Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL
(A constituent unit of MAHE, Manipal)

VI SEMESTER B.TECH.

MAKEUP EXAMINATIONS, JUNE 2019

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

REVISED CREDIT SYSTEM

(20/06/2019)

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

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

1A. Compare **5**

i) Linear regression with locally weighted linear regression
ii) Stochasatic and batch gradient descent algorithm
iii) Parametric and non-parametric algorithm.
iv) Bias with variance

1B. Provide the Maximum likelihood estimate of the parameters in the case of multivariate Bernoulli event model and multinomial event model taking into account that event with zero probability is likely to occur. **3**

1C. Given the input x and output y, derive the expression for the parameter θ in the case of logistic regression. **2**

2A. Given the primal optimization problem for finding the optimal margin classifier as **5**

$$\min_{\gamma,w,b} \quad \frac{1}{2}||w||^2$$
$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1, \quad i = 1, \ldots, m,$$

Get the expression for b(bias) using the optimization problem.

2B. Explain the various cross validation techniques. **3**

2C. How is online learning different from batch learning? How is the predictions and update rules carried out using the perceptron algorithm? **2**

3A. Assume that the uniform convergence result holds in the case of finite hypothesis $\mathcal{H}$. Obtain the relation between the generalization error for the hypothesis with minimum training error and best hypothesis. Compare this with the result that is obtained when d=VC($\mathcal{H}$). **5**

3B. The size of training example is m, input is x and hidden variable z. Find the MLE prameters for the case when z is unknown and EM algorithm is applied. Compare the result with Gaussian distribution and k-means algorithm. **3**

3C. Why do you say k-means is a coordinate descent on the distortion function? Does the parameters of the function converge? If not, how can you find the optimal value? **2**

4A. Specify the preprocessing steps in the case of PCA. Given the unit vector u and $x^{(i)}$ the points in the dataset, how do you maximize the variance of the projection of $x^{(i)}$ on to u. Obtain the relation for the principal eigen vector of the covariance matrix. **5**

4B. Map the parameters b(y), η, T(y) and a(η) of the generalized expression of the exponential family with that of i) Bernoulli's ii) Gaussian iii) Poisson with distribution $p(y|\lambda) = \frac{1}{y!} exp(ylog\lambda - \lambda)$ **3**

4C. State and explain the Jensen's inequality theorm. **2**

5A. Obtain an expression for optimal value function and optimal policy given the tuple (S,A,{Psa},γ,R) as defined in the Markov decision process. **5**

5B. Given the vector valued random variable $x=[x_1\ x_2]^T$ with x_1 and x_2 being joint multivariate Gaussian distributed, find the covariance of x. What is the marginal distribution of x_1 and conditional distribution of x_1 given x_2 ? **3**

5C. What are the problems that are encountered when you try to retrieve the independent component from the mixed signal. Explain if it is possible to recover the independent component from the mixed signal if the data distribution is Gaussian. **2**