Online Research Seminar Syllabus

1. **Overview**

<table>
<tr><td>Title</td><td colspan="3">Distributed Machine Learning: Foundations and Algorithms</td></tr>
<tr><td>Mode</td><td colspan="3">Leading Instructor Sessions & Teaching Fellow Sessions</td></tr>
<tr><td>Hours</td><td colspan="3">4*2 hours lecture +2*2 hours research project +1*2 hours final presentation session+ 6*1.5 hours teaching fellow sessions</td></tr>
<tr><td rowspan="4">Prerequisites</td><td rowspan="2">High School Students</td><td>Required course/Knowledge</td><td>The course is primarily aimed at undergraduate students. The course is generally not recommended for high-school students. Very advanced high-school students may participate, if they meet the same level of prerequisites as undergraduate students detailed below.</td></tr>
<tr><td>Recommended Materials for preparing for the course</td><td>(See below, for undergraduate students.)</td></tr>
<tr><td rowspan="2">College Students</td><td>Required course/Knowledge</td><td>The students should have good grasp of multidimensional calculus, matrix analysis and basic linear algebra at the undergraduate level. Also, basic concepts from undergraduate probability are expected.</td></tr>
<tr><td>Recommended Materials for preparing for the course</td><td>Basic multidimensional calculus at the level of the book by Kreyszig (Engineering Mathematics) is recommended; matrix analysis and linear algebra at the level of Prof. Gilbert Strang's undergraduate course at MIT is recommended (see https://ocw.mit.edu/courses/18-06-linear-algebra-spring-2010/ for lectures and videos).

More sophisticated mathematical concepts will be taught as and when needed in the course; additionally, more advanced prerequisite topics will be often reviewed in the lectures to make the course self-contained.</td></tr>
</table>

2. Program Introduction and Objectives

Course Description	The topic of distributed machine learning and in particular federated learning is of immense practical and theoretical interest in the broad ML community. This is due to the ever-increasing scale and complexity of learning problems as well as applications in which data is naturally distributed across multiple entities. This course focuses on an overview of this emerging research area with a self-contained set of lectures focusing on key technical ingredients such as supervised learning and stochastic optimization for model training. It introduces in a tutorial manner the transition to decentralized ML from centralized paradigms with illustrations and examples. The topic of federated learning, a major paradigm for distributed ML, is studied in depth with formulations, methods and algorithms. The course reviews various topics in performance analysis and design of federated learning algorithms that cater to different operating scenarios, from training in data-parallel data center environments to distributed ML in IoT type applications.
Software/Tools (if any)	Basic familiarity in python scripting; the student should be familiar with at least one high-level programming language, although, prior knowledge in python will be helpful. Some of the assignments and the project will involve programming to validate algorithms and perform experiments on real data; while the students may use another programming language, Python is highly recommended due to the availability of numerous machine learning modules that will be useful. In particular, we will use the “scikit-learn” package from Python.

3. **Program Schedule**

Week		Leading Instructor Session	Teaching Fellow Session (lab/case study, etc.)	Assignment	Reading Materials
1	**Topic**	Training ML Models	Solving Classification and Regression Problems	The student will use real data to formulate basic regression and classification tasks and perform analysis and experiments to validate the theory and methods covered in the lecture and study the influence of various parameters (hyperparameters) in training.	Lecture slides and other notes to be provided.
	Detail	This lecture will introduce the basics of formulating machine learning models for tasks such as classification and regression. The student will be introduced to the (parametric) supervised learning framework and how to systematically abstract the basic ingredients, such as training vs test data, parametric family of predictors, loss functions, required to train ML models. The framework of empirical risk minimization will be introduced as a unified formulation for training ML models such as linear and logistic models, support vector machines, neural networks for performing regression and classification tasks.	In this session, instances of supervised learning framework will be developed to train ML models for regression and classification tasks on real data sets, for example, predicting house prices based on historical data or classifying images. The students will work on standard Python packages to create training examples from real data.		
2	**Topic**	Stochastic Optimization: Basics	Designing Stochastic Gradient DescentType Algorithms	The student will use stochastic gradient descent type algorithms to train regression and classification tasks on real data. The assignment will also focus on understanding various modes of convergence and common variants of stochastic gradient descent used in stochastic optimization relevant to training ML models.	Lecture slides and other notes to be provided.
	Detail	In this lecture, we will first introduce key elements of stochastic optimization that form the core of all ML training, from training simple linear models to complex neural networks. The student will be introduced to stochastic gradient descent type algorithms that power today's large-scale ML training. Several variants of stochastic gradient descent will be introduced along with a discussion of pros and cons and basic techniques in ML training algorithm design.	In this session, the student will learn to analyze and experiment on stochastic gradient descent type algorithms for performing stochastic optimization relevant to training ML models. The role of algorithm hyperparameters such as learning rate and batch size, technical conditions such as convexity or lack thereof and nature of convergence will be illustrated.		
3	**Topic**	Stochastic Optimization: Theory, Convergence and Implementations	Distributed ML: Architectures and Basic Algorithms	A single assignment will be given on content covered in Weeks 3 and 4 after Week 4's lecture.	Lecture slides and other notes to be provided.
	Detail	This lecture first provides theoretical and convergence guarantees for stochastic gradient type algorithms, specifically in the context of training large ML models. sets up the background for distributed ML. Both convex and non-convex learning scenarios will be studied. Finally,	This session will focus on rigorously developing ML training approaches, i.e., how to trade-off between performance metrics such as convergence rate and error floor, via systematic hyperparameter (e.g., learning rate, batch size) optimization		

		the lecture focuses on non-centralized implementations of stochastic optimization and sets up the stage for distributed ML to be studied in the next couple of lectures.	to obtain efficiency. The role of various technical assumptions, such as convexity or lack thereof, will be illustrated and best practices for ML training algorithm design will be explored.		
4	**Topic**	Distributed ML and Federated Learning: Basics	Federated Learning Algorithms	This assignment will focus on distributed training of ML models to perform regression and classification tasks using federated learning algorithms. The focus will be on analyzing performance on different problem classes and data and studying various trade-offs (example accuracy vs runtime)	Lecture slides and other notes to be provided.
	Detail	This lecture sets up the background for distributed ML. The focus will be on different distributed architectures and their motivation; key performance metrics such as latency, convergence issues, ability to continuously learn from online distributed training data will be covered; basic intuition and challenges in setting up distributed ML frameworks will be discussed. Subsequently, this lecture will dive into the basics of federated learning, a decentralized ML paradigm of current interest. Basic training algorithms based on federated stochastic gradient descent and applications will be introduced; design and analysis tools for studying federated learning algorithms will be introduced.	This session will focus on distributed ML model training formulations and algorithms introduced in the lecture. The students will gain experience in state-of-the-art algorithms commonly used for training ML models in the data-parallel distributed setup and, in particular, federated learning algorithms. Various algorithm design issues will be considered.		
5	**Topic**	Federated Learning: Advanced Topics and Algorithms	Federated Learning Algorithms		
	Detail	This lecture will dive into more advanced analysis of federated learning algorithms and also highlight topics of current research interest. Algorithms that tackle issues such as data and computational heterogeneity across computing nodes in distributed ML will be discussed. Other advances that focus on communication efficient design and heterogeneous or multi-task learning will be reviewed.	This session will review different algorithms for federated learning, discuss their applicability and pros and cons based on the operating environment. We will review the performance of algorithms (in terms of accuracy and runtime) both theoretically (as covered in the lecture) and experimentally.		Lecture slides and other notes to be provided.
6	**Topic**	Research Workshop	Research Workshop		
	Detail	This workshop will focus on logistics of the project and discuss sample research topics to be performed by student groups. (See Section 5.)	The mentoring sessions will guide the student groups and provide feedback as they perform their project.		
7	**Final Oral Presentation and Written Reporting**				

4. **Problem Sets/Written Assignments/Quizzes**

Total Number of Assignments	3 times
Submission Deadline	5 days after issued
Will there be Quizzes? How often/how many?	2-3 quizzes focusing on basic concepts during teaching fellow sessions

5. Final Oral and Written Project

Detailed requirements of the final project:

- The final project will focus on developing either:
  - new algorithmic modifications of the basic distributed learning algorithms studied in the course with the objective of improving existing algorithms in at least one of the major performance dimensions, i.e., for example, improving communication-efficiency or runtime or being able to address issues such as node failures;
  - or, an application of distributed (federated) learning in a domain such as traffic monitoring or classifying images or any other application proposed by the research group
  - or, a combination of new methods and applications, e.g., how to modify existing algorithms to address distributed ML training in a particular application.
- Typically, students will work in groups of 3-5 based on their interests.
- Some sample project topics will be provided by the instructor, however, the students are also encouraged to propose their own project topics.

5.1 Final Oral Presentation

- Oral Project Requirements (e.g: if slides needed; Format; Criteria; Deadline):
  - A well prepared presentation by the student group with participation by all the group members.
  - The presentation should be with the help of Powerpoint or a similar presentation tool and should be carefully structured.
  - The presentation will be held in week 7.
  - Each group will have 15 minutes to make the presentation followed by 5 minutes for questions.
  - The presentation should provide the general objective of the project, motivation and application (usefulness), a survey of existing approaches and why they are not adequate. The presentation should describe the key findings of the project, provide analytical or experimental (as applicable) justification of the findings, and discuss improvements over existing approaches or benchmarks.

5.2 Final Written Report

- Written Project Requirements:
  - A well prepared written report by the student group with participation by all the group members.
  - The report should be formatted properly (for example, in latex, or word with legible symbols and mathematical notation).
  - The report will be due a day before the presentation.
  - Each group will prepare a 4 -- 5 page report on their project topic; additional pages, if needed, containing derivations or experiments, may be accommodated in an appendix. However, the main gist of the project together with the key findings should be within the main body of the report.
  - The report should follow the same structure as the presentation, but with more details as appropriate: it should provide the general objective of the project, motivation and application (usefulness), a survey of existing approaches and why they are not adequate. The report should describe the key findings of the project, provide analytical or experimental (as applicable) justification of the findings, and discuss improvements over existing approaches or benchmarks.

6. Suggested Future Research Fields/Direction/Topics

This course will provide the students with the basics to perform research on distributed machine learning that forms the core of today’s sophisticated academic and industrial work on complex learning problems such as in autonomous driving or information processing and decision-making in IoTs. With the tools developed in the course, the student should be able to read current research papers published in top ML conferences and develop their own research or applications.