UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 7 Examination in Engineering: July 2017

Module Number: EE7206 **Module Name: Machine Learning**

[Three Hours]

Index No.: .....

Instructions for Candidates:

1. This question paper consists of two parts; PART-A and PART-B carries 20 and 30 marks respectively.
2. PART-A consists of 20 questions. For PART-A, candidates should answer in the same paper (Use the ☐ space provided for answering).
3. There are 5 or 4 answers for each question. More than one ***correct*** answer or true statement may exist for one question. Candidates should mark ' ✓ ' for the ***correct*** answers and ' * ' for the ***incorrect*** answers. There won't be any negative marks given. 0.2 or 0.25 marks are given for marking a ***correct*** answer as correct. 0.2 or 0.25 marks are given for marking an ***incorrect*** answer as incorrect. Unmarked answers are not given any marks.
4. PART-B consists of three essay questions. Candidates should answer them in the given answer book.

PART - A

Q1. Mark the cases where machine learning techniques could be used effectively for finding solutions.

(a) Finding the shortest path to a destination host in a communication network when the communication protocols are acting optimally ☐

(b) Forecasting the weather ☐

(c) Diagnosing a patient for dengue ☐

(d) Detecting a computer virus using a virus guard ☐

(e) Identifying a new strain of DNA which is benign to HIV where prior knowledge is inadequate ☐

Q2. Select the TRUE statements in relation to the definition of machine learning given below regarding an email spam filter.

"A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E".

(a) Observing the process of labeling an email as spam or not is T ☐

(b) Clustering an email as spam or not is T ☐

(c) Classifying an email as spam or not is T ☐

(d) The probability of an email becoming a spam email is P ☐

(e) The number of emails correctly classified as spam or not is P ☐

Q3. Select machine learning algorithms from the following.

(a) Decision trees ☐

(b) Least square regression ☐

(c) Logistic regression ☐

(d) Dijkstra's ☐

(e) Greedy ☐

Q4. Evaluate the following statements regarding regression and classification.

(a) Predicting the next weeks average rainfall amount (in mm) is a regression problem ☐

(b) Predicting the overall weather (sunny, cloudy or rainy) for tomorrow is a regression problem ☐

(c) Selecting a particular cancer is treatable or not is a classification problem ☐

(d) Predicting the stock value (in LKR) of ABC company for the next seven days in the stock market is a classification problem ☐

(e) The same problem could be solved using both regression or classification approaches ☐

Q5. Select the TRUE statements regarding supervised learning and unsupervised learning techniques.

(a) Developing a score predicting program for cricket matches based on player statistics and statistics at the venue of the match being held is a supervised learning problem ☐

(b) Examining the Central Processing Unit (CPU) usage of a set of programs running in a computer to group the utilization level of the programs is a unsupervised learning problem ☐

(c) A data set of heart patient statistics are examined to identify different treatment methods from grouping is a supervised learning problem ☐

(d) Clustering is an approach under supervised learning ☐

(e) Reinforcement learning is a concept defined under supervised learning ☐

Q6. Select the supervised learning algorithms from the following.

(a) K-means ☐

(b) K- Nearest Neighbor (KNN) ☐

(c) Naïve Bayes ☐

(d) Support Vector Machines (SVM) ☐

(e) Neural Networks (NN) ☐

Q7. Evaluate the following statements regarding Reinforcement Learning (RL).

(a) RL is an area in machine learning which is inspired by behaviorist psychology ☐

(b) A person is receiving an order after time intervals of 5, 10 and 15 minutes every time he visits a restaurant, the reward is based on varying time ☐

(c) The multi-armed bandit problem captures the essence of conflict between exploration and exploitation ☐

(d) A rats reaction (turning left or right) is not affected by the probability of placing the reward (piece of cheese) at either end over successive trials in the T-maze problem ☐

(e) If a person is given a 15% discount after 2 consecutive meals in a restaurant, the reinforcement happens based on the trial ratio rather than the trail times ☐

Q8. Evaluate the following statements regarding Support Vector Machines (SVM).

(a) Support vectors are the data points which lie on the decision hyperplane ☐

(b) SVM uses an optimization approach rather than a greedy search for solution which improves its efficiency ☐

(c) Support vectors are the elements of the training set that would change the position of the decision hyperplane if removed ☐

(d) Lagrangian duality is used for maximizing the margin of the decision hyperplane in SVM ☐

(e) Gaussian kernel is better suited for linear classifications ☐

Q9. Mark the TRUE statements regarding probabilistic classification schemes.

(a) Naïve Bayes assumption is $P(o_1, o_2, ..., o_n | h_j) = \prod_i P(o_i | h_j)$ ☐

(b) Bayes classifier only relies on the prior probabilities for classification ☐

(c) Outputs of several classifiers can be combined with Bayes classifier ☐

(d) Bayes classifier computes the Maximum Likelihood Probability (MLP) ☐

(e) Probability based classifiers perform better with predictive problems ☐

Q10. Suppose there are three clusters with centroids $\mu_1 = (1, 2), \mu_2 = (-3, 0)$ and $\mu_3 = (4, 2)$. What is the cluster being assigned (c_i) for the training example $x_i = (-1, 2)$?

(a) $c_i = 2$ ☐

(b) $c_i = 3$ ☐

(c) $c_i = 1$ ☐

(d) c_i is not assigned ☐

Q11. Which of the following statements are TRUE?

(a) Suppose you have a multi-class classification problem with three classes, trained with a 3 layer network. Let $a_1^{(3)} = (h_\Theta(x))_1$ be the activation of the first output unit, and similarly $a_2^{(3)} = (h_\Theta(x))_2$ and $a_3^{(3)} = (h_\Theta(x))_3$. Then for any input x, it must be the case that $a_1^{(3)} + a_2^{(3)} + a_3^{(3)} = 1$ ☐

(b) A two layer (one input layer, one output layer; no hidden layer) neural network can represent the XOR function ☐

(c) Any logical function over binary-valued (0 or 1) inputs x1 and x2 can be (approximately) represented using some neural network ☐

(d) The activation values of the hidden units in a neural network, with the sigmoid activation function applied at every layer, are always in the range (0, 1) ☐

Q12. Consider the following neural network which takes two binary-valued inputs x_1, $x_2 \in \{0,1\}$ and outputs $h_\Theta(x)$. Which of the following logical functions does it (approximately) compute?

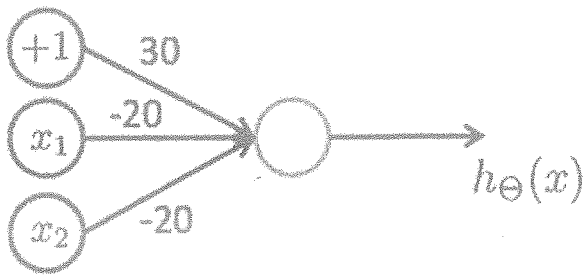

Figure Q12

(a) NAND (meaning "NOT AND") ☐

(b) AND ☐

(c) OR ☐

(d) XOR (exclusive OR) ☐

Q13. Consider the neural network given below. Which of the following equations correctly computes the activation $a_1^{(3)}$? Note: $g(z)$ is the sigmoid activation function.

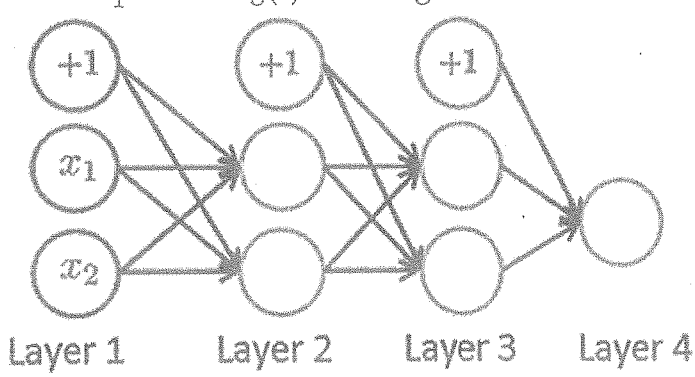

Figure Q13

(a) $a_1^{(3)} = g\,(\Theta_{1,0}^{(2)} a_0^{(2)} + \Theta_{1,1}^{(2)} a_1^{(2)} + \Theta_{1,2}^{(2)} a_2^{(2)})$ ☐

(b) $a_1^{(3)} = g\,(\Theta_{1,0}^{(2)} a_0^{(1)} + \Theta_{1,1}^{(2)} a_1^{(1)} + \Theta_{1,2}^{(2)} a_2^{(1)})$ ☐

(c) $a_1^{(3)} = g\,(\Theta_{1,0}^{(1)} a_0^{(2)} + \Theta_{1,1}^{(1)} a_1^{(2)} + \Theta_{1,2}^{(1)} a_2^{(2)})$ ☐

(d) $a_1^{(3)} = g\,(\Theta_{2,0}^{(2)} a_0^{(2)} + \Theta_{2,1}^{(2)} a_1^{(2)} + \Theta_{2,2}^{(2)} a_2^{(2)})$ ☐

Q14. You have the following neural network:

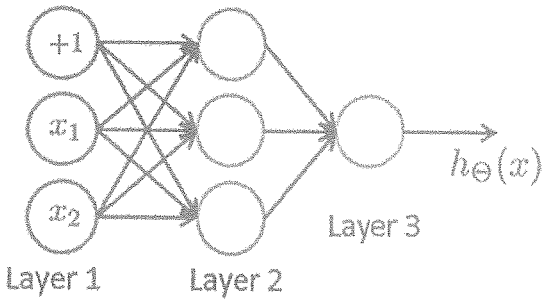

Figure Q14

You'd like to compute the activations of the hidden layer $a^{(2)} \in \mathbb{R}^3$. One way to do so is the following code:

```
% Theta1 is Theta with superscript "(1)" from lecture
% ie, the matrix of parameters for the mapping from layer 1 (input) to layer 2
% Theta1 has size 3x3
% Assume 'sigmoid' is a built-in function to compute 1 / (1 + exp(-z))

a2 = zeros (3, 1);
for i = 1:3
  for j = 1:3
    a2(i) = a2(i) + x(j) * Theta1(i, j);
  end
  a2(i) = sigmoid (a2(i));
end
```

You want to have a vectorizer implementation of this (i.e., one that does not use for loops). Which of the following implementations correctly compute $a^{(2)}$?

(a) a(2) = sigmoid (Theta1 * x);

(b) a(2) = sigmoid (x * Theta1);

(c) a(2) = sigmoid (Theta2 * x);

(d) z = sigmoid(x); a(2) = Theta1 * z;

Q15. You are using the neural network pictured below and have learned the parameters $\Theta^{(1)}=\begin{bmatrix}1 & -1.5 & 3.7\\ 1 & 5.1 & 2.3\end{bmatrix}$ (used to compute $a^{(2)}$) and $\Theta^{(2)}=[1 \quad 0.6 \quad -0.8]$ (used to compute $a^{(3)}$) as a function of $a^{(2)}$). Suppose you swap the parameters for the first hidden layer between its two units so $\Theta^{(1)}=\begin{bmatrix}1 & 5.1 & 2.3\\ 1 & -1.5 & 3.7\end{bmatrix}$ and also swap the output layer so $\Theta^{(2)}= [1 \quad -0.8 \quad 0.6]$. How will this change the value of the output $h_\Theta(x)$?

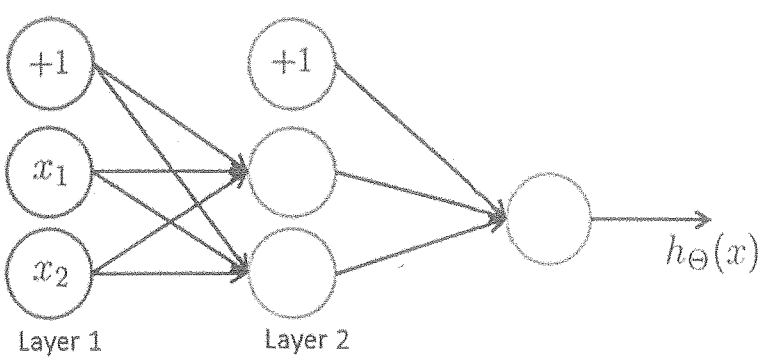

Figure Q15

(a) It will stay the same ☐

(b) It will increase ☐

(c) It will decrease ☐

(d) Insufficient information to tell: it may increase or decrease ☐

Q16. Which of the following statements is the best description of reproduction?

(a) Randomly change a small part of some strings ☐

(b) Randomly change the fitness function ☐

(c) Randomly combine the genetic information from 2 strings ☐

(d) Randomly generate small initial values for the weights ☐

(e) Randomly pick strings to make the next generation ☐

Q17. Which of the following statements is the best description of mutation?

(a) Randomly change a small part of some strings ☐

(b) Randomly change the fitness function ☐

(c) Randomly combine the genetic information from 2 strings ☐

(d) Randomly generate small initial values for the weights ☐

(e) Randomly pick strings to make the next generation ☐

Q18. The weighted roulette wheel is a technique used for,

(a) selecting the best chromosome ☐

(b) randomly selecting the chromosomes ☐

(c) crossing-over the selected chromosomes ☐

(d) mutating the fittest chromosomes ☐

(e) measuring the fitness of the chromosomes ☐

Q19. Ranking is a technique used for,

(a) copying the fittest member of each population into the mating pool ☐

(b) obtaining the selection probabilities for reproduction ☐

(c) allowing many similar individuals to survive into the next generation ☐

(d) deleting undesirable members of the population ☐

Q20. Which of the following applications can be implemented by a genetic algorithm?

(a) Feature selection for pattern recognition problems ☐

(b) Learning the weights in a neural network ☐

(c) Robot path planning ☐

(d) All the above answers ☐

(e) None of the above answers ☐