Question Paper

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



MANIPAL ACADEMY OF HIGHER EDUCATION

SCHOOL OF INFORMATION SCIENCES

FIRST SEMESTER MASTER OF ENGINEERING - ME (BIG DATA AND DATA ANALYTICS/ MEDICAL SOFTWARE/ CLOUD COMPUTING)

Fundamentals of Machine Learning [BDA 611]

Marks: 100

1)

END SEMESTER DEGREE EXAMINATION NOVEMBER 2018

Answer all the questions.

Write the steps involved in choosing the Training Experience in designing a machine learning systems.

- ^{b)} How will you choose the Target Function in designing a machine ⁽⁴⁾ learning system? Explain with example.
- ²⁾ Obtain the most general and most specific hypotheses using Candidate- ⁽¹⁰⁾ Elimination algorithm for the training data 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	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

3)

a.)

- a.)
 b.) Explain K-Fold Cross-validation technique to find the accuracy in training and testing with suitable example.
 (3)
- ⁴⁾ If the Version Space contains only hypotheses with $\epsilon = 0.1$ and δ ⁽⁴⁾
 - a.) = 0.05. Find the number of training examples "m" required for the enjoy sports with 973 semantically distinct hypotheses.
 - ^{b.)} Obtain the equation of hMAP and hML from Bayes theorem. ⁽⁶⁾
- ⁵⁾ What is shattering a set of instances means? Give example. ⁽⁷⁾
- b.) Define Vapnik-Chervonenkis (VC) Dimension. (3)
 6) What is conditional probability? (2)

Duration: 180 mins.

a.) b.)

7)

Consider a medical diagnosis problem in which there are two alternative hypotheses: (1) 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?

Training data given in the table below is classifying days according to whether someone will play tennis or not.

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

This table provides a set of 14 training examples of the target concept PlayTennis, where each day is described by the attributes Outlook, Temperature, Humidity, and Wind.

Use naive Bayes classifier and the training data from this table to classify the following novel instance:

(Outlook = sunny, Temperature = cool, Humidity = high, Wind = strong)

(10)

(8)

Write the principle of Instances Based Learning and Case-Based Reasoning methods. Explain the learning process of Case-Based Reasoning with suitable example.

⁹⁾ Find one principal component (PC) for the data showing ⁽¹⁰⁾ relationship between numbers of hours studied against the marks 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

10)

Write a short note on the following Ensemble methods with ⁽¹⁰⁾ example

(3 Marks)
(3 Marks)
(4 Marks)

-----End-----