Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY, MANIPAL 576104

(Constituent College of Manipal University)

SIXTH SEMESTER B.TECH DEGREE MAKE UP EXAMINATION-JULY 2016
SUBJECT:OPEN ELECIVE-II MACHINE LEARNING (ICT 364)
(REVISED CREDIT SYSTEM)

TIME: 3 HOURS   **8/07/2016**   **MAX. MARKS: 50**

Instructions to candidates

- Answer any **FIVE FULL** questions. All questions carry equal marks.
- Missing data if any, may be suitably assumed.

1A. With probabilistic assumption and interpretation, derive the required relation for least-square regression.

1B. What are the limitations of linear regression? How does locally weighted linear regression overcome those limitations?

1C. 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 the Gaussian distribution is an example of exponential distribution.

[5+3+2]

2A. Suppose you are given a dataset $\{(x^{(i)}, y^{(i)}; i = 1, \ldots, m)\}$ consisting of m independent examples, where $x^{(i)} \in \mathbb{R}^n$ are n-dimensional vectors, and $y^{(i)} \in \{0, 1\}$. You will model the joint distribution of (x, y) according to:

$$p(y) = \phi^y(1-\phi)^{1-y}$$

$$p(x|y=0) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\Big(-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\Big).$$

$$p(x|y=1) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\Big(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\Big)$$

Suppose you have already fit ϕ, μ_0, μ_1, and Σ, and now want to make a prediction at some new query point x. Show that the posterior distribution of the label at x takes the form of a logistic function, and can be written as

$$p(y=1|x;\phi,\Sigma,\mu_0,\mu_1) = \frac{1}{1+\exp(-\theta^T x)},$$

where θ is some appropriate function of $\phi, \Sigma, \mu_0, \mu_1$.

2B. What is the intuition behind *margins* in SVMs? Explain the following margins:

i) Functional margin, and

ii) Geometric margin.

2C. State *Hoeffding inequality* and give its interpretation.

[5+3+2]

Use EM algorithm to derive the expression for μ.

4C. Write k-means clustering algorithm.

[5+3+2]

5A. Consider a learning problem in which you have a finite hypothesis class $\mathcal{H} = \{h_1, \ldots, h_k\}$ consisting of k hypothesis. Derive *uniform convergence* result.

5B. State Jensen's inequality and graphically depict its behavior.

5C. State Mercer's theorem on valid kernels.

[5+3+2]

6A. For the data set given in Table Q.6A, design a polynomial learning machines whose inner product kernel is given by

$$K(\mathrm{x}, \mathrm{x}_i) = (1 + \mathrm{x}^T \mathrm{x}_i)^2.$$

Table: Q.6A

Input Vector, x	Desired Response, d
$(-1, -1)$	-1
$(-1, +1)$	$+1$
$(+1, -1)$	$+1$
$(+1, +1)$	-1

6B. Consider a classification problem in which the response variable y can take on any one of k values, so $y \in \{1, 2, \ldots, k\}$. You can model such distribution as a multinomial distribution. Use GLM for modeling this multinomial distribution and derive the relation for parameters of exponential family distribution.

6C. Briefly explain *Fisher scoring* for maximizing $l(\theta)$.

[5+3+2]