Roll No.												

PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING

TEST 1

Sem: Odd Sem 2019-20
Course Code: ECE 401
Course Name: ARTIFICIAL NEURAL NETWORKS
Program & Sem: B.Tech (ECE) & VII (OE)

Date: 12.10.2019
Time: 1.30 PM to 2.30 PM
Max Marks: 40
Weightage: 20%

Instructions:

(i) Read the question properly and answer accordingly.
(ii) Question paper consists of 3 parts.
(iii) Scientific and Non-programmable calculators are permitted.

Part A [Memory Recall Questions]

Answer all the Questions. Each Question carries six marks. (3Qx6M=18M)

1. Discuss the applications of Neural Networks in detail. (C.O.NO.1)[Knowledge]

2. Explain the Neuron Model with Suitable Diagrams consider bias and activation function. (C.O.NO.1)[Knowledge]

3. What is Signal flow graph? Explain the rules associated with signal flow graph representation of a model. (C.O.NO.1)[Knowledge]

Part B [Thought Provoking Questions]

Answer both the Questions. Each Question carries seven marks. (2Qx7M=14M)

4. a) In a neuron model what are the two functions which can acts as activation function? Name it and explain the same with suitable mathematical and graphical representation. (C.O.NO.1)[Comprehension]

   b) Name the property which can exhibits a graceful balance between linear and nonlinear model and deicide which activation function having this property.

   (C.O.NO.1)[Comprehension]

5. a) Identify the networks which have at least one feedback loop and can capable to distinguish feed forward neural networks, then draw the same network.

Part C [Problem Solving Questions]

Answer the Question. The Question carries eight marks. (1Qx8M=8M)

6. What is the importance of feedback in neural network? Consider 'w' is the initial weight and feedback is 'z^{-1}', determine $Y_K(n)$ for the input $X_K(n)$ where 'k' is neuron. Realize the system and also discuss on its stability (C.O.NO.1)[Application]

SCHOOL OF ENGINEERING

Semester: 2019-2020 (ODD Semester)

Course Code: ECE-401

Course Name: Artificial Neural Networks

Program & Sem: B.Tech(ECE) & 7th Sem

Date: 27-09-2019 (Friday)
Time: 1.00PM to 2.00PM

Max Marks: 40

Weightage: 20%

Extract of question distribution [outcome wise & level wise]

Q.NO.	C.O.NO	Unit/Module Number/Unit /Module Title	Memory recall type [Marks allotted] Bloom's Levels			Thought provoking type [Marks allotted] Bloom's Levels			Problem Solving type [Marks allotted]			Total Marks
			K			C			A			
1	CO1	MODULE-1	K	6								6
2	CO1	MODULE-1	K	6								6
3	CO1	MODULE-1	K	6								6
4	CO1	MODULE-1				C	7					7
5	CO1	MODULE-1				C	7					7
6	CO1	MODULE-1							A	8		8
	Total Marks			18			14			8		40

K =Knowledge Level   C = Comprehension Level, A = Application Level

Note: While setting all types of questions the general guideline is that about 60%

Of the questions must be such that even a below average students must be able to attempt, About 20% of the questions must be such that on above average students must be able to attempt and finally 20% of the questions must be such that only the bright students must be able t attempt.

[I hereby certify that All the questions are set as per the above guide lines.]

Reviewers' Comments

(1) Q5 Timing (2) Q1 → Q is something else, Ans is defin
(Q6) More Control Sys Oriented Less ANN
(Q4) – Check Remark in scheme

Annexure- II: Format of Answer Scheme

SCHOOL OF ENGINEERING

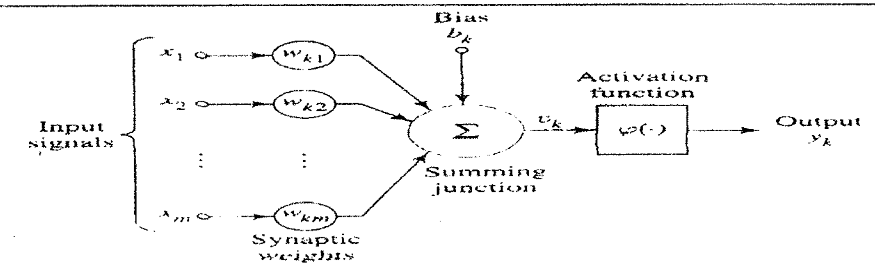
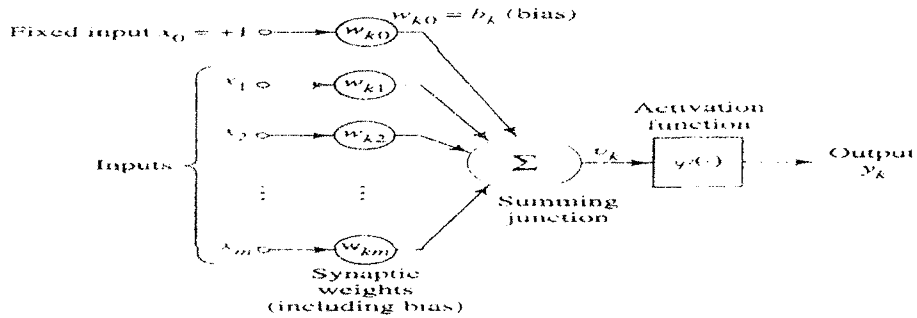
SOLUTION

Semester: 2019-2020 (ODD Semester)
Course Code: ECE-401
Course Name: Artificial Neural Networks
Program & Sem: B.Tech(ECE) & 7th Sem

Date: 27-09-2019 (Friday)
Time: 1.00PM to 2.00PM
Max Marks: 40
Weightage: 20%

Part A (3Q x6 M =18 Marks)

Q No	Solution	Scheme of Marking	Max. Time required

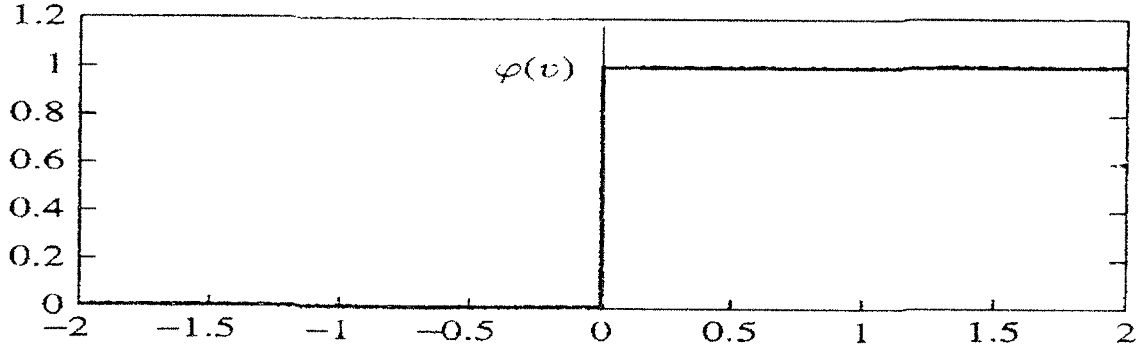
			Question
1	*A neural network is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects.* 1. *Knowledge is acquired by the network from its environment through a learning process.* 2. *Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge.*	3+3	5min
2	 	3+3	10min
3	A *signal-flow graph* is a network of directed *links* (*branches*) that are interconnected at certain points called *nodes*. A typical node j has an associated *node signal* x_j. A typical directed link originates at node j and terminates on node k: it has an associated **Rule 1.** A signal flows along a link only in the direction defined by the arrow on the link. Two different types of links may be distinguished: **Rule 2.** A node signal equals the algebraic sum of all signals entering the pertinent node via the incoming links. **Rule 3.** The signal at a node is transmitted to each outgoing link originating from that node, with the transmission being entirely independent of the transfer functions of the outgoing links.	2+4	5min

Part B (2Q x 7M = 14 Marks

<table>
<tr><th></th><th>Solution</th><th>of Marking</th><th>Time required for each Question</th></tr>
<tr><td>4</td><td>

1. *Threshold Function.* For this type of activation function, described in Fig. 1.8a, we have

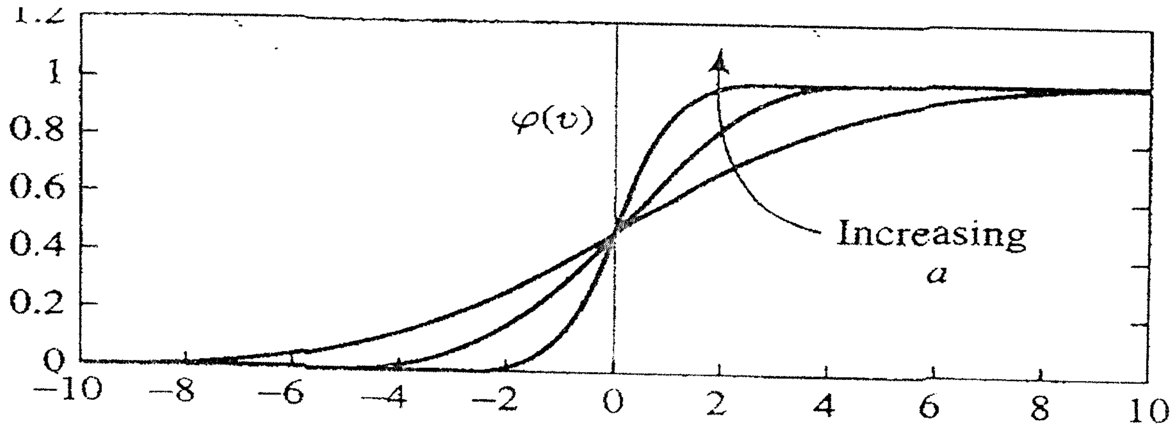
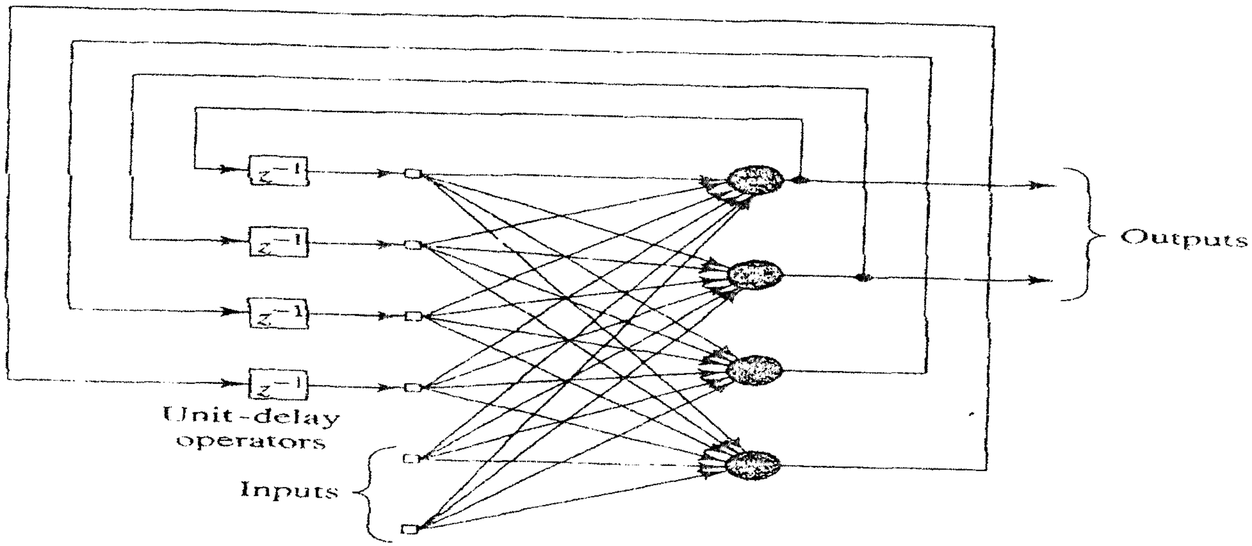
$$\varphi(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{if } v < 0 \end{cases} \qquad (1.8)$$



3. *Sigmoid Function.* The sigmoid function, whose graph is s-shaped, is by far the most common form of activation function used in the construction of artifical neural networks. It is defined as a strictly increasing function that exhibits a graceful balance between linear and nonlinear behavior.[3] An example of the sigmoid function is the *logistic function*,[4] defined by

$$\varphi(v) = \frac{1}{1 + \exp(-av)} \qquad (1.12)$$

</td><td>3+4</td><td>10min</td></tr>
</table>

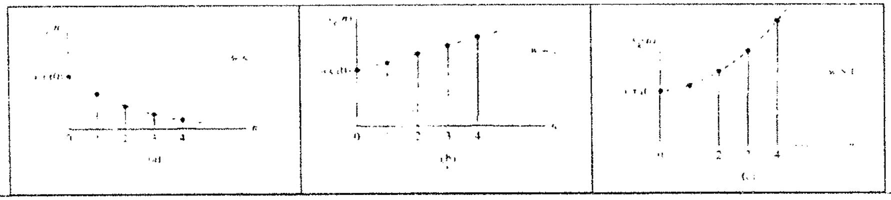
	 Sigmaidol function have the differentiability property where as threshold function not Other Func is also allowed		
5	Recurrent Networks With Hidden layers  	2+5	7

Part C (1Q x8 M = 8Mark

Q No	Solution	Scheme of Marking	Max. Time required for each Question
6	For a single loop feedback system shown in fig. below  a) Find the close loop operator of the system. b) Express the output signal $Y_K(n)$ as an infinite weighted summation of present and past samples of input signal $X_j(n)$ c) How the selection of the values of W will lead to divergence . $\frac{A}{1-AB} = \frac{w}{1-wz^{-1}}$ $= w(1-wz^{-1})^{-1}$ Using the binomial expansion for $(1-wz^{-1})^{-1}$, we may rewrite the closed loop operator of the system as $\frac{A}{1-AB} = w\sum_{l=0}^{\infty} w^{l} z^{-l}$ (1.19) Hence, substituting Eq (1.19) in (1.18), we get $y_k(n) = w\sum_{l=0}^{\infty} w^{l} z^{-l}[x_j(n)]$ (1.20) where again we have included square brackets to emphasize the fact that z^{-1} is an operator. In particular, from the definition of z^{-1} we have $z^{-l}[x_j(n)] = x_j(n-l)$ (1.21)	2+4+2	15min

may express the output signal $y_k(n)$ as an infinite weighted summation of present and past samples of the input signal $x_j(n)$, as shown by

$$y_k(n) = \sum_{l=0}^{\infty} w^{l+1} x_j(n-l) \qquad (1.22)$$



Roll No.												

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING

TEST 2

Sem & AY: Odd Sem 2019-20
Course Code: ECE 401
Course Name: ARTIFICIAL NEURAL NETWORKS
Program & Sem: B.Tech (ECE/CSE/EEE/MEC) & VII Sem

Date: 16.11.2019
Time: 1.00 PM to 2.00 PM
Max Marks: 40
Weightage: 20%

Instructions:

(i) Read the question properly and answer accordingly.
(ii) Question paper consists of 3 parts.
(iii) Scientific and Non-programmable calculators are permitted.

Part A [Memory Recall Questions]

Answer all the Questions. Each question carries six marks. (3Qx6M=18M)

1. Explain the process of supervised learning with example. (C.O.NO.2)[Knowledge]
2. Explain Fuzzy K-means Algorithm with Suitable example. (C.O.NO.2)[Knowledge]
3. Explain the concept of linear seperability with respect to Perceptron Algorithm and mention the limitation of perceptron. (C.O.NO.2)[Knowledge]

Part B [Thought Provoking Questions]

Answer both the Questions. Each question carries seven marks. (2Qx7M=14M)

4. Let D1, D2, D3 are the three data with Y(X,D1), Y(X,D2), Y(X,D3) as actual outputs and E_D[Y(X,D1)], E_D[Y(X,D2)], E_D[Y(X,D3)] are expected outputs respectively. Set the reference value g(x).
   a. Name property which denotes the deviation from g(x) to Y(X,D1), Y(X,D2) and Y(X,D3) .
   b. Name property which denotes the deviation within Y(X,D1), Y(X,D2) and Y(X,D3)
   c. Finally while estimating the target value how we can calculate the expected loss and comment on the result. (C.O.NO.2)[Comprehension]

5. What is Linear Model for regression, extend this Model and find pseudo inverse by using a function which is having the property of "linear combination of fixed non-linear function." (C.O.NO.2)[Comprehension]

Part C [Problem Solving Questions]

Answer the Question. The question carries eight marks. (1Qx8M=8M)

6. Consider a Linear model for regression with design matric $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ -1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ and target vector $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ then identify the number of weights in the model and find maximum likelihood weight vector. (C.O.NO.2)[Application]

Annexure- II: Format of Answer Scheme

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

SCHOOL OF ENGINEERING

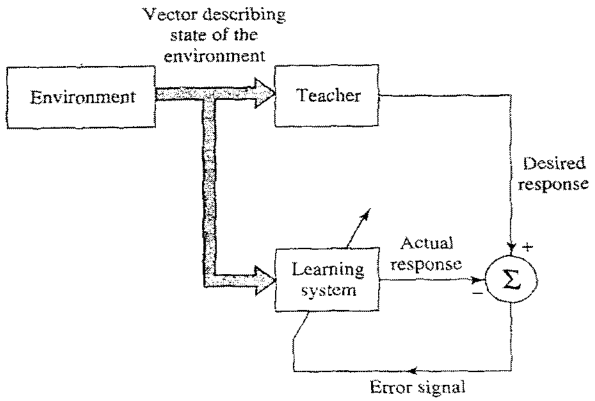
SOLUTION

Semester: 2019-2020 (ODD Semester)
Course Code: ECE-401
Course Name: Artificial Neural Networks
Program & Sem: B.Tech(ECE) & 7th Sem

Date: 27-09-2019 (Friday)
Time: 1.00PM to 2.00PM
Max Marks: 40
Weightage: 20%

Part A (3Q x6 M =18 Marks)

Q No	Solution	Scheme of Marking	Max. Time required for each Question
1	**Learning with teacher is refered to superwised learning .** **Explanation**	**2+2+2**	**5min**

			
2		**4(steps)+2(diagram)**	**10min**

Step 1: Initialize the membership value (M_{ik}) randomly

Step 2: Compute the centroids of the clusters

Eg: 6 clusters

C_1 = [vector 1 * (membership value of vector 1 belongs to cluster 1)^2 i.e. M_{11}^2
+ vector 2 * (" " " " 2 " " ")^2 i.e. M_{12}^2
+ ... + vector 100 × M_{1100}^2] / (Sum of the squared values of the membership values belongs to the cluster 1)

Slly All centroids are obtained

Step 3: Update the membership values M using ~~the~~ new (current) values of centroids

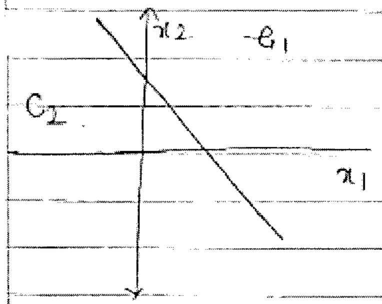
3	If two mutually independent variables are there then perceptron will give 100% guarantee to classify.	**2(statement)+1+2+1**	**5min**

The Simplest form of peceptron.

The two decision regions Seperated by hyperplane, and is defined by

$$\sum_{i=1}^{m} w_i x_i + b = 0$$

The fig explains the case of two input votiable x_1 and x_2



(A) $x_1 x_2$ lies above the boundry lies in class C_1

(B) $x_1 x_2$ lies below the boundry is assigned to C_2.

(C) The role of bias is to Shift decision boundry away from origine.

The weights $w_1, w_2 - - - w_n$ of the perceptron are adopted on an iteration by Iteration basis.

For this a error- Correction rule known as perceptron Convergence algoritham is used.

	If the hyperplane that seperates the inputs into 2 different class by drawing a hyperplane clearly then this is called as linearly seperable class of data . if the perceptron receive this data it will converge or else it will fail to classify this data. **Perceptron has limitation that it can take only 2 class problems.**		

Part B (2Q x 7M = 14 Marks)

Q No	**Solution**	**Scheme of Marking**	**Max. Time required for each Question**
5	Name of the function is Basis Function	2 for the definition +4 for the derivation +1 for the comment	10min

$$X = \{x_1, x_2, x_3 \cdots\cdots x_N\}$$

$x_1 \longrightarrow t_1$

$x_2 \longrightarrow t_2$

$\vdots$

$x_N \longrightarrow t_N$

$$t = y(x, w)$$

$$y = (x, w) = w_0 x_0 + w_1 x_1 + w_2 x_2 + \cdots\cdots w_{m-1} x_N$$

$$y(x, w) = w_0 + w_1 x_1 + w_2 x_2 + \cdots\cdots + w_{m-1} x_N$$

The above eqn is called Linear model. we can extend this Class of model by considering "Basis Function"

* Basis function is linear combination of fixed non-linear function.

$$y(x, w) = w_0 + \sum_{j=1}^{m} w_j \phi_j(X)$$

w_0 → Fixed offset (or) Basis parameter

$\phi_j(X)$ → Basis fun^c

$$y(x, w) = \sum_{j=0}^{m-1} w_j \phi_j(X)$$

$\phi_0(X) = 1$ → dummy basis fun^c

$$t = \sum_{j=0}^{m-1} w_j \phi_j(X)$$

In matrix form,

$$\phi w = t$$

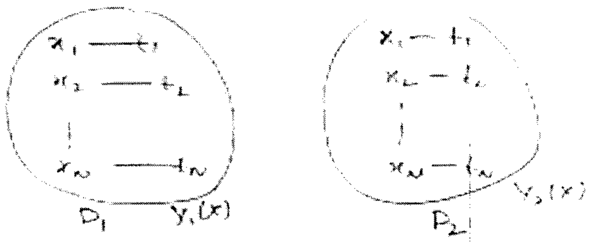
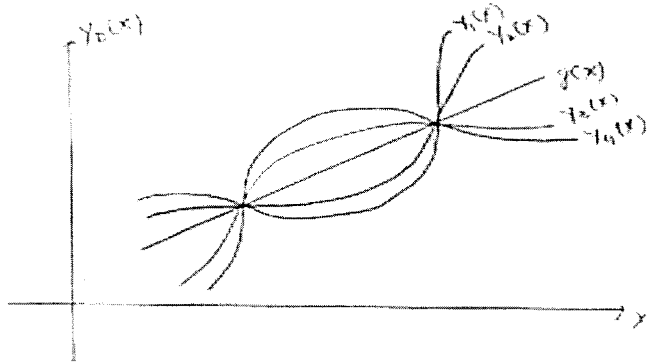
$$\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix}_{N\times 1} = \begin{bmatrix} \phi_0(x_1) & \phi_1(x_1) & \phi_2(x_1) & \cdots & \phi_{m-1}(x_1) \\ \phi_0(x_2) & \phi_1(x_2) & \phi_2(x_2) & \cdots & \phi_{m-1}(x_2) \\ \vdots & & & & \\ \phi_0(x_N) & \phi_1(x_N) & \phi_2(x_N) & \cdots & \phi_{m-1}(x_N) \end{bmatrix}_{N\times m} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{m-1} \end{bmatrix}_{m\times 1}$$

$$\phi w = t$$

$$\phi^T \phi \, w = \phi^T t$$

$$w = \underbrace{(\phi^T \phi)^{-1} \phi^T}_{\text{pseudo inverse}} t$$

w → eqⁿ for Least Square

4	Bais, Variance decomposition:  similarly D_3, D_4 . . . . D $Y_1(x)$, $Y_2(x)$, $Y_3(x)$, $Y_4(x)$ are considered as '4' datasets D_1, D_2, D_3, D_4 estimating t as $\bar{T} = y(x)$ $y(x)$ = reference only 	2(graphical)+5(mathematical)	10min

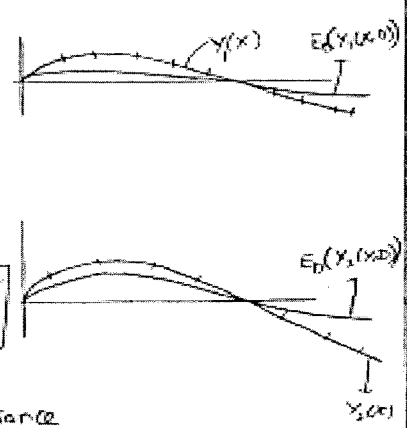
While estimating the target value.

$$E\left[(Y(x) - g(x))^2\right]$$

$$E_D\left[(Y(x,D) - g(x))^2\right]$$

By adding & substracting with $E_D(Y(x,D))$

$$E_D\left[\left(\underbrace{Y(x,D) - E_D(Y(x,D))}_{a} + \underbrace{E_D(Y(x,D)) - g(x)}_{b}\right)^2\right]$$



Ist term $E_D\left[\left(Y(x,D) - E_D(Y(x,D))\right)^2\right] \longrightarrow$ Variance

IInd term $E_D\left[\left(E_D(Y(x,D)) - g(x)\right)^2\right] \longrightarrow$ Bias^2

~~IIIrd term $2\left[E_D(Y(x,D) - E_D(Y(x,D)))\right]\left[E_D(E_D(Y(x,D)) - g(x))^2\right]$~~

IIIrd term $2E_D\left[\left(Y(x,D) - E_D(Y(x,D))\right)\left(E_D(Y(x,D)) - g(x)\right)\right] \longrightarrow$ Error

Finally, while estimating the target value

Expected loss = Variance + $(\text{bias})^2$ + Noise

Part C (1Q x8 M = 8Marks)

Q No	Solution	Scheme of Marking	Max. Time required for each Question
6	Solution is obtained by maximizing likelihood func $w_{ML} = (\phi^T \phi)^{-1} \phi^T t$ w_{ML}: $m \times 1$; ϕ^T: $m \times N$, ϕ: $N \times m$ → $m \times m$; ϕ^T: $m \times N$ → $m \times N$; t: $N \times 1$ → $m \times 1$ Design Matrix = $\begin{bmatrix} 1 & 0 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ Target vector= $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ Maximum likelihood weight vector W_{ML}= $\begin{bmatrix} 0 \\ 0.6 \\ 0.8 \end{bmatrix}$	2(no. of. Weights)+5(ML Weight vector)	15min

Roll No												

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING

END TERM FINAL EXAMINATION

Semester: Odd Sem. 2019 - 20 — **Date**: 23 December 2019

Course Code: ECE 401 — **Time**: 9:30 AM to 12:30 PM

Course Name: ARTIFICIAL NEURAL NETWORKS — **Max Marks**: 80

Program & Sem: B.Tech (All Programs) & VII (OE–I) — **Weightage**: 40%

Instructions:

(i) *Read the question properly and answer accordingly.*
(ii) *Question paper consists of 3 parts.*
(iii) *Scientific and Non-programmable calculators are permitted.*

Part A [Memory Recall Questions]

Answer all the Questions. Each Question carries 4 marks. **(5Qx4M=20M)**

1. What is the use of activation function in neuron model? Explain any two activation functions. (C.O.No.1) [Knowledge]
2. What are the methods used to estimate target variables and weight parameters. (C.O.No.2) [Knowledge]
3. Explain the steepest descent algorithm with diagrams. (C.O.No.3) [Comprehension]
4. What is Design matrix? Explain with an Example. (C.O.No.1) [Knowledge]
5. Explain Perceptron convergence theorem. (C.O.No.3) [Knowledge]

Part B [Thought Provoking Questions]

Answer all the Questions. Each Question carries 8 marks. **(3Qx8M=24M)**

6. How support vector machine treated as a 'Minimization problem'? Consider two classes and Explain with suitable diagrams. (C.O.No.3) [Application]
7. Which algorithm is used to reduce data from higher dimension to lower dimension? Will the same algorithm is used for classification also? Identify that algorithm and write the step by step procedure involved in it. (C.O.No.3) [Comprehension]
8. For three class problem (Say C1, C2, C3), Consider two scatter matrix S_B (between class Scatter matrix) and S_W (Within Class Scatter Matrix). (C.O.No.3) [Comprehension]
   Statement: trace of S_B will be maximized and trace of S_W will be minimized.

   i) The statement is correct or wrong, if it is correct, why? If it is wrong, why?
   ii) Formulate those scatter matrix S_B and S_W
   iii) Formulate the Objective function.

Part C [Problem Solving Questions]

Answer all the Questions. Each Question carries 12 marks. **(3Qx12M=36M)**

9. Implement Back Propagation Model and find $P_{ij}(t+1)$, $q_{jk}(t+1)$, by Minimizing the objective function "J" (where i=1,2,3,4; j=1,2; k=1,2, t= target variable) from the following specifications (C.O.No.3) [Application]

   i) Number of inputs are 4(say i_1, i_2, i_3, i_4) Hidden Layer nodes are 2(say h_1, h_2) and outputs are2(say O_1, O_2)

   ii) Hidden layer biasing elements are U_1, U_2 and output layer biasing elements are V_1 , V_2

10. Using Discriminant Approach and prove that the minimum value for objective function will be zero if the target value is equals to the actual output. For minimization of objective function you may use steepest descent algorithm. (C.O.No.2) [Comprehension]

11. Consider three data sets C_1= [X_{11}, X_{12}, X_{13}, ...X_{1n1}], C_2= [X_{21}, X_{22},X_{23}, ...X_{2n2}], C_3= [X_{31}, X_{32},X_{33}, ...X_{3n3}]. Expand the three datasets (C.O.No.3) [Comprehension]

    i) as a Gram Matrix

    ii) as a Gram Matrix with basis function $\Phi(x)$

    iii) as a Gram Matrix with Kernel

SCHOOL OF ENGINEERING

END TERM FINAL EXAMINATION

Date: 23rd December 2019
Time: 9.30 AM to 12.30 PM
Max Marks: 80
Weightage: 40%

Semester: Odd Sem. 2019-20
Course Code: ECE401
Course Name: ARTIFICIAL NEURAL NETWORKS
Program & Sem: B.Tech (ECE, EEE, CSE, MEC) & 7th Sem

Extract of question distribution [outcome wise & level wise]

Q.NO.	C.O.NO (% age of CO)	Unit/Module Number/Unit /Module Title	Memory recall type [Marks allotted] Bloom's Levels	Thought provoking type [Marks allotted] Bloom's Levels	Problem Solving type [Marks allotted]	Total Marks
			K	C	A	
1	1	Module 1	4			4
2	2	Module 2	4			4
3	3	Module 3	4			4
4	1	Module 1	4			4

5	3	Module 3	4			4
6	3	Module 3			8	8
7	3	Module 3		8		8
8	3	Module 3		8		8
9	3	Module 3			12	12
10	2	Module 2		12		12
11	3	Module 3		12		12
	Total Marks		20	40	20	80

K =Knowledge Level C = Comprehension Level, A = Application Level

Note: While setting all types of questions the general guideline is that about 60%

Of the questions must be such that even a below average students must be able to attempt, About 20% of the questions must be such that only above average students must be able to attempt and finally 20% of the questions must be such that only the bright students must be able to attempt.

I hereby certify that all the questions are set as per the above guidelines.

Faculty Signature:

Reviewer Commend: Dr. M. Lavy.

1. Faculty not signed
2. Thought provoking Question — ?
3. Answer scheme step marks —?

Format of Answer Scheme

SCHOOL OF ENGINEERING

SOLUTION

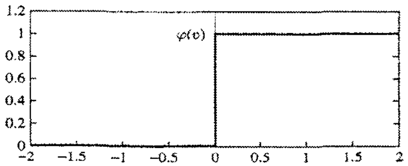
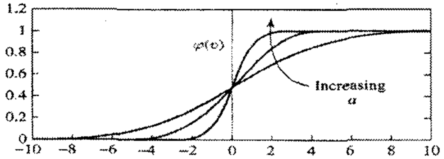
GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

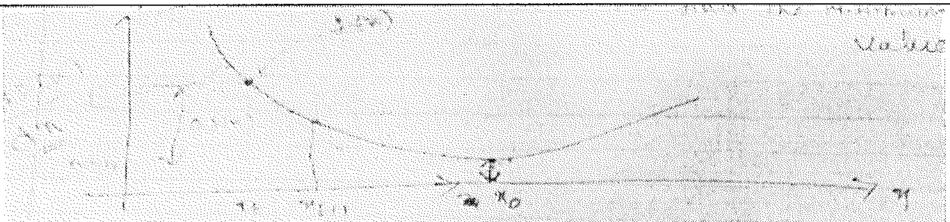
Semester: Odd Sem. 2019-20
Course Code: ECE401
Course Name: ARTIFICIAL NEURAL NETWORKS
Program & Sem: B.Tech (ECE, EEE, CSE, MEC) & 7th Sem

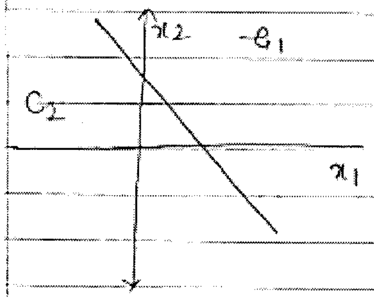
Date: 23rd December 2019
Time: 9.30 AM to 12.30 PM
Max Marks: 80
Weightage: 40%

Part A (5Q x4M =20Marks)

Q No	Solution	Scheme of Marking	Max. Time required for each Question
1	Activation function defines o/p of neuron interms of the induced local field. **1.** *Threshold Function*. For this type of activation function, described in Fig. 1.8a, we have $$\varphi(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{if } v < 0 \end{cases} \quad (1.8)$$	2M+2M=4M	8min

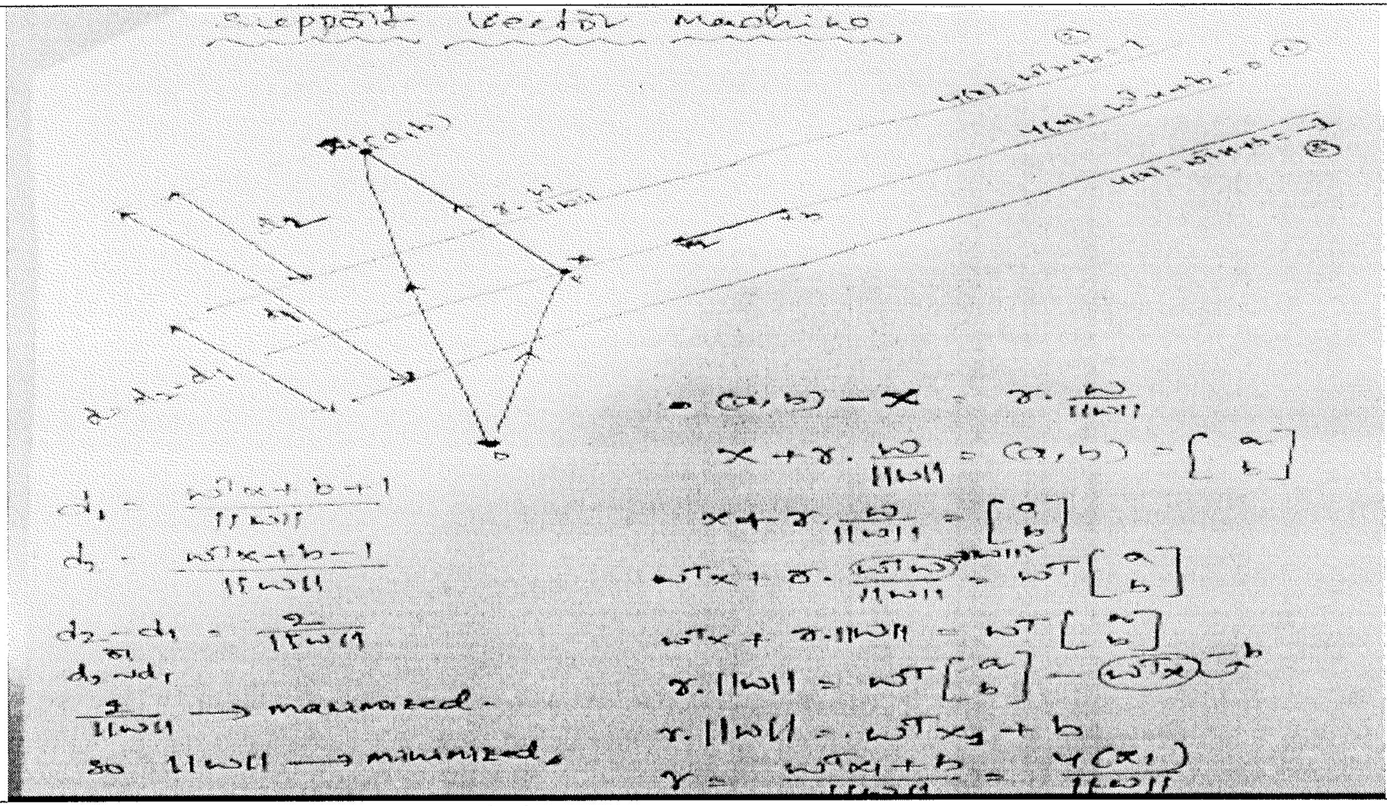
	  **3.** *Sigmoid Function*: The sigmoid function, whose graph is s-shaped, is by far the most common form of activation function used in the construction of artifical neural networks. It is defined as a strictly increasing function that exhibits a graceful balance between linear and nonlinear behavior.[3] An example of the sigmoid function is the *logistic function*,[4] defined by $$\varphi(v) = \frac{1}{1 + \exp(-av)} \quad (1.12)$$		
2	$f_1(t)$ ⟶ Prior density $f_2(t/X)$ ⟶ Aposterior density $f_3(X/t)$ ⟶ likelyhood , for estimating weight 'w' parameter $f_1(w)$ → Prior estimation $f_2(w/t, X)$ → Aposterior estimation $f_3(t/w, X)$ → Likelyhood estimation	2M+2M=4M	8min

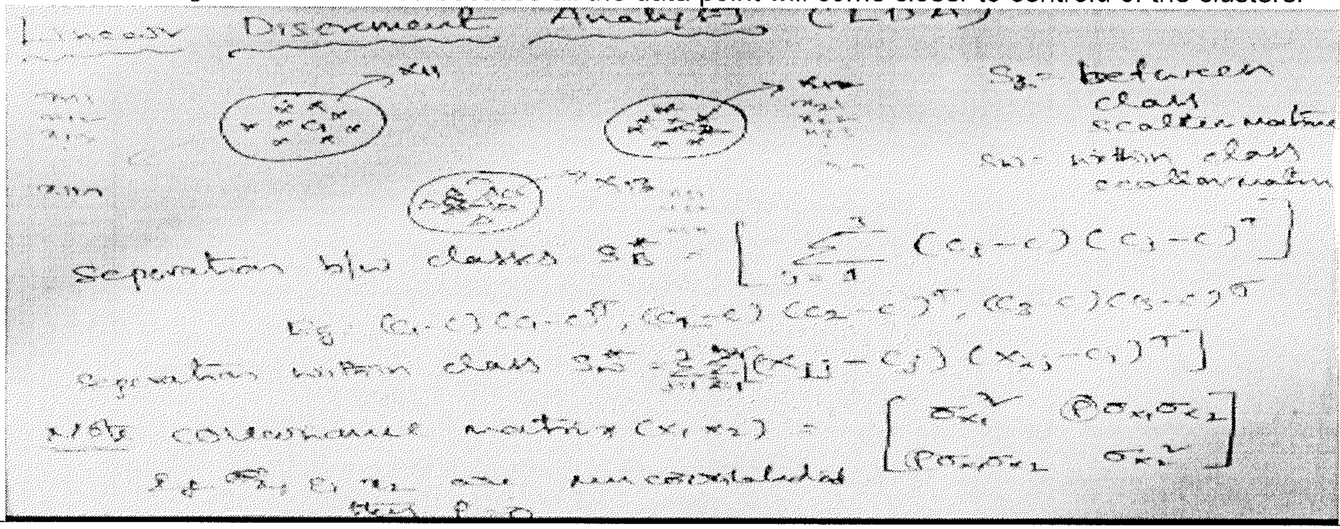
3	 $x_{t+1} = x_t - \alpha \frac{df(x)}{dx}\Big\|_{x_t}$ $\frac{df}{dx} \rightarrow$ 2D → sl $\nabla f \rightarrow$ 3D,4D → g $w(t+1) = w(t) - \alpha \nabla J$ → objective function → Regularization factor $x_{t+1} = x_t - \alpha \nabla f(x)$ if $f(x)$ is 3D, 4D, ... $x_{t+1} = x_t - \alpha \frac{df(x)}{dx}\Big\|_{x=t}$ if $f(x)$ is 2D gradient → indicates slope steepest descent method α $w(t+1) = w(t) - \eta \nabla J$ Area application	4M	8min
4	In matrix form, $\phi w = t$ $\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix}_{N\times 1} = \begin{bmatrix} \phi_0(x_1) & \phi_1(x_1) & \phi_2(x_1) & \cdots & \phi_{m-1}(x_1) \\ \phi_0(x_2) & \phi_1(x_2) & \phi_2(x_2) & \cdots & \phi_{m-1}(x_2) \\ \vdots \\ \phi_0(x_N) & \phi_1(x_N) & \phi_2(x_N) & \cdots & \phi_{m-1}(x_N) \end{bmatrix}_{N\times m} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{m-1} \end{bmatrix}_{m\times 1}$	4M	8min

5	The Simplest form of peceptron. The two decision regions Seperated by hyperplane, and is defined by $\sum_{i=1}^{m} w_i x_i + b = 0$ The fig explains the case of two input votiable x_1 and x_2  (A) $x_1 x_2$ lies above the boundry lies in class C_1 (B) $x_1 x_2$ lies below the boundry is assigned to C_2. (C) The role of bias is to Shift decision boundry away from origine. The weights $w_1, w_2 \cdots w_n$ of the perceptron are adopted on an iteration by iteration basis. For this a error-correction rule known as perceptron Convergence algoritham is used.	4M	8min

Part B

(3Q x 8M =24 Marks)

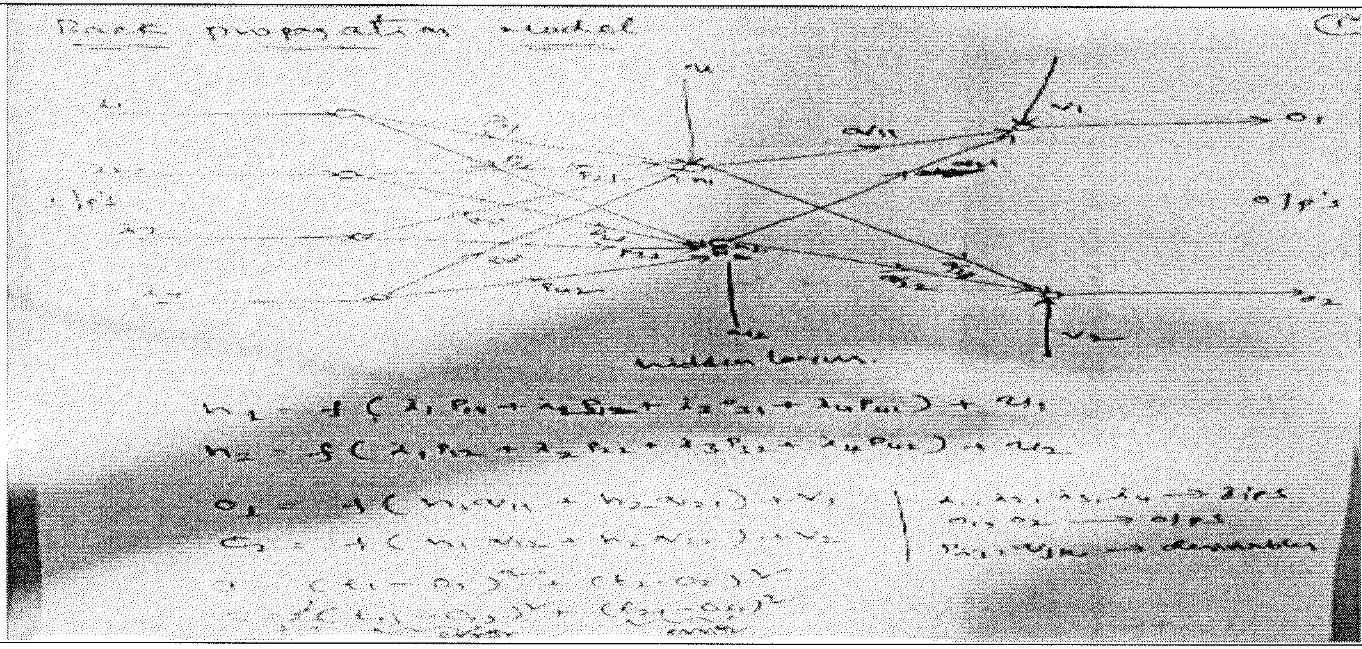
Q No	Solution	Scheme of Marking	Max. Time required for each Question
6		8M	19min
7	Algorithm is Principle Component Analysis, Yes, PCA we can use as a dimensionality reduction technique.	1M+1M+6M	19min

	Principal component analysis (PCA) is one of the most popular techniques for dimensionality reduction. Starting from an original set of l samples (features), which form the elements of a vector $x \in R^l$ the goal is to apply a linear transformation to obtain a new set of samples: $y = A^T x$ *so that the components of 'y' are uncorrelated. In a second stage, one chooses the most significant of these components. The steps are summarized bellow:* • Estimate the autocorrelation matrix for N feature vectors $R \approx \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T$ • Perform the eigen decomposition of S and compute the l e n vectors, λ_i, $a_i \in R^l$, $i = 0, 2, \ldots, l-1$. • Arrange the eigenvalues in descending order, $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{l-1}$. • Choose the m largest eigenvalues. Usually m is chosen so that the gap between λ_{m-1} and λ_m is large. Eigenvalues $\lambda_0, \lambda_1, \ldots, \lambda_{m-1}$ are known as the *m principal components*. • Use the respective (column) eigen vectors a_i, $i = 0, 1, 2, \ldots, m-1$ to form the transformation matrix $A = [a_0\ a_1\ a_2 \cdots a_{m-1}]$ • Transform each l-dimensional vector x in the original space to an m-dimensional vector y via the transformation $y = A^T x$. In other words, the ith element $y(i)$ of y is the *projection* of x on a_i $y(i) = a_i^T x$		
8	i) Yes, The given statement is correct because if we maximizing the trace of S_w we can get good classification of data sets. S_B will be minimized because all the data point will come closer to centroid of the clusters. 	1M+2M+5M	19min

	Similarly [illegible] (2.5) we can write $\tilde{S}_W = w^T S_W w$ $= w^T Q w$ $\tilde{S}_B = w^T S_B w$ Then the objective function J, [illegible] maximize function J = $\frac{\text{trace}[w^T S_B w]}{\text{trace}[w^T S_W w]}$ [illegible] the objective function [illegible] λ and take single [illegible] w_1. $\lambda w_1^T S_W w_1 = w_1^T S_B w_1$ $S_B w_1 = \lambda S_W w_1$ → Eigenvalue $S_W^{-1} S_B w_1 = \lambda w_1$ → EIGEN vectors		

Part C (3Q x 12M = 36Marks)

Q No	Solution	Scheme of Marking	Max. Time required for each Question

9		6M+6M	24min

$$J = \left(t_1 - f(h_1 w_{11} + h_2 w_{21} + v_1)\right)^2 + \left(t_2 - f(h_1 w_{12} + h_2 w_{22} + v_2)\right)^2$$

This is minimization problem so, J should ↓ (minimized)

using steepest descent algorithm.

$$w_{11}(t+1) = w_{11}(t) - \frac{\eta}{2} \frac{dJ}{dw_{11}}\Big|_t \qquad p_{11}(t+1) = p_{11}(t) - \frac{\eta}{2}\frac{dJ}{dp_{11}}$$

$$w_{11}(t+1) = w_{11}(t) + \eta e_1 h_1 \qquad p_{12}(t+1) = p_{12}(t) - \frac{\eta}{2}\frac{dJ}{dp_{12}}$$

$$w_{12}(t+1) = w_{11}(t) + \eta e_2 h_1$$

||ly $p_{21}, p_{22}, p_{31}, p_{32}, p_{41}, p_{42}$

$$w_{21}(t+1) = w_{21}(t) + \eta e_1 h_2$$

$$w_{22}(t+1) = w_{22}(t) + \eta e_2 h_2$$

$$J = \left(t_1 - f\left(f(i_1 p_{11} + i_2 p_{21} + i_3 p_{31} + i_4 p_{41} + u_1) w_{11} + h_2 w_{21} + v_1\right)\right)^2$$
$$+ \left(t_2 - f\left(f(i_1 p_{11} + i_2 p_{21} + i_3 p_{31} + i_4 p_{41} + u_1) w_{12} + h_2 w_{22} + v_2\right)\right)^2$$

$$\frac{dJ}{dp_{11}} = 2 e_1 (-i_1 w_{11}) + 2 e_2 (-i_1 w_{12})$$
$$= -2 e_1 i_1 w_{11} - 2 e_2 i_1 w_{12}$$
$$= -2 i_1 (e_1 w_{11} + e_2 w_{12})$$

$$p_{11}(t+1) = p_{11}(t) + \eta i_1 (e_1 w_{11} + e_2 w_{12})$$

||ly $p_{21}, p_{22}, \ldots$

10	Discriminant approach (for two class problems) using probabilistic methods: $x_1 — t_1$, $x_2 — t_2$, …, $x_N — t_N$ $p(c_1 \mid x) = \frac{p(x \mid c_1)\, p(c_1)}{p(x \mid c_1)\, p(c_1) + p(x \mid c_2)\, p(c_2)}$ $p(c_1 \mid x) = \frac{e^{s_1(x)}}{e^{s_1(x)} + e^{s_2(x)}} = y_1(x)$ Here $e^{s_1(x)} = p(x \mid c_1) \cdot p(c_1)$ $e^{s_2(x)} = p(x \mid c_2)\, p(c_2)$ Ily $p(c_2 \mid x) = \frac{p(x \mid c_2) \cdot p(c_2)}{p(x \mid c_1)\, p(c_1) + p(x \mid c_2) \cdot p(c_2)} = \frac{e^{s_2(x)}}{e^{s_1(x)} + e^{s_2(x)}} = y_2(x)$ $s_1(x) = \ln p(x \mid c_1) + \ln p(c_1)$ $s_2(x) = \ln p(x \mid c_2) + \ln p(c_2)$ [illegible]	6M+6M	24min

$$J = \prod_{k=1}^{N} P(C_1|x_k)^{t_k} \cdot P(C_2|x_k)^{(1-t_k)}$$

$$J = \prod_{k=1}^{N} y_k^{t_k} \cdot (1-y_k)^{(1-t_k)}$$

note: Taking 'ln' and "–" will leads to minimi[ze] the objective function

$$J = -\sum_{k=1}^{N} t_k \ln y_k - (1-t_k) \ln(1-y_k)$$

$$\nabla J = -\sum_{k=1}^{N} \frac{t_k}{y_k} \frac{dy_k}{dw} + \frac{(1-t_k)}{(1-y_k)} \cdot \frac{dy_k}{dw}$$

$$y_k = \frac{e^{s_1(x_k)}}{e^{s_1(x_k)} + e^{s_2(x_k)}} \qquad \frac{dy_k}{dw} = \frac{dy_k}{ds_1} \cdot \frac{ds_1}{dw}$$

$$\frac{dy_k}{ds_1} = \frac{[e^{s_1(x_k)} + e^{s_2(x_k)}] \cdot e^{s_1(x_k)} - e^{s_1(x_k)} \cdot e^{s_1(x_k)}}{[e^{s_1(x_k)} + e^{s_2(x_k)}]^2}$$

$$\frac{dy_k}{ds_1} = y_k - y_k^2 = y_k(1-y_k)$$

$$\frac{ds_1}{dw} = \frac{d}{dw}\left[w_1^T \phi(x_k)\right] = \phi(x_k)$$

	$= -\sum_{k=1}^{N} \frac{t_k}{y_k} y_k(1-y_k) + \sum_{k=1}^{N} \frac{(1-t_k)}{(1-y_k)} y_k(1-y_k)\phi(x_k)$ $= \sum_{k=1}^{N} -t_k\phi(x_k) + t_k y_k \phi(x_k) + y_k\phi(x_k) - t_k y_k \phi(x_k)$ $= \sum_{k=1}^{N} (y_k - t_k)\,\phi(x_k)$ $\nabla J = \phi^T(x)\,(y-t)$		
11	$\begin{bmatrix} x_{11}^T x_{11} & x_{11}^T x_{12} & \cdots & x_{11}^T x_{1n_1} & x_{11}^T x_{21} & x_{11}^T x_{22} & \cdots & x_{11}^T x_{2n_2} & x_{11}^T x_{31} & \cdots & x_{11}^T x_{3n_3} \\ x_{12}^T x_{11} & x_{12}^T x_{12} & \cdots & x_{12}^T x_{1n_1} & x_{12}^T x_{21} & & \cdots & x_{12}^T x_{2n_2} & x_{12}^T x_{31} & \cdots & x_{12}^T x_{3n_3} \\ \vdots & & & & & & & & & & \\ x_{1n_1}^T x_{11} & & \cdots & x_{1n_1}^T x_{1n_1} & x_{1n_1}^T x_{21} & & \cdots & x_{1n_1}^T x_{2n_2} & x_{1n_1}^T x_{31} & \cdots & x_{1n_1}^T x_{3n_3} \\ x_{21}^T x_{11} & x_{21}^T x_{12} & \cdots & x_{21}^T x_{1n_1} & x_{21}^T x_{21} & & \cdots & x_{21}^T x_{2n_2} & x_{21}^T x_{31} & \cdots & x_{21}^T x_{3n_3} \\ \vdots & \vdots & & \vdots & \vdots & & & & & & \\ x_{2n_2}^T x_{11} & x_{2n_2}^T x_{12} & \cdots & x_{2n_2}^T x_{1n_1} & x_{2n_2}^T x_{21} & & \cdots & x_{2n_2}^T x_{2n_2} & x_{2n_2}^T x_{31} & \cdots & x_{2n_2}^T x_{3n_3} \\ x_{31}^T x_{11} & x_{31}^T x_{12} & \cdots & x_{31}^T x_{1n_1} & x_{31}^T x_{21} & & \cdots & x_{31}^T x_{2n_2} & x_{31}^T x_{31} & \cdots & x_{31}^T x_{3n_3} \\ \vdots & & & & & & & & & & \\ x_{3n_3}^T x_{11} & x_{3n_3}^T x_{12} & \cdots & x_{3n_3}^T x_{1n_1} & x_{3n_3}^T x_{21} & & \cdots & x_{3n_3}^T x_{2n_2} & x_{3n_3}^T x_{31} & \cdots & x_{3n_3}^T x_{3n_3} \end{bmatrix}$ $f = \ldots$	4M+4M+4M	24min

∴ Gram matrix is,

$$G = \begin{bmatrix} \phi^T(x_{11})\,\phi(x_{11}) & \cdots & \phi^T(x_{11})\,\phi(x_{3n_3}) \\ \vdots & & \vdots \\ \phi^T(x_{3n_3})\,\phi(x_{11}) & \cdots & \phi^T(x_{3n_3})\,\phi(x_{3n_3}) \end{bmatrix}$$

$(n_1+n_2+n_3) \times (n_1+n_2+n_3)$

No need to calculate $\phi(x_{11})$, instead replaced with kernel matrix.

$$G = \begin{bmatrix} K(x_{11}, x_{11}) & \cdots & K(x_{11}, x_{3n_3}) \\ K(x_{12}, x_{11}) & & K(x_{12}, x_{3n_3}) \\ \vdots & & \vdots \\ K(x_{3n_3}, x_{11}) & \cdots & K(x_{3n_3}, x_{3n_3}) \end{bmatrix}$$

$(n_1+n_2+n_3) \times (n_1+n_2+n_3)$

60 × 60.

• Inner product matrix.