Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
(A constituent unit of MAHE, Manipal)

VI SEMESTER B.TECH.

END SEMESTER EXAMINATION, APRIL /MAY 2019

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

REVISED CREDIT SYSTEM

(6/5/2019)

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

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

1A.	Given the training set of size $m \times n$ and the output vector y, derive the expression for parameter θ with minimum cost using LMS algorithm. Obtain the value for θ using probabilistic interpretation.	5
1B.	Obtain the MLE parameters in the case of Naïve Bayes. How do you carry out the discretization process if the data is continuous?	3
1C.	Given initial value of θ, derive an expression for the updated value of θ using Newtons method. Generalize this to obtain it's equivalent in the case of Newton-Raphson method.	2
2A.	Bring out the relation between geometric and functional margin. Given optimization problem to maximize geometric margin γ having linear constraints and $\|w=1\|$. Obtain the opimization problem with convex quadratic objective with only linear constraints.	5
2B.	Why is Backward search computationally expensive? How is it overcome using Filter featue selection method?	3
2C.	SVM can learn in the high dimensional feature space without having to explicitly find the feature vectors $\emptyset(x)$. Substantiate on this using relevant example.	2
3A.	Given the training error and generalization error, using the Chernouff bound and the union bound obtain the uniform convergence result. Find the training set size, by chosing probability of $1-\delta$, for all hypothesis $h \in \mathcal{H}$ in the finite hypothesis class having k hypotheses.	5
3B.	The hidden variable z, has different mixture of Gaussians distribution and x is obtained from this distribution. The number of training examples is m. The joint distribution of $p(x^{(i)}, z^{(i)})$ with log likelihood is given as $\ell(\emptyset,\mu,\Sigma) = \sum_{i=1}^{m} log\ p(x^{(i)} \mid z^{(i)};\ \mu,\Sigma) + \log p(z^{(i)};\emptyset)$. Write the EM algorithm to compute the parameters and compare it with Gaussian distribution where the values of $z^{(i)}$ are known.	3
3C.	Write the k-means algorithm. What is the role of a cluster centroid when no training example are associated with it? Incorporte this in the algorithm.	2
4A.	Obtain the KKT dual complementarity conditions in the case of SVM's with datastes which is non separable or having outliers. Explain the Coordinate ascent algorithm which is essential to solve the dual problem.	5

4B. In order to derive the Generelized Linear Model for prediction or classification what are the three assumptions about the conditional distribution of output label given input training data? 3

4C. Specify an example where the principle component analysis is used. Provide the vector value for the output y when k of available n principle features are selected using the PCA method. 2

5A. Explain with an example, the significance of the each elements of the tuples in the Markov decision process. Write the algorithm for value iteration and policy iteration. 5

5B. The joint distribution of (x, z) has z as the latent random variable such that $z \sim N(0, I)$, $\varepsilon \sim N(0, \psi)$, $x = \mu + \Lambda z + \varepsilon$ the vector $\mu \in \mathbb{R}^n$, the matrix $\Lambda \in \mathbb{R}^{n\times k}$, the diagonal matrix $\psi \in \mathbb{R}^{n\times n}$, ε and z are independent and k<n. Obtain the expression for $\ell(\mu, \Lambda, \psi)$. 3

5C. Obtain the updated value for the inverse of the mixing matrix which is required to obtain the independent components. Bring out the relation for the probability distribution function and cumulative distribution function . 2