Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

(A constituent institution of MAHE, Manipal)

VI SEMESTER B.TECH. END SEMESTER EXAMINATIONS, APRIL 2018

SUBJECT: OPEN ELECTIVE-II: MACHINE LEARNING [ICT 3285]

REVISED CREDIT SYSTEM
(28/04/2018)

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

- Answer ALL the questions.
- Missing data if any, may be suitably assumed.

1A. Explain probabilistic interpretation with respect to regression problem. Show that maximizing log likelihood is same as minimizing the cost function $J(\theta)$. 5

1B. A generalized linear model assumes that the response variable y (conditioned on x) is distributed according to a member of the exponential family:

$$p(y; \eta) = b(y) \exp(\eta^T T(y) - a(\eta))$$

Show that Bernoulli distribution is an example of exponential distribution. 3

1C. The logistic function is given by $\sigma(x) = 1/(1+e^{-x})$. Show that $\sigma'(x) = \sigma(x)(1- \sigma(x))$. 2

2A. Suppose you have a classification problem in which the input features x are continuous-valued random variables and $y^{(i)} \in \{0,1\}$. You will use the Gaussian Discriminant Analysis (GDA) model, to model $p(x|y)$ using a multivariate normal distribution. The model is:

$$y \sim \text{Bernoulli}(\varphi)$$
$$x|y = 0 \sim N(\mu_0, \Sigma)$$
$$x|y = 1 \sim N(\mu_1, \Sigma)$$

Find the maximum likelihood estimate and derive the parameters φ, Σ, μ_0 and μ_1. How are you going to make a prediction, when a new value of x is posited to you? 5

2B. What are the limitations of linear regression? How does locally weighted linear regression overcome these limitations? 3

2C. Consider the problem of binary classification using the Naive Bayes classifier. You are given two dimensional features (X_1, X_2) and the categorical class conditional distributions in the Table Q.2C below. The entries in the tables correspond to $P(X_1 = x_1|C_i)$ and $P(X_2 = x_2|C_i)$ respectively. The two classes are equally likely.

2

Table Q.2C

X_1 = \ Class	C_1	C_2
−1	0.2	0.3
0	0.4	0.6
1	0.4	0.1

X_2 = \ Class	C_1	C_2
−1	0.4	0.1
0	0.5	0.3
1	0.1	0.6

Given a data point (−1, 1), using Bayes rule and conditional independence assumption of Naive Bayes, calculate posterior probabilities $P(C_1|X_1 = -1, X_2 = 1)$ and $P(C_2|X_1 = -1, X_2 = 1)$.

3A. Suppose, you have an estimation problem in which you have a training set $\{x^{(1)}, \ldots, x^{(m)}\}$ consisting of m independent examples. You wish to fit the parameters of a model p(x, z) to the data, where the likelihood is given by

$$\ell(\theta) = \sum_{i=1}^{m} \log p(x;\theta)$$
$$= \sum_{i=1}^{m} \log \sum_{z} p(x, z;\theta)$$

But, explicitly finding the maximum likelihood estimates of the parameters θ may be hard. Here, the $z^{(i)}$'s are the latent random variables. For such a setting, use EM algorithm which gives an efficient method for maximum likelihood estimation. 5

3B. Explain the intuition behind large margin classifiers in SVM's. 3

3C. Bias and variance are the twin evils of machine learning. With appropriate diagrams, explain the bias-variance trade off and behavior of the model. 2

4A. Given a training set $\{x^{(1)}, \ldots, x^{(m)}\}$ in the unsupervised learning setting, you wish to model the data by specifying a joint distribution as:

$p(x^{(i)}, z^{(i)}) = p(x^{(i)}|z^{(i)})p(z^{(i)})$.

$z^{(i)} \sim \text{Multinomial}(\phi)$

where,

$\phi_j \geq 0, \sum_{j=1}^{k} \phi_j = 1$

$x^{(i)}|z^{(i)} = j \sim \mathcal{N}(\mu_j, \Sigma_j)$

Explain, how density estimation can be done on the mixture of Gaussians? 5

4B. Suppose you are given a dataset $\{x^{(i)} | i = 1, \ldots, m\}$ of attributes of m different types of automobiles, such as their maximum speed, turn radius, and so on. Let $x^{(i)} \in R^n$ for each i ($n \ll m$). (But unknown to you are two different attributes - some $x^{(i)}$ and $x^{(j)}$ - respectively give a car's maximum speed measured in miles per hour, and the maximum speed measured in kilometers per hour. These two attributes are therefore almost linearly dependent. Thus, the data really lies approximately on an $n - 1$ dimensional subspace). How dimensionality reduction can be done using PCA for $n << m$ case? 3

4C. State Jenson's inequality for the case of the function being concave and graphically depict its behavior. 2

5A. Consider MDPs with finite state and action spaces ($|S| < \infty$, $|A| < \infty$). Write the algorithm for value iteration and policy iteration. Compare and contrast value iteration and policy iteration for the following cases: 5

i) Small MDP's

ii) MDP's with large state spaces.

5B. State and explain Chernoff bound in learning theory. 3

5C. A random variable X is conditionally independent of Y given Z if and only if: $P(X|Y, Z) = P(X|Z)$. Prove that if $P(X,Y|Z)=P(X|Z)P(Y|Z)$, then X is conditionally independent of Y given Z. 2