COMP 7180
Quantitative Methods for Data Analytics and Artificial Intelligence

Lecture 0 – Course Introduction

Course Lecturers: Dr. LIU, Yang and Dr. HAN, Bo
Teaching Assistant: Mr. LI, Minghao

Outline

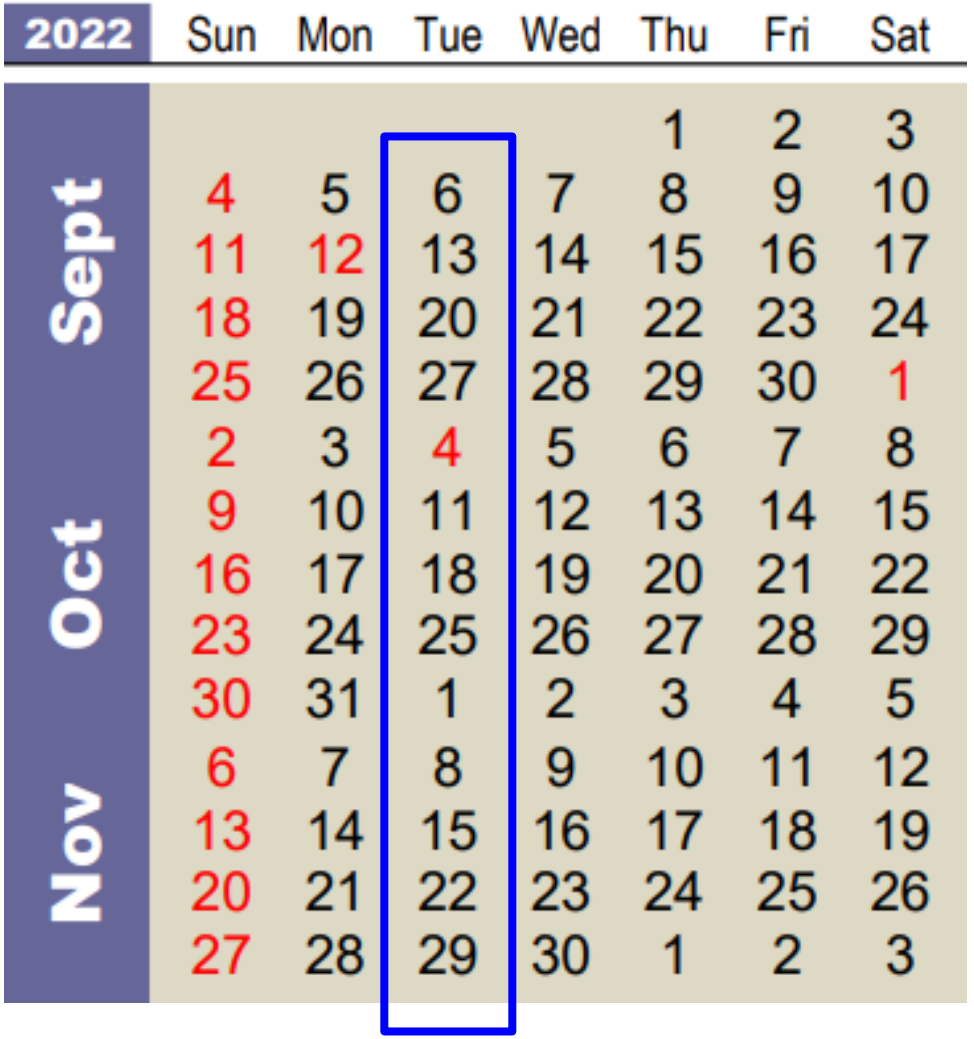
- Introduction to COMP7180
  - General information
  - Course syllabus
  - Expected learning outcomes
  - Assessment methods
- Introduction to Quantitative Methods in AI and Machine Learning

Outline

- Introduction to COMP7180
  - **General information**
  - Course syllabus
  - Expected learning outcomes
  - Assessment methods
- Introduction to Quantitative Methods in AI and Machine Learning

Timetable

- Time of Our Class
  - 12 Lectures: *Sept. 6 – Nov. 29 (every Tuesday) 6:30pm – 9:30pm*
  - 1 Public Holiday: *Oct. 4*
- Classroom of our lectures
  - OEE1017



2022	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Sept					1	2	3
	4	5	6	7	8	9	10
	11	12	13	14	15	16	17
	18	19	20	21	22	23	24
	25	26	27	28	29	30	1
Oct	2	3	4	5	6	7	8
	9	10	11	12	13	14	15
	16	17	18	19	20	21	22
	23	24	25	26	27	28	29
	30	31	1	2	3	4	5
Nov	6	7	8	9	10	11	12
	13	14	15	16	17	18	19
	20	21	22	23	24	25	26
	27	28	29	30	1	2	3

About the Lecturers

- Dr. LIU, Yang (Me)
  - Assistant Professor in Department of Computer Science
  - Postdoc in **Statistics** @Yale University; Visiting Scholar in **Robotics** @ Carnegie Mellon University; PhD in **Machine Learning** @HK PolyU
  - Brain modeling, infectious disease modeling, visual data analysis, linear algebra, multilinear algebra, …
- Course duty
  - Subject design; lecture teaching; quiz and final examination grading; overall performance evaluation
- Contact information
  - Email: csygliu@comp.hkbu.edu.hk
  - Office: **RRS729**
  - Office Hours: Monday 2:30pm – 4:30pm or by appointment

About the Lecturers

- Dr. HAN, Bo
  - Assistant Professor in Department of Computer Science
  - Visiting Scientist @ RIKEN Center for **Advanced Intelligence** Project; PhD in **Machine Learning** @ University of Technology Sydney
  - Deep learning, representation learning, probabilistic reasoning, optimization, …
- Course duty
  - Subject design; lecture teaching; quiz and final examination grading; overall performance evaluation
- Contact information
  - Email: bhanml@comp.hkbu.edu.hk
  - Office: **DLB640**

About the Teaching Assistant

- Mr. LI, Minghao
  - PhD student in Department of Computer Science
  - Artificial Intelligence, Machine Learning, etc.
- Course duty
  - Assignment grading
  - Tutorial delivering
- Contact information
  - Email: csmhli@comp.hkbu.edu.hk
  - Office: **FSC1000**

Outline

- Introduction to COMP7180
  - General information
  - **Course syllabus**
  - Expected learning outcomes
  - Assessment methods
- Introduction to Quantitative Methods in AI and Machine Learning

Course Syllabus

- Part I (first six weeks, Dr. LIU, Yang)
  - Linear Algebra
  - Univariable and Multivariable Differentiation and Calculus
- Part II (second six weeks, Dr. HAN, Bo)
  - Probability and Statistics
  - Optimization

Course Content of Part I

- Linear algebra
  - Basic vector and matrix operations
  - Matrix properties: trace, rank, range, and determinant
  - Eigenvalues and eigenvectors
  - Principal Component Analysis
  - Singular Value Decomposition

Course Content of Part I

- Univariable and Multivariable Differentiation and Calculus
  - Introduction to artificial intelligence and machine learning
  - Partial derivatives and gradients
  - Multivariable chain rule
  - Jacobian and Hessian matrices

Course Content of Part II

- Probability and Statistics
  - Conditional probability and independence
  - Discrete and continuous random variables
  - Expectation and variance
  - Multiple random variables
  - Maximum likelihood estimation
  - Regression analysis

Course Content of Part II

- Optimization
  - Mathematical optimization
  - Convex sets and convex functions
  - Least squares and convex optimization
  - Gradient descent methods

Resources

- Online materials
  - MIT: Mathematics of Machine Learning: https://ocw.mit.edu/courses/mathematics/18-657-mathematics-of-machine-learning-fall-2015/
  - University of Maryland: Math for Machine Learning http://users.umiacs.umd.edu/~hal/courses/2013S_ML/math4ml.pdf
- Books
  - Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong. *Mathematics for Machine Learning*, 2019
  - Christopher M. Bishop, *Pattern Recognition and Machine Learning*, 2006.
  - Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*, 2003.

Outline

- Introduction to COMP7180
  - General information
  - Course syllabus
  - **Expected learning outcomes**
  - Assessment methods
- Introduction to Quantitative Methods in AI and Machine Learning

Expected Learning Outcomes

- Knowledge
  - Describe the essential concepts in linear algebra for DA&AI
  - Understand fundamental univariable and multivariable differentiation and calculus for DA&AI
  - Explain the essential concepts in probability and statistics for DA&AI
  - Understand the essential concepts in optimization for DA&AI

Expected Learning Outcomes

- Professional Skill
  - Determine suitable quantitative methods for effective data analytics
  - Apply suitable quantitative methods for real-world problem solving

Outline

- Introduction to COMP7180
  - General information
  - Course syllabus
  - Expected learning outcomes
  - **Assessment methods**
- Introduction to Quantitative Methods in AI and Machine Learning

Assessment Methods

- Continuous assessment (40%)
  - Assignments (2-4) + Quizzes (2)
  - Assess your mastery of the quantitative methods and their applications in DA&AI
- Examination (60%)
  - Final examination
  - Measure the extent to which you have reached all of the learning outcomes

Important Notices

- Plagiarism: Students who **plagiarized** and who **were plagiarized** will be given **zero mark and reported to the departmental exam committee for further penalty**
- Final Exam: In order to pass this course, students should attain **at least 30% of the final examination mark**
- This course will be **useful** but require **very hard work** ☺

Outline

- Introduction to COMP7180
  - General information
  - Course syllabus
  - Expected learning outcomes
  - Assessment methods
- **Introduction to Quantitative Methods in AI and Machine Learning**

Artificial Intelligence

- Artificial intelligence (AI) is the broad science of mimicking human abilities.
  - It aims to build smart machines capable of performing tasks that typically require human intelligence.
  - Interdisciplinary research: brain science, cognitive science, psychology, computer science, mathematics, …

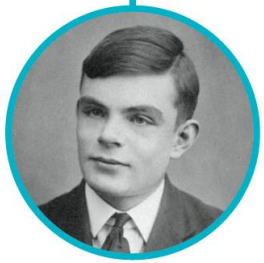
https://qbi.uq.edu.au/brain/intelligent-machines/history-artificial-intelligence

1642 — First mechanical calculating machine built by French mathematician and inventor Blaise Pascal.

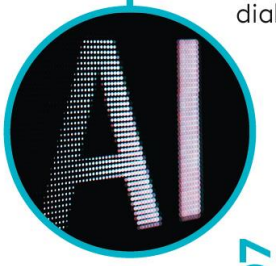
1837 — First design for a programmable machine, by Charles Babbage and Ada Lovelace.

1943 — Foundations of neural networks established by Warren McCulloch and Walter Pitts, drawing parallels between the brain and computing machines.

1950 — Alan Turing introduces a test—the Turing test—as a way of testing a machine's intelligence.



1955 — 'Artificial intelligence' is coined during a conference devoted to the topic.



1965 — ELIZA, a natural language program, is created. ELIZA handles dialogue on any topic; similar in concept to today's chatbots.

1980s — Edward Feigenbaum creates expert systems which emulate decisions of human experts.

1997 — Computer program Deep Blue beats world chess champion Garry Kasparov.

2002 — iRobot launches Roomba, an autonomous vacuum cleaner that avoids obstacles.

2009 — Google builds the first self-driving car to handle urban conditions.

2011 — IBM's Watson defeats champions of US game show Jeopardy!

2011–2014 — Personal assistants like Siri, Google Now, Cortana use speech recognition to answer questions and perform simple tasks.

2014 — Ian Goodfellow comes up with Generative Adversarial Networks (GAN).

2016 — AlphaGo beats professional Go player Lee Sedol 4-1.

2018 — Most universities have courses in Artificial Intelligence.

Machine Learning

- Machine learning (ML) is a specific and modern subset of AI that trains a machine to learn.
- Learning = Improving with experience at some tasks

-- Tom Mitchell, Machine Learning, 1997

  - Improve over task *T*
  - With respect to performance measurement *P*
  - Based on experience *E*

Why AI and ML

In one day, the entire content generated by the Internet can be engraved with **168 million** DVDs; there are as many as **294 billion** emails

Everyday, more than **500 million** images will be uploaded; every minute, **20 hours** of video will be shared in the internet

The Shanghai medical information platform has accumulated over **1400 TB** of electronic medical and health data covering **39 million** people

The Guangzhou service platform has recorded more than **1.2 billion** new traffic operations per day, the daily data volume has reached **150-300 GB**

Why AI and ML

- Big Data ≠ Big Knowledge/Value ?
  - Data heterogeneity, redundancy, noise, …
- Have to extract information from big data in an effective way
  - Human being: powerful learning capacity but limited computational and storage capacity
  - Computer/Machine: powerful computational and storage capacity
    - **So the question is: how to make it capable of learning as human being?**

Real Applications of AI and ML

- Clothing (衣): clothing brand recommendation
  - Task (T): recommending suitable apparel brands to users
  - Performance Measurement (P): user satisfaction/rating
  - Experience (E): user preferences, browsing/purchase records; apparel appearance, material; brand style, popularity, etc.

Real Applications of AI and ML

- Food (食): food safety monitoring
  - Task (T): assessing the food safety rating
  - Performance Measurement (P): false-alarm rate
  - Experience (E): restaurant credit, quality inspection information, food ingredients, etc.

Real Applications of AI and ML

- Housing (住): forecast of house price trends
  - Task (T): forecasting the trend of housing prices in the next 3-5 years
  - Performance Measurement (P): The difference between the predicted price and the actual price
  - Experience (E): House prices and related information for the past 10-20 years

Real Applications of AI and ML

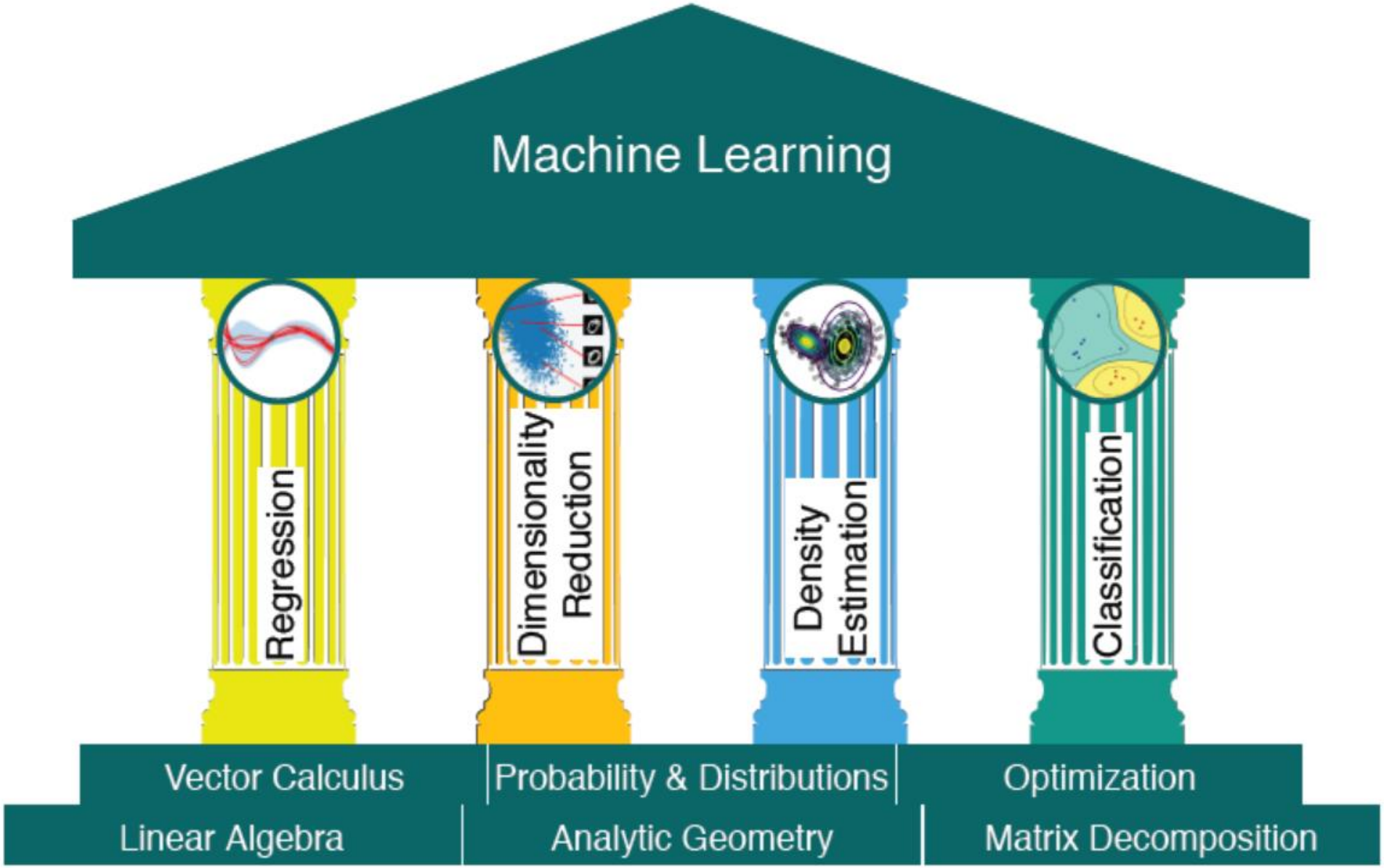
- Transportation (行于): self-driving system
  - Task (T): autonomous driving based on the vision sensors
  - Performance Measurement (P): continuous safe driving distance/time
  - Experience (E): Road condition information and the corresponding operation instructions when a human driver drives

More Examples

- Multimedia content understanding
  - Face/fingerprint recognition, movie editing, music therapy
- Natural language processing
  - Machine translation, information retrieval, smart input
- Bioinformatics
  - Gene function prediction, protein sequence alignment, gene regulation
- Network security
  - Spam filtering, virus detection

How Does AI/ML Work

- Given a real-world problem/task
- Describe/formulate the problem using computer logic/language
- Select/build an AI/ML model to optimize the objective function (usually the performance measurement)
- Use **quantitative/mathematical methods** to find the solution of the model, based on the data/experience
- Refine the model until satisfactory
- Apply the model in the real-world problem/task



Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong. Mathematics for Machine Learning, 2019

Why Worry About Math

- There are lots of easy-to-use machine learning packages out there !

HOWEVER

- To get really useful results, you need good mathematical intuitions about certain general machine learning principles, as well as the inner workings of the individual algorithms.

Why Worry About Math

- These intuitions will allow you to:
  - Choose the right algorithm(s) for the problem
  - Make good choices on parameter settings, validation strategies
  - Recognize over- or underfitting
  - Troubleshoot poor / ambiguous results
  - Put appropriate bounds of confidence / uncertainty on results
  - Do a better job of coding algorithms or incorporating them into more complex analysis pipelines

Fundamental Role of Math in AI and ML

Differentiation

Stock Price Forecast

$$Q = \sum_{i=1}^{n} \left(Y_i - \hat{Y}\right)^2 = \sum_{i=1}^{n} \left(Y_i - a - bX_i\right)^2$$

$$\frac{\partial Q}{\partial a} = \sum_{i=1}^{n} -2\left(Y_i - a - bX_i\right) = 0$$

$$\frac{\partial Q}{\partial b} = \sum_{i=1}^{n} -2X_i\left(Y_i - a - bX_i\right) = 0$$

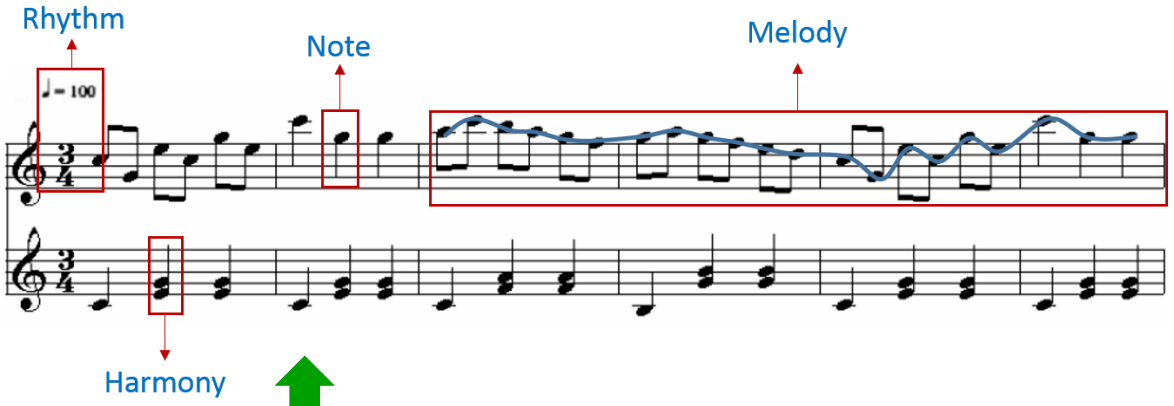
$$a = \overline{Y} - b\overline{X} \qquad b = \frac{\sum_{i=1}^{n}\left(X_i Y_i\right) - n\overline{X}\,\overline{Y}}{\sum_{i=1}^{n}\left(X_i^2\right) - n\overline{X}^2}$$

https://images.app.goo.gl/Yd1hSAQGENYPc6N1A

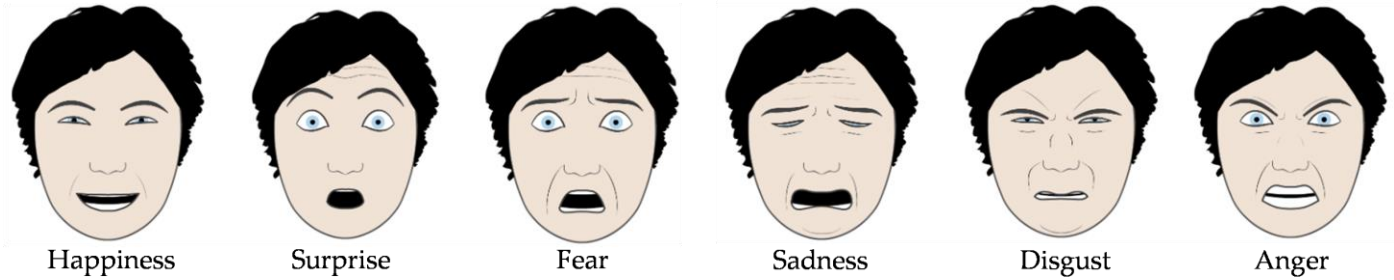
Fundamental Role of Math in AI and ML

Linear Algebra and Matrix Analysis

Music Therapy via Emotion Analysis



Music Emotion Recognition

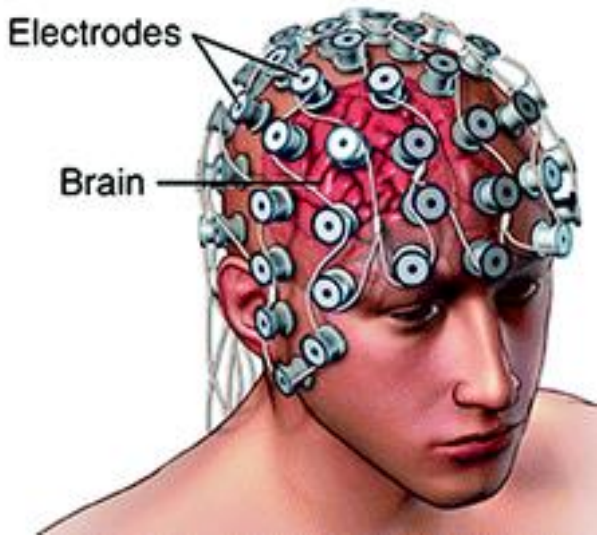


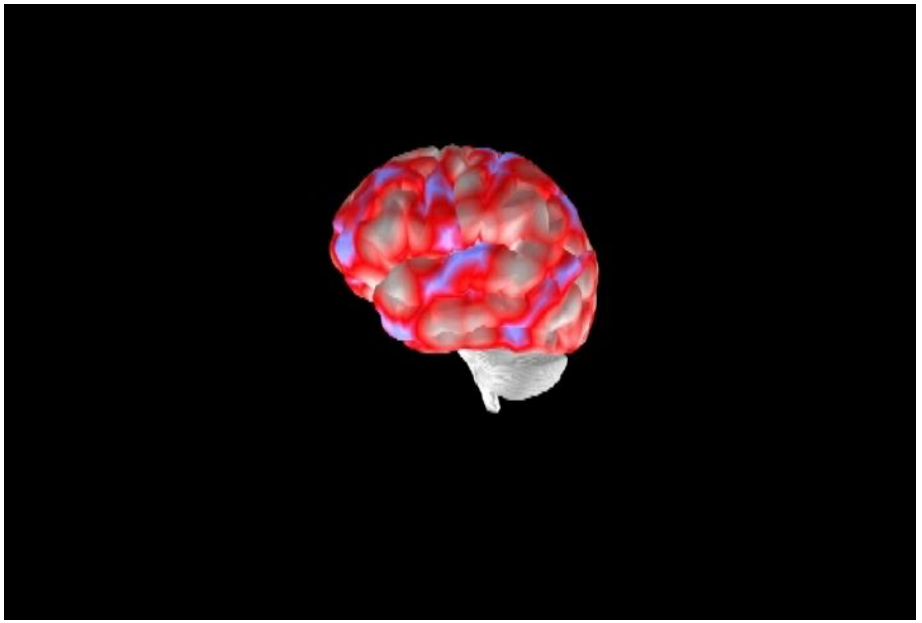
$$\mathbf{U} = \arg\min_{\mathbf{U}} \sum_{i=1}^{n} \sum_{j=1}^{n} ||\mathbf{U}^T\mathbf{x}_i - \mathbf{U}^T\mathbf{x}_j||^2 \cdot \langle \mathbf{y}_i/||\mathbf{y}_i||, \mathbf{y}_j/||\mathbf{y}_j|| \rangle$$

$$\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{u}_i = \lambda_i \mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{u}_i$$

https://images.app.goo.gl/Sm24P5mSEezhNJvA8

Our Experience





https://images.app.goo.gl/cSfAuvCiUSUzsjZK9

Our Experience

- Participants
  - Suffered from chronic pain, defined as any non-cancer pain that persists for at least 3 months
- With regularly listening to the Chinese *Pipa* music (春江花月夜) for 4 weeks
  - Statistically significant decrease in anxiety score
  - Less analgesic use was demonstrated

Fundamental Role of Math in AI and ML

Calculus and Probabilistics

Malaria (疟疾) Elimination

- One of the most endemic and life-threatening global health problems
  - ~50% of the world's population at risk
  - 200+ million cases of malaria
  - ~400,000 deaths

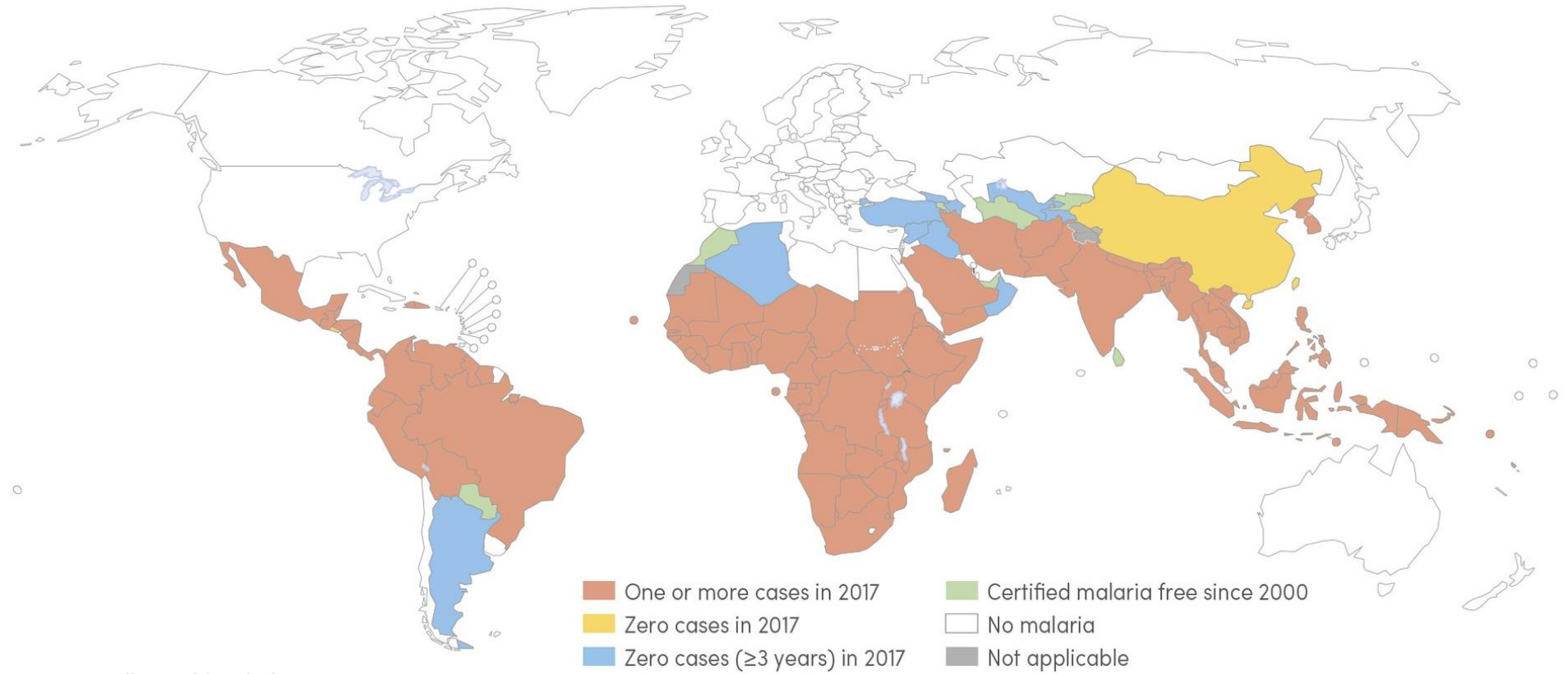
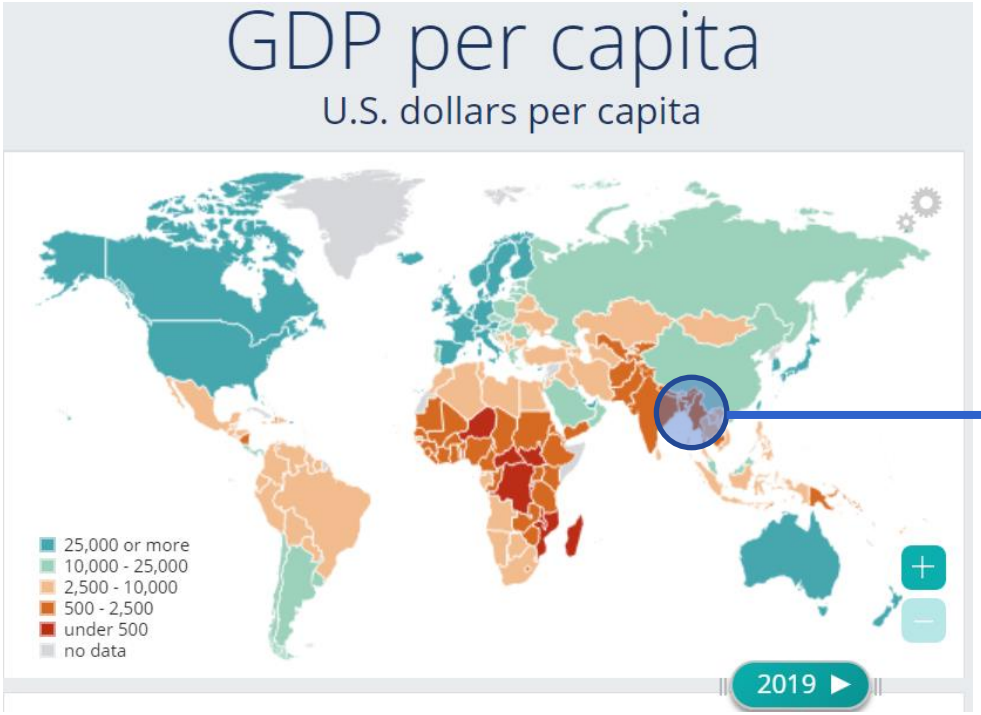
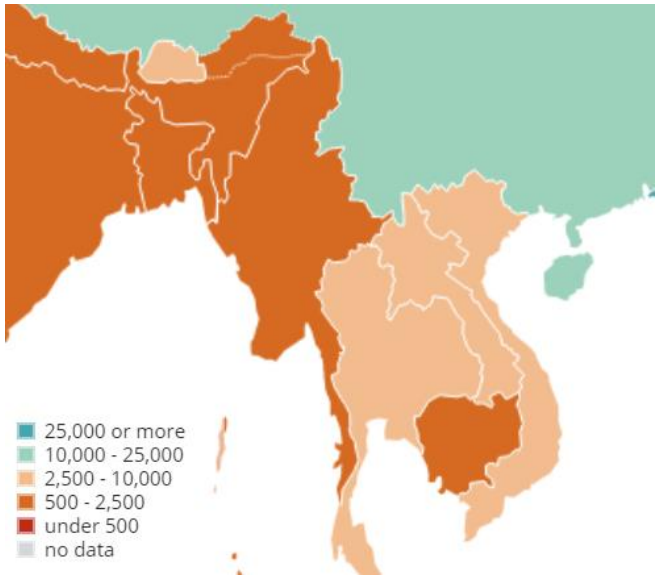


Image Credit: World Malaria Report 2018

Our Focus: Malaria Control and Prevention in GMS/SEA

under Extremely Limited Resources





https://www.imf.org/external/datamapper/NGDPDPC@WEO/OEMDC/ADVEC/WEOWORLD/NPL

Extremely Limited Resources

Example: Tengchong, a city located in China-Myanmar border

- **5845 square kilometers** = 5 times bigger than HK
- Border: **148 kilometers** with irregular and frequent migrant mobility (more than **1 million** per year)
- Total population: **682,700**
- Most of the areas are hard-to-reach areas. Takes more than **2 hours from one town to another**.
- In **18 counties**, **4 local CDC staffs** possess the know-how to perform complete disease diagnosis, case investigation, and epidemic treatment

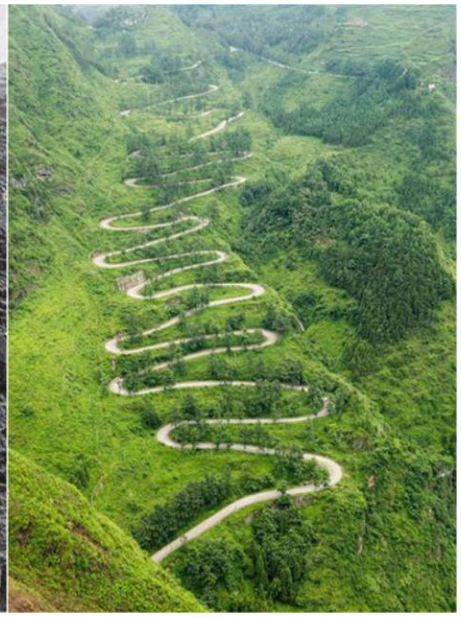
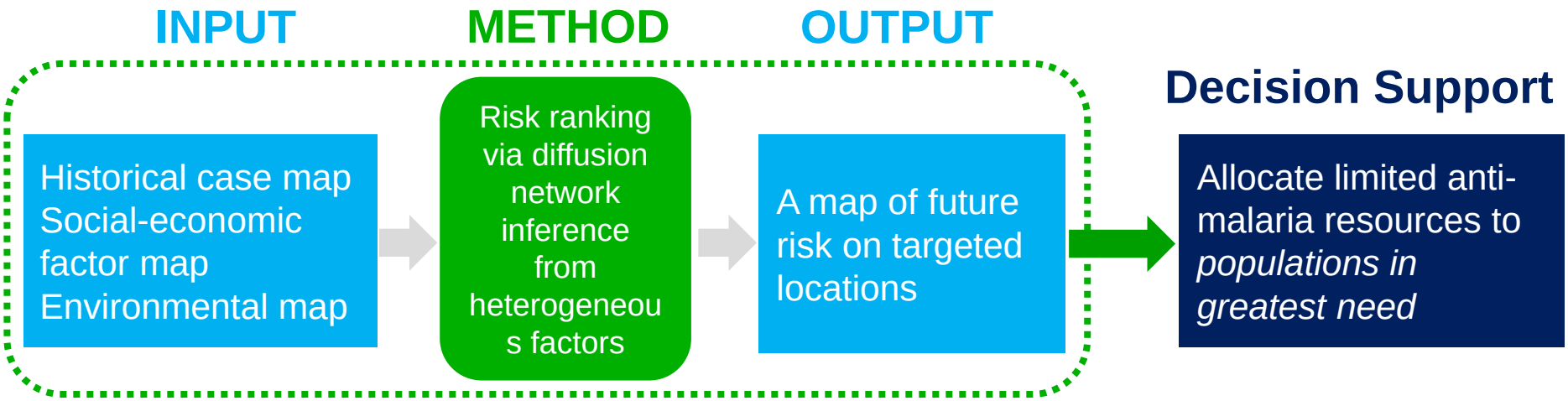


Image Credit: http://slide.history.sina.com.cn/w/slide_61_40602_61631.html/d/1#p=1
http://www.aminoacid-jirong.com/news/9j2zqfq9szs29c9.html

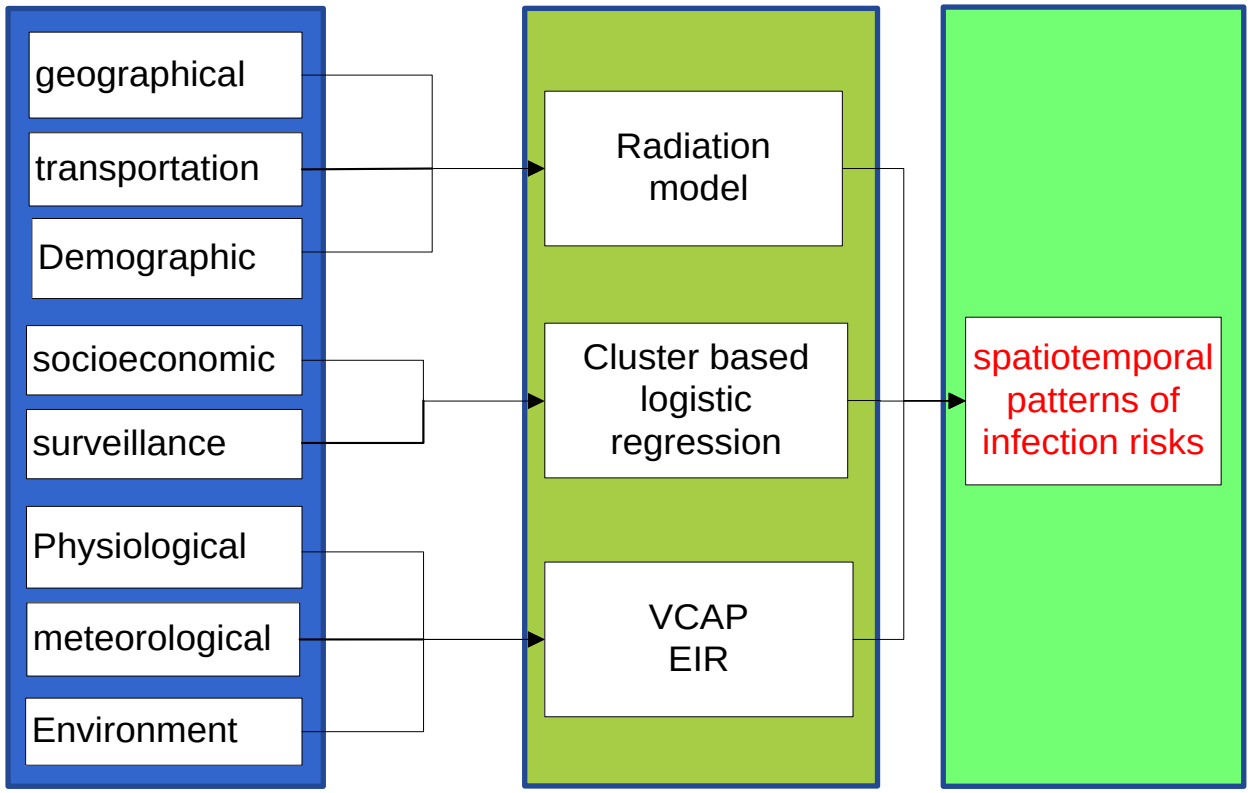
How to *Effectively* Control?

Objective

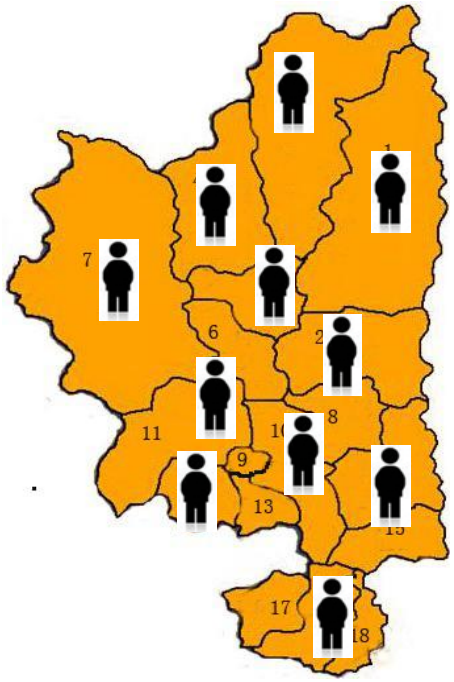
To predict geographical locations and population groups with greatest incidence of malaria cases



Infection Risk Prediction with Heterogeneous Data



When and Where



$$p(\boldsymbol{y}_*|\boldsymbol{\Phi}_*^{\boldsymbol{s}}, \boldsymbol{y}, \boldsymbol{\Phi}) = \int p(\boldsymbol{y}_*|\boldsymbol{\Phi}_*^{\boldsymbol{s}}, \boldsymbol{s}) p(\boldsymbol{s}|\boldsymbol{y}, \boldsymbol{\Phi}) d\boldsymbol{s}$$

Our Experience

Deployed and Implemented our AI-enabled malaria control and prevention strategy in Tengchong, one of the most malaria endemic counties in China

Before

During 2010-2012, the number of reported cases and the incidence rate in Tengchong are the **highest** in China

After

In 2016, Tengchong was the **first** in the 18 China-Myanmar border counties to obtain the malaria **elimination** certification.

Have been applied by the provincial malaria elimination office to other China-Myanmar border cities, among others.