# BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
## WORK INTEGRATED LEARNING PROGRAMMES

# COURSE HANDOUT

# Part A: Content Design

| | |
|---|---|
| **Course Title** | Mathematical Foundations for Machine Learning |
| **Course No(s)** | AIMLC ZC416 |
| **Credit Units** | 4 |
| **Course Author** | Srinath Naidu |
| **Version No** | 3 |
| **Date** | 01.08.2022 |
| **Lead Instructor** | Srinath Naidu |

**Course Description**

Vector and matrix algebra, systems of linear algebraic equations and their solutions; Eigenvalues, eigenvectors and diagonalization of matrices; multivariate calculus, vector calculus, Jacobian and Hessian, multivariate Taylor series, gradient descent, unconstrained optimization, constrained optimization, nonlinear optimization, stochastic gradient descent, dimensionality reduction and PCA, optimization for support vector machines.

**Course Objectives**

| No | Objective- The course aims to |
|---|---|
| **CO1** | Introduce concepts in linear algebra and to use it as a platform for machine learning. |
| **CO2** | Provide techniques for analytical and numerical solutions of linear equations. |
| **CO3** | Introduce concepts in multivariate calculus and vector calculus, and introduce multivariate Taylor series. |
| **CO4** | Introduce concepts of gradient descent, stochastic gradient descent and the concepts of unconstrained, constrained and nonlinear optimization. |
| **CO5** | Introduce the concepts of dimensionality reduction, principal components analysis, and optimization for support vector machines. |

**Text Book(s)**

| No | Author(s), Title, Edition, Publishing House |
|---|---|
| T1 | M.P. Diesenroth, A. Aldo Faisal, Cheng Soon Ong, Mathematics for Machine Learning, Cambridge University Press, 2020. |
| T2 | Charu C. Aggarwal, Linear Algebra and Optimization, Springer Nature Switzerland AG, 2020 |

**Reference Book(s) & other resources**

| No | Author(s), Title, Edition, Publishing House |
|---|---|
| R1 | K Hoffman and R Kunze, Linear Algebra, Pearson Education, 2nd Edition, 2005. |

| | R2 | Erwin Kreyszig, Advanced Engineering Mathematics, Wiley India, 10th Edition, 2015 (earlier editions are also okay) |
|---|---|---|

## Content Structure

| No | Title of the module | References |
|---|---|---|
| M1 | Solution of linear systems – systems of linear equations, matrices, solving systems of linear equations. | T1: Sec 2.1, 2.2, 2.3 |
| M2 | Vectors Spaces -  linear independence, basis and rank, affine spaces, Norms, inner products, Lengths and distances, Angles and orthogonality, Orthonormal basis | T1: Sec 2.4, 2.5, 2.6, 2.7, 2.8, 3.1, 3.2, 3.3, 3.4, 3.5 |
| M3 | Matrix Decomposition methods - Determinant and Trace, Eigenvalues and Eigenvectors, Cholesky decomposition, Eigen-decomposition and diagonalization, singular value decomposition, matrix approximation | T1: Sec 4.1, 4,2, 4.3, 4.4, 4.5, 4.6 |
| M4 | Vector Calculus  - Differentiation of univariate functions, Partial differentiation and gradients, Gradients of vector-valued functions , Gradients of matrices, Some useful identities for computing gradients, Backpropagation and automatic differentiation | T1: Sec 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8 |
| M5 |  Continuous Optimization - Optimization using gradient descent, Constrained optimization and Lagrange multipliers, Convex optimization | T1: 7.1, 7.2, 7.3 |
| M6 | Nonlinear Optimization - Minutiae of Gradient Descent – learning rate decay, initialization,  Properties of optimization in learning – typical objective functions, stochastic gradient descent, how optimization in machine learning is different, tuning hyperparameters, importance of feature pre-processing, Challenges in Gradient-based optimization, local optima and flat regions, differential curvature, examples of difficult topologies like cliffs and valleys, adjusting first-order derivatives for descent, momentum-based learning, AdaGrad, RMSProp, Adam | T2: 4.4, 4.5, 5.2, 5.3 |
| M7 | Dimensionality reduction and PCA – problem  setting, maximum variance perspective, projection perspective, eigenvector and low-rank approximations, PCA in high dimensions, key steps of PCA in practice, latent variable perspective, Mathematical preliminaries of SVM, primal/dual perspective for SVM, nonlinear SVM - kernels. | T1: 10.1, 10.2, 10.3, 10.4, 10.5, 10.6, 10.7, 12.1, 12.2, 12.3, 12.4, 12.5, 12.6, T2: Sec 6.4 <br> Class Notes |

## Learning Outcomes:

| No | Learning Outcomes |
|---|---|
| LO1 | Students will be able to effectively use matrix algebra tools to analyze and solve systems of linear equations. |
| LO2 | Students will be able to learn concepts of linear algebra that form the foundation of data science problems. |

| | |
|---|---|
| LO3 | Students would be able to learn the basic techniques of constrained and unconstrained optimization and understand how these are applied in the context of support vector machines. |
| LO4 | Students will be able to learn the concepts of dimensionality reduction, and principal components analysis. |

# Part B: Contact Session Plan

| | |
|---|---|
| **Academic Term** | Second Semester 2022-2023 |
| **Course Title** | Mathematical Foundations for Machine Learning |
| **Course No** | AIMLC ZC416 |
| **Lead Instructor** | Srinath Naidu |

## *Course Contents*

| Contact Hours | List of Topic Title | Text/Ref Book/external resource |
|---|---|---|
| 1 | Solution of linear systems – systems of linear equations, matrices, solving systems of linear equations | T1: Sec 2.1, 2.2, 2.3 |
| 2 | Vectors Spaces, linear independence, basis and rank, affine spaces | T1: Sec 2.4, 2.5, 2.6, 2.8 |
| 3 | Analytic Geometry – norms, inner products, lengths and distances, angles and orthogonality, orthonormal basis | T1: Sec 3.1, 3.2, 3.3, 3.4, 3.5 |
| 4 | Matrix Decomposition – I<br>Determinant and Trace<br>Eigenvalues and Eigenvectors<br>Cholesky Decomposition | T1: Sec 4.1, 4.2, 4.3 |
| 5 | Matrix Decomposition – II | T1: Sec 4.4, 4.5, 4.6 |

| | | |
|---|---|---|
| | Eigen-decomposition and Diagonalization<br>Singular Value Decomposition<br>Matrix Approximation | |
| 6 | Vector Calculus – I<br>    Differentiation of univariate functions<br>    Partial differentiation and gradients<br>    Gradients of vector-valued functions | T1: Sec 5.1, 5.2, 5.3 |
| 7 | Vector Calculus – II<br>    Gradients of matrices<br>    Some useful identities for computing gradients<br>    Backpropagation and automatic differentiation | T1: Sec 5.4, 5.5, 5.6 |
| 8 | Vector Calculus – III<br>    Higher-order derivatives<br>    Linearization and multivariate Taylor's series<br>    Computing maxima and minima for unconstrained optimization | T1: 5.7, 5.8<br>Class Notes |
| 9 | Continuous Optimization<br>    Optimization using gradient descent<br>    Constrained optimization and Lagrange multipliers<br>    Convex optimization | T1: Sec 7.1, 7.2, 7.3, |
| 10 | Nonlinear Optimization- I<br>    Minutiae of Gradient Descent – learning rate decay, initialization<br>    Properties of optimization in learning – typical objective functions, stochastic gradient descent, how optimization in machine learning is different, tuning hyperparameters, importance of feature pre-processing | T2: Sec 4.4, Sec 4.5 |
| 11 | Nonlinear Optimization- II<br>    Challenges in Gradient-based optimization, local optima and flat regions, differential curvature, examples of difficult topologies like cliffs and valleys.<br>    Adjusting first-order derivatives for descent, Momentum-based learning, AdaGrad, RMSProp, Adam | T2: Sec  5.2, 5.3 |

| 12 | Dimensionality Reduction and PCA – I<br>    Problem setting<br>    Maximum variance perspective<br>    Projection perspective | T1: Sec 10.1, 10.2, 10.3 |
|---|---|---|
| 13 | Dimensionality Reduction and PCA – II<br>    Eigenvector Computation and low-rank approximation<br>    PCA in high dimensions<br>    Key steps of PCA in practice<br>    Latent variable perspective | T1: 10.4, 10.5, 10.6, 10.7 |
| 14 | Mathematical preliminaries for SVM<br>    Karash-Kuhn-Tucker conditions | T2: Sec 6.4 |
| 15 | Primal/dual perspective for linear SVM | T1: 12.1, 12.2, 12.3, 12.4 |
| 16 | Nonlinear SVM (Expert lecture)<br>    kernels<br>    examples | Class Notes |

*# The above contact hours and topics can be adapted for non-specific and specific WILP programs depending on the requirements and class interests.*

## *Lab Details*

| Title | Access URL |
|---|---|
| Lab Setup Instructions | Not applicable |
| Lab Capsules | Not applicable |
| Additional References | Not applicable |

## Select Topics and Case Studies from business for experiential learning

| Topic No. | Select Topics in Syllabus for experiential learning | Access URL |
|-----------|------------------------------------------------------|------------|
| 1 | Assignment - linear algebra topics | |
| 2 | Assignment- Vector Calculus/Optimization based topics | |

## Evaluation Scheme

Legend: EC = Evaluation Component

| No | Name | Type | Duration | Weight | Day, Date, Session, Time |
|----|------|------|----------|--------|--------------------------|
| 1 | Assignment 1 | Online | | 10% | To be announced |
| 2 | Assignment 2 | Online | | 10% | To be announced |
| 3 | Quiz 1 | Online | * | 5% | To be announced |
| 4 | Quiz 2 | Online | * | 5% | To be announced |
| 5 | Mid-Semester Exam | Closed book | 90 min | 30% | To be announced |
| 6 | Comprehensive Exam | Open book | 150 min | 40% | To be announced |

## Important Information

Syllabus for Mid-Semester Test (Closed Book): Topics in Weeks 1-8

Syllabus for Comprehensive Exam (Open Book): All topics (in sessions 1 to 16) given in plan of study

Evaluation Guidelines:

1. EC-1 consists of two Assignments and two Quizzes (best two out of the three would be taken for grading). Announcements regarding the same will be made in a timely manner.

2. For Closed Book tests: No books or reference material of any kind will be permitted. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.

3. For Open Book exams: Use of prescribed and reference text books, in original (not photocopies) is permitted. Class notes/slides as reference material in filed or bound form is

permitted. However, loose sheets of paper will not be allowed. Use of calculators is permitted in all exams. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.

4. If a student is unable to appear for the Regular Test/Exam due to genuine exigencies, the student should follow the procedure to apply for the Make-Up Test/Exam. The genuineness of the reason for absence in the Regular Exam shall be assessed prior to giving permission to appear for the Make-up Exam. Make-Up Test/Exam will be conducted only at selected exam centres on the dates to be announced later.

It shall be the responsibility of the individual student to be regular in maintaining the self-study schedule as given in the course handout, attend the lectures, and take all the prescribed evaluation components such as Assignment/Quiz, Mid-Semester Test and Comprehensive Exam according to the evaluation scheme provided in the handout.