Registration no: | | | | | | | | | | |

Total Number of Pages: 02 **M.TECH**
CSPE209

2nd Semester MTech Regular/Back Examination – 2014-15
PATTERN RECOGNITION
BRANCH(S): CSE
Time: 3 Hours
Max marks: 70
Q.CODE: T411

Answer Question No.1 which is compulsory and any five from the rest.
The figures in the right hand margin indicate marks.

Q1 Answer the following questions: (2 x 10)

a) What do you mean by Cluster validation?
b) What is meant by a pseudoinverse? How is it used in MSE?
c) Distinguish between supervised and unsupervised learning.
d) Define Delta Rule.
e) Distinguish between Top down and Bottom up parsing techniques.
f) What is the significance of activation function in Back Propagation Algorithm?
g) What is meant by the likelihood ratio**?**
h) Explain Design Principles of Pattern Recognition System.
i) Distinguish between Parametric and Non-parametric pattern recognition methods.
j) What is stochastic grammar? Explain with an example.

Q2 a) Obtain an expression for Linear Discriminate functions for (5)
(i) Two category case (ii) Multi Category Case.

b) Explain the Parzen Window approaches for Density Estimation. (5)

Q3 a) Explain the concept of feature extraction in pattern recognition system with examples. (5)

b) Consider a hierarchical clustering procedure in which clusters are merged so as to produce the smallest increase in the sum-of-squared error at each step. If the *i* th cluster contains n_i samples with sample mean $\mathbf{m}_i$, show that the smallest increase results from merging the pair of clusters for which (5)

$$\frac{n_i n_j}{n_i + n_j}\left\|m_i - m_j\right\|^2$$

is minimum.

Q4 a) Show that if our model is poor, the maximum-likelihood classifier we derive is not the best-even among our (poor) model set- by exploring (5) the following example:-

Suppose we have two equally probable categories (i.e., $P(\omega_1) = P(\omega_2) = 0.5$). Furthermore, we know that $p(x|\omega_1) \sim N(0,1)$ but assume that $p(x|\omega_2) \sim N(\mu,1)$. (That is, the parameter θ we seek by maximum-likelihood techniques is the mean of the second distribution). Imagine, however, that the *true* underlying distribution is $p(x|\omega_2) \sim N(1, 10^6)$.

I. What is the maximum-likelihood estimate $\hat{\mu}$ in our poor model, given a large amount of data?

II. What is the decision boundary arising from this maximum-likelihood estimate in the poor model?

b) Show that if the activation function of hidden units is linear, a three-layer network is equivalent to a two-layer one. Use your result to explain why a three-layer network with linear hidden units cannot solve a nonlinearly separable problem such as XOR or d-bit parity. (5)

Q5 a) Explain the nearest neighbour approach for multicategory classification. Give suitable example. (5)

b) State the Bayes Rule and explain how it is applied to pattern classification problems. Show that in a multiclass classification task the Bayes decision rule minimizes the error probability. (5)

Q6 a) Consider a grammar with A={a,b,c}, S=S, I={A,B}, and (5)

$$P=\begin{Bmatrix} S \rightarrow \text{aSBA OR aBA} & \text{AB} \rightarrow \text{BA} \\ \text{bB} \rightarrow \text{bb} & \text{bA} \rightarrow \text{bc} \\ \text{cA} \rightarrow \text{cc} & \text{aB} \rightarrow \text{ab} \end{Bmatrix}$$

Prove that this grammar generates the language $L(G)=\{a^n b^n c^n \mid n \geq 1\}$.

b) Briefly explain about grammatical inference with suitable example. (5)

Q7 a) Briefly explain about Ho-Kashyap procedures. (5)

b) Consider a d-n_H–c network trained with n patterns for m_e epochs. (5)

i) What is the space complexity in this problem? (Consider both the storage of network parameter as well as patterns but not the program itself).

ii) Suppose the network is trained stochastic mode. What is the time complexity?

iii) Suppose the network is trained in batch mode. What is the time complexity?

Q8 Write short notes on any two of the following (5 x 2)

a) Hopfield Network
b) Valiant's framework
c) K-means clustering
d) Pattern associators