Question Paper

Exam Date & Time: 20-Apr-2018 (10:00 AM - 01:00 PM)



MANIPAL ACADEMY OF HIGHER EDUCATION

SCHOOL OF INFORMATION SCIENCES (SOIS) FIRST SEMESTER MASTER OF ENGINEERING- ME (Big Data and Data Analytics) DEGREE EXAMINATION-APRIL 2018 Friday, April 20, 2018 Time: 10:00am to 1:00pm Fundamentals of Machine Learning [BDA 611]

Marks: 100

Duration: 180 mins.

Answer all the questions.

- Write the rule for Estimating Training Values in the design of (4)
 machine learning system.
 - ^{B)} Write the approaches involved in adjusting the weights to best fit ⁽⁶⁾ the set of training examples.
- ²⁾ Implement Candidate-elimination algorithm to obtain most general and most ⁽¹⁰⁾ specific hypotheses for the training examples given in the following table

	Exampl	e Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySp	ort
	1 2 3 4	Sunny Sunny Rainy Sunny	Warm Warm Cold Warm	Normal High High High	Strong Strong Strong Strong	Warm Warm Warm Cool	Same Same Change Change	Yes Yes No Yes	
3)		State and determine this proba	prove the the numb bility of fai	e-Exhaust er of traini lure below	ed Versio ng exam some de	on Space ples req esired le	theorem to the theorem to the theorem to the theorem tensor to th	to duce	(10)
1)	A)	What is Ag training ex	gnostic Lea xamples "r	arning? Obt n" required	tain the e d in this o	equation case.	for numbe	er of	(7)
	B)	Define Va	pnik-Cherv	onenkis (V	C) Dime	nsion			(3)
5)		Consider a medical diagnosis problem in which there are two alternative hypotheses: (1) the patient has a particular disease (denoted by <i>cancer</i>) (2) the patient does not (denoted by <i>¬cancer</i>)						(10)	
		Prior know have this laboratory negative). of the dise	vledge ove disease. Th test with Furtherme ease. The t	r the entire ne available two possib ore, the lab est returns	e populat e data is le outcor o test is c a correc	tion of pe from a p mes (pos only an in t positiv	eople only particular sitive and mperfect in ve result in	0.008 ndicator 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?*

Derive Bayes Optimal Classifier

(4)

A) B)

6)

7)

Consider a hypothesis space containing three hypotheses: h1, h2, ⁽⁶⁾ and h3. Posterior probabilities of h1, h2, and h3 given the training data are 0.4, 0.3, and 0.3 respectively. New instance x is encountered, which is classified positive by h1, but negative by h2 and h3. Use Bayes Optimal Classifier to obtain the most probable classification of the new instance given the training data?

Expert assigned some basic outcomes to the nodes given in the ⁽¹⁰⁾ table below for Bayesian Networks shown in the following figure. What's the probability that it's raining when the yard is wet?



YARD			
Hose	Rain	True	False
False	False	0.0	1.0
False	True	0.8	0.2
True	False	0.9	0.1
True	True	0.99	0.01

HOSE		
Value of Rain Node	True	False
False	0.4	0.6
True	0.01	0.99

RAIN	AIN	
True	False	
0.2	0.8	

- ⁸⁾ Write the algorithm and explain Distance-Weighted Nearest ⁽¹⁰⁾ Neighbor Learning algorithms for learning the discrete-valued and real-valued target functions. (5+5 Marks)
- ⁹⁾ Explain the learning concept of Case-Based Reasoning with ⁽¹⁰⁾ suitable example.
- ¹⁰⁾ Find the covariance matrix and principal components (PCs) for the ⁽¹⁰⁾ data showing relationship between number of hours studied against the mark received.

	Hours(H)	Mark(M)
Data	9	39
	15	56
	25	93
	14	61
	10	50
	18	75
	0	32
	16	85
	5	42
	19	70
	16	66
	20	80

-----End-----