

INTRODUCTION TO THE COURSE

LECTURE 0
SECTION 1
JUNE 1ST



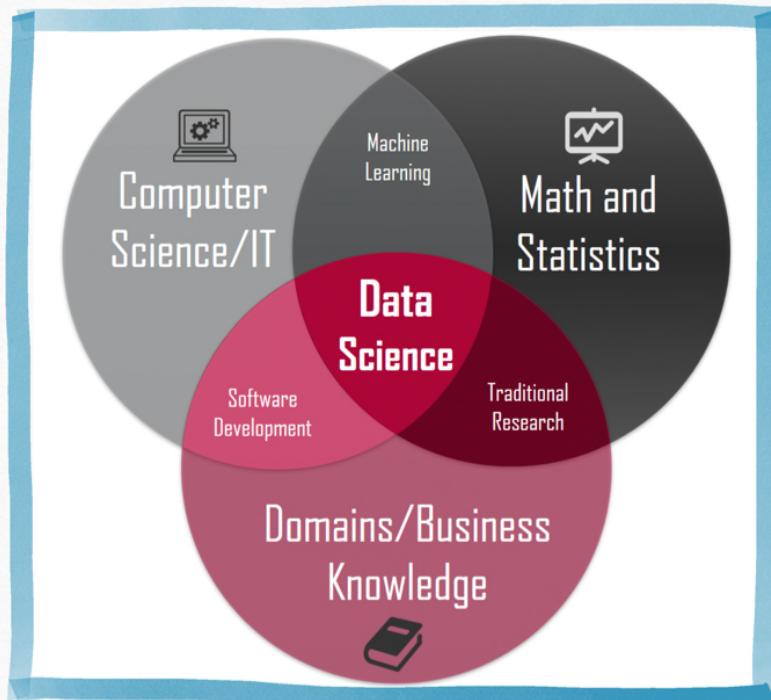
IACS
INSTITUTE FOR APPLIED
COMPUTATIONAL SCIENCE
AT HARVARD UNIVERSITY



UNIVERSITY *of*
RWANDA

WHAT IS MACHINE LEARNING & DATASCIENCE?

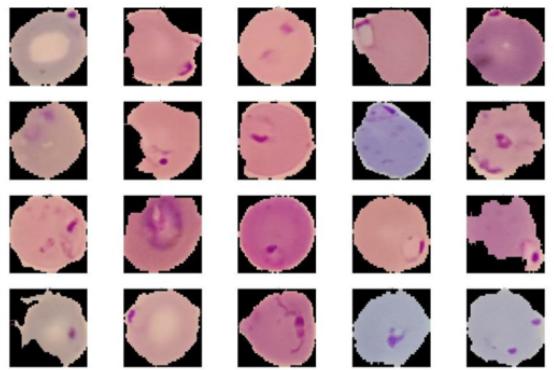
In both data science and machine learning we extract patterns and insights from data.



- Inter-disciplinary
- Data and task focused
- Resource aware
- Adaptable to changes in the environment & needs

THE POTENTIAL OF MACHINE LEARNING:

Disease Diagnosis



Detecting malaria from blood smears

Agriculture



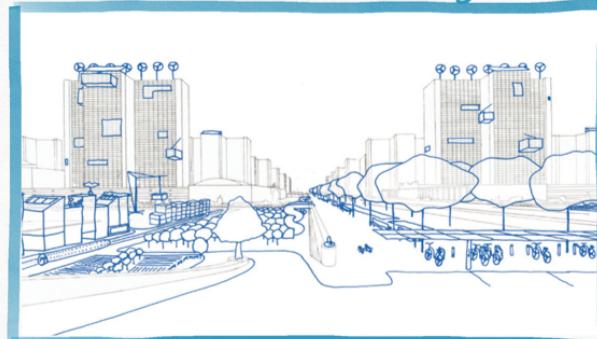
Precision agriculture for changing Climate

Drug Discovery



Quickly discovering new drugs for COVID

Urban Planning



Predicting and planning for resource needs

THE POTENTIAL OF MACHINE LEARNING:

Gender Bias



Some ML models for evaluating job applications show bias in favour of male candidates.

Racial Bias



Risk models used in U.S. courts have been shown to be biased against non-white defendants.

GOAL OF THE COURSE:

Theory	Practice	Impact
<ol style="list-style-type: none">1. Key ML concepts2. Important metrics for evaluation3. Handling different kinds of data4. Extracting insights from analysis of ML models	<ol style="list-style-type: none">1. Implementing ML models using Python libraries2. Using free online tools & resources for ML and data science	<ol style="list-style-type: none">1. Mapping real-life problems to ML concepts2. Evaluating the social impact of ML

WHO ARE YOUR INSTRUCTORS?



Weiwei Pan
Post-Doctoral Fellow
Harvard University



Melanie P. Fernandez
Post-Doctoral Fellow
Harvard University

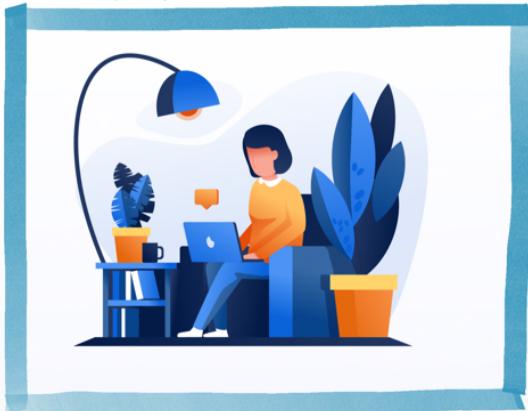


Pavlos Protopapas
Scientific Director
Institute of Applied
Computational Sciences,
Harvard University

FORMAT OF THE COURSE:

This is a **flipped-classroom** course.

Before Class



- watch lecture videos
- study course readings

During Class



- Discuss concepts from videos and readings in groups
- Complete coding exercises in groups

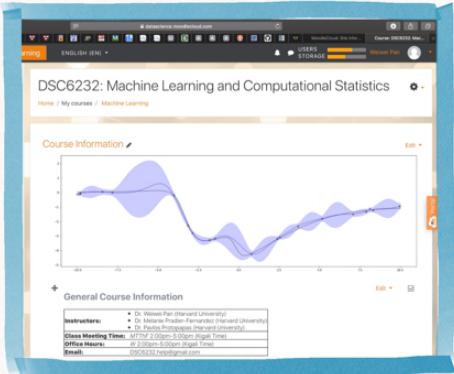
COURSE COMPONENTS:

There are 4 graded components of the course.

Before Class	During Class	Outside Class
<ul style="list-style-type: none">• Three concept quizzes (Each day)	<ul style="list-style-type: none">• Participation in discussions (Each day)• In-class practical exercise (Each day)	<ul style="list-style-type: none">• Homework (2 total)

TOOLS FOR THE COURSE:

Moodle



- Syllabus, schedule
- Lecture videos
- Readings
- Quizzes
- Assignments
- Grades

Piazza

Solutions for Homework 4 and Homework 5

Dear all,

Please find attached the solutions for HW4 and HW5. Let us know if you have any questions!

Best regards!

M.

Followup discussions: An ongoing questions and comments.

- Questions and answers about course related topics
- Discussions with instructors and peers

colab

```
fig, ax = plt.subplots(1, 2, figsize=(10, 5))
scater=plt.scatter(Xtrain, Y_train, ax=ax[0])
ax[0].set_title('Training Data')
plt_decision_boundary(Xtrain, Y_train, model, ax=ax[1], poly_degree=1, plot_boundary_only=False)
plt.show()
```

Before neural network models became wildly popular in machine learning, a common method for building non-linear classifiers is to first map the data, in the input space $\mathbb{R}^{n_{\text{input}}}$, into a feature space $\mathbb{R}^{n_{\text{feature}}}$, such that the classes are well-separated in the feature space. Then, a linear classifier can be used in the feature space.

If we ignore the output node of our neural network classifier, we are left with a function $\mathbb{R}^d \rightarrow \mathbb{R}^1$, mapping the data from the input space to a 2-dimensional feature space. The transformed data (and in general, the output from a hidden layer in a neural network) is called a representation of the data.

- Implementing and running your ML models