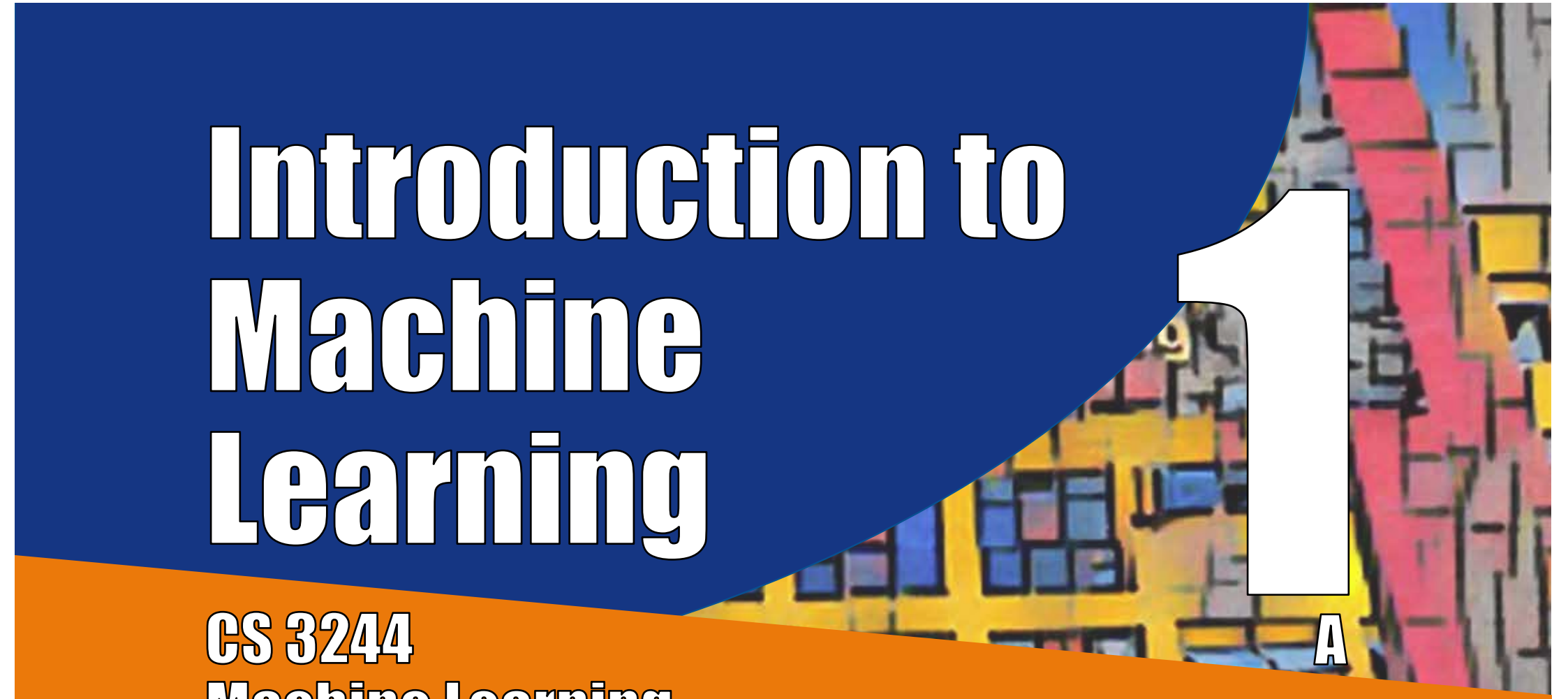
Introduction to Machine Learning



1A

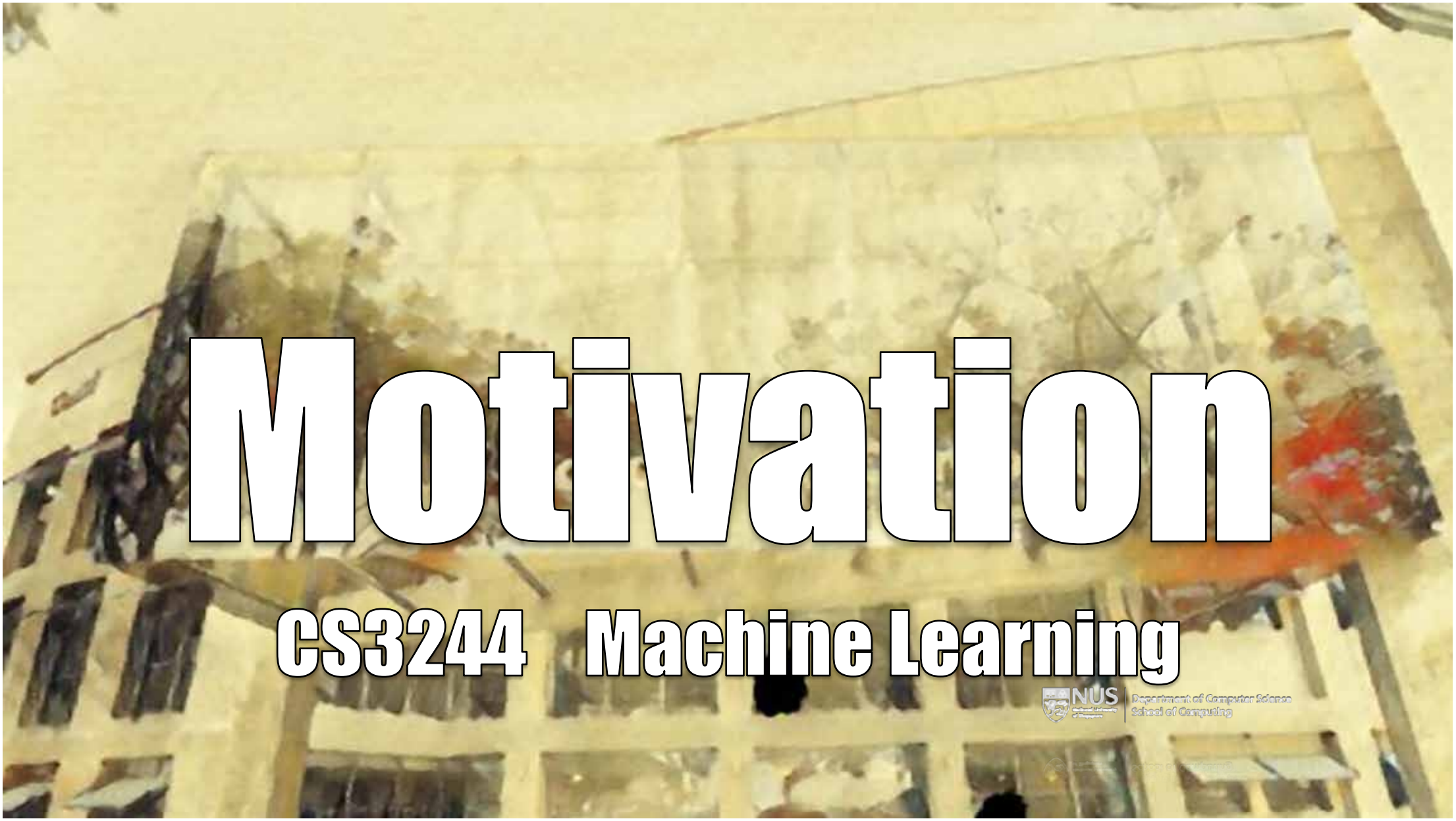
CS 3244
Machine Learning

NUS | Computing

Forecast

Learning Outcomes for this week:

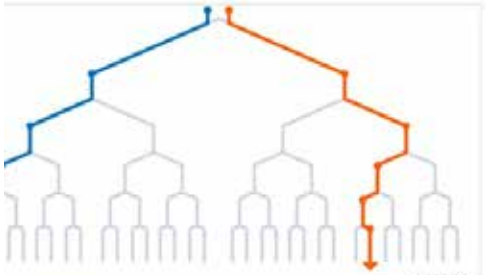
- Understand the rationale for our course's structure, the why of our module
- Recall the components of the course and their weightages and general deadlines
- Execute a basic machine learning workflow in Google Colab
- Conceptualize the division of labor in a typical machine learning project


Motivation
CS3244 Machine Learning
NUS
Department of Computer Science
School of Computing

fference Between AI, Machine Learning, and
earning
ne learning, and deep learning are terms that are often used
eably. But they are not the same things.
IDIA.COM

Harvard Business Review · October 22, 2015 · Like Page · Sign Up

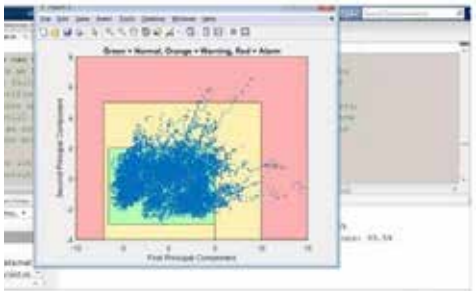
learning, explained.



Machines Learn (And You Win)
smart predictions with statistical models

ATLAB · 19 at 10:55pm · Like Page · Message

fundamentals of machine learning with MATLAB in this video!



ne Learning Made Easy

Get ready for this year's AI-centric EmTech Digital, starting with our Q+A with the leader of Google Brain.

Machine Learning for Everyone
Recent advances are making machine learning useful outside the tech industry, says the leader of the Google Brain research group.
TECHNOLOGYREVIEW.COM

World Economic Forum · June 18 · Like Page · Sign Up

"You may not know it, but it's all around you."

Why you need to understand machine learning
From picking stocks to finding partners, machine learning plays a part in every stage of our lives - yet most of us know little about these powerful tools.
WEFORUM.ORG

The Economist · May 14, 2015 · Like Page · Sign Up

The standard joke about artificial intelligence (AI) is that, like nuclear fusion, it has been the future for more than half a century now. AI is in the news again, for there has been impressive progress in the past few years in a particular subfield of AI called machine learning. But what exactly is that? The Economist explains how machine learning works
http://econ.st/1IA3T6V

NOW,
Everyone
needs to
understand it.

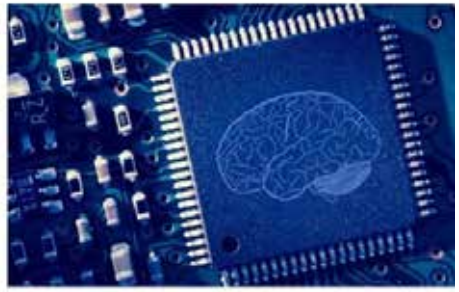
to the group: DataScience SG.

Bloomberg Launches "Foundations of Machine Learning" Course for Those with Strong Math Backgrounds

BLOOMBERG.GITHUB.IO
Foundations of Machine Learning

Futurism · July 20 at 3:30pm · Like Page

Machines can now teach themselves without human interventio
how it works—and how it will transform our society.



Understanding Machine Learning [INFOGRAPI
FUTURISM.COM | BY TODD JAQUITH

Popular Science · July 8 at 3:15am

Achieve the impossible by mastering artificial intelligence

Launch Into the Innovative Field of Machine Lea
10 Courses & 63.5 Hours of Training

It is far reaching and transformative, disruptive.

The Verge
August 6 at 7:30 PM

Like Page

Some of Dota 2's most beloved names got owned by OpenAI bots.

THEVERGE.COM

The OpenAI Dota 2 bots just defeated a team of former pros

And it wasn't even close

Forbes
October 28, 2015

Like Page · Watch Video

Machine learning started as far back as the 1950s, when computer scientists figured out how to teach a computer to play checkers.

5 Ways Machine Learning Is Reshaping Our World

Machine learning is really just the very advanced application of statistics to learning to identify patterns in data and then make predictions from those patterns.

WWW.FORBES.COM | BY BERNARD MARR

Singularity Hub
May 27

Like Page · Sign Up

"For overwhelmed researchers navigating the growing stack of science literature — the value isn't in having so much new information, but finding relevant insights when they need them."

Machine Learning's Next Trick Will Transform How Research Is Done

Though research is a slow moving and rigid process, one study shows that the rate of scientific study has exploded in the last 50 years. According to the paper,...

SINGULARITYHUB.COM

Advertising Age
Yesterday at 6:05am

Like Page

Here are five industries where machine learning will shape customer experience and marketing programs. http://bit.ly/2a7Kosd

When Machines Get Smarter, So Do CMOs

The proliferation of data and intelligent platforms has become increasingly prevalent and is beginning to shape the way businesses must transform their own organizations, writes Evan Rowe.

VentureBeat
July 20 at 9:15pm

Like Page · Message

How machine learning and A.I. are transforming commerce, according to Pinterest

Pinterest's commerce head explains how A.I. and machine learning will transform shopping

Aunkur Arya, Braintree's mobile general manager, and Michael Yamartino, Pinterest's head of commerce, took the stage at VentureBeat's MobileBeat 2016...

calculate how pollution will spread across the city.

How IBM is trying to use AI to reduce pollution in China

As China's air quality sinks, IBM's research group is using AI techniques to predict pollution levels for every square kilometer 10 days in advance.

TECHNOLOGYREVIEW.COM

Digital Trends · July 20 at 8:02am · Like Page · Sign Up

Heading straight to the source.

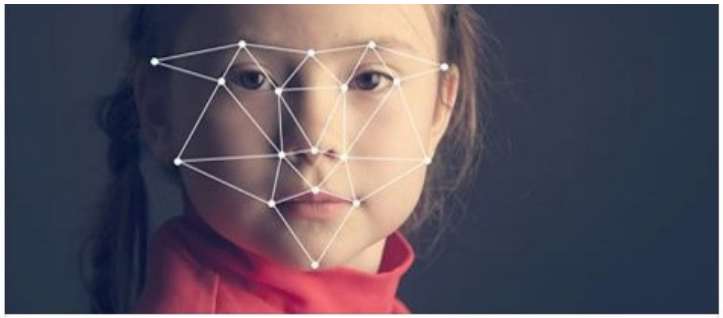
The key to stopping Ebola? Using machine learning to track infected bats

A team of scientists have used machine learning to create a "bat map" which may help quell the spread of Ebola

DIGITALTRENDS.COM

Scientific American magazine · July 14 at 8:44am · Like Page · Learn More

Although Germany is the world's poster child for renewable energy, its grids cannot yet cope with the erratic nature of wind and solar power.



How Human-Machine Learning Partnerships Can Reduce Unconscious Bias

ENTREPRENEUR.COM

It can potentially save the world and save us from ourselves.

New Scientist · 11 hrs · Like Page · Sign Up

With US policing in the spotlight, the police are turning to AI to predict which officers are at risk of cracking under stress

Machine learning is going to revolutionize the way you use your phone

It'll order you an Uber when you're too drunk, for one.

TECHINSIDER.IO

Futurism · April 5 · Like Page · Sign Up

As more and more devices vie for precious bandwidth space, DARPA is looking for creative ways to manage the radio spectrum.

Managing Bandwidth Through Machine Learning

FUTURISM.COM

NOVA | PBS · July 25 at 5:51am · Like Page · Sign Up

The same techniques that inspired AlphaGo could help Google combat climate change.

Machine Learning Could Dramatically Reduce Google's

Corporates are all-in.

Fortune Magazine · April 14 · Like Page · Sign Up

Facebook now uses machine learning to translate 2 billion News Feed items per day.

Inside Facebook's Biggest Artificial Intelligence Project Ever

Facebook's "machine learning" platform is powering your cat videos and family pics.

FOR.TN

http://www.forbes.com/.../how-microsoft-is-making-big-impac.../...

How Microsoft Is Making Big Impact With Machine Learning

During the last two years, Microsoft has upped the ante on Machine Learning

FORBES.COM | BY JANAKIRAM MSV

TechCrunch · February 20 · Like Page · Sign Up

Salesforce plans to use some of the tech to build out its own machine learning capabilities as part of SalesforceIQ

Salesforce Acquires PredictionIO To Build Up Its Machine Learning Muscle

Salesforce has made another acquisition to build out its technology in machine learning and big data analytics: the company has acquired PredictionIO, a..

TCRN.CH

Forbes · July 22 at 1:00 PM · Like Page

The company is breaking down internal barriers and showing how other companies can do the same.

The Wall Street Journal · May 19 · Like Page · Use App

Google is betting it can outsmart rivals on new products—from virtual assistants to messaging apps—with advances in artificial intelligence.

Business Insider: Tech · July 26 at 5:45am · Like Page · Sign Up

Machine learning is a major priority of tech giants like Google, Facebook and Amazon.

Do not ignore machine learning — Google, Facebook, and Amazon are all betting big on it

This is why we can ignore machine learning. It's not about what it can do, but the effects of its prioritization.

BUSINESSINSIDER.COM

In more ways than one.

Series A funding and acquisitions

I've written exactly, oh, about five machine learning scripts. Anyone want to create a Machine Learning startup in our virtual garage? I'm kidding, almost, but not 100%. The ML world is starting to feel like a mini dot-com 1990's bubble. It's not as easy as putting ".com" on the end of a domain name and writing an ambiguous business plan, not at all, but it has the faint smell of the dot-com bubble.

Activity

Record For AI Startups

In Q1'16, there were over 140 deals to startups focused on AI. Khosla Ventures emerges as most active VC investor in AI

NEA - New Enterprise Associates

GraphLab is now Dato!
Oh, and they've raised $18.5M.
Congratulations!

Machine Learning Startup GraphLab Gets A New Name And An $18.5M Check

Looking to bring machine learning to the masses -- or at least the masses of businesses that could benefit from embedding artificial intelligence in their...

TECHCRUNCH.COM

VentureBeat

Machine vision is advancing, but still early stages.

Google buys machine learning startup Moodstock to help your phone's camera identify objects

Google announced today that it has acquired Paris-based Moodstock, a startup that has developed machine learning technology to bolster the image recognition...

VENTUREBEAT.COM

TechCrunch

Google is next to acquire a machine learning startup

Two years ago today, I wrote the first lines of code at Starbucks while unemployed and waiting for my UK visa. Happy St Patrick's Day!

WSJ

Inspired by Astrophysics, Machine Learning Startup Wise.io Raises $2.5M - Venture Capital...

The startup's technology analyzes interactions among businesses and their customers.

BLOGS.WSJ.COM | BY DEBORAH GAGE

VentureBeat

Self-driving cars may not need pre-loaded detailed 3D maps of the world's roads. This startup promises to map the world in real time as it goes along.

U.K. machine-learning startup FiveAI raises $2.7 million to create the smart autonomous cars of tomo

FiveAI, a U.K. startup leveraging artificial intelligence (A.I.) to build the brains for the autonomous cars of the future, has announced a $2.7 million funding round...

The Wall Street Journal

Magic Pony's technology allows for better-quality video compression, essentially allowing for better-looking videos that use less data and are less choppy.

Twitter to Buy Machine-Learning Company Magic Pony

Twitter said Monday that it bought Magic Pony Technology, a London-based machine-learning company, in a move to help bolster its live and streaming video content.

You might be able to use it today…

TechCrunch
April 10, 2015 ·

Like Page Sign Up

AWS announced a new Machine Learning platform today that it says puts machine learning into the hands of ordinary developers.

AWS Wants To Put Machine Learning In Reach Of Any Developer

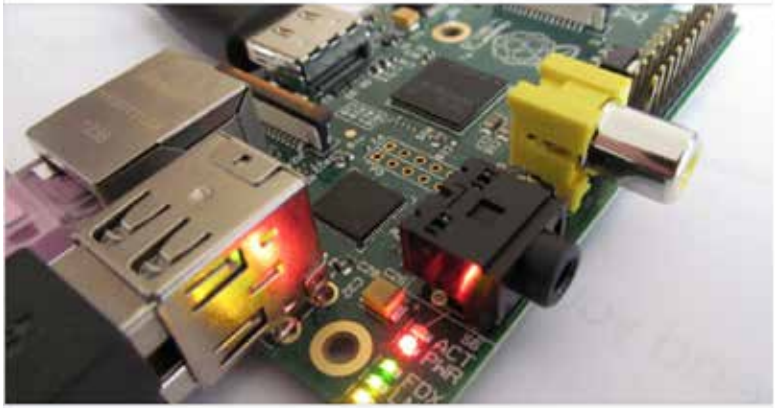
Andy Jassy, senior vice president at Amazon Web Services, announced a new machine learning platform today at the AWS Summit in San Francisco. The…

TCRN.CH

…red a link to the group: Data Artificial Intelligence.

August 5 at 1:32 AM ·

When TensorFlow was first launched in 2015, we wanted it to be an "open source machine learning framework for everyone". To do that, we need to run on as many of the platforms that people are using as possible. We've long supported Linux, MacOS, Windows, iOS, and Android, but despite the heroic efforts of many contributors, running TensorFlow on a Raspberry Pi has involved a lot of work. Thanks to a collaboration with the Raspberry Pi Foundation, we're now happy to say that the latest 1.9 release of TensorFlow can be installed from pre-built binaries using Python's pip package system! If you're running Raspbian 9 (stretch), you can install it by running these two commands from a terminal:



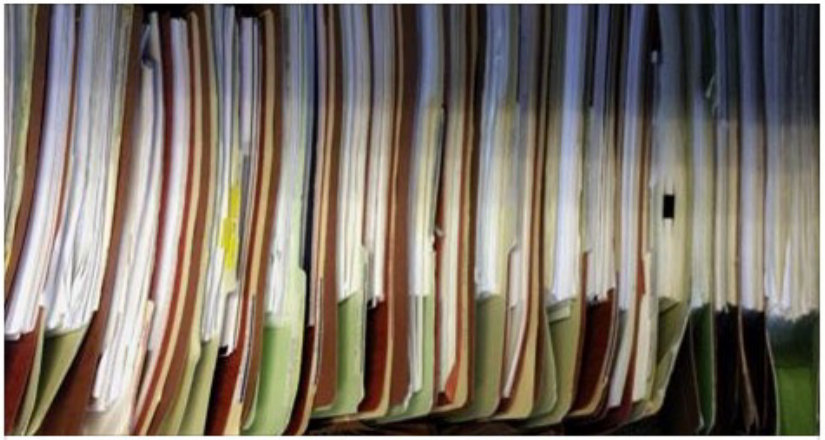
MEDIUM.COM

TensorFlow 1.9 Officially Supports the Raspberry Pi

… but perhaps only a few understand it.

Like Page

"Machine learning will likely advance in a punctuated equilibrium fashion: in a few, hard-won, big leaps. The reason for this is deep learning software is nearly impossible to debug because we don't fully understand how it works."



The Business Implications of Machine Learning

It's not about what it can do, but the effects of its prioritization

MEDIUM.COM I BY DREW BREUNIG

WIRED

December 13, 2015 ·

Learn More

It is the guiding principle of many applied mathematicians that if something mathematical works really well, there must be a good underlying mathematical reason for it, and we ought to be able to understand it.

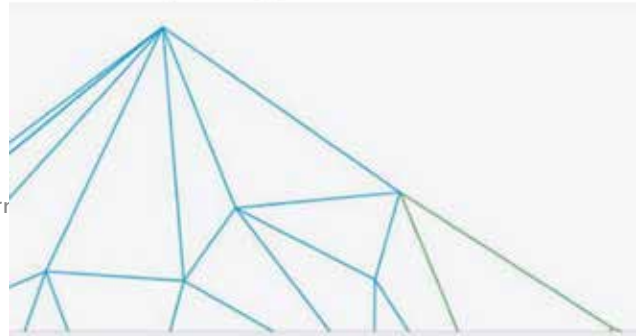
Machine Learning Works Great—Mathematicians Just Don't Know Why

Our current mathematical understanding of many techniques that are central to the ongoing big-data revolution is inadequate, at best.

WIRED.COM I BY INGRID DAUBECHIES

link to the group: DataScience SG.

Bloomberg Launches "Foundations of Machine Learning" Course for Those with Strong Math Backgrounds



BLOOMBERG.GITHUB.IO

Why are you here? 🤔

Please help fill out the Pre-Course Survey in LumiNUS
by Tue night 10 Aug, 23:59.

Help us understand your background and your goals.

We'll recap your answers to these questions over the next week or so.

Instruction
Staff
CS3244 Machine Learning
NUS
Department of Computer Science
School of Computing

Instructor:
Kan Min-Yen

kanmy@comp.nus.edu.sg

- Easy to remember: $min(x)$
- Ph.D./B.S. in CS, Columbia University

Research Interests

- Web
- Information Retrieval
- Natural Language Processing

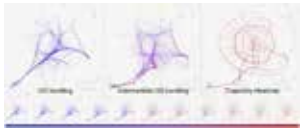
IoT Sensors | Health Behavior Change | Data Analytics

NUS Ubicomp Lab

Apps and Analytics for Smart Cities and Healthcare

http://ubiquitous.comp.nus.edu.sg

Explainable Artificial Intelligence



Interactive Data Visualization

[Instructor] Brian Lim

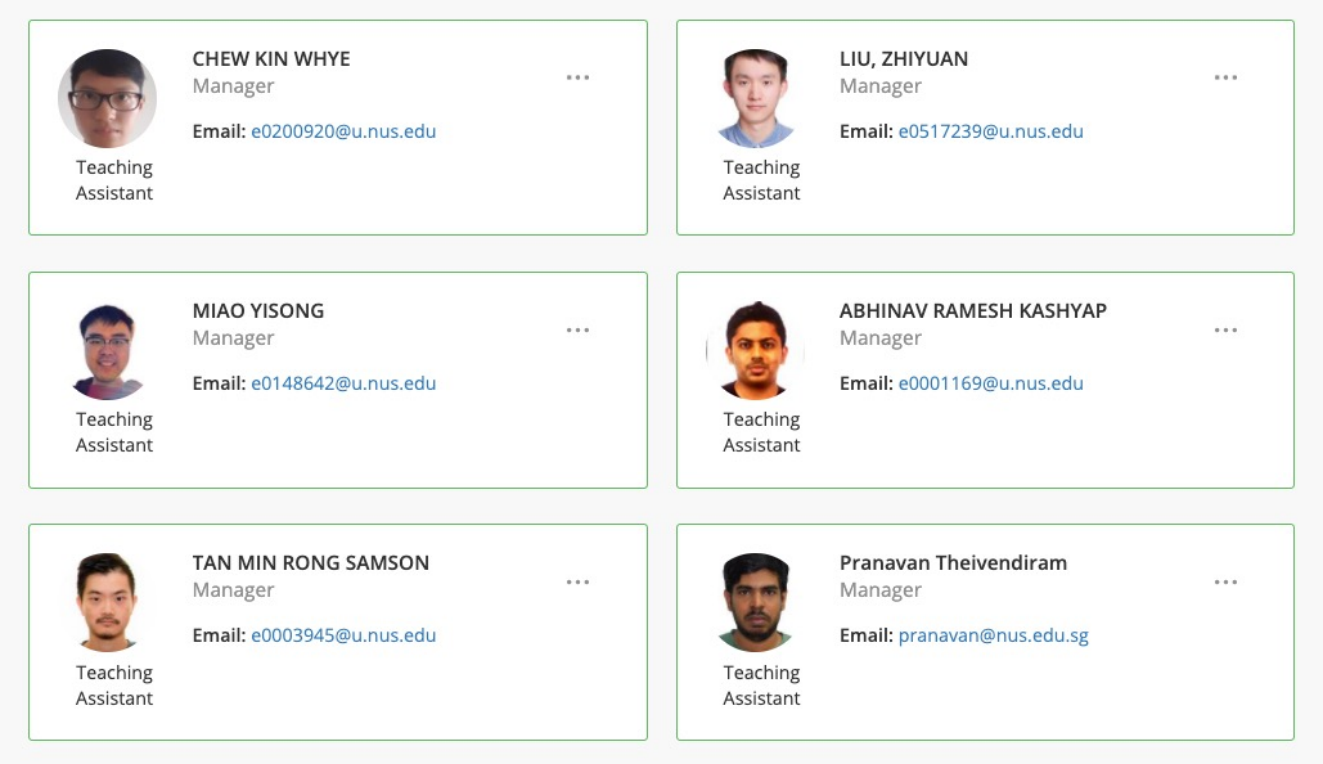
brianlim@comp.nus.edu.sg

- Asst. Prof. in Computer Science
- Ph.D. in HCI, Carnegie Mellon University
- B.S. in Engineering Physics, Cornell University

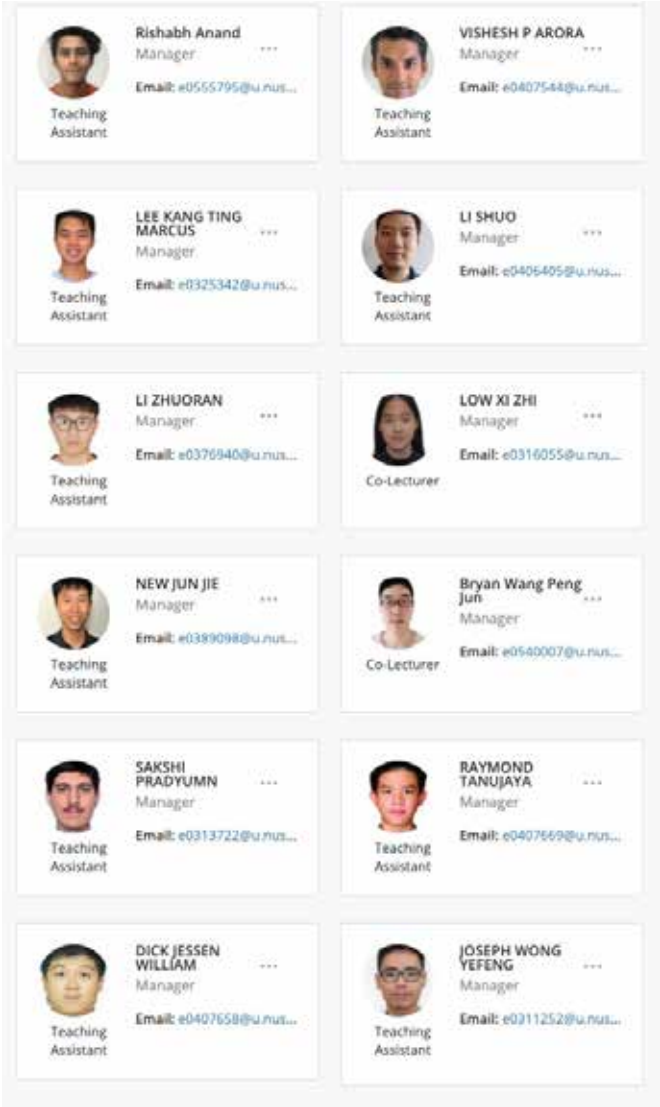
Research Interests

- HCI: understand people with tech, help people with tech
- Explainable Artificial Intelligence
- Ubiquitous Computing
- Data analysis and visualization
- Smart Health and Smart Cities

Graduate Teaching Assistants



Undergraduate Teaching Assistants

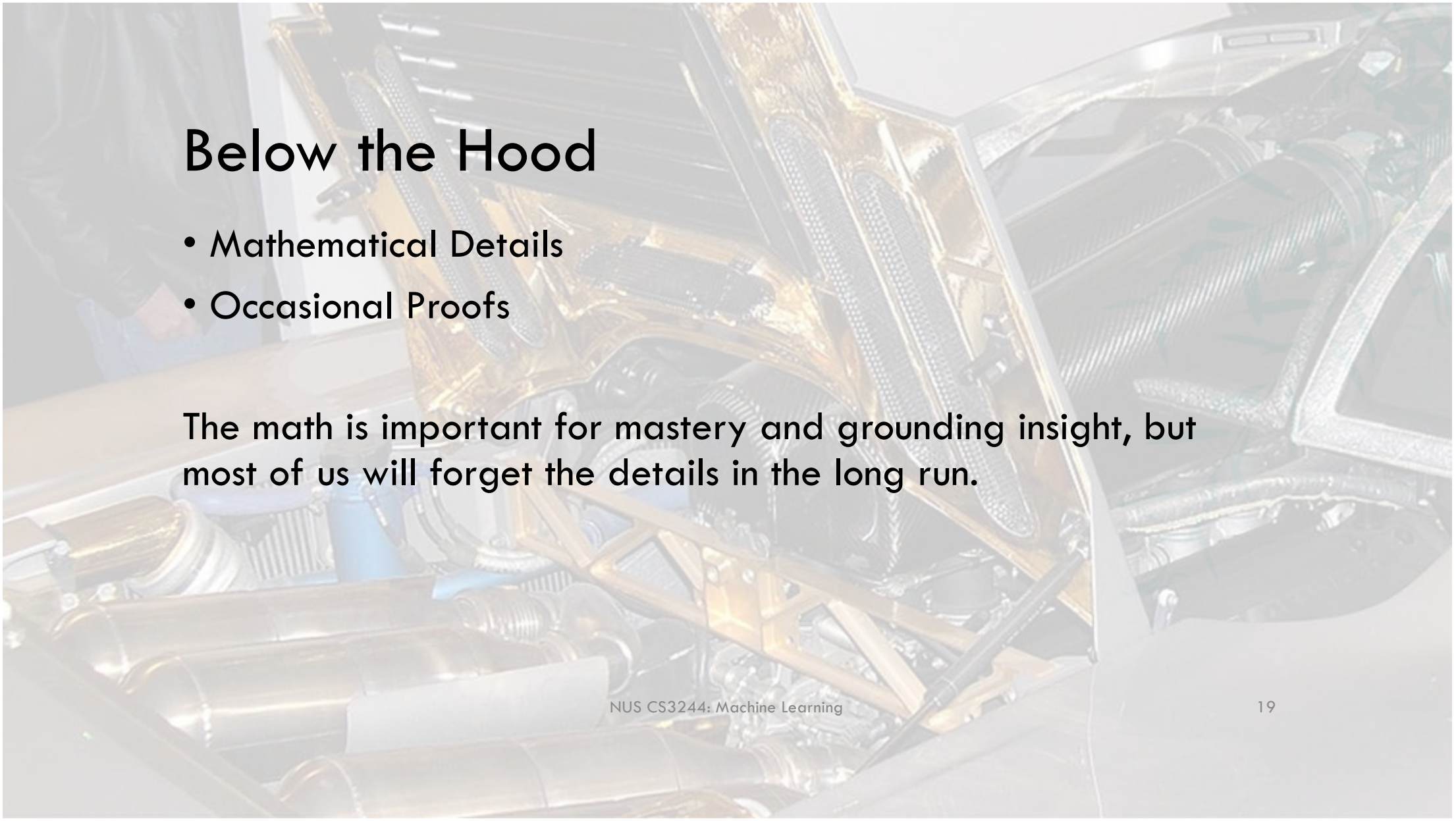


Course
Characteristics
CS3244 Machine Learning
NUS
Department of Computer Science
School of Computing

Above the Hood

- "Big Picture" applications
- Set up and go
- Hands-on exercises
- Connection with Industry
- Mock Job Interview

For executive and continuing lifelong education.



Below the Hood

- Mathematical Details
- Occasional Proofs

The math is important for mastery and grounding insight, but most of us will forget the details in the long run.

Barriers to success in this course

- Serious Math.
  - Probability and Statistics
  - Linear Algebra
- Programming competence.
- E Learning course implementation.

Will need your support and understanding – please offer critical feedback to us too.

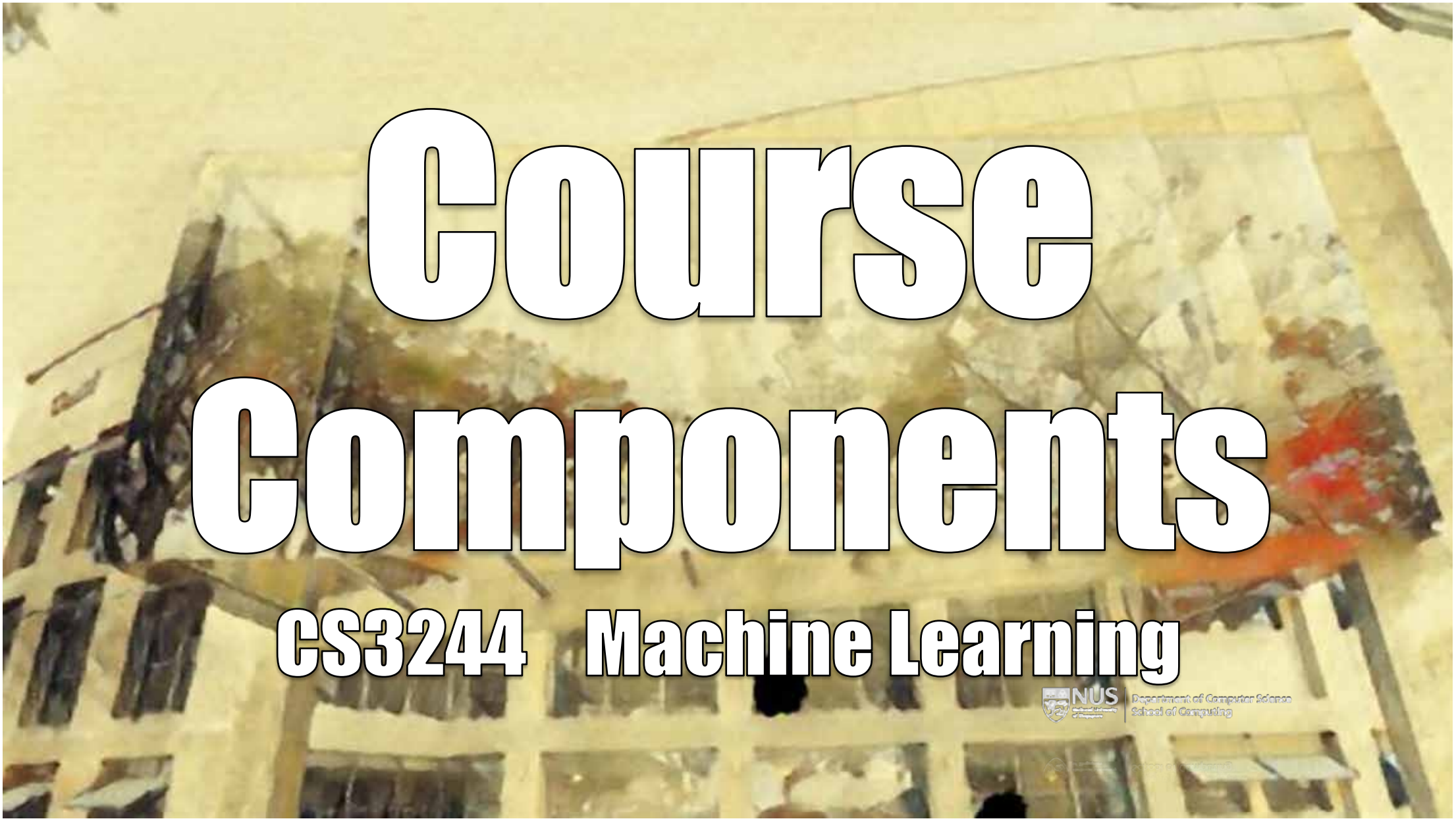
Python & Jupyter

We'll be using the Python programming language for our entire class homework assignments, in particular the 3.7.11 version (currently default version in Colab).

We'll also be using Jupyter / iPython Notebooks to show and simulate experiments.

Google has a very nice, convenient Python-in-a-browser experience, Colab, which we'll take advantage of.

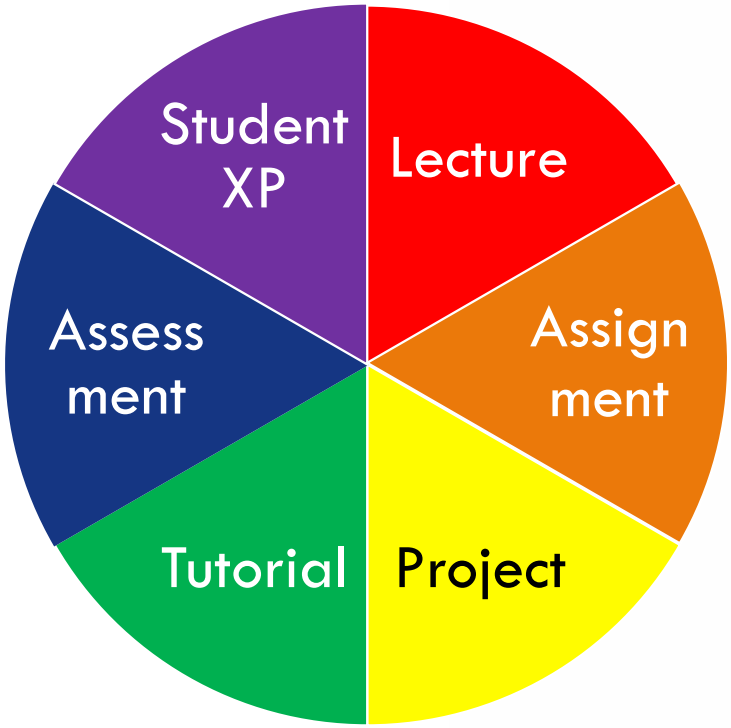
Python is an easy language to transition to for most programmers and it has numerous libraries that work well with machine learning as well as data analysis.


Course
Components
CS3244 Machine Learning
NUS
Department of Computer Science
School of Computing

Components

55% **Assessments**
1. Final 35%
2. Midterm 20%

10% **Participation**
1. Tutorial (Colab Notebooks and Participation)
2. Pre-lecture Activities
3. In-Lecture Activities



35% **Assignments**
1. Individual Assignment 10%
2. Group Project 25%
   2.1 Proposal
   2.2 Presentation
   2.3 Peer Review

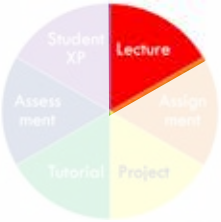
1. Lectures – Synchronous Zoom

(Better learning outcomes when you are here with us)

	Week t	Week $t+1$
Mon	Lecture [1]	Tutorials for Week t
Tue		
Wed		
Thu	Lecture [2]	
Fri		
Sat		
Sun		

Most lecture activity will occur during our Monday 2-hour slot.

Every week you will:



Contributes towards the 10% participation grade

1. Do the pre-lecture activity and discuss within your tutorial group before Week t's Monday lecture.
2. Attend lecture (in person or online) and participate in the in-lecture activity on Week t.
3. Reinforce concepts in the tutorial on Week $t+1$.

cs3244-2110.slack.com

Install on your computer **and** smartphone.

https://bit.ly/cs3244-join-slack

Use your NUS @nus.edu.sg or @u.nus.edu address.

Optional In-class lecture

Come down to the LT for lecture!

Lecture[1] Mon 16:00–18:00: i3 Aud
Lecture[2] Thu 11:00–12:00: LT 15

You must be one of the first n students to emote on the Week t lecture post in Slack to secure a seat.

Make sure you do your temperature declaration before coming to campus to attend lecture.

Thu Lecture: other activities and AMA

(Get answers when you have questions)

On lighter weeks we'll use the Thursday lecture slot for auxiliary activities. Occasionally there will be no activities, so we'll schedule an AMA session.

Ask Me Anything (AMA) sessions are purely voluntary. We'll be recapping what was covered in the previous lectures or tutorials first. Then cover your questions and get to know you better.

2. Assessment

10% of your total grade

(Practice before ~~Perfect~~ more practice)

Individually done. Class integrity policy applies.

Executed in Weeks 03–05.

"Classic" tabular machine learning assignment via Google Colab.

Simplified evaluation.

Run through the whole cycle of a machine learning application.

More formative than summative.

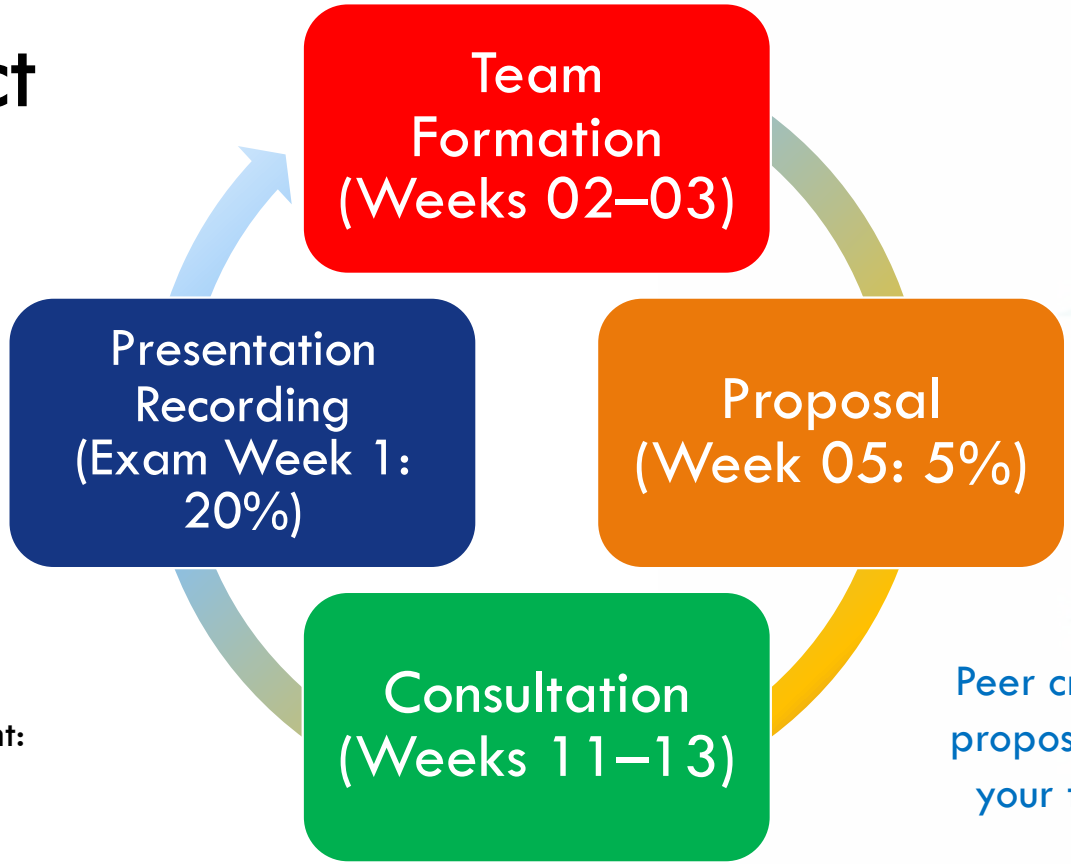
3. Project

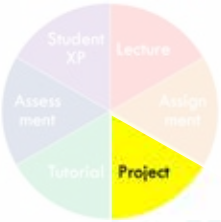
(In-depth, group self-determined immersive project)

Weeks 05–15.

Peer and Instruction Staff Review.

Emphasis on summative assessment: Can you apply what you've learned?





25% of your total grade

Peer critique of the proposal as part of your final grade.

Computing Support

We are rounding up computation support from both Google Education and School of Computing to support your projects.

- Dedicated GPU compute hosts from SoC's cluster
- Google Compute Cloud credits (usable for Colab VM)

Computing Support – SoC

Non-SoC students should visit https://mysoc.nus.edu.sg/~newacct to create/re-enable their SoC account.

We have reserved xgpf0-9 for CS3244 from 4 Oct (Week 8) to 12 Nov (Week 13). [We may try to extend this lease]

xgpf0 - xgpf9	10	Ubuntu	2 x Xeon Silver 4116 (2.1Ghz)	256GB DDR4	960GB SSD & 3.84TB SSD	10GbE	1x NVIDIA Tesla T4 GPU	2019	Asus

Computing Support – Google

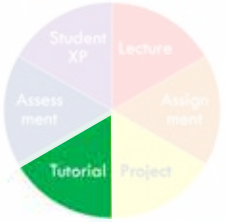
Google Workspace for Education is sponsoring USD $50 per student for compute.

More details to be announced later.

Intra-team Peer Evaluation: TEAMMATES

- Exam Week 2.
- 5-6 students per group.
- Depending on your peer feedback, you may get the same, lower, or higher grade than your group grade.

4. Tutorials via Zoom



(Face to face check on your understanding in groups)

Tutorials start in Week 03 and will be likely be entirely online in Zoom.

Have another connected device (e.g., phone) ready for warm-up exercises.

	1000	1100	1200	1300	1400	1500	1600	1700	1800
Mon			Tutorial [06]	Tutorial [01]	Tutorial [04]	Tutorial [03]	Lecture [1]		
Tue			Tutorial [11]	Tutorial [12]		Tutorial [09]	Tutorial [10]		
Wed	Tutorial [08]	Tutorial [05]				Tutorial [07]	Tutorial [02]		
Thu		Lecture [2]							
Fri									

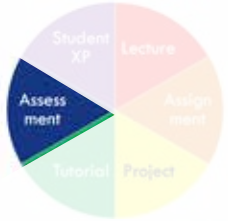
Tutorial Activities

(Let's mix it up)

Tutorials will vary in their activities, to be led by the tutorial leader:

- "Standard" tutorial: recap of lecture materials and practice questions;
- Hands-on activities, akin to "labs" via Google Colab.
- Later in the semester, project consultations.
- Last tutorial: exam (p)review.

5. Assessments



55% of your grade in total

(Primary assessment of the basic concepts)

Midterm (20%)

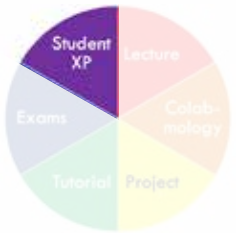
- Mon 27 Sep (Week 07) 16:00–17:00
- Covers Weeks 01–06

Final (35%)

- Wed 24 Nov (Exam Week 1; Week 15) 17:00–19:00
- Emphasis on Weeks 07–13, but also touches on earlier foundations

Both Open Book

6. Student XP

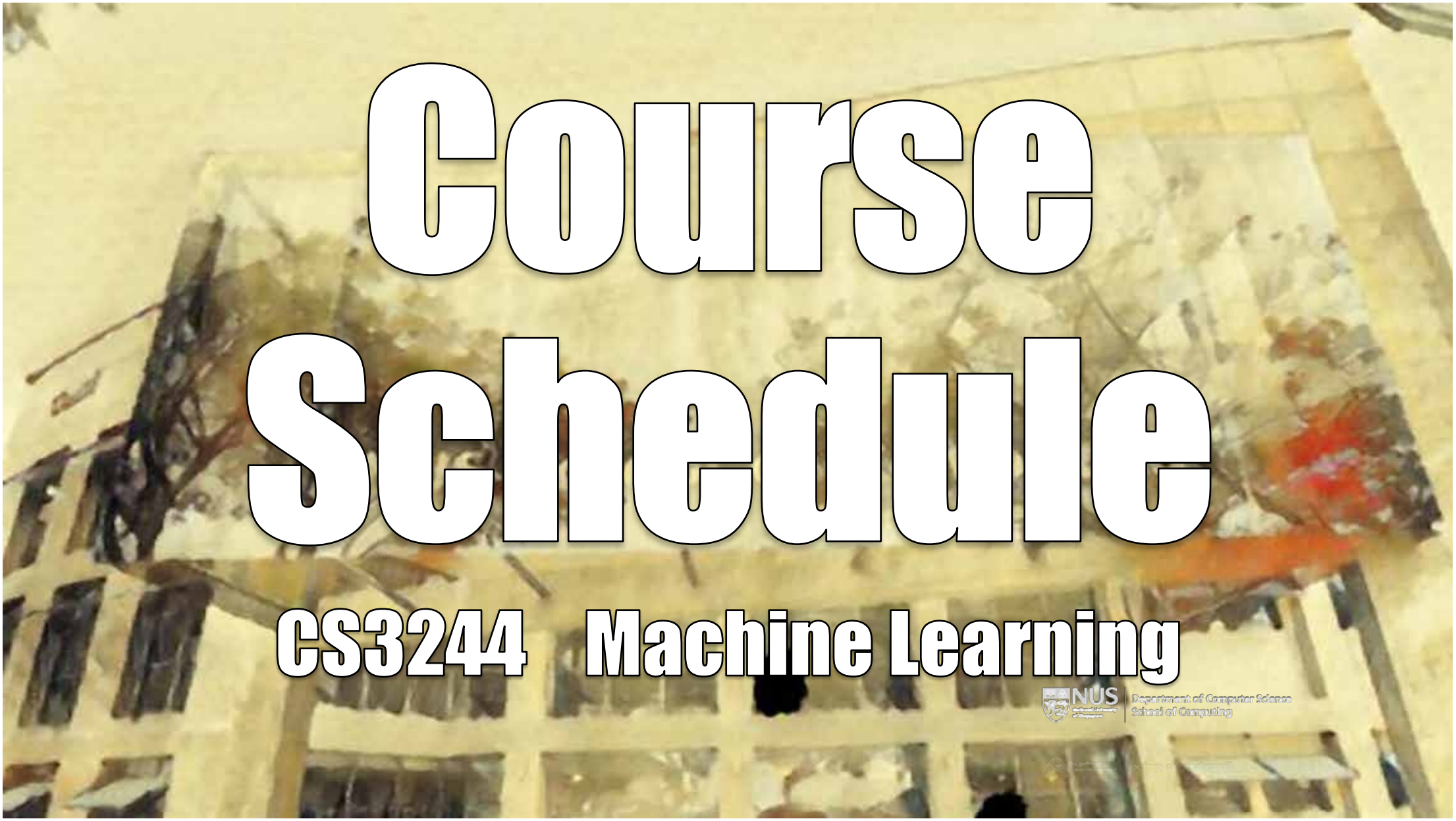


(Keeping tabs on your personalized progress in the course)

Clear student learning outcomes for each component and subcomponent.

We'll be tracking your progress through your pre- and in-lecture activities, along with your tutorial participation.

Occasional interviews with machine learning experts in our school and community.


Course
Schedule
CS3244 Machine Learning
NUS
Department of Computer Science
School of Computing

Student Learning Outcomes (SLOs)

1. Understand the **basic** concepts of machine learning.

2. Apply an appropriate machine learning **algorithm** for a given problem.

3. **Evaluate** the performance of a machine learning solution.

4. Use a machine learning tool to carry out machine learning **experiments**.

<table>
<tr><th>Weeks</th><th>Lecture Topics</th><th colspan="2">Deliverables (90%)</th></tr>
<tr><td><u>The Basics</u>
(Weeks 01–06)</td><td>• Intro to the ML Pipeline (SLO1)
• Fundamental Theory (SLO1)
• Basic Models (SLO2)</td><td></td><td>Assignment (10%)</td></tr>
<tr><td>Recess</td><td colspan="3"></td></tr>
<tr><td rowspan="2"><u>Widening and Deepening</u>

(Weeks 07–13)</td><td rowspan="2">• Preprocessing (SLO1)
• Visualization (SLO1)
• Evaluation (SLO3)
• Explanation (SLO3)

• Neural Networks (SLO2)
• Deep Learning (SLO2)

• Unsupervised & Semi-Supervised (SLO1)</td><td>Midterm (20%)</td><td rowspan="3">Project (25%)</td></tr>
<tr><td></td></tr>
<tr><td>Reading</td><td colspan="2"></td></tr>
<tr><td>Exam 1</td><td></td><td>Final (35%)</td></tr>
</table>

Week	Lecture Topics	Tutorial Topics	Deliverables	
01	ML Pipeline			
02	Paradigms of ML and kNN			Mini-Team Formation (due Sun, 22 Aug)
03	Decision Trees	Paradigms and kNN	Assignment (Sun, 5 Sep)	Team Formation (by staff; Sun, 29 Aug)
04	Linear Models	Decision Trees		
05	Bias and Variance	Linear Models		Project Proposal (Sun, 12 Sep)
06	Regularization and Validation	Bias and Variance, Regularization and Validation (2 topics)		Proposal Peer Critique (Sun, 19 Sep)
Recess				
07	Evaluation Metrics		1-hour Midterm (Mon, 27 Sep)	*Project*

Week	Lecture Topics	Tutorial Topics	Deliverables	
08	Preprocessing and Visualization	Evaluation Metrics		*Project*
09	Perceptron and Neural Networks	Preprocessing and Visualization		*Project*
10	Deep Learning	Perceptron and Neural Networks		Individual Team Project Consulting (3 weeks, by appointment)
11	Unsupervised ML	Deep Learning		
12	Open Project Consulting	Unsupervised ML		
13	ML Ethics and Revision	Practice Exams		*Project*
Reading				
Exam 1			2-hour Final Assessment (Wed, 24 Nov)	Recorded Project Presentation (Sun, 28 Nov)

Course
Policy
CS3244 Machine Learning
NUS
Department of Computer Science
School of Computing

Academic Integrity:
Freedom of Information Rule

Collaboration is acceptable and encouraged.

To assure that all collaboration is on the level,
you must always fill in the name(s) of your collaborators on your assignments and project work.

You will be assessed for the parts for which you claim is your own contribution.

Academic Independence

You are free to meet with fellow students(s) and discuss assignments with them.

Writing on a board or shared piece of paper is acceptable during the meeting; however, you **may not take any written (electronic or otherwise) record away from the meeting**.

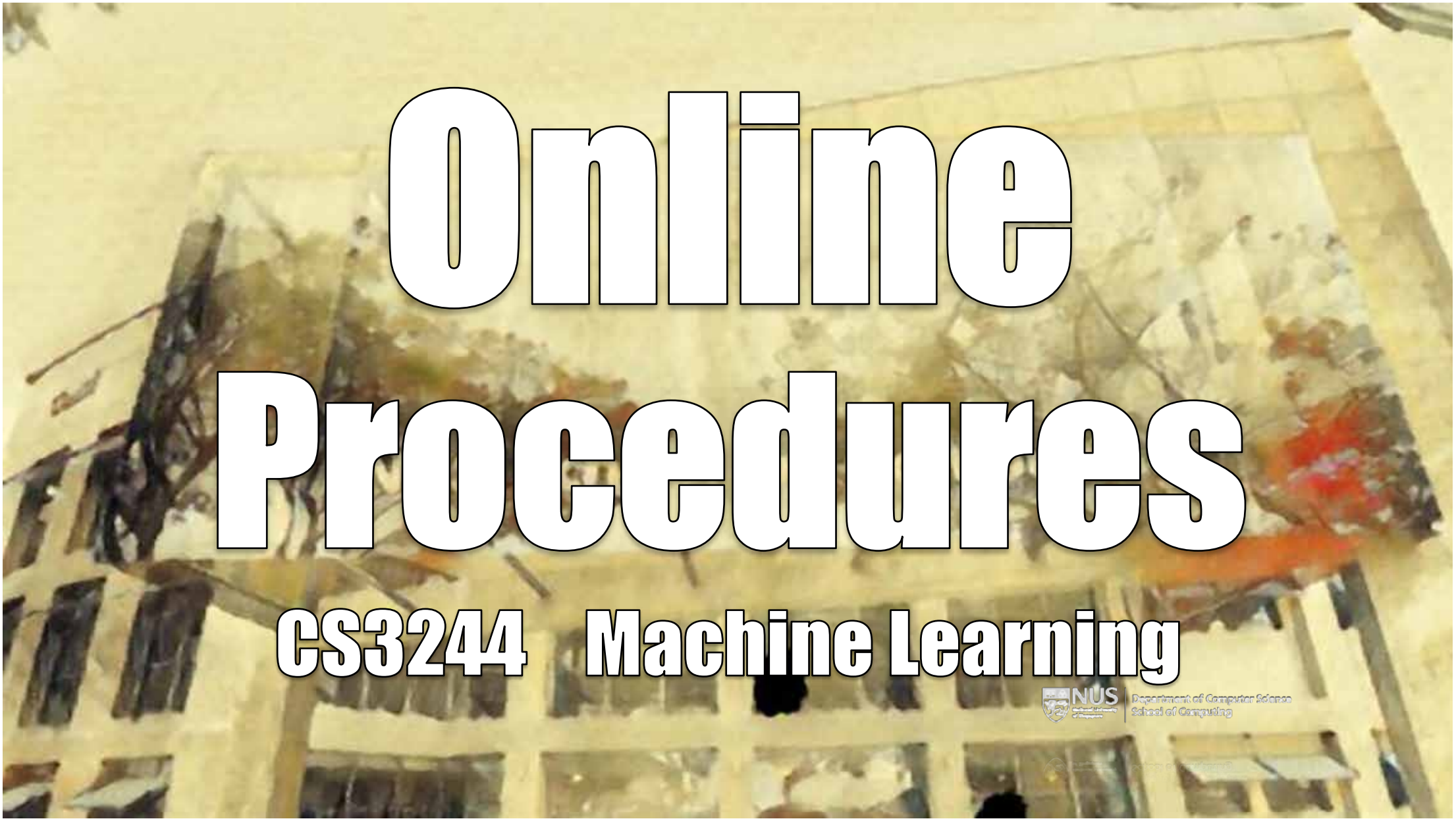
After the meeting, do something else for at least a half-hour (nap, TikTok, or doing an assignment for a different class), before working on the assignment.

> This will assure that you are able to reconstruct what you learned from the meeting, by yourself, using your own brain.
>
> You will be asked to certify that you meet this requirement per assessment.

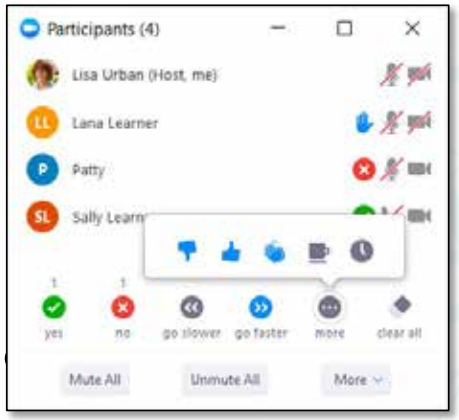
Punctuality (Lateness)

5% of your grade for every day of lateness; starts at midnight by LumiNUS' records; capped to 40% as a maximum deduction.

This is a large class, and grading your assessments takes significant time. For us to release feedback on time to help you troubleshoot accordingly, we need you to be timely.


Online
Procedures
CS3244 Machine Learning
NUS
Department of Computer Science
School of Computing

Zoom Etiquette for Tutorial



Each tutorial will have a tutorial leader.

Have your video on where possible, especially durin
breakout activities.

Leaders will be giving the tutorial, but occasionally other tutors and staff will sit in to observe the tutor (not you :D)

Be familiar with polling, changing your name, screen annotation, and chat.

Zoom Exam Procedure

Midterm on Monday of Week 07;
Final on Wednesday of Exam Week 1 (Week 15).

Each exam is multipart, where you may not revisit the previous parts.
Short washroom break between parts.

Proctoring by standard e-learning protocol: Zoom gallery and ffmpeg desktop recording.

- Zoom: set up a side camera or external webcam and connect to the proctor.
- ffmpeg: record your single screen desktop and upload to LumiNUS.
- There'll be a physical option in i3 Auditorium.