Question Paper

Exam Date & Time: 24-Nov-2018 (10:00 AM - 01:00 PM)



MANIPAL ACADEMY OF HIGHER EDUCATION

SCHOOL OF INFORMATION SCIENCES

SECOND SEMESTER MASTER OF ENGINEERING - ME (BIG DATA AND DATA ANALYTICS)

Machine Learning [BDA 602]

Marks: 100

Duration: 180 mins.

END SEMESTER DEGREE EXAMINATION NOVEMBER 2018

Answer all the questions.

- ¹⁾ Write the definitions of Machine Learning given by Arthur ⁽¹⁰⁾ Samuel and Tom Mitchell. Briefly describe about any three applications for which machine learning approaches seem appropriate.
- ²⁾ Obtain the most general and most specific hypotheses using
 ⁽¹⁰⁾ Candidate-Elimination algorithm for the training data contains errors as given in the following table.

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	No (error
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

- ³⁾ Write the principle of Occam Razor and describe the problems of Occam Razor principle with size of the hypothesis.
 ^{b)} Explain K-Fold Cross-validation technique in training and testing with suitable example.
- ⁴⁾ Solve for number of training examples "m" sufficient for ⁽¹⁰⁾ successful PAC learning of any target concept in the finite hypotheses space?
- ⁵⁾ What is shattering a set of instances means? Give ⁽⁶⁾ _{a)} example.
 - ^{b)} Define Vapnik-Chervonenkis (VC) Dimension. ⁽⁴⁾
- ⁶⁾ Consider a medical diagnosis problem in which there are two ⁽¹⁰⁾

alternative hypotheses:

the patient has a particular disease (denoted by cancer)

(2) the patient does not (denoted by ¬cancer)

Prior knowledge over the entire population of people only 0.008 have this disease. The available data is from a particular laboratory test with two possible outcomes (positive and negative). Furthermore, the lab test is only an imperfect indicator of the disease. The test returns a correct positive result in only 98% of the cases in which the disease is actually present and a correct negative result in only 97% of the cases in which the disease is not present. In other cases, the test returns the opposite result.

Suppose, a new patient is observed for whom the lab test returns a positive result. Should you diagnose the patient as having cancer or not?

- Write the algorithm and explain Distance-Weighted (10)
 Nearest Neighbor Learning algorithms for learning the discrete-valued and real-valued target functions.
- ⁸⁾ What is confusion matrix? Define the following ⁽¹⁰⁾ terms
 - a). True Positive
 - b). True Negative
 - c). False Positive
 - d). False Negative
- ⁹⁾ What are the activation functions used in Artificial Neural ⁽¹⁰⁾ Network (ANN)? Explain each with suitable example.
- ¹⁰⁾ Generate the output of logic XOR function using McCulloch- ⁽¹⁰⁾ Pitts neuron model.

-----End-----

(1)