# STAT3009 Recommender Systems

## Lec0: Course Information

by Ben Dai (The Chinese University of Hong Kong)
on Department of Statistics

# » Key features

- **New** topic
  - RS was raised by industry around 2000, we will cover methods from **2010** to **now**.
- **Hybrid**: interdisciplinary + teaching mode
  - **Interdisciplinary**: Statistics (STAT) + Machine Learning (ML) + Python Programming + Mathematics (MATH)
  - **Teaching mode**: lecture **slides** + real time competitions in **Kaggle** + executable **Jupyter** notebook

# » Key features

- **New** topic
  - RS was raised by industry around 2000, we will cover methods from **2010** to **now**.
- **Hybrid**: interdisciplinary + teaching mode
  - **Interdisciplinary**: Statistics (STAT) + Machine Learning (ML) + Python Programming + Mathematics (MATH)
  - **Teaching mode**: lecture **slides** + real time competitions in **Kaggle** + executable **Jupyter** notebook
- **Useful**: RS are almost everywhere
  - e-commerce + retail + media + finance + health care
- Very **different** with other STAT courses
  - Intensive workload for programming
  - Good math background is better, but far from enough

## » Basic information

- **Q**: How is your course situated in the major curriculum?
  - **A**: STAT3009 is an **elective** course for **Statistics Major**.

## » Basic information

- **Q**: How is your course situated in the major curriculum?
  - **A**: STAT3009 is an **elective** course for **Statistics Major**.
- **Q**: Any prerequisite course?
  - **A**: At least you should know: (i) **linear regression** and **ridge regression**; (ii) **Python programming**: Data Manipulation with Numpy and Pandas (NumPy and Pandas Tutorial is also provided); basic usage of **sklearn**.

## » Basic information

- **Q**: How is your course situated in the major curriculum?
  - **A**: STAT3009 is an **elective** course for **Statistics Major**.
- **Q**: Any prerequisite course?
  - **A**: At least you should know: (i) **linear regression** and **ridge regression**; (ii) **Python programming**: Data Manipulation with Numpy and Pandas (NumPy and Pandas Tutorial is also provided); basic usage of **sklearn**.
- **Q**: Class size? Background of students?
  - **A**: Up to Sep 1: enroll/capacity = 37/60
  - **A**: Statistics + Risk Management Science + Quantitative Finance + Natural Sciences + Computer Science

## » Basic information

- **Q**: How is your course situated in the major curriculum?
  - **A**: STAT3009 is an **elective** course for **Statistics Major**.
- **Q**: Any prerequisite course?
  - **A**: At least you should know: (i) **linear regression** and **ridge regression**; (ii) **Python programming**: Data Manipulation with Numpy and Pandas (NumPy and Pandas Tutorial is also provided); basic usage of **sklearn**.
- **Q**: Class size? Background of students?
  - **A**: Up to Sep 1: enroll/capacity = 37/60
  - **A**: Statistics + Risk Management Science + Quantitative Finance + Natural Sciences + Computer Science
- **Q**: Form of the lecture?
  - **A**: Statistical model + Interpretation + **Real-time** Python coding

## » Basic information

- **Q**: How is your course situated in the major curriculum?
  - **A**: STAT3009 is an **elective** course for **Statistics Major**.
- **Q**: Any prerequisite course?
  - **A**: At least you should know: (i) **linear regression** and **ridge regression**; (ii) **Python programming**: Data Manipulation with Numpy and Pandas (NumPy and Pandas Tutorial is also provided); basic usage of **sklearn**.
- **Q**: Class size? Background of students?
  - **A**: Up to Sep 1: enroll/capacity = 37/60
  - **A**: Statistics + Risk Management Science + Quantitative Finance + Natural Sciences + Computer Science
- **Q**: Form of the lecture?
  - **A**: Statistical model + Interpretation + **Real-time** Python coding
- **More info**
  - Homepage + GitHub

## » Basic information

- **Q**: How is your course situated in the major curriculum?
  - **A**: STAT3009 is an **elective** course for **Statistics Major**.
- **Q**: Any prerequisite course?
  - **A**: At least you should know: (i) **linear regression** and **ridge regression**; (ii) **Python programming**: Data Manipulation with Numpy and Pandas (NumPy and Pandas Tutorial is also provided); basic usage of **sklearn**.
- **Q**: Class size? Background of students?
  - **A**: Up to Sep 1: enroll/capacity = 37/60
  - **A**: Statistics + Risk Management Science + Quantitative Finance + Natural Sciences + Computer Science
- **Q**: Form of the lecture?
  - **A**: Statistical model + Interpretation + **Real-time** Python coding
- **More info**
  - Homepage + GitHub

# » Grading policy

- Homework (15%)
- Inclass Kaggle Competition (Open-book InClass Kaggle Competition) (50%)
- Final InClass Quiz (coding and exercise) (35%):
  Basic Python programming and implementation of recommender systems models (during the final lecture of the semester)

# » Preparation

- **Kaggle**
  - Data competition platform
- **Colab**
  - Online Python computing platform
  - Jupyter notebook
  - By adding " ! " to use it as a terminal

# » Preparation

- Python package installation
  - All packages will be installed under a virtual environment
  - Using **pip** in terminal
  - `$ pip install <package name>`
  - Basic packages: `numpy` + `pandas` + `seaborn` + `scipy` + `scikit-learn`
- Jupyter notebook
  - create and **share documents** that contain live code, equations, visualizations and narrative text
  - Install packages in **Jupyter notebook** via `!pip install <package name>`

# » To-do list

- ☐ Explore **Colab** and **Kaggle**
- ☐ Participate one competition in **Kaggle**, and try to make a submission
- ☐ Install “numpy”, “pandas”, “seaborn” in your colab
- ☐ Initialize a Jupyter notebook, load “numpy”, print the numpy version in your notebook

Thank you!