf{y}$$

.

x	[3, 2]	[4, 5]	[5.5, 3]	[8, 7]
y	[5]	[4]	[3]	[1]

Based on the least square regression, please select the correct option below.

I) The estimated w* minimize the sum of squared errors between the predicted y' and the target output y for all training samples.

II) A squared error between the predicted y' and the target output y for one sample is called loss function.

III) This is learning of vectored function, instead of scalar function

IV) A squared error between the predicted y' and the target output y for one sample is called objective function.

A. I &III
✓B. I &II
C. I, III &IV
D. III &IV
E. None of the above

Item ID: 247869 / 1
Item Description: Copy of S2-Midterm-Part2M-MCQ4
Item Weight: 3.0

Question #: 16

A set of equations is written as

$$\mathbf{X}\mathbf{w} = \mathbf{y}$$

, where

$\mathbf{X} \in \mathbf{R}^{4\times5}$

and

$\mathbf{y} \in \mathbf{R}^{4\times1}$

. How many simultaneous equations are there in this set of equations?

A. 3
✓B. 4
C. 5
D. 6

Item ID: 247868 / 1
Item Description: Copy of S2-Midterm-Part2M-MCQ3
Item Weight: 3.0

Question #: 17

Select the correct options regarding derivatives.

I) Suppose g(**x**) is a *scalar* function of *b* variables where **x** is a *bx*1 vector. Without taking data points into consideration, the outcome of differentiation of g(**x**) w.r.t. **x** is a 1*xb* vector.

II) Suppose f(**x**) is a *scalar* function of *d* variables where **x** is a *dx*1 vector. Without taking data points into consideration, the outcome of differentiation of f(**x**) w.r.t. **x** is a scalar.

III) If **f**(**x**) is a vector function of *size b*x1 and **x** is a *d*x1 vector, then differentiation of **f**(**x**) w.r.t. **x** is a *dx1* vector.

IV) If **f**(**x**) is a vector function of *size b*x1 and **x** is a *d*x1 vector, then differentiation of **f**(**x**) w.r.t. **x** is a *bxd* matrix.

A. I
B. I &IV
✓C. IV
D. II & III
E. None of the above

Item ID: 247867 / 2
Item Description: Copy of S2-Midterm-Part2M-MCQ2
Item Weight: 3.0

Question #: 18

Which of the following is/are true about the matrix below?

$$\mathbf{X} = \begin{bmatrix} 4 & -3 & 6 \\ 1 & 0 & 10 \end{bmatrix}$$

I) **X** is left invertible

II) **X** is right invertible

III) **X** has determinant

IV) Consider **X** is the input data matrix to the linear system **Xw=Y**, the system has more equations than unknowns

V) Consider **X** is the input data matrix to the linear system **Xw=Y**, the system is an under-determined system

A. I &V
✓B. II &V
C. I, III &IV
D. II, III &V
E. None of the above

Item ID: 247866 / 2
Item Description: Copy of S2-Midterm-Part2M-MCQ1
Item Weight: 3.0

Question #: 19

This question is related to determination of types of system where an appropriate solution can be found subsequently. With **Xw=Y**, the following matrix has a right inverse.

$$\mathbf{X} = \begin{bmatrix} 5 & 0 \\ 6 & -6 \\ 8 & 11 \\ 4 & 1 \end{bmatrix}$$

A. True
✓B. False

Item ID: 247865 / 2
Item Description: Copy of S2-Midterm-Part2-TF4
Item Weight: 2.0

Question #: 20

This question is related to the determination of types of system where an appropriate solution can be found subsequently. The linear system below is an under-determined system and has unique constrained solution.

$$\begin{bmatrix} 2 & 5 & 6 & 7 \\ -1 & 3 & -3 & 8 \\ 7 & 9 & 8 & 2 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \\ 6 \end{bmatrix}$$

✓A. True
B. False

Item ID: 247864 / 1
Item Description: Copy of S2-Midterm-Part2-TF2
Item Weight: 2.0

Question #: 21

We would like to design a linear regression function to predict future cognitive performance and depressive severity in the elderly. We have collected two clinical features of 5 participants (x_1, x_2) that are related to future cognitive decline and depression. We also have the five-year follow-up data of cognitive scores (1st columne of **Y**) and depressive scores (2nd column of **Y**) of these 5 participants (see below).

$$\mathbf{X} = \begin{bmatrix} 45 & 9 \\ 50 & 10 \\ 63 & 12 \\ 70 & 8 \\ 80 & 4 \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} 6 & 5 \\ 9 & 6 \\ 8 & 9 \\ 3 & 2 \\ 2 & 4 \end{bmatrix}$$

i) Perform least square estimation based on the existing training data. What is the mean of squared error of the estimated model for cognitive scores? __1__ (up to 4 decimal places) What is the mean of squared error of the estimated model for depressive scores? __2__ (up to 4 decimal places)

ii) If we have a new participant with the two clinical features (x_1, x_2) =(63, 9). What is his/her five year follow-up cognitive score? __3__ (up to 4 decimal places) What is his/her five year follow-up depressive score? __4__ (up to 4 decimal places)

1. Range - Min:1.40 Max:1.41
2. Range - Min:2.65 Max:2.66
3. Range - Min:5.76 Max:5.77
4. Range - Min:5.52 Max:5.53

Item ID: 247863 / 1
Item Description: Copy of S2-Midterm-Part2-FIB1
Item Weight: 8.0

Question #: 22

If 60 older adults have completed 40 brain measures, this is an over-determined system, i.e. fewer equations than unknowns. To do the prediction, we need to use left inverse.

A. True
✓B. False

Item ID: 247862 / 3
Item Description: Copy of S2-Midterm--Part2-TF1
Item Weight: 2.0