DATA SCIENCE AND DECISION TECHNOLOGY

Jesse Chang, 2023

COURSE OUTLINE

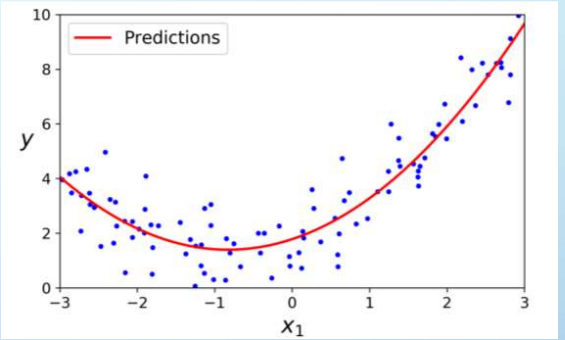
- PART I: DATA SCIENCE/MACHINE LEARNING/MODEL BUILDING
- Broad coverage of modern modeling, Data Science, Machine Learning (ML) methods used for data analytics, regression and classification.
- Complement fundamental-based engineering modeling.

- PART II: OPTIMIZATION → OPTIMAL DECISION MAKING USING MODELS
- Provides a foundation of optimal decision making that can be applied to multiple business and engineering disciplines such as planning, profit maximization, and business strategy setting.

PART I COURSE OBJECTIVES

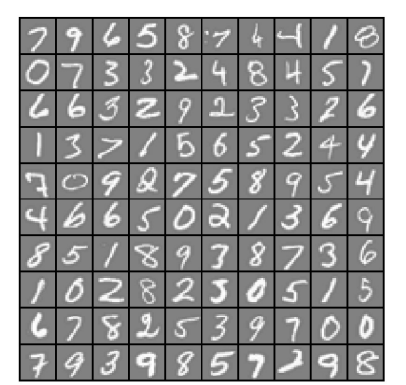
- Introduces the modern machine learning methods used for numeric prediction and classification. Key algorithms, performance metrics, and modeling issues will be covered. Course also includes deep learning neural networks for image and sequence data processing.
- Actual and potential applications will be discussed whenever possible to inspire attendees to discover applications in their own fields.

SYLLABUS: 1/3



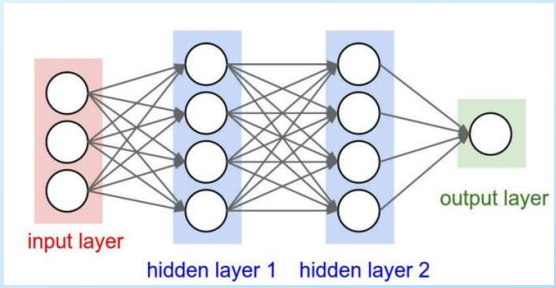
- Introduction, Regression and Common Modeling Issues
  - What is Machine Learning (ML) and types of ML
  - Machine Learning landscape and state of art
  - Regression methods using statistics and ML
  - Performance measures of regression models
  - Common modeling issues:
    - Data preparation, standardization, and split into train/validate/test sets.
    - Underfit (bias) and Overfit (variance), Regularization in ML.
    - Feature Engineering: selection, correlated inputs, transformation, etc.,
    - Variations of regression (Ridge, Lasso, stepwise and relationship to ML)
- Time permits: Use of Regression Models in Refinery and Petrochem planning.

SYLLABUS: 2/3



- Classification Methods and applications.
  - Wide range of classification applications
  - Classification methods: K-nearest neighbors, Decision trees, Support Vector Machines, Ensemble methods using Random forest, Bagging and Boosting.
  - Classification's performance metrics.
- Dimension reduction methods – Principal Component Analysis (PCA)
  - Use of PCA for Process Multivariable Monitoring
- Unsupervised ML and clustering

SYLLABUS: 3/3



- Neural Network and Deep Learning,
  - Neural Network's basic structure and computation methods
- Deep Learning Neural Nets:
  - Image/Vision: Convolutional Neural Network
  - Time-series data, sound, text (information in sequence): Recurrent Neural Network
  - Potential Applications e.g. Predictive Maintenance of plant machinery.
- Software and hardware platforms
- Applications in Refining, Chemical and Biomedical Areas

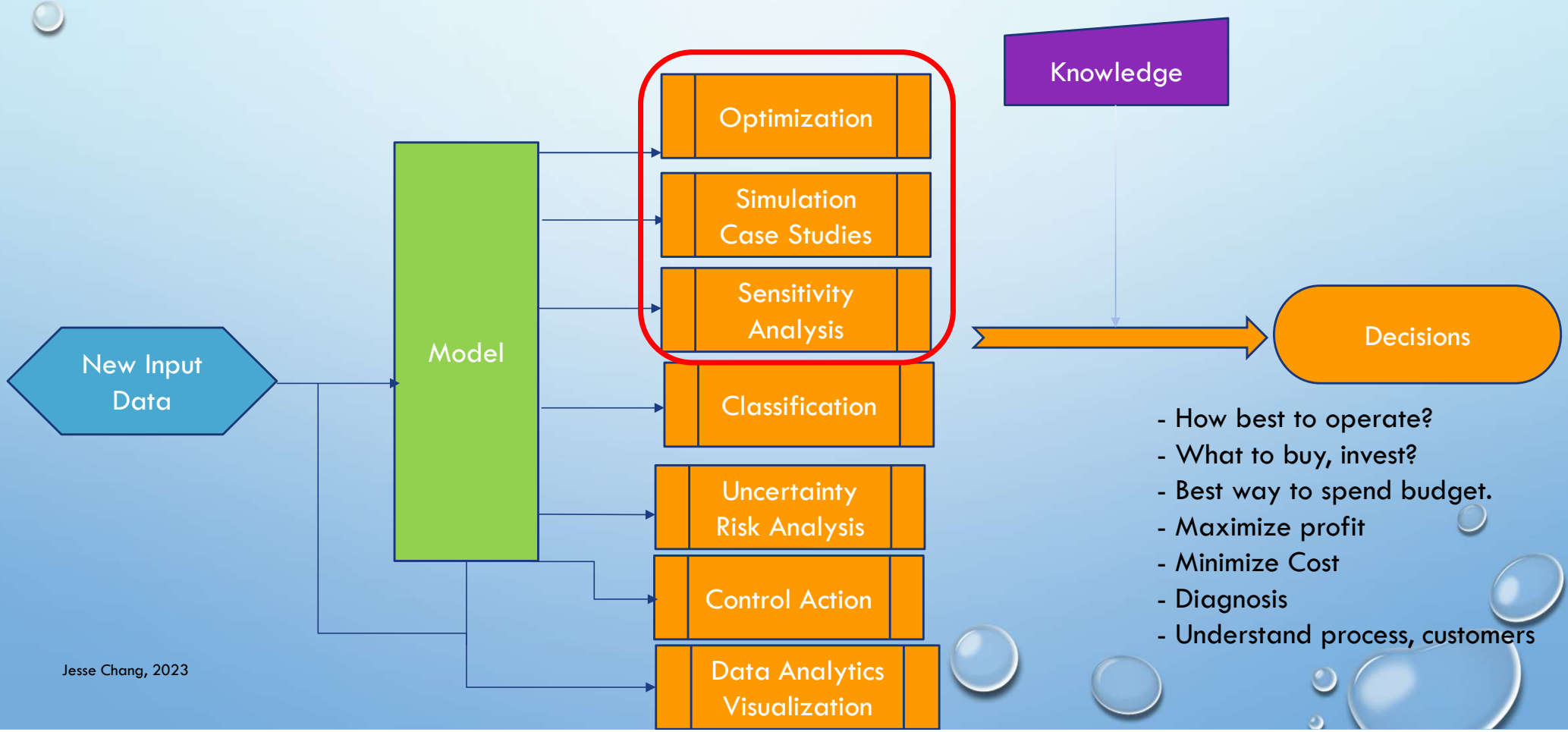
AT THE END OF THIS PART YOU SHOULD..

- Have an overview of the state of art of AI/Machine learning and their real-Life applications.
- Have learned the main algorithms for regression and classification and practiced through the homework exercises.
- Have knowledge of the typical issues and considerations in data preprocessing , 'feature engineering' and performance measures.
- Have learned the basic elements of deep learning network esp. convolutional Neural net for image processing and time series data modeling

PART II COURSE OBJECTIVES

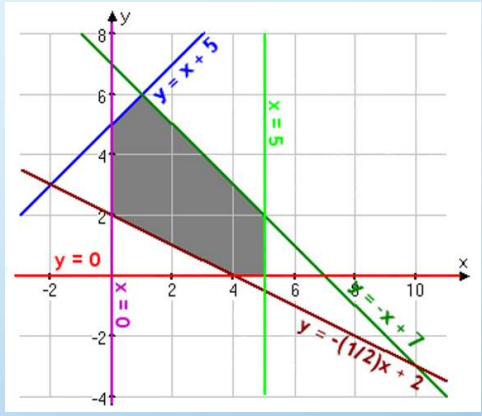
- This part introduces key optimization methods for business and engineering decision making to improve profitability. The technology can be used in a wide range of applications and multiple examples will be discussed
- Several important applications in refinery and petrochemical plants will be mentioned as examples: refinery planning, model predictive control, model tuning, and real-time optimization.
- Note: Past experience was that there was not enough time to cover all the intended subjects. However an overview and application examples of areas not covered will be given.

USAGE OF MODEL AND DATA



Jesse Chang, 2023

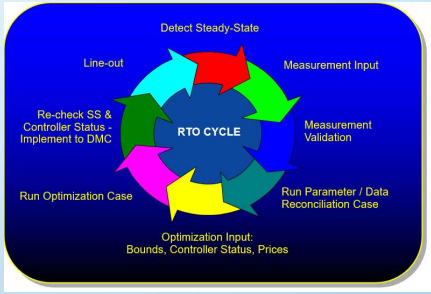
OPTIMIZATION SYLLABUS: 1/3



- Mathematical foundation of Optimization with Constraints.
- Linear Programming (LP)
  - LP applications in both business and engineering.
  - Network LP model in standard Form
    - Refinery and Petrochem Planning model, Product distribution.
- Brief introduction of Model Predictive Control (MPC) and Roles of Optimization in MPC Controllers

Jesse Chang, 2023

TIME PERMITS ONLY: OPTIMIZATION SYLLABUS: 2/3



- Nonlinear Programming (NLP)
  - General NLP business applications.
  - Applications of NLP in engineering.
- Brief introduction of Real-Time Optimization (RTO) in Refinery and Chemical processes.

TIME PERMITS ONLY: OPTIMIZATION SYLLABUS: 3/3

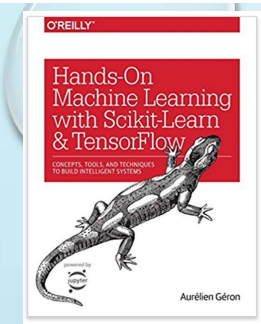
- Integer Programming and Mixed Integer Programming
  - Applications in production and distribution decision, project planning decision.

- Uncertainty and Monte-Carlo Simulation …. Only if time allows
  - Include distribution of uncertain variables and optimize expected objective values.
  - Monte-Carlo simulation of problems with uncertainty.

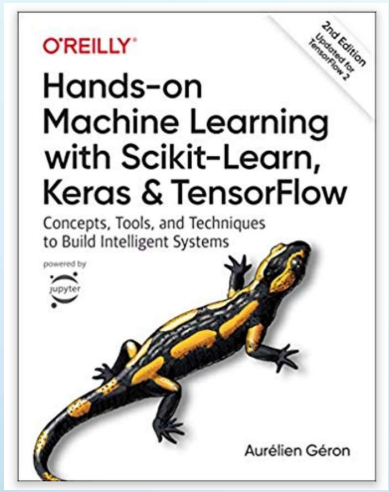
TEXTBOOK, SOFTWARE, GRADING

Jesse Chang, 2023

Part I Textbook

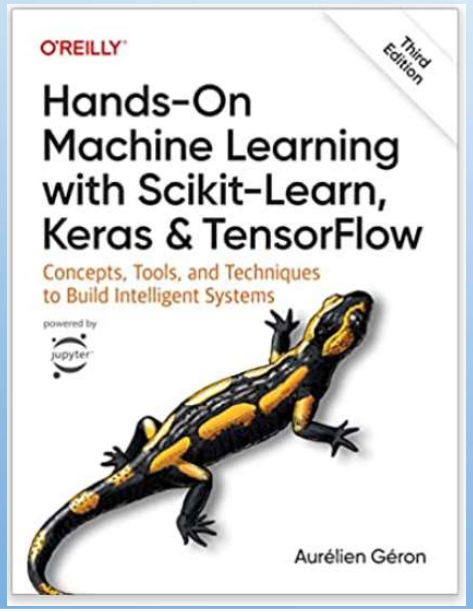


1st Edition



2nd Edition released in 10/2019.

Chinese edition also available



3rd Edition released in 11/2022.

- Book was chosen for its clearer explanation of Machine Learning algorithms and more complete coverage.
- Contents may be re-arranged to a sequences better suited for the course.
- Course is "software neutral".
- For exercises: Welcome to use Python or may use Matlab Apps to minimize programming prerequisite.

- Python and related tools are the mainstream of ML. You are encouraged to explore further using the Jupyter notebooks that come with the book.

Figure 1.13 Tool usage across the machine learning and data science industry (Source: www.kaggle.com/kaggle-survey-2020)

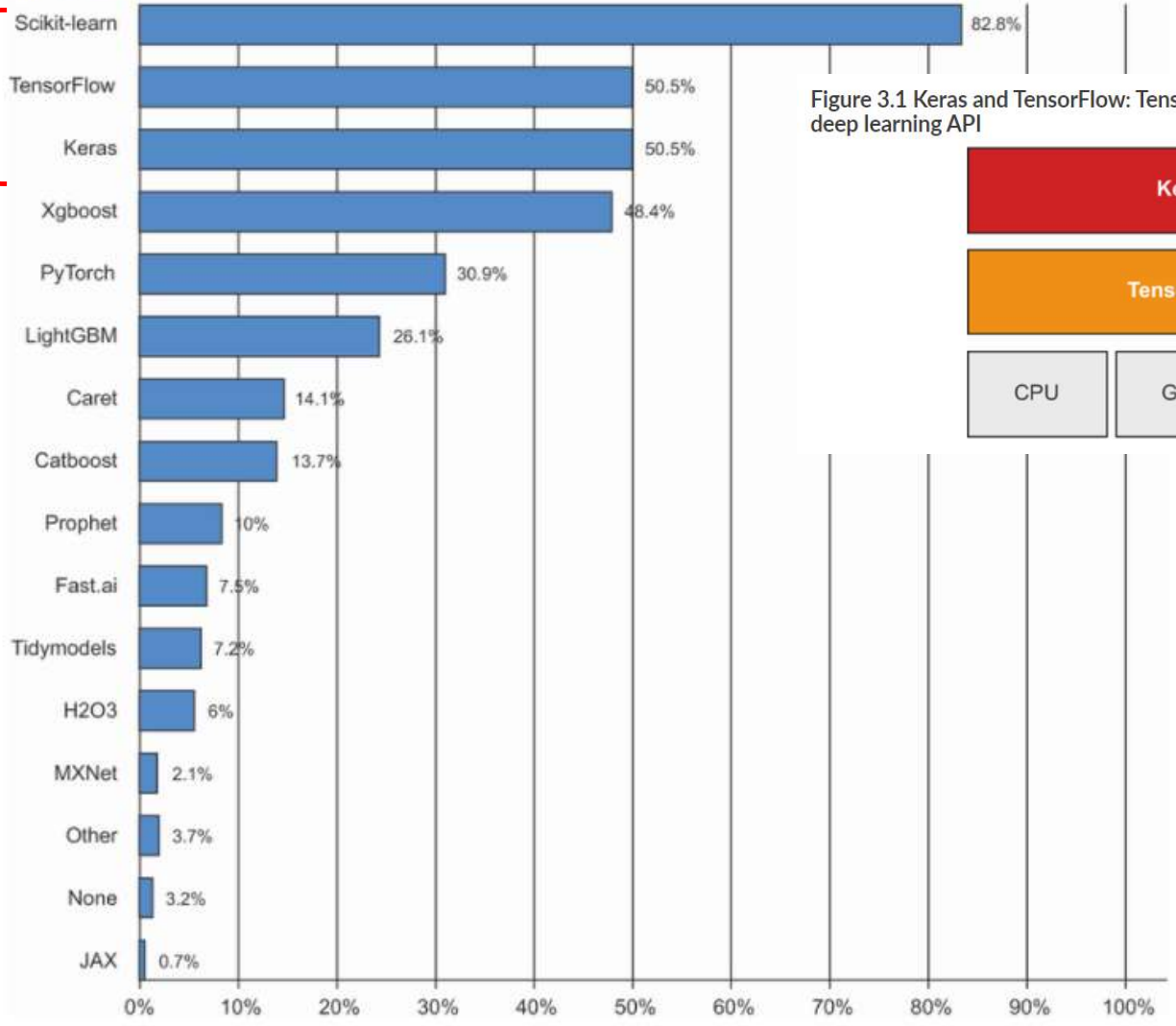
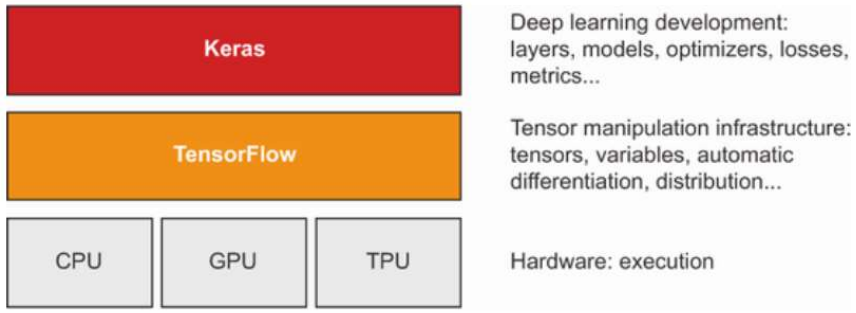
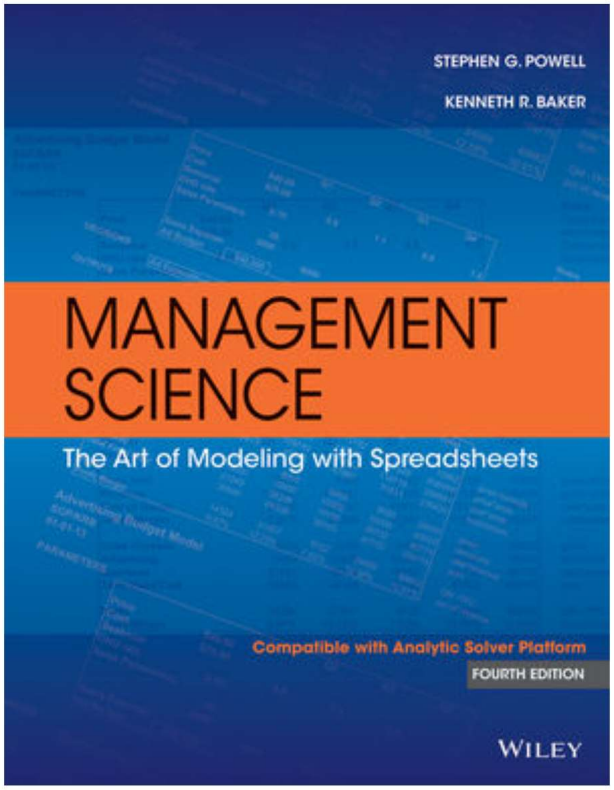


Figure 3.1 Keras and TensorFlow: TensorFlow is a low-level tensor computing platform, and Keras is a high-level deep learning API



Chollet 2nd ed.

PART II TEXTBOOK (OPTIONAL)
ALSO CAN BE A REFERENCE FOR PART I



- MAIN TEXTBOOK: "MANAGEMENT SCIENCE" 4ED, S.G. POWELL AND K.R. BAKER, 2014.
  - A BOOK USED IN MBA PROGRAM AT UNIVERSITY OF CHICAGO AND DARTMOUTH.
  - COVERS A BROAD RANGE OF MODELS AND APPLICATIONS WHICH IS CONSISTENT WITH THE COURSE OBJECTIVE – THE MAIN REASON IT WAS CHOSEN.
  - http://faculty.tuck.dartmouth.edu/management-science/ --- Book site that contains slides and some spreadsheets
  - HOWEVER FOR ENGINEERS IT'S RELATIVELY WEAK IN MATH DEPTH→SUPPLEMENT WITH OTHER MATERIAL
- 5TH EDITION IS NAMED "BUSINESS ANALYTICS: THE ART OF MODELING WITH SPREADSHEETS", 2016

OTHER SOURCES

- My Own Experience
- Additional books/resource cited:
  - "Deep Learning using Python", F. Chollet, Manning, 2018/2nd 2021 (Note: Chollet→Keras)
  - "Practical Statistics for Data Scientists", P. Bruce and A. Bruce, O'Reilly, 2017
  - "Predictive Analytics", E. Siegel, Wiley, 2017
  - Coursera: Deep Learning Specialization, Andrew Ng, 2019.
  - Machine Learning, Stanford University thru Coursera, Andrew Ng
  - Intel Nirvana AI Academy, 2019
  - "Deep Learning Illustrated", J. Krohn et.al, Addison-Wesley, 2020.
  - "Data Science for Business", F Provost and T. Fawcett, O'Reilly, 2013
  - "Machine Learning with Matlab", J. Smith, ?
  - "An Introduction to Statistical Learning", 7th ed, G. James, D. Witten, T, Hastie, R. Tibshirani, Springer, 2017. (note: use R in examples)
  - ….

GRADING

- HOMEWORK: 45%

  12% EACH: 1 (REGRESSION) ,2 (CLASSIFICATION) ,3 (NEURAL NET)

  9% : 4 (LP)
- TEST 1: 25%
- TEST 2: 25%
- CLASS INTERACTION (Q&A, SPECIAL CONTRIBUTIONS): 5%

- USE EMAIL: LJCHANG20@GMAIL.COM TO COMMUNICATE.