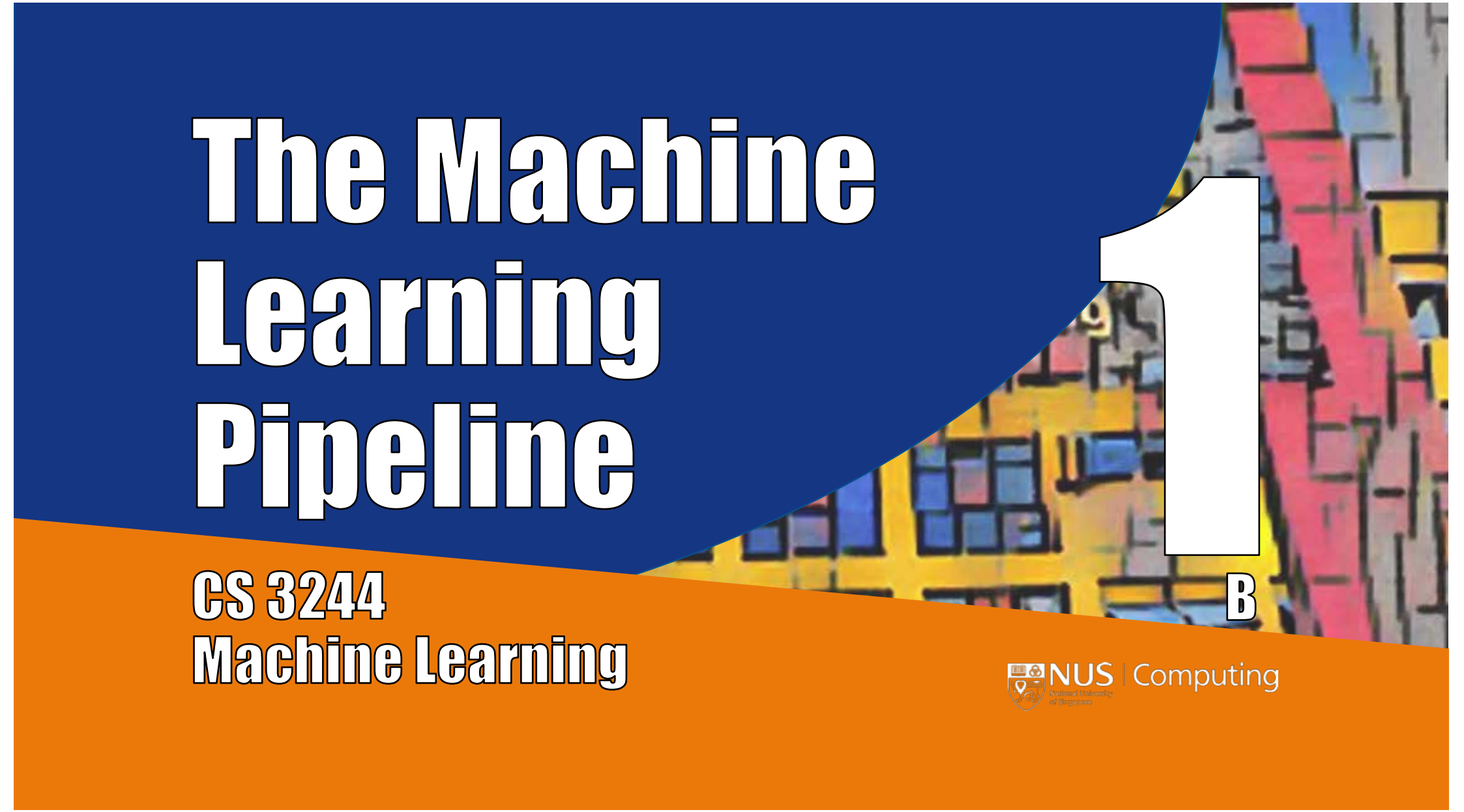
The Machine Learning Pipeline



1B

CS 3244
Machine Learning

NUS | 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

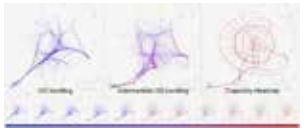
IoT Sensors | Health Behavior Change | Data Analytics

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

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

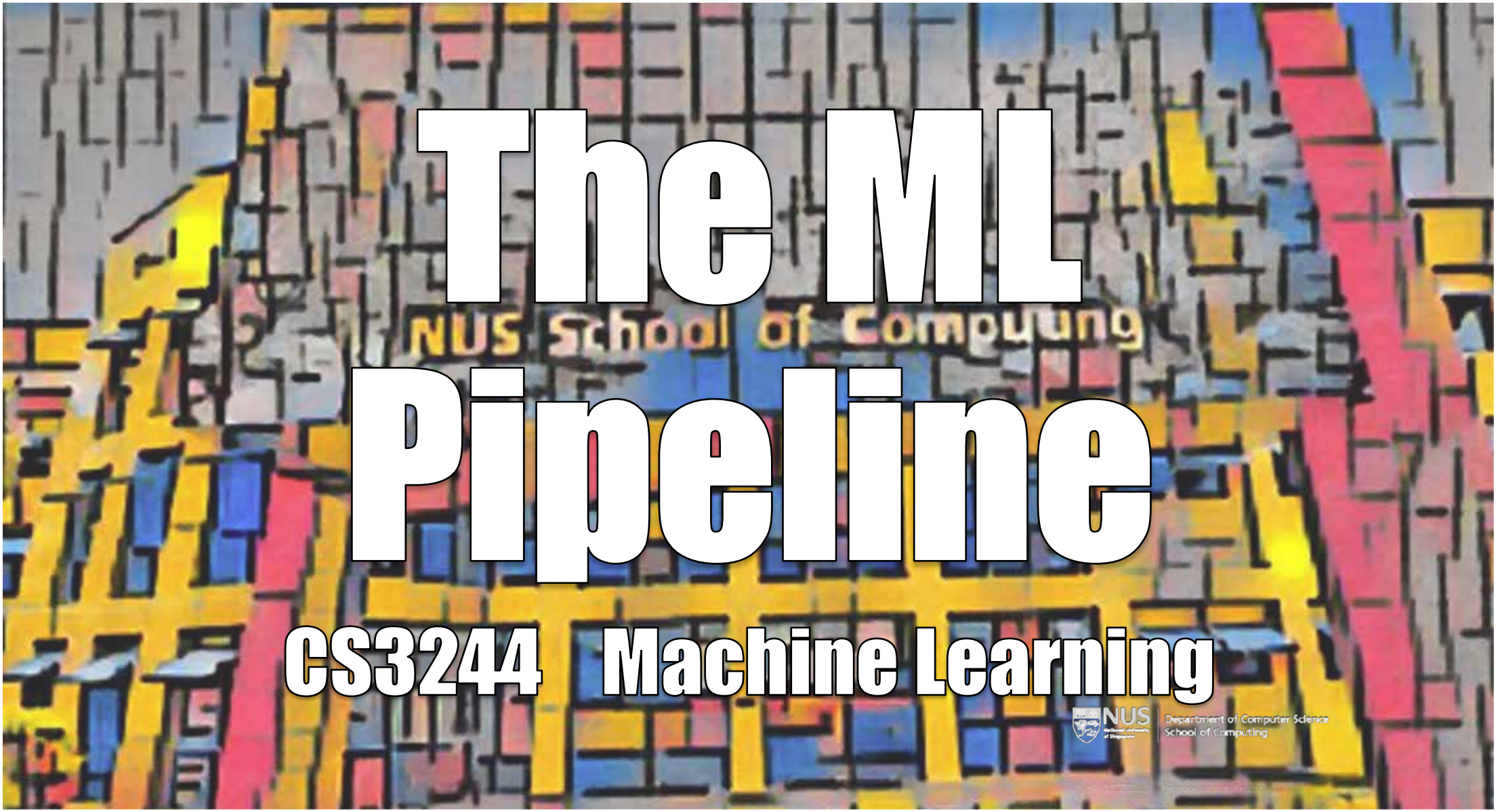
slack cs3244-2110.slack.com

You'll need this for this lecture.

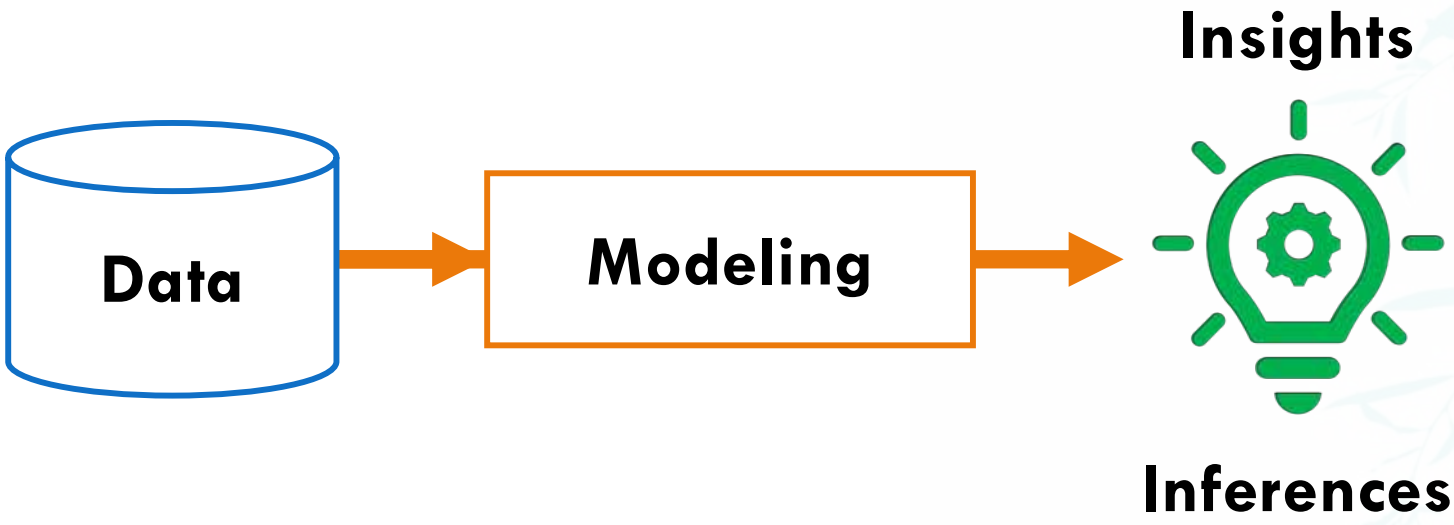
https://bit.ly/cs3244-join-slack

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

👉 Use your LumiNUS name for your display name.


The ML
Pipeline
CS3244 Machine Learning
NUS School of Computing
NUS
Department of Computer Science
School of Computing

Machine Learning Pipeline



What do you think a ML Pipeline is? What is its purpose?

Go to Slack, on the #general channel

Post an answer to our thread. A short 1-2 sentence or phrase will do.

If you'd prefer just to upvote another's answer, feel free to do that (too).

In-Lecture Activity: Noteboard

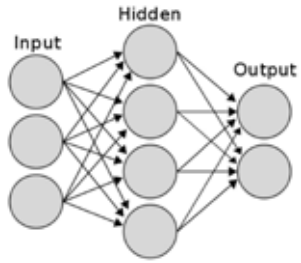
What is a model?

$y = ax + b$

Equation (Line)

```
If Rain, then Stay Home
Else Go Out
```

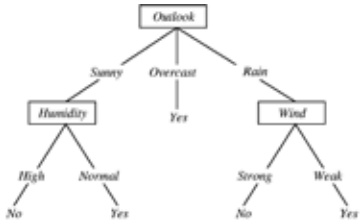
Code? Rules



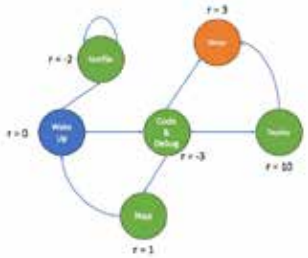
Neural Network (NN)

$y = ax^2 + bx + c$

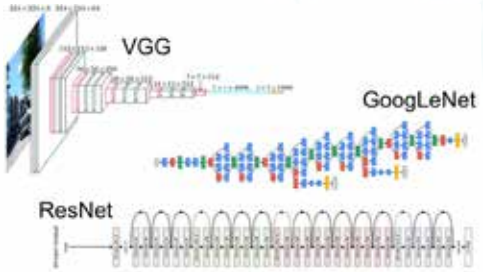
Equation (Nonlinear)



Decision Tree



Bayesian Graph



Deep NN

A **model** is a mathematical representation of a behavior

Machine Learning Pipeline

1. Data collection
2. Data extraction (Feature engineering)
3. Data understanding (with Visualization)
4. Data pre-processing
5. Model choice / design
6. Model training
7. Model validation (Evaluation)
8. Model understanding (Visualization / Explainability)
9. Model deployment

Machine Learning Pipeline

1. Data collection
2. Data extraction (Feature engineering)
3. Data understanding (with Visualization)
4. Data pre-processing
5. Model choice / design
6. Model training
7. Model validation (Evaluation)
8. Model understanding (Visualization / Explainability)
9. Model deployment

Machine Learning Pipeline

1. Data collection
2. Data extraction (Feature engineering)
3. Data understanding (with Visualization) [W08]
4. Data pre-processing [W08]
5. Model choice / design [W01–06, 10–12]
6. Model training [with Step 5]

"Machine learning"

7. Model validation (Evaluation) [W07]
8. Model understanding (Explainability [W10] / Visualization [W08])
9. Model deployment

Machine Learning Pipeline

5. Model choice / design [W01–06, 10–12]
   - Paradigms of ML
     - Supervised Learning
     - Unsupervised Learning
     - ~~Reinforcement Learning~~
   - Bias and Variance

Machine Learning Pipeline

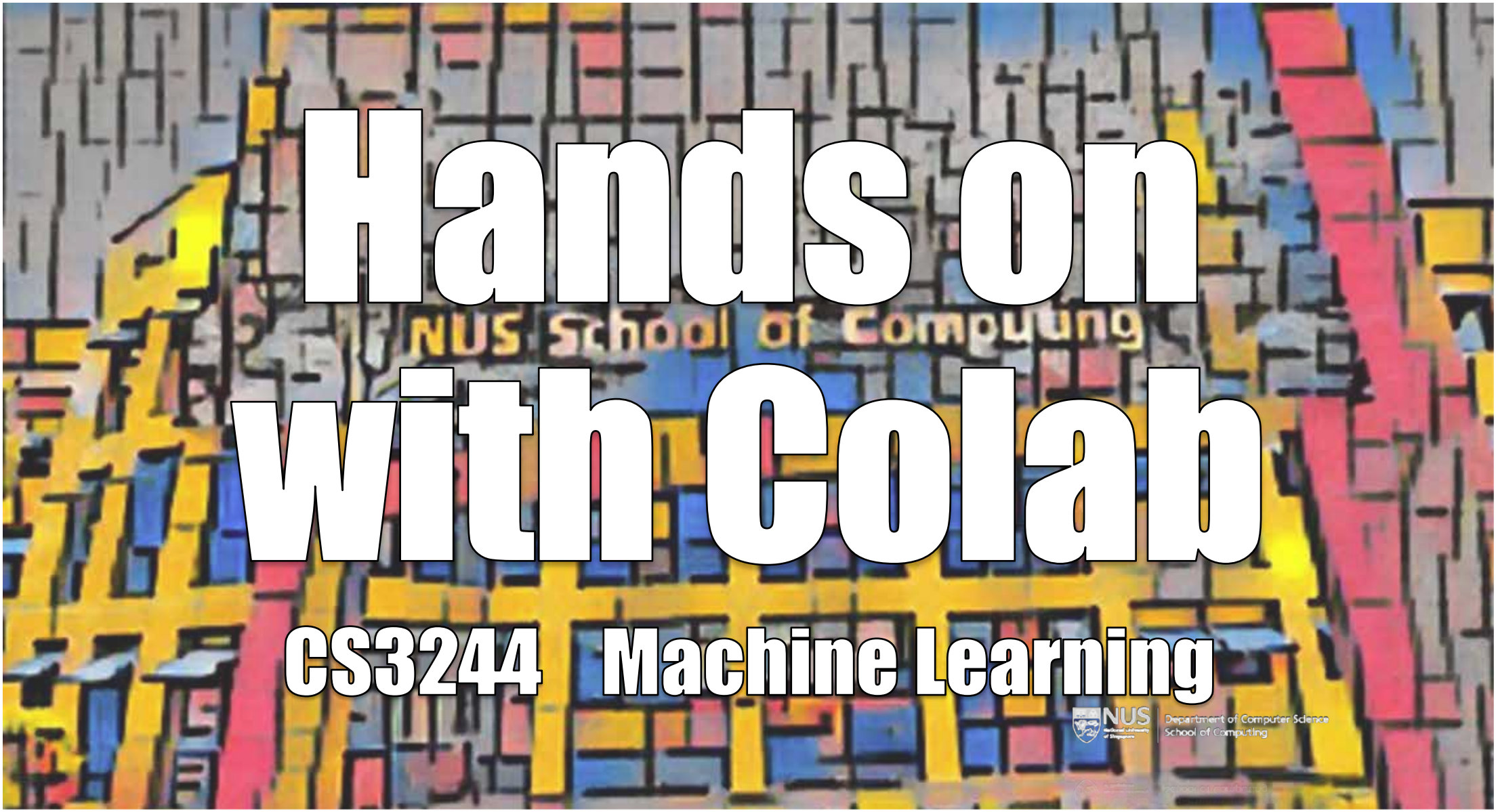
5. Model choice / design
   - Paradigms of ML [W02]
     - Supervised Learning [W02…]
     - Unsupervised Learning [W11]
     - ~~Reinforcement Learning~~
   - Bias and Variance [W05]

Machine Learning Pipeline

5. Model choice / design
    - Paradigms of ML [W02]
        - Supervised Learning
            - Decision trees [W03]
            - Linear models [W04]
            - Support Vector Machines [W04]
            - Perceptron and Neural Networks [W09]
            - Deep Learning [W10]
        - Unsupervised Learning [W11]
    - Bias and Variance [W05]

(Iterative) Machine Learning Pipeline

1. Data collection
2. Data extraction (Feature engineering)
3. Data understanding (with Visualization)
4. Data pre-processing
5. Model choice / design
6. Model training
7. Model validation (Evaluation)
8. Model understanding (Visualization / Explainability)
9. Model deployment


Hands on
with Colab
CS3244 Machine Learning
NUS School of Computing
NUS
Department of Computer Science
School of Computing

Hands-On ML Model w/ Colab (30 minutes)

Let's Go!

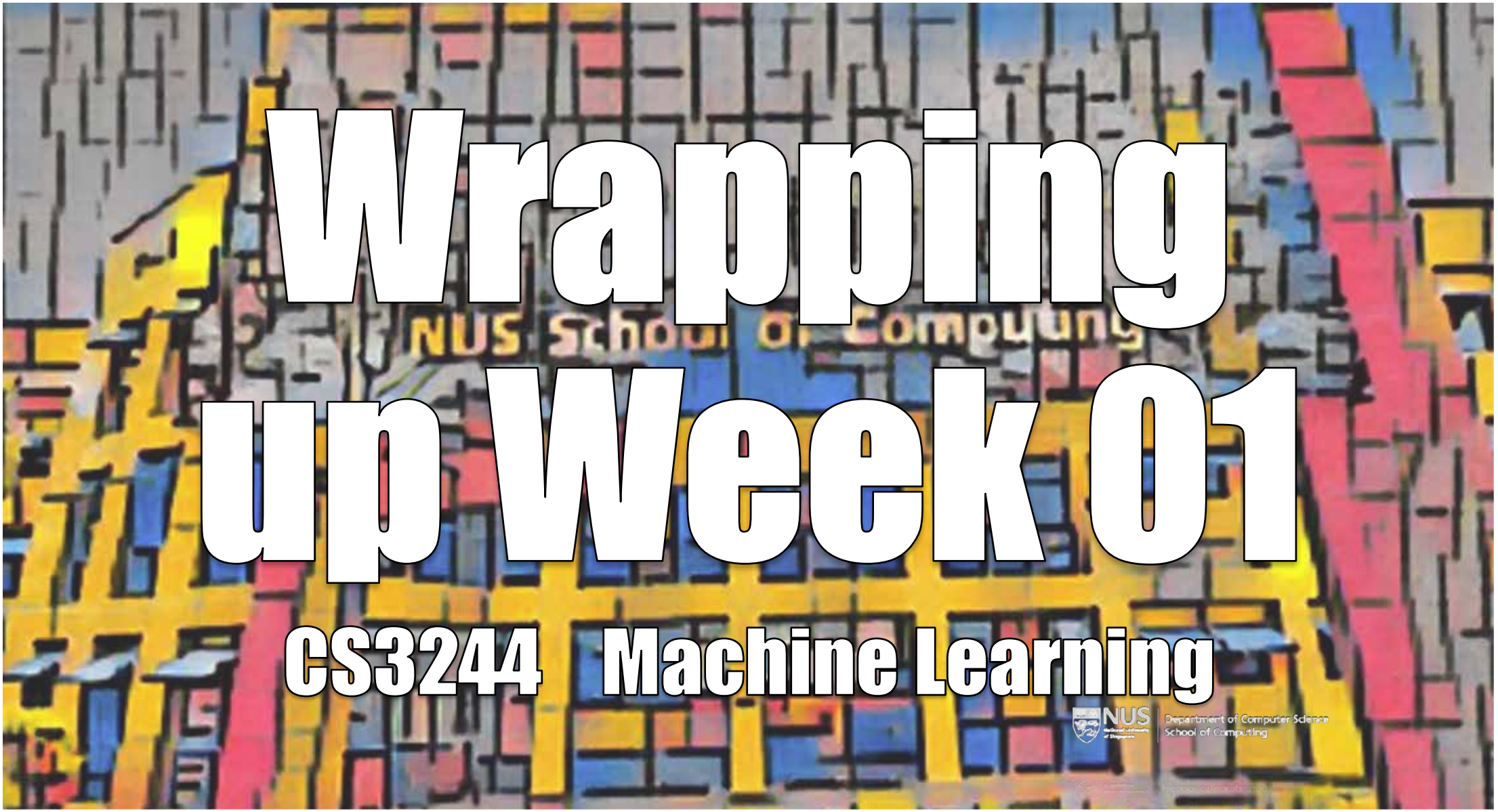
http://www.comp.nus.edu.sg/~cs3244/AY2122S1/01.colab.html

What did we learn?

We ran a ML classifier over a famous dataset, MNIST, for digit classification.

Procedurally, we also had a first try at:

- Using Google Colab.
- Using LumiNUS Quiz to submit your answers to in-Colab exercises.


Wrapping
up Week 01
CS3244 Machine Learning
NUS
Department of Computer Science
School of Computing

What did we learn 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

Outlook for next week

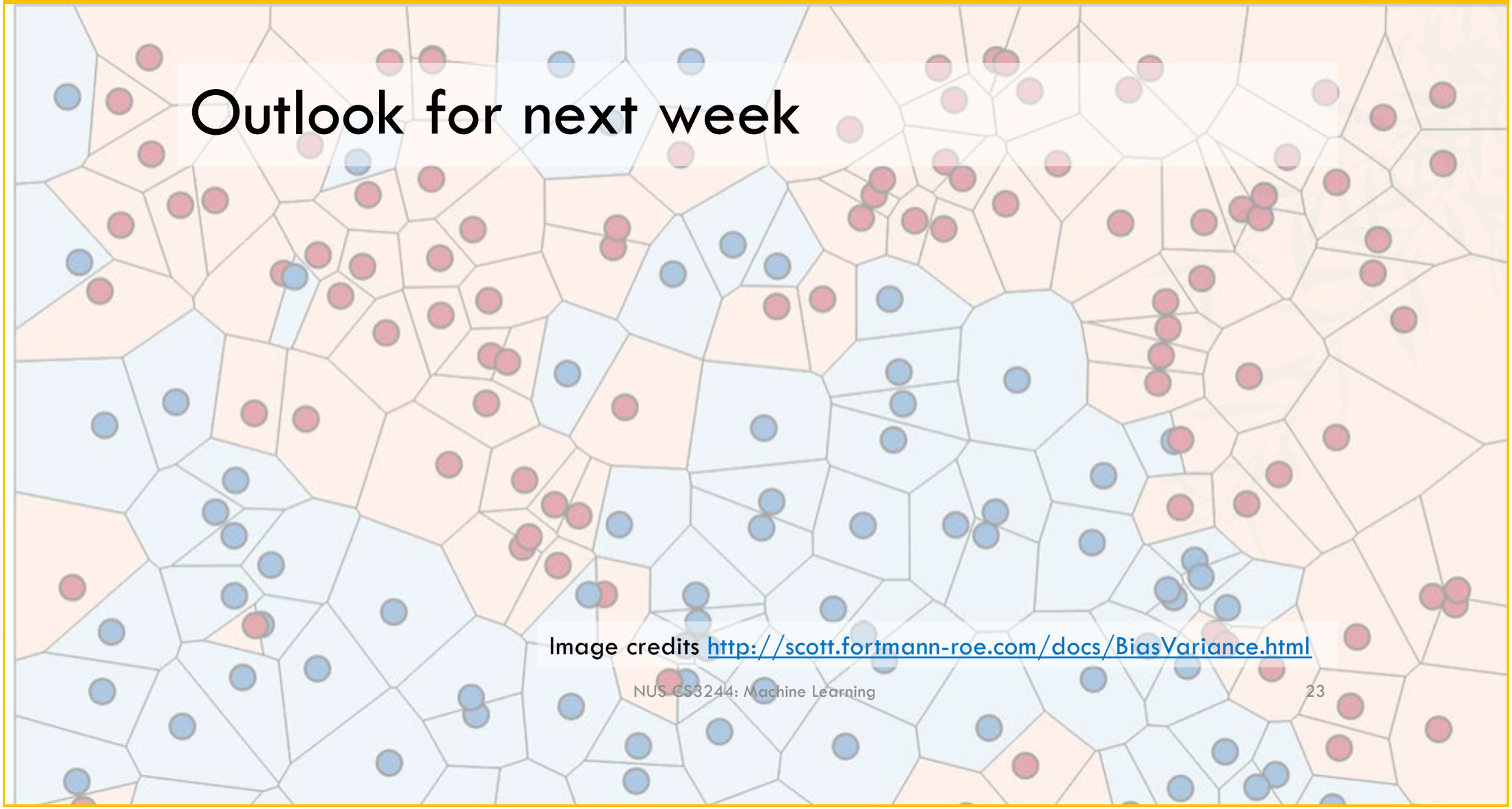


Image credits http://scott.fortmann-roe.com/docs/BiasVariance.html

Assigned Readings (due before next Mon)

First, let's do some reading to set the stage:

- Paradigms of ML: https://medium.datadriveninvestor.com/learning-paradigms-in-machine-learning-146ebf8b5943 (5 minutes)

- 5 Tribes of ML: https://medium.com/dummykoders/5-tribes-of-ai-the-perspective-6b9e0334638c (10 minutes)

Assigned Task (due before next Mon)

Let's intersect the two readings together!

Post a 1-2 sentence (not long!) answer to the below question in your Slack tutorial group #tg-xx or #tg-na (if you don't have a tutorial group yet).

Give an example of an ML application (a URL to it is good).
Describe which paradigm and tribe you think it belongs to.

[Don't worry if you're not sure, we'll cover this again in Week 02.]