NANYANG TECHNOLOGICAL UNIVERSITY

SEMESTER 2 EXAMINATION 2022-2023

MA6514 – MACHINE LEARNING AND DATA SCIENCE

April/May 2023 Time Allowed: 3 hours

INSTRUCTIONS

1. This paper contains **FOUR (4)** and comprises **SIX (6)** pages.
2. **COMPULSORY** to answer **ALL** questions.
3. Marks for each question are as indicated.
4. This is a **CLOSED-BOOK** examination.

---

1. An attempt to describe Smart Manufacturing and Intelligent Manufacturing is as follows:

Smart Manufacturing: Smart manufacturing is a broad category of manufacturing with the goal of optimizing concept generation, production, and product transaction to create flexible manufacturing processes that respond rapidly to changes in demand at low cost without damage to the environment.

Intelligent Manufacturing: Intelligent manufacturing is implementing artificial intelligence-based production that can automatically adapt to changing environments and varying process requirements, with minimal intervention from humans with the focus on artificial intelligence (AI), advanced sensing and control, optimization, and knowledge management.

(a) Based on these descriptions, write a short essay (at least 100 words) on whether there is (or is not) any difference between smart manufacturing and intelligent manufacturing.
(10 marks)

(b) Rank and describe four (4) of the most significant benefits of Industry 4.0. Briefly describe each of these benefits and explain your rationale for the ranking.
(8 marks)

(c) It is said that both Smart Manufacturing and Intelligent Manufacturing rely heavily on Big Data.

(i) Briefly explain what big data is.
(2 marks)

(ii) State and write one sentence each for five key benefits of big data to manufacturing.
(5 marks)

2. (a) Define briefly what is Dimensionality Reduction (DR). Discuss the role of Pearson correlation coefficient in dimensionality reduction and compare it with another unsupervised machine learning method to achieve the same goal.

(5 marks)

(b) Discuss feature reduction and its key steps in LASSO.

(5 marks)

(c) Explain Kernel Trick in Support Vector Machines and its importance to computation reduction. Does it apply to other Machine Learning algorithms?

(5 marks)

(d) The three training vectors are:

$$x_1 = [0.0\ 0.0]^T, y_1 = -1,$$
$$x_2 = [-1.0\ 0.0]^T, y_2 = +1,$$
$$x_3 = [0.0\ 1.0]^T, y_3 = +1,$$

The objective function is:

$$L(\alpha) = \sum_{1}^{3} \alpha_i - \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3} \alpha_i \alpha_j y_i y_j (x_i^T x_j)$$

Subject to $\alpha_i \geq 0$ and $\sum_{i=1}^{3} \alpha_i y_i = 0$.

Assuming equality constraints, i.e., $\sum_{i=1}^{3} \alpha_i y_i = 0$, a Lagrange multiplier, say "b" can be introduced to derive the new Lagrangian function.

(i) State the new Lagrangian function
(ii) Derive the decision boundary

(10 marks)

3. (a) Explain Highest-Posterior Density (HPD) Interval and ROPE (Region of Practical Equivalence). What is the relationship between the Two?

(4 marks)

Note: Question No. 3 continues on Page 3.

(b) The following program uses Bayesian Statistics:

```
1 import arviz as az
2 from scipy import stats
3 import os
4 import pandas as pd
5 import numpy as np
6 import matplotlib.pyplot as plt
7 usecols = [ 'LOA', 'LBP', 'BEAM', 'DEPTH','DRAFT', 'PROP
POWER', 'AUX POWER', 'TEU', 'SPEED']
8 cwd =
os.path.abspath('/content/gdrive/MyDrive/Container_Ship_Data.xls
x')
9 ships_data2 = pd.read_excel(cwd, usecols = ['LBP','TEU',
'SPEED'])
10 ships_data2.head()
11 list1 = ships_data2['LBP']
12 list2 = ships_data2['TEU']
13 x_3 = list1.values
14 y_3 = list2.values
15 x_3 = x_3 - x_3.mean()
16 _, ax = plt.subplots(1, 2, figsize=(10, 5))
17 beta_c, alpha_c = stats.linregress(x_3, y_3)[:2]
18 import pymc3 as pm
19 with pm.Model() as ship_model_t:
20     α = pm.Normal('α', mu=y_3.mean(), sd=2)
21     β = pm.HalfNormal('β', sd = 1)
22     ϵ = pm.HalfNormal('ϵ', 5)
23     y_pred = pm.Normal('y_pred', mu=α + β * x_3,
                         sd=ϵ, observed=y_3)
24     trace_t = pm.sample(2000)
```

Explain the following in the above program:

(i) Line 9: ships_data2 = pd.read_excel(cwd, usecols = ['LBP', 'TEU', 'SPEED'])
(ii) Line 19: with pm.Model() as ship_model_t:
(iii) Line 20: α = pm.Normal('α', mu=y_3.mean(), sd=2)
(iv) Line 22: ϵ = pm.HalfNormal('ϵ', 5)
(v) Line 23: y_pred = pm.Normal('y_pred', mu=α + β * x_3, sd=ϵ, observed=y_3)
(vi) Line 24: trace_t = pm.sample (2000)

(12 marks)

Note: Question No. 3 continues on Page 4.

(c) A certain company called 'Self-Driving Car Innovation', involved in the development of autonomous car project considers Reinforcement Learning as the promising solution to achieve learning of its environment in the navigation of the vehicle. Knowing that you have taken a course in Machine Learning, you are tasked to explain the SARSA algorithm to the management. The SARSA algorithm is given below:

Step 1. Initialize the Q-factors, $Q(l,u)$, for all (l,u) pairs. Let n denote the number of state transitions (iterations) of the algorithm. Initialize n_{max} to a large number. Start a simulation. Set $n = 1$.

Step 2. Let the current system state in the simulation be i. Set s to any state from S, and set w to any action from $\mathcal{A}(s)$. Let $rimm = 0$.

Step 2a. Simulate action $a \in \mathcal{A}(i)$ in state i using an exploratory policy in which the exploration must be decayed gradually after every iteration.

Step 2b. Let the next state encountered in the simulator be j. Let $r(i,a,j)$ be the immediate reward earned in the transition from state i to state j. Update α.

Then update $Q(s,w)$ using:

$$Q(s,w) \leftarrow Q(s,w) + \alpha[rimm + \lambda Q(i,\ a) - Q(s,w)]\, I(n \neq 1)$$

where $I(.)$, the indicator function equals 1 when the condition inside the brackets is satisfied and 0 otherwise.

Step 2c. Set $w \leftarrow a$, $s \leftarrow i$, and $rimm \leftarrow r(i,a,j)$.

Step 2d. Increment n by 1. If $n < n_{max}$, set $i \leftarrow j$ and return to Step 2a. Otherwise go to Step 3.

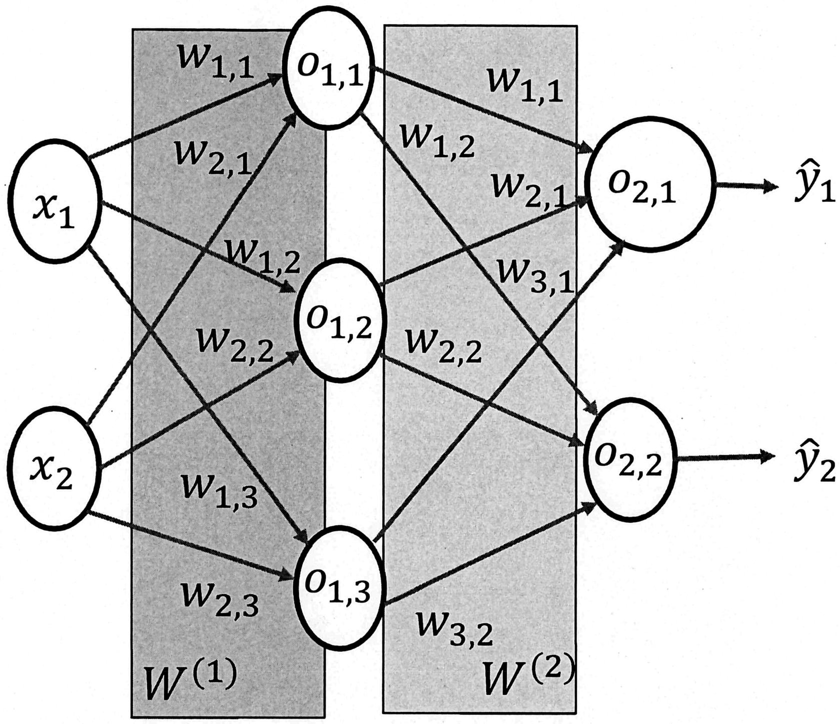
Step 3. For each $l \in S$, select $d(l) \in \arg\max_{b \in \mathcal{A}(i)} Q(l,\ b)$. The policy (solution) generated by the algorithm is $\hat{d}$. Stop.

You are to explain clearly but succinctly to the management the following terms used in the algorithm. They are as follows:

(i) Q-factors, $Q\ (l,u)$ for all (I,u) pairs,
(ii) $r(i,a,j)$,
(iii) $rimm$,
(iv) $I(.)$,
(v) The role of α in the SARSA algorithm.
(vi) Specify the strategy to update α and the reason for its choice.

(9 marks)

4. (a) Figure 1 shows a simple Artificial Neural Network (ANN) with two input neurons, a single hidden layer with three neurons and an output layer with two neurons. The desired output is $y_1 = [1 \;\; 0]$.



Input Layer $\in \mathbb{R}^2$ Hidden Layer $\in \mathbb{R}^3$ output Layer $\in \mathbb{R}^2$

Figure 1

The randomly initialized network weights are as follows:

$$\boldsymbol{W}^{(1)} = \begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,3} \\ w_{2,1} & w_{2,2} & w_{2,3} \end{bmatrix} = \begin{bmatrix} -0.168 & 0.441 & -0.100 \\ -0.401 & -0.706 & -0.81593281 \end{bmatrix}$$

$$\boldsymbol{W}^{(2)} = \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} = \begin{bmatrix} -0.630 & -0.320 \\ -0.220 & 0.079 \\ -0.165 & 0.375 \end{bmatrix}$$

and the input $\boldsymbol{X} = [7.220, 6.204]$

(i) Calculate the $z^{(1)}$ at each neuron in the hidden layer.

(ii) $z^{(1)}$ is passed through the sigmoid function to obtain $o(\cdot)$. Calculate $o_1\left(z^{(1)}\right)$.

(iii) If the predicted outcome is $\hat{y}_1 = [0.475 \;\; 0.505]$, calculate Mean Squared Error (MSE).

Note: Question No. 4 continues on Page 6.

(iv) The loss, L given by the partial derivative, $\frac{\partial L}{\partial W^{(2)}}$ of the backward pass of the backpropagation of the Artificial Neural Network is given by the following:

$$\frac{\partial L}{\partial \boldsymbol{W}^{(2)}} = \frac{\partial L}{\partial \boldsymbol{o}_2} \cdot \frac{\partial \boldsymbol{o}_2}{\partial \boldsymbol{z}^{(2)}} \cdot \frac{\partial \boldsymbol{z}^{(2)}}{\partial \boldsymbol{W}^{(2)}}$$

where,

$$\frac{\partial L}{\partial \boldsymbol{o}_2} = -(\boldsymbol{y} - \boldsymbol{o}_2)$$

$$\frac{\partial \boldsymbol{o}_2}{\partial \boldsymbol{z}^{(2)}} = \boldsymbol{o}_2(1 - \boldsymbol{z}^{(2)})$$

$$\frac{\partial \boldsymbol{z}^{(2)}}{\partial \boldsymbol{W}^{(2)}} = \boldsymbol{o}_1$$

$$\text{The sigmoid function is given as } f(x) = \frac{1}{1 + e^{-x}}$$

Explain the significance of each of the partial derivatives on the right-hand side of the above equation in the Backpropagation ANN algorithm.

(12 marks)

(b) Discuss the similarity and differences between K-Mean and Expectation-Maximization.

(3 marks)

(c) Describe the key steps in the Expectation-Maximization algorithm.

(4 marks)

(d) (i) If the datasets named car-attributes have missing values in several of its features, suggest how you would remove or add to these values by writing one or two statements to remove or add to these missing values. Explain what the function of these statements on the missing values are.

(ii) Instead of removing the missing values that follow a certain probability distribution, explain clearly how you would use Expectation Maximization algorithm to fill in these values.

(6 marks)

END OF PAPER

MA6514 MACHINE LEARNING & DATA SCIENCE

Please read the following instructions carefully:

1. **Please do not turn over the question paper until you are told to do so. Disciplinary action may be taken against you if you do so.**

2. You are not allowed to leave the examination hall unless accompanied by an invigilator. You may raise your hand if you need to communicate with the invigilator.

3. Please write your Matriculation Number on the front of the answer book.

4. Please indicate clearly in the answer book (at the appropriate place) if you are continuing the answer to a question elsewhere in the book.