APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

CET campus, Thiruvananthapuram - 695 016

Ph: 0471 2598122; Fax: 2598522 www. ktu.edu.in Email: university@ktu.edu.in

ACADEMIC SECTION

U.O.No. 1111/2021/KTU Thiruvananthapuram, Dated: 14.07.2021

Read:-Minutes of the seventh meeting of the Board of Studies of Computer Science and Engineering, held on 02/07/2021 vide Agenda item no. 1.

ORDER

Complaints were received from faculties handling CST294-Computational Fundamentals for Machine Learning and CST284-Mathemaics for Machine Learning courses, included in the Fourth semester of B.Tech Honours and Minor Programmes respectively. The major apprehension of the concerned faculties were that time allotted as per the teaching plan were not sufficient to complete the topics.

The complaints on the above matter were placed before the Board of Studies (Computer Science & Engineering) for consideration. The seventh meeting of Board of (CSE) held on 02.07.2021 considered the matter, analysed and discussed in detail with Board of Studies members of Mathematics. After much deliberation, the Board of Studies (CSE) vide paper read above, recommended the following;

1. **Modules 5 of the existing syllabi of CST284 and CST294 are to be removed.** Some of the topics of Module 5, will be learned in semester 5 and 6, as the Minor and Honors courses are having continuity in topics, in higher semesters. Remaining topics can be included as part of 7th semester, Minor and Honors courses.
2. As the courses are running currently in the 4th semester, **no major modification is suggested for syllabus, from module 1 to module 4.**

Considering the urgency in the matter, the above resolution and revised syllabi of the aforesaid two courses, recommended by the Board of Studies(CSE), has been approved by the Hon'ble Vice Chancellor **with effect from 11.07.2021** by invoking clauses under Chapter-III, section 14(5) of the APJ Abdul Kalam Technological University Act 2014. The matter shall be reported to the next Academic Council.

The revised syllabi, approved as above, are appended herewith. The Curriculum and syllabi of B.Tech Computer Science and Engineering (2019 scheme) stands modified to the above extent.

Orders are issued accordingly.

Sd/-

Dr. Sadiq A. *

Dean (Academic)

Copy to:-

1. Principals of all colleges.
2. AD(IT) to publish in the website and website updation.
3. PS to VC/PVC, Controller, Dean(Acad), IQAC Co-ordinator/JR(Acad)/JD(Acad).
4. CE (for noting the Model Question Paper in the attached syllabi)
5. SF/FC.

Forwarded / By Order

Section Officer

* This is a computer system (Digital File) generated letter. Hence there is no need for a physical signature.

CODE CST 284	Mathematics for Machine Learning	CATEGORY	L	T	P	CREDIT
		MINOR	3	1	0	4

Preamble: This is the foundational course for awarding B. Tech. Honours in Computer Science and Engineering with specialization in ***Machine Learning***. The purpose of this course istointroducemathematicalfoundationsofbasicMachineLearningconceptsamonglearners, on which Machine Learning systems are built. This course covers Linear Algebra, Vector Calculus, Probability and Distributions, Optimization and Machine Learning problems. Concepts in this course help the learners to understand the mathematical principles in Machine Learning and aid in the creation of new Machine Learning solutions, understand & debug existing ones, and learn about the inherent assumptions & limitations of the current methodologies.

Prerequisite: A sound background in higher secondary school Mathematics.

Course Outcomes: After the completion of the course the student will be able to

CO 1	Make use of the concepts, rules and results about linear equations, matrix algebra, vector spaces, eigenvalues & eigenvectors and orthogonality & diagonalization to solve computational problems (Cognitive Knowledge Level: **Apply**)
CO 2	Perform calculus operations on functions of several variables and matrices, including partial derivatives and gradients (Cognitive Knowledge Level: **Apply**)
CO 3	Utilize the concepts, rules and results about probability, random variables, additive & multiplicative rules, conditional probability, probability distributions and Bayes' theorem to find solutions of computational problems (Cognitive Knowledge Level: **Apply**)
CO 4	Train Machine Learning Models using unconstrained and constrained optimization methods (Cognitive Knowledge Level: **Apply**)

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
CO 1	√	√	√	√								√
CO 2	√	√	√									√
CO 3	√	√	√	√								√
CO 4	√	√	√	√		√						√

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination
	1	2	
Remember	20%	20%	20%
Understand	40%	40%	40%
Apply	40%	40%	40%
Analyse			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	**50**	**100**	**3 hours**

Continuous Internal Evaluation Pattern:

Attendance : 10 marks

Continuous Assessment Tests : 25 marks

Continuous Assessment Assignment : 15 marks

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks

First Internal Examination shall be preferably conducted after completing the first half of the syllabus and the Second Internal Examination shall be preferably conducted after completing remaining part of the syllabus.

There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly covered module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly covered module), each with 7 marks. Out of the 7 questions in Part B, a student should answer any 5.

End Semester Examination Pattern: There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which student should answer anyone. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module 1

LINEAR ALGEBRA: Systems of Linear Equations – Matrices, Solving Systems of Linear Equations. Vector Spaces –Vector Spaces, Linear Independence, Basis and Rank. Linear Mappings – Matrix Representation of Linear Mappings, Basis Change, Image and Kernel.

Module 2

ANALYTIC GEOMETRY, MATRIX DECOMPOSITIONS: Norms, Inner Products, Lengths and Distances, Angles and Orthogonality, Orthonormal Basis, Orthogonal Complement, Orthogonal Projections – Projection into One Dimensional Subspaces, Projection onto General Subspaces, Gram-Schmidt Orthogonalization.

Determinant and Trace, Eigenvalues and Eigenvectors, Cholesky Decomposition, Eigen decomposition and Diagonalization, Singular Value Decomposition, Matrix Approximation.

Module 3

VECTOR CALCULUS : Differentiation of Univariate Functions - Partial Differentiation and Gradients, Gradients of Vector Valued Functions, Gradients of Matrices, Useful Identities for Computing Gradients. Back propagation and Automatic Differentiation – Gradients in Deep Network, Automatic Differentiation.Higher Order Derivatives-Linearization and Multivariate TaylorSeries.

Module 4

Probability and Distributions : Construction of a Probability Space - Discrete and Continuous Probabilities, Sum Rule, Product Rule, and Bayes' Theorem. Summary Statistics and Independence – Gaussian Distribution - Conjugacy and the Exponential Family - Change of Variables/Inverse Transform.

Module 5

Optimization : Optimization Using Gradient Descent - Gradient Descent With Momentum, Stochastic Gradient Descent. Constrained Optimization and Lagrange Multipliers - Convex Optimization - Linear Programming **-** Quadratic Programming.

Text book:

1. Mathematics for Machine Learning by Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong published by Cambridge University Press (freely available at https:// mml - book.github.io)

Reference books:

1. Linear Algebra and Its Applications, 4th Edition by Gilbert Strang
2. Linear Algebra Done Right by Axler, Sheldon, 2015 published bySpringer
3. Introduction to Applied Linear Algebra by Stephen Boyd and Lieven Vandenberghe, 2018 published by Cambridge UniversityPress
4. Convex Optimization by Stephen Boyd and Lieven Vandenberghe, 2004 published by Cambridge UniversityPress
5. Pattern Recognition and Machine Learning by Christopher M Bishop, 2006, published bySpringer
6. Learning with Kernels – Support Vector Machines, Regularization, Optimization, and Beyond by Bernhard Scholkopf and Smola, Alexander J Smola, 2002, published by MIT Press
7. Information Theory, Inference, and Learning Algorithms by David J. C MacKay, 2003 published by Cambridge UniversityPress
8. Machine Learning: A Probabilistic Perspective by Kevin P Murphy, 2012 published by MITPress.
9. The Nature of Statistical Learning Theory by Vladimir N Vapnik, 2000, published by Springer

Sample Course Level Assessment Questions.

Course Outcome 1 (CO1):

1. FindthesetSofallsolutionsinxofthefollowinginhomogeneouslinearsystems$\boldsymbol{Ax}$ $= \boldsymbol{b}$, where $\boldsymbol{A}$ and $\boldsymbol{b}$ are defined as follows:

$$\boldsymbol{A} = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 \\ 1 & 1 & 0 & -3 & 0 \\ 2 & -1 & 0 & 1 & -1 \\ -1 & 2 & 0 & -2 & -1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 3 \\ 6 \\ 5 \\ -1 \end{bmatrix}$$

2. Determine the inverses of the following matrix ifpossible

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

3. Find the characteristic equation, eigenvalues, and eigenspaces corresponding to each eigenvalue of the following matrix

$$\begin{bmatrix} 2 & 0 & 4 \\ 0 & 3 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

4. Diagonalize the following matrix, if possible

$$\begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 3 \end{bmatrix}$$

5. Find the singular value decomposition (SVD) of the following matrix

$$\begin{bmatrix} 0 & 1 & 1 \\ \sqrt{2} & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

Course Outcome 2 (CO2):

1. For a scalar function $f(x, y, z) = x^2 + 3y^2 + 2z^2$, find the gradient and its magnitude at the point **(1, 2, -1)**.
2. Find the maximum and minimum values of the function $f(x,y)=4x+4y-x^2-y^2$ subject to the condition $x^2 + y^2 <= 2$.
3. Suppose you were trying to minimize $f(x, y) = x^2 + 2y + 2y^2$. Along what vector should you travel from (5,12)?
4. Find the second order Taylor series expansion for $f(x, y) = (x + y)^2$ about **(0 ,0)**.
5. Find the critical points of $f(x, y) = x^2 - 3xy + 5x - 2y + 6y^2 + 8$.
6. Compute the gradient of the Rectified Linear Unit (ReLU) function $ReLU(z) = max(0 , z)$.
7. Let $L = ||Ax - b||^2_2$, where A is a matrix and x and b are vectors. Derive dL in terms of dx.

Course Outcome 3 (CO3):

1. Let J and T be independent events, where $P(J)=0.4$ and $P(T)=0.7$.

   i. Find $P(J \cap T)$

   ii. Find $P(J \cup T)$

   iii. Find $P(J \cap T')$

2. Let A and B be events such that $P(A)=0.45$, $P(B)=0.35$ and $P(A \cup B)=0.5$. Find $P(A|B)$.

3. A random variable **R** has the probability distribution as shown in the followingtable:

r	1	2	3	4	5
P(R=r)	0.2	a	b	0.25	0.15

   i. Given that $E(R)=2.85$, find a and b.

   ii. Find $P(R>2)$.

4. A biased coin (with probability of obtaining a head equal to $p> 0$) is tossed repeatedly and independently until the first head is observed. Compute the probability that the first head appears at an even numbered toss.

5. Two players A and B are competing at a trivia quiz game involving a series of questions. On any individual question, the probabilities that A and B give the correct answer are p and q respectively, for all questions, with outcomes for different questions being independent. The game finishes when a player wins by answering a question correctly. Compute the probability that A winsif

   i. A answers the firstquestion,

   ii. B answers the first question.

6. A coin for which $P(heads) = p$ is tossed until two successive tails are obtained. Find the probability that the experiment is completed on the n^{th} toss.

Course Outcome 4(CO4):

1. Find the extrema of $f(x, y) = x$ subject to $g(x, y) = x^2 + 2y^2 = 3$.
2. Maximize the function $f(x, y, z) = xy + yz + xz$ on the unit sphere $g(x, y, z) = x^2 + y^2 + z^2 = 1$.
3. Provide necessary and sufficient conditions under which a quadratic optimization problem be written as a linear least squaresproblem.
4. Consider the univariate function $f(x) = x^3 + 6x^2 - 3x - 5$. Find its stationary points and indicate whether they are maximum, minimum, or saddlepoints.
5. Consider the update equation for stochastic gradient descent. Write down the update when we use a mini-batch size ofone.
6. Consider the function

$$f(x) = (x_1 - x_2)^2 + \frac{1}{1 + x_1^2 + x_2^2}.$$

   i. Is $f(x)$ a convex function? Justify youranswer.

   ii. Is (1 , -1) a local/global minimum? Justify youranswer.

7. Is the function $f(x, y) = 2x^2 + y^2 + 6xy - x + 3y - 7$ convex, concave, or neither? Justify youranswer.
8. Consider the following convex optimization problem

$$\text{minimize } \frac{x^2}{2} + x + 4y^2 - 2y$$

   Subject to the constraint $x + y >= 4, x, y >= 1$.

   Derive an explicit form of the Lagrangian dual problem.

9. Solve the following LP problem with the simplexmethod.

$$max \;\; 5x_1 + 6x_2 + 9x_3 + 8x_4$$

   subject to the constraints

$$\begin{array}{ccccccccc} x_1 & + & 2x_2 & + & 3x_3 & + & x_4 & \leq & 5 \\ x_1 & + & x_2 & + & 2x_3 & + & 3x_4 & \leq & 3 \end{array}$$

$$x_1, x_2, x_3, x_4 \geq 0$$

Model Question paper

	QP Code :		**Total Pages : 5**
Reg No.:_______________		Name:___________________________	

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
IV SEMESTER B.TECH (HONOURS) DEGREE EXAMINATION, MONTH and YEAR

Course Code: CST 294

Course Name: MATHEMATICS FOR FOR MACHINE LEARNING

Max. Marks: 100		Duration: 3 Hours

PART A

		Answer all questions, each carries3 marks.	Marks
1		Show that with the usual operation of scalar multiplication but with addition on reals given by $x \# y = 2(x + y)$ is not a vector space.	
2		Are the following sets of vectors linearly independent? Explain your answer. $x_1 = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \quad x_3 = \begin{bmatrix} 3 \\ -3 \\ 8 \end{bmatrix}$	
3		Find the angle between the vectors $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $y = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.	
4		Find the eigen values of the following matrix in terms of k. Can you find an eigen vector corresponding to each of the eigen values? $\begin{bmatrix} 1 & k \\ 2 & 1 \end{bmatrix}$	
5		Let $f(x, y, z) = xye^r$, where $r = x^2+z^2-5$. Calculate the gradient of f at the point **(1, 3, -2).**	
6		Compute the Taylor polynomials T_n, $n = 0, \ldots, 5$ of $f(x) = \sin(x) + \cos(x)$ at $x_0 = 0$.	
7		Let X be a continuous random variable with probability density function on $0 <= x <= 1$ defined by $f(x) = 3x^2$. Find the pdf of $Y = X^2$.	
8		Show that if two events A and B are independent, then A and B' are independent.	
9		Explain the principle of the gradient descent algorithm.	
10		Briefly explain the difference between (batch) gradient descent and stochastic gradient descent. Give an example of when you might prefer one over the other.	

PART B

		Answer any one Question from each module. Each question carries 14 Marks	
11	a)	i.Find all solutions to the system of linearequations $$-4x + 5z = -2$$ $$-3x - 3y + 5z = 3$$ $$-x + 2y + 2z = -1$$	(4)
		ii. Prove that all vectors orthogonal to $\mathbf{[2,-3,1]^T}$ forms a subspace **W** of $\boldsymbol{R^3}$. What is ***dim (W)*** and why?	(4)
	b)	A set of $\boldsymbol{n}$ linearly independent vectors in $\boldsymbol{R^n}$ forms a basis. Does the set of vectors **(2, 4,–3),(0, 1, 1),(0, 1,–1**) form a basis for $\boldsymbol{R^3}$? Explain yourreasons.	(6)
		OR	
12	a)	Find all solutions in $x = \begin{bmatrix} x1 \\ x2 \\ x3 \end{bmatrix} \in R^3$ of the equation system $Ax = 12x$, where $A = \begin{bmatrix} 6 & 4 & 3 \\ 6 & 0 & 9 \\ 0 & 8 & 0 \end{bmatrix}$ and $\sum_{i=1}^{3} x_i = 1$.	(7)
	b)	Consider the transformation ***T (x, y) = (x + y, x + 2y, 2x + 3y)***. Obtain ***ker T*** and use this to calculate the nullity. Also find the transformation matrix for***T***.	(7)
13	a)	Use the Gramm-Schmidt process to find an orthogonal basis for the column space of the following matrix. $$\begin{bmatrix} 2 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 3 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$	(7)
	b)	Find the SVD of the matrix. $$\begin{bmatrix} 2 & 2 \\ -1 & -1 \end{bmatrix}$$	(7)
		OR	
14	a)	i. Let $\boldsymbol{L}$ be the line through the origin in $\boldsymbol{R^2}$ that is parallel to the vector $\mathbf{[3, 4]^T}$. Find the standard matrix of the orthogonal projection onto L. Also find the point on $\boldsymbol{L}$ which is closest to the point **(7, 1)** and find the point on $\boldsymbol{L}$ which is closest to the point ***(-3 , 5)***.	(6)
		ii. Find the rank-1 approximation of	

		$\begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$	
	b)	i. Find an orthonormal basis of R^3 consisting of eigenvectors for the following matrix. $\begin{bmatrix} 1 & 0 & -2 \\ 0 & 5 & 0 \\ -2 & 0 & 4 \end{bmatrix}$	(8)
		ii. Finda3×3orthogonalmatrix***S***anda3×3diagonalmatrix***D*** such that $A = SDS^T$	
15	a)	Askierisonamountainwithequation$z=100–0.4x^2–0.3y^2$,where$z$ denotes height. i. The skier is located at the point with xy-coordinates **(1, 1),** and wants to ski downhill along the steepest possible path. In which direction (indicated by a vector **(a, b**) in the xy-plane) should the skier beginskiing. **ii.** The skier begins skiing in the direction given by the xy-vector (**a, b)** you found in part (i), so the skier heads in a direction in space given by the vector **(a, b, c**). Find the value of **c**.	(8)
	b)	Find the linear approximation to the function $f(x,y) = 2 - \sin(-x - 3y)$ at the point $(0 , \pi)$, and then use your answer to estimate $f(0.001 , \pi)$.	(6)
		OR	
16	a)	Let g be the function given by $$g(x,y) = \begin{cases} \dfrac{x^2y}{x^2+y^2} & \text{if } (x,y) \neq (0,0); \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$ i. Calculate the partial derivatives of g at **(0,0)**. ii. Show that g is not differentiable at **(0,0)**.	(8)
	b)	Find the second order Taylor series expansion for $f(x,y) = e^{-(x2+y2)} \cos(xy)$ about **(0 , 0)**.	(6)
17	a)	There are two bags. The first bag contains four mangos and two apples; the second bag contains four mangos and four apples. We also have a biased coin, which shows “heads” with probability 0.6 and “tails” with probability 0.4. If the coin shows “heads”. we pick a fruitat random from bag 1;	(6)

<table>
<tr><td></td><td></td><td>otherwise we pick a fruit at random from bag 2. Your friend flips the coin (you cannot see the result), picks a fruit at random from the corresponding bag, and presents you a mango.
What is the probability that the mango was picked from bag 2?</td><td></td></tr>
<tr><td></td><td>b)</td><td>Suppose that one has written a computer program that sometimes compiles and sometimes not (code does not change). You decide to model the apparent stochasticity (success vs. no success) x of the compiler using a Bernoulli distribution with parameter μ:
$p(x \mid \mu) = \mu^x (1-\mu)^{1-x}, \quad x \in \{0,1\}$
Choose a conjugate prior for the Bernoulli likelihood and compute the posterior distribution $\boldsymbol{p(\mu \mid x_1, \ldots, x_N)}$.</td><td>(8)</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td>OR</td><td></td></tr>
<tr><td>18</td><td>a)</td><td>Two dice are rolled.
A = ‘sum of two dice equals 3’
B = ‘sum of two dice equals 7’
C = ‘at least one of the dice shows a 1’
i. What is P(A|C)?
ii. What is P(B|C)?
iii. Are A and C independent? What about B and C?</td><td>(6)</td></tr>
<tr><td></td><td>b)</td><td>Consider the following bivariate distribution p(x,y) of two discrete random variables X and Y .
<table>
<tr><td rowspan="3">Y</td><td>y_1</td><td>0.01</td><td>0.02</td><td>0.03</td><td>0.1</td><td>0.1</td></tr>
<tr><td>y_2</td><td>0.05</td><td>0.1</td><td>0.05</td><td>0.07</td><td>0.2</td></tr>
<tr><td>y_3</td><td>0.1</td><td>0.05</td><td>0.03</td><td>0.05</td><td>0.04</td></tr>
<tr><td></td><td></td><td>x_1</td><td>x_2</td><td>x_3</td><td>x_4</td><td>x_5</td></tr>
<tr><td></td><td></td><td colspan="5">X</td></tr>
</table>
Compute:
i. The marginal distributions p(x) and p(y).
ii. The conditional distributions $p(x|Y = y_1)$ and $p(y|X = x_3)$.</td><td>(8)</td></tr>
<tr><td>19</td><td>a)</td><td>Find the extrema of $\boldsymbol{f(x,y,z) = x - y + z}$ subject to $\boldsymbol{g(x,y,z) = x^2 + y^2 + z^2}$</td><td>(8)</td></tr>
</table>

		=2.	
	b)	Let $$P = \begin{bmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{bmatrix}, \; q = \begin{bmatrix} -22.0 \\ -14.5 \\ 13.0 \end{bmatrix}, \; \text{and} \;\; r = 1.$$ Show that $\boldsymbol{x^* = (1, 1/2, -1)}$ is optimal for the optimization problem $$\begin{array}{ll} \text{min} & \frac{1}{2}x^\top Px + q^\top x + r \\ \text{s.t.} & -1 \le x_i \le 1, \; i = 1, 2, 3. \end{array}$$	(6)
		OR	
20	a)	Derive the gradient descent training rule assuming that the target function is represented as $\boldsymbol{o_d = w_0 + w_1x_1 + \ldots + w_nx_n}$. Define explicitly the cost/ error function $\boldsymbol{E}$, assuming that a set of training examples $\boldsymbol{D}$ is provided, where each training example $\boldsymbol{d\ D}$ is associated with the target output $\boldsymbol{t_d}$.	(8)
	b)	Find the maximum value of $\boldsymbol{f(x,y,z) = xyz}$ given that $\boldsymbol{g(x,y,z) = x + y + z = 3}$ and $\boldsymbol{x,y,z>=0}$.	(6)
***			

Teaching Plan		
No	**Topic**	**No. of Lectures (49)**
	Module-I (LINEAR ALGEBRA)	**8**
1.1	Matrices, Solving Systems of Linear Equations	1
1.2	Vector Spaces	1
1.3	Linear Independence	1
1.4	Basis and Rank (Lecture – 1)	1
1.5	Basis and Rank (Lecture – 2)	1
1.6	Linear Mappings	1
1.7	Matrix Representation of Linear Mappings	1
1.8	Images and Kernel	1
	Module-II (ANALYTIC GEOMETRY, MATRIX DECOMPOSITIONS)	**11**
2.1	Norms, Inner Products	1
2.2	Lengths and Distances, Angles and Orthogonality	1
2.3	Orthonormal Basis, Orthogonal Complement	1
2.4	Orthogonal Projections – Projection into One Dimensional Subspaces	1
2.5	Projection onto General Subspaces.	1
2.6	Gram-Schmidt Orthogonalization	1
2.7	Determinant and Trace, Eigen values and Eigenvectors.	1

2.8	Cholesky Decomposition	1
2.9	Eigen decomposition and Diagonalization	1
2.10	Singular Value Decomposition	1
2.11	Matrix Approximation	1
	Module-III (VECTOR CALCULUS)	**9**
3.1	Differentiation of Univariate Functions, Partial Differentiation and Gradients	1
3.2	Gradients of Vector Valued Functions (Lecture 1)	1
3.3	Gradients of Vector Valued Functions (Lecture 2)	1
3.4	Gradients of Matrices	1
3.5	Useful Identities for Computing Gradients	1
3.6	Backpropagation and Automatic Differentiation – Gradients in deep Netwok	1
3.7	Automatic Differentiation	1
3.8	Higher Order Derivatives	1
3.9	Linearization and Multivariate Taylor Series	1
	Module-IV (Probability and Distributions)	**10**
4.1	Construction of a Probability Space	1
4.2	Discrete and Continuous Probabilities (Probability Density Function, Cumulative Distribution Function)	1
4.3	Sum Rule, Product Rule	1
4.4	Bayes' Theorem	1
4.5	Summary Statistics and Independence (Lecture 1)	1
4.6	Summary Statistics and Independence (Lecture 2)	1
4.7	Bernoulli, Binomial, Uniform (Discrete) Distributions	1
4.8	Uniform (Continuous), Poisson Distributions	1
4.9	Gaussian Distribution	1

4.10	Conjugacy and the Exponential Family (Beta – Bernoulli, Beta – Binomial Conjugacies)	1
	Module-V (Optimization)	7
5.1	Optimization Using Gradient Descent.	1
5.2	Gradient Descent With Momentum, Stochastic Gradient Descent	1
5.3	Constrained Optimization and Lagrange Multipliers (Lecture 1)	1
5.4	Constrained Optimization and Lagrange Multipliers (Lecture 2)	1
5.5	Convex Optimization	1
5.6	Linear Programming	1
5.7	Quadratic Programming	1

CODE CST 294	Computational Fundamentals for Machine Learning	CATEGORY	L	T	P	CREDIT
		HONOURS	3	1	0	4

Preamble: This is the foundational course for awarding B. Tech. Honours in Computer Science and Engineering with specialization in ***Machine Learning***. The purpose of this course istointroducemathematicalfoundationsofbasicMachineLearningconceptsamonglearners, on which Machine Learning systems are built. This course covers Linear Algebra, Vector Calculus, Probability and Distributions, Optimization and Machine Learning problems. Concepts in this course help the learners to understand the mathematical principles in Machine Learning and aid in the creation of new Machine Learning solutions, understand & debug existing ones, and learn about the inherent assumptions & limitations of the current methodologies.

Prerequisite: A sound background in higher secondary school Mathematics.

Course Outcomes: After the completion of the course the student will be able to

CO 1	Make use of the concepts, rules and results about linear equations, matrix algebra, vector spaces, eigenvalues & eigenvectors and orthogonality & diagonalization to solve computational problems (Cognitive Knowledge Level: **Apply**)
CO 2	Perform calculus operations on functions of several variables and matrices, including partial derivatives and gradients (Cognitive Knowledge Level: **Apply**)
CO 3	Utilize the concepts, rules and results about probability, random variables, additive & multiplicative rules, conditional probability, probability distributions and Bayes' theorem to find solutions of computational problems (Cognitive Knowledge Level: **Apply**)
CO 4	Train Machine Learning Models using unconstrained and constrained optimization methods (Cognitive Knowledge Level: **Apply**)

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
CO 1	√	√	√	√								√
CO 2	√	√	√									√
CO 3	√	√	√	√								√
CO 4	√	√	√	√		√						√

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination
	1	2	
Remember	20%	20%	20%
Understand	40%	40%	40%
Apply	40%	40%	40%
Analyse			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	**50**	**100**	**3 hours**

Continuous Internal Evaluation Pattern:

Attendance : 10 marks

Continuous Assessment Tests : 25 marks

Continuous Assessment Assignment : 15 marks

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks

First Internal Examination shall be preferably conducted after completing the first half of the syllabus and the Second Internal Examination shall be preferably conducted after completing remaining part of the syllabus.

There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly covered module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly covered module), each with 7 marks. Out of the 7 questions in Part B, a student should answer any 5.

End Semester Examination Pattern: There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which student should answer anyone. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module 1

LINEAR ALGEBRA: Systems of Linear Equations – Matrices, Solving Systems of Linear Equations. Vector Spaces –Vector Spaces, Linear Independence, Basis and Rank. Linear Mappings – Matrix Representation of Linear Mappings, Basis Change, Image and Kernel.

Module 2

ANALYTIC GEOMETRY, MATRIX DECOMPOSITIONS: Norms, Inner Products, Lengths and Distances, Angles and Orthogonality, Orthonormal Basis, Orthogonal Complement, Orthogonal Projections – Projection into One Dimensional Subspaces, Projection onto General Subspaces, Gram-Schmidt Orthogonalization.

Determinant and Trace, Eigenvalues and Eigenvectors, Cholesky Decomposition, Eigen decomposition and Diagonalization, Singular Value Decomposition, Matrix Approximation.

Module 3

VECTOR CALCULUS : Differentiation of Univariate Functions - Partial Differentiation and Gradients, Gradients of Vector Valued Functions, Gradients of Matrices, Useful Identities for Computing Gradients. Back propagation and Automatic Differentiation – Gradients in Deep Network, Automatic Differentiation.Higher Order Derivatives-Linearization and Multivariate TaylorSeries.

Module 4

Probability and Distributions : Construction of a Probability Space - Discrete and Continuous Probabilities, Sum Rule, Product Rule, and Bayes' Theorem. Summary Statistics and Independence – Gaussian Distribution - Conjugacy and the Exponential Family - Change of Variables/Inverse Transform.

Module 5

Optimization : Optimization Using Gradient Descent - Gradient Descent With Momentum, Stochastic Gradient Descent. Constrained Optimization and Lagrange Multipliers - Convex Optimization - Linear Programming **-** Quadratic Programming.

Text book:

1. Mathematics for Machine Learning by Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong published by Cambridge University Press (freely available at https://mml - book.github.io)

Reference books:

1. Linear Algebra and Its Applications, 4th Edition by Gilbert Strang
2. Linear Algebra Done Right by Axler, Sheldon, 2015 published bySpringer
3. Introduction to Applied Linear Algebra by Stephen Boyd and Lieven Vandenberghe, 2018 published by Cambridge UniversityPress
4. Convex Optimization by Stephen Boyd and Lieven Vandenberghe, 2004 published by Cambridge UniversityPress
5. Pattern Recognition and Machine Learning by Christopher M Bishop, 2006, published bySpringer
6. Learning with Kernels – Support Vector Machines, Regularization, Optimization, and Beyond by Bernhard Scholkopf and Smola, Alexander J Smola, 2002, published by MIT Press
7. Information Theory, Inference, and Learning Algorithms by David J. C MacKay, 2003 published by Cambridge UniversityPress
8. Machine Learning: A Probabilistic Perspective by Kevin P Murphy, 2012 published by MITPress.
9. The Nature of Statistical Learning Theory by Vladimir N Vapnik, 2000, published by Springer

Sample Course Level Assessment Questions.

Course Outcome 1 (CO1):

1. FindthesetSofallsolutionsinxofthefollowinginhomogeneouslinearsystems$\boldsymbol{Ax} = \boldsymbol{b}$, where $\boldsymbol{A}$ and $\boldsymbol{b}$ are defined as follows:

$$\boldsymbol{A} = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 \\ 1 & 1 & 0 & -3 & 0 \\ 2 & -1 & 0 & 1 & -1 \\ -1 & 2 & 0 & -2 & -1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 3 \\ 6 \\ 5 \\ -1 \end{bmatrix}$$

2. Determine the inverses of the following matrix ifpossible

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

3. Find the characteristic equation, eigenvalues, and eigenspaces corresponding to each eigenvalue of the following matrix

$$\begin{bmatrix} 2 & 0 & 4 \\ 0 & 3 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

4. Diagonalize the following matrix, if possible

$$\begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 3 \end{bmatrix}$$

5. Find the singular value decomposition (SVD) of the following matrix

$$\begin{bmatrix} 0 & 1 & 1 \\ \sqrt{2} & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

Course Outcome 2 (CO2):

1. For a scalar function $f(x, y, z) = x^2 + 3y^2 + 2z^2$, find the gradient and its magnitude at the point **(1, 2, -1)**.
2. Find the maximum and minimum values of the function $f(x,y) = 4x + 4y - x^2 - y^2$ subject to the condition $x^2 + y^2 <= 2$.
3. Suppose you were trying to minimize $f(x, y) = x^2 + 2y + 2y^2$. Along what vector should you travel from (5,12)?
4. Find the second order Taylor series expansion for $f(x, y) = (x + y)^2$ about **(0 ,0)**.
5. Find the critical points of $f(x, y) = x^2 - 3xy + 5x - 2y + 6y^2 + 8$.
6. Compute the gradient of the Rectified Linear Unit (ReLU) function $ReLU(z) = max(0, z)$.
7. Let $L = \|Ax - b\|^2_2$, where A is a matrix and x and b are vectors. Derive dL in terms of dx.

Course Outcome 3 (CO3):

1. Let J and T be independent events, where $P(J)=0.4$ and $P(T)=0.7$.

    i. Find $P(J \cap T)$

    ii. Find $P(J \square T)$

    iii. Find $P(J \cap T')$

2. Let A and B be events such that $P(A)=0.45$, $P(B)=0.35$ and $P(A \cup B)=0.5$. Find $P(A|B)$.

3. A random variable **R** has the probability distribution as shown in the followingtable:

r	1	2	3	4	5
P(R=r)	0.2	a	b	0.25	0.15

    i. Given that $E(R)=2.85$, find a and b.

    ii. Find $P(R>2)$.

4. A biased coin (with probability of obtaining a head equal to $p> 0$) is tossed repeatedly and independently until the first head is observed. Compute the probability that the first head appears at an even numbered toss.

5. Two players A and B are competing at a trivia quiz game involving a series of questions. On any individual question, the probabilities that A and B give the correct answer are p and q respectively, for all questions, with outcomes for different questions being independent. The game finishes when a player wins by answering a question correctly. Compute the probability that A winsif

    i. A answers the firstquestion,

    ii. B answers the first question.

6. A coin for which $P(heads) = p$ is tossed until two successive tails are obtained. Find the probability that the experiment is completed on the n^{th} toss.

Course Outcome 4(CO4):

1. Find the extrema of $f(x, y) = x$ subject to $g(x, y) = x^2 + 2y^2 = 3$.
2. Maximize the function $f(x, y, z) = xy + yz + xz$ on the unit sphere $g(x, y, z) = x^2 + y^2 + z^2 = 1$.
3. Provide necessary and sufficient conditions under which a quadratic optimization problem be written as a linear least squaresproblem.
4. Consider the univariate function $f(x) = x^3 + 6x^2 - 3x - 5$. Find its stationary points and indicate whether they are maximum, minimum, or saddlepoints.
5. Consider the update equation for stochastic gradient descent. Write down the update when we use a mini-batch size ofone.
6. Consider the function

$$f(x) = (x_1 - x_2)^2 + \frac{1}{1 + x_1^2 + x_2^2}.$$

   i. Is $f(x)$ a convex function? Justify youranswer.

   ii. Is (1 , -1) a local/global minimum? Justify youranswer.

7. Is the function $f(x, y) = 2x^2 + y^2 + 6xy - x + 3y - 7$ convex, concave, or neither? Justify youranswer.
8. Consider the following convex optimization problem

$$\text{minimize } \frac{x^2}{2} + x + 4y^2 - 2y$$

   Subject to the constraint $x + y >= 4, x, y >= 1$.

   Derive an explicit form of the Lagrangian dual problem.

9. Solve the following LP problem with the simplexmethod.

$$max \;\; 5x_1 + 6x_2 + 9x_3 + 8x_4$$

   subject to the constraints

$$\begin{array}{ccccccccc} x_1 & + & 2x_2 & + & 3x_3 & + & x_4 & \leq & 5 \\ x_1 & + & x_2 & + & 2x_3 & + & 3x_4 & \leq & 3 \end{array}$$

$$x_1, x_2, x_3, x_4 \geq 0$$

Model Question paper

	QP Code :		Total Pages : 5
Reg No.:_______________		Name:___________________________	

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY IV SEMESTER B.TECH (HONOURS) DEGREE EXAMINATION, MONTH and YEAR
Course Code: CST 294
Course Name: COMPUTATIONAL FUNDAMENTALS FOR MACHINE LEARNING

Max. Marks: 100		Duration: 3 Hours

PART A

		Answer all questions, each carries3 marks.	Marks
1		Show that with the usual operation of scalar multiplication but with addition on reals given by ***x # y = 2(x + y)*** is not a vector space.	
2		Are the following sets of vectors linearly independent? Explain your answer. $x_1 = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \quad x_3 = \begin{bmatrix} 3 \\ -3 \\ 8 \end{bmatrix}$	
3		Find the angle between the vectors $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $y = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.	
4		Find the eigen values of the following matrix in terms of k. Can you find an eigen vector corresponding to each of the eigen values? $\begin{bmatrix} 1 & k \\ 2 & 1 \end{bmatrix}$	
5		Let $f(x, y, z) = xye^{r}$, where $r = x^2+z^2-5$. Calculate the gradient of f at the point **(1, 3, -2).**	
6		Compute the Taylor polynomials T_n, $n = 0, \dots, 5$ of $f(x) = \sin(x) + \cos(x)$ at $x_0 = 0$.	
7		Let X be a continuous random variable with probability density function on $0 <= x <= 1$ defined by $f(x) = 3x^2$. Find the pdf of $Y = X^2$.	
8		Show that if two events A and B are independent, then A and B' are independent.	
9		Explain the principle of the gradient descent algorithm.	
10		Briefly explain the difference between (batch) gradient descent and stochastic gradient descent. Give an example of when you might prefer one over the other.	

PART B

		Answer any one Question from each module. Each question carries 14 Marks	
11	a)	i.Find all solutions to the system of linearequations $$-4x + 5z = -2$$ $$-3x - 3y + 5z = 3$$ $$-x + 2y + 2z = -1$$	(4)
		ii. Prove that all vectors orthogonal to $\mathbf{[2,-3,1]^T}$ forms a subspace **W** of $\boldsymbol{R^3}$. What is ***dim (W)*** and why?	(4)
	b)	A set of $\boldsymbol{n}$ linearly independent vectors in $\boldsymbol{R^n}$ forms a basis. Does the set of vectors **(2, 4,–3),(0, 1, 1),(0, 1,–1**) form a basis for $\boldsymbol{R^3}$? Explain yourreasons.	(6)
		OR	
12	a)	Find all solutions in $x = \begin{bmatrix} x1 \\ x2 \\ x3 \end{bmatrix} \in R^3$ of the equation system $Ax = 12x$, where $A = \begin{bmatrix} 6 & 4 & 3 \\ 6 & 0 & 9 \\ 0 & 8 & 0 \end{bmatrix}$ and $\sum_{i=1}^{3} x_i = 1$.	(7)
	b)	Consider the transformation ***T (x, y) = (x + y, x + 2y, 2x + 3y)***. Obtain ***ker T*** and use this to calculate the nullity. Also find the transformation matrix for***T***.	(7)
13	a)	Use the Gramm-Schmidt process to find an orthogonal basis for the column space of the following matrix. $\begin{bmatrix} 2 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 3 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	(7)
	b)	Find the SVD of the matrix. $$\begin{bmatrix} 2 & 2 \\ -1 & -1 \end{bmatrix}$$	(7)
		OR	
14	a)	i. Let $\boldsymbol{L}$ be the line through the origin in $\boldsymbol{R^2}$ that is parallel to the vector $\mathbf{[3, 4]^T}$. Find the standard matrix of the orthogonal projection onto L. Also find the point on $\boldsymbol{L}$ which is closest to the point **(7, 1)** and find the point on $\boldsymbol{L}$ which is closest to the point ***(-3 , 5)***.	(6)
		ii. Find the rank-1 approximation of	

		$\begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$	
	b)	i. Find an orthonormal basis of $\boldsymbol{R^3}$ consisting of eigenvectors for the following matrix. $\begin{bmatrix} 1 & 0 & -2 \\ 0 & 5 & 0 \\ -2 & 0 & 4 \end{bmatrix}$	(8)
		ii. Finda3×3orthogonalmatrix***S***anda3×3diagonalmatrix***D*** such that $A = SDS^T$	
15	a)	Askierisonamountainwithequation$z=100-0.4x^2-0.3y^2$,where$z$ denotes height. i. The skier is located at the point with xy-coordinates **(1, 1),** and wants to ski downhill along the steepest possible path. In which direction (indicated by a vector **(a, b**) in the xy-plane) should the skier beginskiing. **ii.** The skier begins skiing in the direction given by the xy-vector (**a, b)** you found in part (i), so the skier heads in a direction in space given by the vector **(a, b, c**). Find the value of **c**.	(8)
	b)	Find the linear approximation to the function $f(x,y) = 2 - \sin(-x - 3y)$ at the point $(0 , \pi)$, and then use your answer to estimate $f(0.001 , \pi)$.	(6)
		OR	
16	a)	Let g be the function given by $g(x,y) = \begin{cases} \dfrac{x^2y}{x^2+y^2} & \text{if } (x,y) \neq (0,0); \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$ i. Calculate the partial derivatives of g at **(0,0)**. ii. Show that g is not differentiable at **(0,0)**.	(8)
	b)	Find the second order Taylor series expansion for $f(x,y) = e^{-(x2+y2)} \cos(xy)$ about **(0 , 0)**.	(6)
17	a)	There are two bags. The first bag contains four mangos and two apples; the second bag contains four mangos and four apples. We also have a biased coin, which shows "heads" with probability 0.6 and "tails" with probability 0.4. If the coin shows "heads". we pick a fruitat random from bag 1;	(6)

<table>
<tr><td></td><td></td><td>otherwise we pick a fruit at random from bag 2. Your friend flips the coin (you cannot see the result), picks a fruit at random from the corresponding bag, and presents you a mango.
What is the probability that the mango was picked from bag 2?</td><td></td></tr>
<tr><td></td><td>b)</td><td>Suppose that one has written a computer program that sometimes compiles and sometimes not (code does not change). You decide to model the apparent stochasticity (success vs. no success) x of the compiler using a Bernoulli distribution with parameter μ:
$p(x \mid \mu) = \mu^{x}(1-\mu)^{1-x}, \quad x \in \{0,1\}$
Choose a conjugate prior for the Bernoulli likelihood and compute the posterior distribution $\boldsymbol{p(\mu \mid x_1, \ldots, x_N)}$.</td><td>(8)</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td>OR</td><td></td></tr>
<tr><td>18</td><td>a)</td><td>Two dice are rolled.
A = ‘sum of two dice equals 3’
B = ‘sum of two dice equals 7’
C = ‘at least one of the dice shows a 1’
i. What is P(A|C)?
ii. What is P(B|C)?
iii. Are A and C independent? What about B and C?</td><td>(6)</td></tr>
<tr><td></td><td>b)</td><td>Consider the following bivariate distribution p(x,y) of two discrete random variables X and Y .
<table>
<tr><td>Y</td><td>y_1</td><td>0.01</td><td>0.02</td><td>0.03</td><td>0.1</td><td>0.1</td></tr>
<tr><td></td><td>y_2</td><td>0.05</td><td>0.1</td><td>0.05</td><td>0.07</td><td>0.2</td></tr>
<tr><td></td><td>y_3</td><td>0.1</td><td>0.05</td><td>0.03</td><td>0.05</td><td>0.04</td></tr>
<tr><td></td><td></td><td>x_1</td><td>x_2</td><td>x_3</td><td>x_4</td><td>x_5</td></tr>
<tr><td></td><td></td><td></td><td></td><td>X</td><td></td><td></td></tr>
</table>
Compute:
i. The marginal distributions p(x) and p(y).
ii. The conditional distributions $p(x \mid Y = y_1)$ and $p(y \mid X = x_3)$.</td><td>(8)</td></tr>
<tr><td>19</td><td>a)</td><td>Find the extrema of $\boldsymbol{f(x,y,z) = x - y + z}$ subject to $\boldsymbol{g(x,y,z) = x^2 + y^2 + z^2}$</td><td>(8)</td></tr>
</table>

		=2.	
	b)	Let $$P = \begin{bmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{bmatrix}, \quad q = \begin{bmatrix} -22.0 \\ -14.5 \\ 13.0 \end{bmatrix}, \quad \text{and} \quad r = 1.$$ Show that $x^* = (1, 1/2, -1)$ is optimal for the optimization problem $$\begin{array}{ll} \min & \frac{1}{2}x^{\mathsf{T}}Px + q^{\mathsf{T}}x + r \\ \text{s.t.} & -1 \leq x_i \leq 1, \ i = 1, 2, 3. \end{array}$$	(6)
		OR	
20	a)	Derive the gradient descent training rule assuming that the target function is represented as $o_d = w_0 + w_1x_1 + \ldots + w_nx_n$. Define explicitly the cost/ error function E, assuming that a set of training examples D is provided, where each training example d D is associated with the target output t_d.	(8)
	b)	Find the maximum value of $f(x,y,z) = xyz$ given that $g(x,y,z) = x + y + z = 3$ and $x,y,z>=0$.	(6)
***			

Teaching Plan		
No	**Topic**	**No. of Lectures (49)**
	Module-I (LINEAR ALGEBRA)	**8**
1.1	Matrices, Solving Systems of Linear Equations	1
1.2	Vector Spaces	1
1.3	Linear Independence	1
1.4	Basis and Rank (Lecture – 1)	1
1.5	Basis and Rank (Lecture – 2)	1
1.6	Linear Mappings	1
1.7	Matrix Representation of Linear Mappings	1
1.8	Images and Kernel	1
	Module-II (ANALYTIC GEOMETRY, MATRIX DECOMPOSITIONS)	**11**
2.1	Norms, Inner Products	1
2.2	Lengths and Distances, Angles and Orthogonality	1
2.3	Orthonormal Basis, Orthogonal Complement	1
2.4	Orthogonal Projections – Projection into One Dimensional Subspaces	1
2.5	Projection onto General Subspaces.	1
2.6	Gram-Schmidt Orthogonalization	1
2.7	Determinant and Trace, Eigen values and Eigenvectors.	1

2.8	Cholesky Decomposition	1
2.9	Eigen decomposition and Diagonalization	1
2.10	Singular Value Decomposition	1
2.11	Matrix Approximation	1
	Module-III (VECTOR CALCULUS)	**9**
3.1	Differentiation of Univariate Functions, Partial Differentiation and Gradients	1
3.2	Gradients of Vector Valued Functions (Lecture 1)	1
3.3	Gradients of Vector Valued Functions (Lecture 2)	1
3.4	Gradients of Matrices	1
3.5	Useful Identities for Computing Gradients	1
3.6	Backpropagation and Automatic Differentiation – Gradients in deep Netwok	1
3.7	Automatic Differentiation	1
3.8	Higher Order Derivatives	1
3.9	Linearization and Multivariate Taylor Series	1
	Module-IV (Probability and Distributions)	**10**
4.1	Construction of a Probability Space	1
4.2	Discrete and Continuous Probabilities (Probability Density Function, Cumulative Distribution Function)	1
4.3	Sum Rule, Product Rule	1
4.4	Bayes' Theorem	1
4.5	Summary Statistics and Independence (Lecture 1)	1
4.6	Summary Statistics and Independence (Lecture 2)	1
4.7	Bernoulli, Binomial, Uniform (Discrete) Distributions	1
4.8	Uniform (Continuous), Poisson Distributions	1
4.9	Gaussian Distribution	1

4.10	Conjugacy and the Exponential Family (Beta – Bernoulli, Beta – Binomial Conjugacies)	1
	Module-V (Optimization)	7
5.1	Optimization Using Gradient Descent.	1
5.2	Gradient Descent With Momentum, Stochastic Gradient Descent	1
5.3	Constrained Optimization and Lagrange Multipliers (Lecture 1)	1
5.4	Constrained Optimization and Lagrange Multipliers (Lecture 2)	1
5.5	Convex Optimization	1
5.6	Linear Programming	1
5.7	Quadratic Programming	1