Curriculum for
M. Tech in Data Science (MDS)

Overview

We are proposing a curriculum below which will be a modern effective and sustainable course for M. Tech degree in Data Science (MDS).

The program is going to be jointly contributed by faculty members from CSE, EE, MA, CE, ME, HSE. The curriculum contains core courses and electives along with capstone project. Electives will help students to pick appropriate track for their specialization. Broad distribution of credits are as follows.

Serial Number	**Category** (Number of items)	**Credits**
1	Core Theory (7)	22
2	Core Lab (3)	7
3	Electives (*)	9
4	Projects	20
5	HSS	0
	Total	58

*Can be divided in number of courses with 1 to 4 credits

Semester	**Credits**
First	20
Second	15
Third	13
Fourth	10
Total	58

Curriculum

Semester 1

Course	**Credit**	**Type**	**Faculty**
Probability and Statistics	4	Core	MA
Linear Algebra	3	Core	EE
Machine Learning	3	Core	CS
Machine Learning Lab	2	Core lab	CS
Data Engineering	3	Core	CS
Data Engineering Lab	2	Core Lab	CS
Optimization	3	Core	CS
Total	20		

Semester 2

Course	**Credit**	**Type**	**Faculty**
Big Data Lab	3	Core Lab	CS
Deep Learning	3	Core	CS
Cyber Security	3	Core	CS
Electives	6	Elective	CS/EE/MA/CE/ME
Total	15		

Semester 3

Course	**Credit**	**Type**	**Faculty**
Communication Skill and Technical Writing	0	HS Core	HSE
Project (phase 1)	10	Core	CS/EE/MA/CE
Electives	3	Elective	CS/EE/MA/CE
Total	13		

Semester 4

Course	**Credit**	**Type**	**Faculty**
Research Methodologies and Professional Ethics	0	HS Core	HSE
Project (phase 2)	10	Core	CS/EE/MA/CE
Total	10		

List of Electives

Semester-2	**Semester-3**
Reinforcement Learning (DS)	Digital Signal Processing (EE)
Information Theory (EE)	Causal Reasoning (CS)
Econometrics (HSS)	Natural Language Processing (CS)
Bayesian Models (DS)	Computer Vision (CS)
Kernel Methods (DS)	
Detection Theory (EE)	
AI on Edge (CS)	

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New)

Course Code and Title	EE5007 Linear Algebra for Engineers				
Programme	M.Tech	Year of study		Semester	
Course credit	3-0-0-3				
Course category	PMT				
Prerequisite, if any	No				
Consent of teacher, if required	Required				
Date of proposal		Date of Senate Approval			
Proposing faculty	Lakshmi Narasimhan T				

Course Content

S/N	Topic	Lecture (hours)
1	Vectors and vector spaces : Notion of vector space, subspaces, dimensionality, span, inner product, norms, orthogonality, and geometric examples.	6
2	Linear transformations : Linear transformation of vectors, projection, geometric interpretations, and representation by matrices.	6
3	Bases and matrices : Basis, rank-nullity, orthogonal bases, determinant, projection, symmetric matrices, definiteness, normal matrices, nilpotent matrices and matrix norms.	8
4	Eigen values and eigen vectors : Diagonalization of matrices, eigen values, geometric interpretation, and eigen value problems.	7
5	Singular value decomposition: Existence of singular values, pseudo inverse, condition number and geometric interpretation.	7
6	Matrix decompositions : QR, LDU, Cholesky, Schur decomposition, and Jordan normal form	8
7	Engineering applications and examples	0(include in topics)
	TOTAL	42

Learning Outcomes:
Upon successful completion, the students shall be able to

- Understand basic mathematical tools required for engineering.
- Apply mathematical concepts to research problems in engineering.
- Model engineering systems mathematically.
- Analyze a mathematical model and deduce physical interpretations.

Learning Objective:

- To introduce the fundamentals of mathematical concepts required for engineers.
- To study and acquire the knowledge of various mathematical tools which are used in engineering applications.
- To mathematically model and analyze practical engineering systems using mathematical tools and concepts.
- To solve linear systems of the form Ax = b

Teaching Methodology **:** Lecture based

Assessment Methods **:** Exam based

Text Books

1. G. Strang, "Introduction to linear algebra", Fifth edition, Wellesley-Cambridge Press, ISBN-13: 978-0980232776.
2. S. Axler, "Linear Algebra Done Right", Ninth edition, Pearson Education, ISBN-13: 978-0387982588.

References

This course has previously been approved as EE5004 - Graduate Engineering Mathematics.

Changes from previous courses are

1. Removed: "Linear Programming" and "Statistics" components.
2. Added topics: symmetric matrices, definitiness, normal matrices, nilpotent matrices, matrix norms, and condition number.
3. Added components: "Matrix Decompositions" - LDU, Cholesky, Schur and Jordan canonical form

Probability & Statistics

1. Course code and Title : MA5xxx**:** Probability and Statistics
2. Course category : ERC
3. Course credit : 4-0-0-4
4. Prerequisite course :
5. Consent of Teacher : Required
6. Learning Objective: The objectives of this course are to

- Acquire the mathematical foundations of probability and statistics.
- Learn probability with underlying motivation being statistics
- Get some hands-on on generating random numbers and programming for statistics

7. Learning Outcome: Upon successful completion of the course, students will be able to

(a) make probabilistic models for problems arising from physical settings
(b) apply the mathematical tools learnt to model and draw conclusions in a more precise manner
(c) use a computer to solve statistical problems

8. **Content:**
Statistics: Review of statistics, describing data – frequency tables, graphs, and summarizing data – measures of central tendency and variation, multivariate data and correlation coefficient.

Probability: Axioms of probability, conditioning and Bayes' rule, independence, random variables, standard probability density functions - binomial, poisson, normal, exponential, etc., expected value, Chebhyshev's inequality, moment generating function, covariance, correlation, functions of random variables, properties of normal distribution, law of large numbers, central limit theorem, conditional expectation.

Statistical Inference: Point estimation – maximum likelihood estimation (mle), method of moments, Bayes' estimator, distributions of sampling statistics, interval estimation, tests of hypotheses – tests for mean and variance, t-test, chi-square test, evaluation of point estimators – unbiasedness, mean squared error (mse), Cramer-Rao bound.

Simulation: Generating random numbers – inverse transform, and Box-Muller methods. Importance sampling and Monte Carlo simulation.

Text books:

1. Sheldon M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*, Academic Press, SBN:978-0-12-370483-2.
2. A. Popoulis and S. Pillai, *Probability, Random Variables and Stochastic Processes*, McGraw Hill Education; 4 edition, SBN: 978-0070486584
3. Robert V. Hogg, Allen Craig, and Joseph W. McKean, *Introduction to Mathematical Statistics*, Pearson, ISBN 978-81-775-8930-6

References :

1. W. Feller, *An Introduction to Probability Theory and its Applications Volume-I*, Third Edition, John Wiley & Sons, ISBN: 978-81-265-1805-0
2. P.L. Meyer, *Introductory Probability and Statistical Applications*, Oxford and IBH Publishers, ISBN: 0-201-04710-1.
3. R.E. Walpole and R.H. Myers, *Probability & Statistics for Engineers and Scientists*, Macmillan, ISBN: 9788131715529
4. Freedman, Pisani and Purves, *Statistics*, Viva Books; 4th ed. (2011), ISBN: 978-0393929720
5. P. Billingsley, *Probability and Measure*, John Wiley & Sons Inc; Third Ed., ISBN: 978-81-265-1771-8

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New)

Course Code and Title	CS5512 : Machine Learning				
Programme	M.Tech/MS/PhD	Year of study		Semester	
Course credit	3-0-0-3				
Course category	PMT				
Prerequisite, if any	Familiarity with Algorithms, Probability, Linear Algebra, Programming				
Consent of teacher, if required	required				
Date of proposal		Date of Senate Approval			
Proposing faculty	Sahely Bhadra				

<u>Course Content</u>

S/N	Topic	Lectres (hours)
1	Introduction to the course, revision of linear algebra and probability.	3
2	Regression: linear regression, ridge regression	3
3	Classification: Linear discriminant analysis, logistic regression, perceptrons, support vector machines, Bayes classifier, decision tree. Nonparametric methods: k-nearest neighbours, Parzen window.	9
4	Principal component analysis, Canonical correlation analysis	3
5	Evaluation and Model Selection: ROC Curves, Evaluation Measures, Cross validation, Significance tests.	6
6	Ensemble methods: boosting, bagging, random forests.	3
7	Clustering: k-means, hierarchical, density based clustering , Gaussian mixture model	9
8	Sequential Learning : hidden Markov model	3
9	Feed forward NN : Tensorflow	3
	TOTAL	42

Learning Outcomes:

- State definitions, theorems/results, algorithms related to key concepts
- Apply standard techniques to solve known problems
- Given a task, derive a learning model by defining appropriate loss function, regulariser, optimization problem and stating the best possible solution.
- Analyse and compare models and algorithms with respect to their complexity, performance and applicability
- Develop models/algorithms with small modifications of existing standard techniques for a modification of known task

Learning Objectives:

- To introduce classical and foundational concepts, results, methodologies and applications in machine learning
- To develop abilities for developing a solution for a given problem starting from problem and data to presenting results
- To develop abilities for acquiring knowledge about existing work, criticise them, and improve them

Teaching Methodology : Lecture based

Assessment Methods : Exam based

Text Books

1. Richard Duda, Peter Hart, David Stork, Pattern Classification, 2nd Ed, John Wiley & Sons, 2001. ISBN 9788126511167
2. Christopher Bishop. Pattern Recognition and Machine Learning. ISBN 0387310738.
3. Trevor Hastie, Robert Tibshirani, Jerome Friedman. Elements of Statistical Learning. ISBN 0387952845.

References

1. Tom Mitchell. Machine Learning. McGraw-Hill. ISBN 0070428077.
2. Shai Shalev-Shwartz, and Shai Ben-David, Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press, 2014. ISBN 978-1-107-05713-5.

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New)

Course Code and Title	CS5101 : Machine Learning Lab				
Programme	M.Tech/MS/PhD	Year of study		Semester	
Course credit	0-0-3-2				
Course category	PMP				
Prerequisite, if any	co-requisite CS5512				
Consent of teacher, if required	required				
Date of proposal		Date of Senate Approval			
Proposing faculty	Sahely Bhadra				

<u>Course Content</u>

S/N	Topic	Lab (hours)
1	Introduction to NumPy	3
2	Regression: linear regression, ridge regression using scipy	6
3	Introduction to Matplotlib	3
4	Gradient descent method for optimization	3
5	Various classification methods using scikitlearn	6
6	Principal component analysis, Canonical correlation analysis	3
7	Ensemble methods: boosting, bagging, random forests.	3
8	Clustering using scikitlearn	6
9	Sequential Learning : hidden Markov model	3
10	Feed forward NN : Tensorflow	6
	TOTAL	42

Learning Outcomes:

- Given a task, derive a learning model by defining appropriate loss function, regulariser, optimization problem and stating the best possible solution.
- Analyse and compare models and algorithms with respect to their complexity, performance and applicability
- Develop models/algorithms with small modifications of existing standard techniques for a modification of known task

Learning Objectives:

- To introduce classical and foundational concepts, results, methodologies and applications in machine learning
- To develop abilities for developing a solution for a given problem starting from problem and data to presenting results

Teaching Methodology **:** Lab based

Assessment Methods **:** Exam based

Text Books

4. Richard Duda, Peter Hart, David Stork, Pattern Classification, 2nd Ed, John Wiley & Sons, 2001. ISBN 9788126511167
5. Christopher Bishop. Pattern Recognition and Machine Learning. ISBN 0387310738.
6. Trevor Hastie, Robert Tibshirani, Jerome Friedman. Elements of Statistical Learning. ISBN 0387952845.

References

3. Tom Mitchell. Machine Learning. McGraw-Hill. ISBN 0070428077.
4. Shai Shalev-Shwartz, and Shai Ben-David, Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press, 2014. ISBN 978-1-107-05713-5.

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New)

Course Code and Title	CS5015 : Data Engineering				
Programme	M.Tech/MS/PhD	Year of study		Semester	
Course credit	3-0-0-3				
Course category	PMT				
Prerequisite, if any	Familiarity with Algorithms, Probability, Linear Algebra, Programming				
Consent of teacher, if required	required				
Date of proposal		Date of Senate Approval			
Proposing faculty	Mrinal Kanti Das				

Course Content

S/N	Topic	Lecture (hours)
1	**Data Collection:** Various sources and types of data: text, video, audio, biology etc	3
2	**Data Preprocessing:** Cleaning data, missing data imputation, noise elimination, feature selection and dimensionality reduction, normalization	6
3	**Data Storage:** Database, Schema, ER diagram, SQL, functions, stored procedures, indexing B+tree, MongoDB, Client-Server Architecture [3	9
4	**Information Retrieval:** index construction, scoring models, complete search engine mechanism, evaluation methods.	6
5	**Data Processing:** *Data structures.* Stack, Queue, Linked List, Associated memory, Graphs. *Algorithms***.** Searching, Sorting, Graph traversal, Complexity	9
6	**Data Analysis:** regression, principal component analysis, canonical correlation analysis, analysis of variance	6
7	**Data Visualization:** table, graph, histogram, pie-chart, area-plot, box-plot, scatter-plot, bubble-plot, waffle charts, word clouds.	3
	TOTAL	42

Learning Outcomes: To be able to state and analyse

- Preprocessing techniques for various datasets,
- Standard database systems concepts like tables, relations, query
- Information retrieval techniques such as indexing, scoring, ranking, evaluation
- Data processing algorithms and data structures
- Visualization techniques

Learning Objectives:

- To be able to learn about the entire pipeline of a typical system involving data, collection, preprocessing, storage, retrieval, processing, analysis, and visualization.

Teaching Methodology **:** Lecture based

Assessment Methods **:** Exam based

Text Books

1. Introduction to Algorithms. Cormen, Leiserson, Rivest, Stein. MIT Press 3ed. ISBN-13: 978-0262533058
2. Database System Concepts. Silberschatz, Korth, Sudarshan. McGraw Hill Education; Sixth edition.ISBN-13: 978-9332901384
3. Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools. Cielen, Meysman,Ali. Dreamtech Press. ISBN-13: 978-9351199373

References

1. Data Engineering: A Novel Approach to Data Design. Brian Shive. Technics Publications. ISBN-13: 978-1935504603
2. Python Data Science Handbook: Essential Tools for Working with Data. Joel Grus. O’Reilly. ISBN-13**:** 978-9352134915

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New)

Course Code and Title	CS5103 : Data Engineering Lab				
Programme	M.Tech/MS/PhD	Year of study		Semester	
Course credit	0-0-3-2				
Course category	PMT				
Prerequisite, if any	Co-requisite of Data Engineering				
Consent of teacher, if required	required				
Date of proposal		Date of Senate Approval			
Proposing faculty	Mrinal Kanti Das				

Course Content

S/N	Topic	Lab (hours)
1	**Data Collection:** Various sources and types of data: text, video, audio, biology etc	3
2	**Data Preprocessing:** Cleaning data, missing data imputation, noise elimination, feature selection and dimensionality reduction, normalization	3
3	**Data Storage:** Database, Schema, ER diagram, SQL, functions, stored procedures, indexing B+tree, MongoDB, Client-Server Architecture [3	9
4	**Information Retrieval:** index construction, scoring models, complete search engine mechanism, evaluation methods.	6
5	**Data Processing:** *Data structures.* Stack, Queue, Linked List, Associated memory, Graphs. *Algorithms***.** Searching, Sorting, Graph traversal, Complexity	12
6	**Data Analysis:** regression, principal component analysis, canonical correlation analysis, analysis of variance	6
7	**Data Visualization:** table, graph, histogram, pie-chart, area-plot, box-plot, scatter-plot, bubble-plot, waffle charts, word clouds.	3
	TOTAL	42

Learning Outcomes: To be able to state and analyse

- Preprocessing techniques for various datasets,
- Standard database systems concepts like tables, relations, query
- Information retrieval techniques such as indexing, scoring, ranking, evaluation
- Data processing algorithms and data structures
- Visualization techniques

Learning Objectives:

- To be able to learn about the entire pipeline of a typical system involving data, collection, preprocessing, storage, retrieval, processing, analysis, and visualization.

Teaching Methodology **:** Lab based

Assessment Methods **:** Exam based

Text Books

1. Introduction to Algorithms. Cormen, Leiserson, Rivest, Stein. MIT Press 3ed. ISBN-13: 978-0262533058
2. Database System Concepts. Silberschatz, Korth, Sudarshan. McGraw Hill Education; Sixth edition.ISBN-13: 978-9332901384
3. Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools. Cielen, Meysman,Ali. Dreamtech Press. ISBN-13: 978-9351199373

References

1. Data Engineering: A Novel Approach to Data Design. Brian Shive. Technics Publications. ISBN-13: 978-1935504603
2. Python Data Science Handbook: Essential Tools for Working with Data. Joel Grus. O'Reilly. ISBN-13**:** 978-9352134915

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New)

Course Code and Title	CS5011 : Optimisation				
Programme	B.Tech/M.Tech/MS/PhD	Year of study		Semester	
Course credit	3-0-0-3				
Course category	PMT				
Prerequisite, if any					
Consent of teacher, if required	required				
Date of proposal		Date of Senate Approval			
Proposing faculty	Chandra Sekhar Lakshminarayanan				

<u>Course Content</u>

S/N	Topic	Lecture (hours)
1	Introduction: Motivation and examples	3
2	Basics: R^d, vectors, matrices, norm, sequences & convergence, functions in one and several variables, Taylor series, derivatives, gradient, sub-gradient, Hessian, properties of symmetric operators, contours, affine functions, hyper-planes, convex functions, minima: local and global, subspaces, affine spaces, half-spaces, convex sets	9
3	Unconstrained Optimisation: gradient descent, line search, rates for various classes of convex functions, steepest descent, Newton's method, conjugate gradient method, quasi-Newton method, linear least-squares regression: rates	9
4	Constrained Optimisation: linear and convex constraint sets, linear programming, simplex, interior-point method, duality theory: primal/dual programs, weak, strong duality, KKT conditions	12
5	Stochastic Optimisation: stochastic gradient descent, step-size conditions, Keifer-Wolwowitz method, simultaneous perturbation stochastic approximation (SPSA) method, smoothed functional method	9
	TOTAL	42

Learning Outcomes: As a result of this course, the student should be able to

- Pose a given optimisation problem by identifying the objective and constraints.
- Draw level sets and graphs of functions, identify constraint regions described by set of function.
- Choose stepsize and identify rates of convergence of optimisation algorithms.
- Use stochastic optimisation techniques to derive data driven algorithms.

Learning Objectives: To

- Look at the regions and functions in `d' dimensions.
- Build algorithms that find minima using first and second order information.
- Look at constraint optimisation and duality theory with linear programming as a special case.
- Introduce basics of stochastic optimisation

Teaching Methodology **:** Lecture based

Assessment Methods **:** Exam based

Text Books

1. Introduction to Optimization: Edwin K. P. Chong and Stanislaw H. Zak, Wiley-Interscience Series in Discrete Mathematics and Optimization (ISBN-13: 978-0471089490, ISBN-10: 0471089494).

References

1. Boyd, Stephen, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004. (ISBN-13: 978-0521833783, ISBN-10: 0521833787)
2. Fletcher, Roger. *Practical methods of optimization*. John Wiley & Sons, 2013. (ISBN: 978-0-471-49463-8)

Deep Learning

1. Course code and Title : CS5007**:** Deep Learning
2. Course category : ERC
3. Course credit : 3-0-0-3
4. Prerequisite course : Familiarity with Algorithms, Probability, Linear Algebra, Programming
5. Consent of Teacher : Required

6. Learning Objectives: Deep learning covers the theory and practice of a big family of very effective techniques today in the domain of machine learning. The objective of the course is to enable students to get familiar with this area and to gain adequate knowledge to apply the techniques in solving real world problems.

7. Learning Outcomes: At the end of the course, the students should be able to precisely state the classical algorithms, models, and theories in the area. Students should be able to identify appropriate algorithm given a practical task. Students should also be able to implement and solve the tasks using deep learning techniques.

8. Content:

1. Revision of Linear Algebra, Probability. Machine learning Basics. [1 week]
2. Perceptron, Neural network, deep feedforward networks. [2 weeks]
3. Optimization techniques for deep networks. Back propagation, gradient descent, sampling techniques. [2 weeks]
4. Regularization, dropout. [1 week]
5. Case studies using tensorflow, pytorch (will be spread across the semester) [2 weeks]
6. Convolutional networks with application in computer vision. [2 weeks]
7. Recurrent networks, Long Short Term Memory networks with application in natural language processing. [2 weeks]
8. Autoencoders, representation learning. [1 week]
9. Variational autoencoders, Generative adversarial networks [1 week]

9. Text books:

1. Deep Learning. Ian Goodfellow and Yoshua Bengio and Aaron Courville. MIT Press. 2016.ISBN-13**:** 978-0262035613.

10. **References:**

1. Pattern Recognition and Machine Learning. Christopher Bishop. Springer. 2006. ISBN-13 978-0-387-31073-2.
2. Deep Learning with Python. Francois Chollet. Publisher: Manning Publications; 1 edition. ISBN-13: 978-1617294433
3. Hands-On Machine Learning with Scikit-Learn and TensorFlow. Aurélien Géron. Publisher: O'Reilly Media; 1 edition. ISBN-13: 978-1491962299.

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New)

Course Code and Title	CS5104: Big Data Lab				
Programme	B.Tech/M.Tech/MS/PhD	Year of study		Semester	
Course credit	1-0-3-3				
Course category	PMT				
Prerequisite, if any					
Consent of teacher, if required	required				
Date of proposal		Date of Senate Approval			
Proposing faculty	Satyajit Das				

Course Content

S/N	Topic	Lecture (hours)	Lab (hours)
1	**Lab on set up :** amp; manipulating files in HDFS	2	9
2	**Basic programs of Hadoop MapReduce:** Driver code, Mapper code, Reducer code, RecordReader, Combiner, Partitioner	4	9
3	**Pig :** Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators	4	12

4	**Big data analytics in Spark using PySpark:** Installing Apache Spark, Spark Ecosystem, Resilient Distributed Dataset (RDD) in Spark, building machine learning model using PySpark	4	12
	TOTAL	14	42

Learning Outcomes:

- Preparing for data summarization, query, and analysis.
- Applying data modelling techniques to large data sets
- Creating applications for Big Data analytics
- Building a complete business data analytic solution

Learning Objectives:

- The primary objective of this course is to optimize business decisions and create a competitive advantage with Big Data analytics. This course will introduce the basics required to develop map reduce programs, derive business benefit from unstructured data. This course will also give an overview of the architectural concepts of Hadoop and introducing map reduce paradigm. Another objective of this course is to introduce programming tools PIG & HIVE in Hadoop ecosystem.

Teaching Methodology **:** Lecture and Lab based

Assessment Methods **:** Exam based

Text Books

1. Big Java 4th Edition, Cay Horstmann, Wiley John Wiley & Sons, INC, ISBN: 9780470509487
2. Hadoop: The Definitive Guide by Tom White, 3 rd Edition, O'reilly, ISBN: 9781449328917

References

1. Hadoop MapReduce Cookbook,Srinath Perera, Thilina Gunarathne, O'reilly, ISBN: 9781849517287
2. Hadoop for Dummies by Dirk deRoos, Paul C.Zikopoulos, Roman B.Melnyk,Bruce Brown, Rafael Coss, John Wiley & Sons, 2014, ISBN: 1118607554
3. Hadoop in Practice by Alex Holmes, MANNING Publication, ISBN: 9351197425

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New)

Course Code and Title	CS5012 : AI for Cyber Security				
Programme	B.Tech/M.Tech/MS/PhD	Year of study		Semester	
Course credit	3-0-0-3				
Course category	PMT				
Prerequisite, if any	Familiarity with Probability, Machine Learning				
Consent of teacher, if required	Required				
Date of proposal		Date of Senate Approval			
Proposing faculty	Vivek Chaturvedi				

Course Content

S/N	Topic	Lecture (hours)
1	Overview on Machine Learning with use cases from cybersecurity, classification of threats, attacks, vulnerabilities, malware, trojans etc.	6

2	Classification of malware using supervised/unsupervised learning based on signatures and profiling. Decision Tree and context based malicious event detection	9
3	Time Series Analysis and Ensemble modelling to detect deviation from normal behaviour, case studies in Reconnaissance detection	9
4	Efficient Network Anomaly detection; familiarize with various stages of network attack and address using deep neural networks, develop intrusion detection systems	9
5	Adversarial attacks on ML systems, model poisoning, black box attacks, white box attacks, state-of-art research paper reading on deep learning systems	9
	TOTAL	42

Learning Outcomes:

- Students will be able to develop ML models to classify malwares.
- Able to implement simple intrusion detection systems using deep neural networks.
- They will be able to demonstrate the vulnerabilities in ML systems and state methods to address adversarial attacks.

Learning Objectives:

- Machine Learning (ML) is increasingly used in sensitive and time-critical systems such as autonomous driving, cyber physical systems etc. to deliver higher performance and protect the confidentiality of the systems. Though ML based systems can be used to classify various malware attacks and develop intrusion detection systems, these systems are also susceptible to several adversarial attacks. This course covers a systematic approach on developing ML based cybersecurity methodologies. It will also cover adversarial attacks which intentionally forces ML systems to behave unexpectedly.

Teaching Methodology **:** Lecture based

Assessment Methods **:** Exam based

Text Books

1. A. Hands-on Machine Learning for Cyber Security by Soma Halder, ISBN139781788992282

References

1. Machine Learning and Security by David Freeman, Clarence Chio Publisher: O'Reilly Media, Inc. Release Date: February 2018 ISBN: 9781491979891

2. Malware Data Science by Joshua Saxe with Hillary Sanders, ISBN-10: 1-59327-859-4 ISBN-13: 978-1-59327-859-5 Publisher: William Pollock

Proposal for New Course

1. **Course code and Title** : CS5xxx: Reinforcement Learning
2. **Course category** : PME/Research
3. **Course credit** : (3-0-0-3)
4. **Prerequisite course** : Probability, Linear Algebra, Data Structures and Algorithms
5. **Consent of Teacher**: Required
6. **Learning Objective**: Reinforcement learning (RL) is a paradigm of learning via interactions with the environment. RL algorithms are at the frontier of current success of AI: AlphaGo, the computer program that beat humans is a RL algorithm. The objective is to provide a bottom up approach: starting from foundation in Markov decision processes (MDP), the course builds up to the state-of-the-art RL algorithms.
7. **Learning Outcomes**: The student should be able to

a) model a control task in the framework of MDPs.
b) Should be able to choose appropriate algorithm based on size of the problem, approximation architecture, and the sampling model.
c) Identify stability/convergence and approximation properties of RL algorithms.
d) Use deep learning methods to RL problems in practice.

8. **Course Content**:

0)Introduction: State of the art applications in Atari, AlphaGo, relation to other problems in artificial intelligence [1 Week]

1) Markov Decision Processes (model based): Formulation, Value Iteration (VI), Policy Iteration (PI), Linear Programming (LP) [2 Weeks]
2) Approximate Dynamic Programming (approximate model based): curse-of-dimensionality, representations, Approximate value iteration, approximate policy iteration, approximate linear program, approximation and convergence guarantees [2 Weeks]
3) Stochastic Approximation: Single and multi-timescale stochastic approximation, introduction to ordinary differential equation based convergence results. [1 Week]
4) Value function learning (approximate model-free): Temporal difference (TD learning, TD(0), TD(lambda), Q-learning, State-Action-Reward-State Algorithm (SARSA) , TD with function approximation, on/off-policy learning, gradient temporal difference learning [2 weeks]
5) Actor-Critic: Policy gradient, Deterministic Policy Gradient Theorem, Natural Actor-Critic [2 Weeks]
6) Deep RL [2 Weeks]
7) Exploration vs Exploitation: Upper Confidence Bound (UCB), Upper Confidence Reinforcement Learning (UCRL) [2 Weeks]

9. **Text books**:

1. Richard S. Sutton and Andrew G. Barto, Introduction to Reinforcement Learning, 2nd Edition, MIT Press. 2017. ISBN-13 978-0262039246.
2. Dimitri Bertsekas and John G. Tsitsiklis, Neuro Dynamic Programming, Athena Scientific. 1996. ISBN-13: 978-1886529106

10. **References**:

1. V. S. Borkar, Stochastic Approximation: A Dynamical Systems Viewpoint, Hindustan Book Agency, 2009. ISBN-13: 978-0521515924
2. Deep Learning. Ian Goodfellow and Yoshua Bengio and Aaron Courville. MIT Press. 2016.ISBN-13: 978-0262035613.
3. The course might require reading outside of the prescribed book. In such cases, supplementary handouts (or links to the same) will either be provided.

Proposing Faculty : Dr. Chandra Shekar Lakshminarayanan
Department / Centre : Computer Science and Engineering
Proposal Type : New Course
Programme : B.Tech/M.Tech/Research