# Introduction

## to MAE301 Applied Experimental Statistics

Max Yi Ren

School for Engineering of Matter, Transport  Energy
Arizona State University

August 20, 2015

# What is statistics?

- Informally, statistics is about *collecting* data and *making sense* of it.

# What is statistics?

- Informally, statistics is about *collecting* data and *making sense* of it.
- Statistics is useful for describing and understanding *variability*. Note: variability can take different forms.

# What is statistics?

- Informally, statistics is about *collecting* data and *making sense* of it.
- Statistics is useful for describing and understanding *variability*. Note: variability can take different forms.
- Statistics gives us a framework for describing this variability and for learning about potential *sources of variability*.

# Applications of statistics

- Preference learning / recommender systems: To predict consumer choice probabilities based on their profile and previous choices. (Example: Netflix, Amazon, Google, etc.)

# Applications of statistics

- Preference learning / recommender systems: To predict consumer choice probabilities based on their profile and previous choices. (Example: Netflix, Amazon, Google, etc.)
- Predictive control in hybrid vehicles: To predict future driving condition and optimize control policy (engine and motor usage) accordingly.

# Applications of statistics

- Preference learning / recommender systems: To predict consumer choice probabilities based on their profile and previous choices. (Example: Netflix, Amazon, Google, etc.)
- Predictive control in hybrid vehicles: To predict future driving condition and optimize control policy (engine and motor usage) accordingly.
- Grading without correct answers (seriously?): A Microsoft study showed that a machine can grade the SAT test almost correctly without using true answers. Why?

# Applications

Design for market: Styling or performance?

Figure: The design of the windshield angle on H3 faced a dilemma: A sloping design is more fuel efficient but looks less like a classical Hummer. Image:imganuncios.mitula.net

# Applications of statistics

## Material Design: Polymer nanocomposites

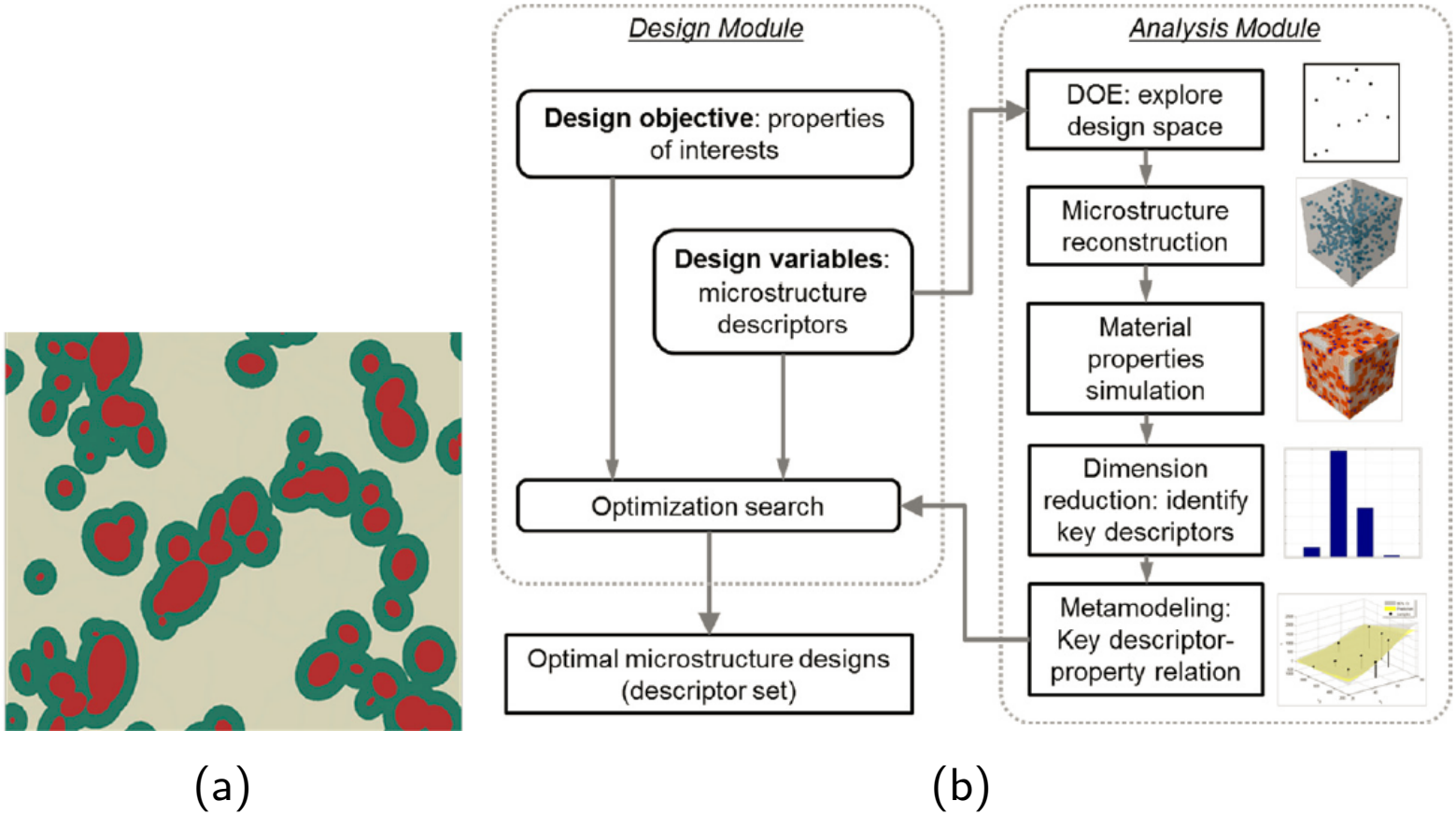



Figure: Figures from Xu et al. A Descriptor-based Design Methodology for Developing Heterogeneous Microstructural Materials System. (a) Microstructure with filler (red), matrix (yellow) and interphase around filler (green). (b) The metamodeling and optimization flow.

# Applications of statistics

Material Design: Stability of crystal structures

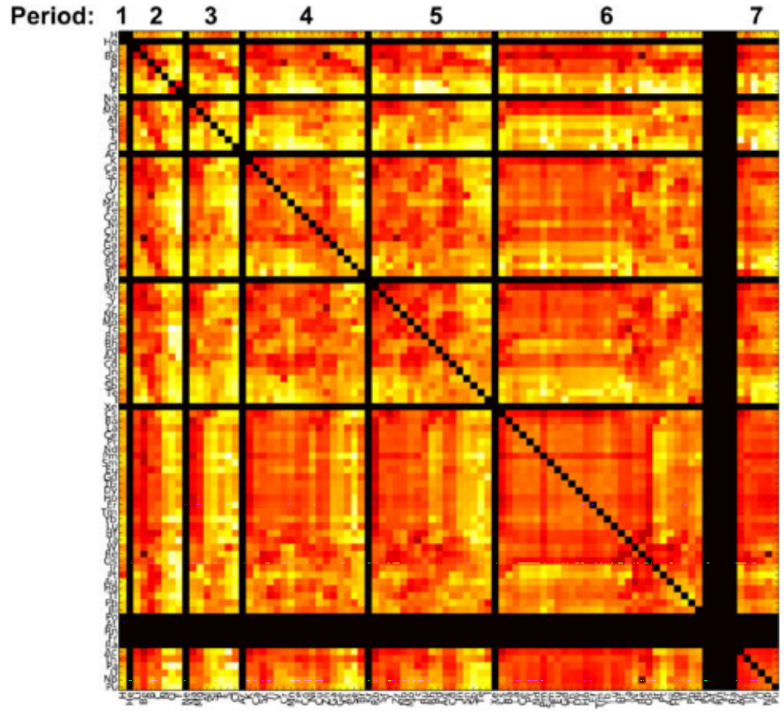
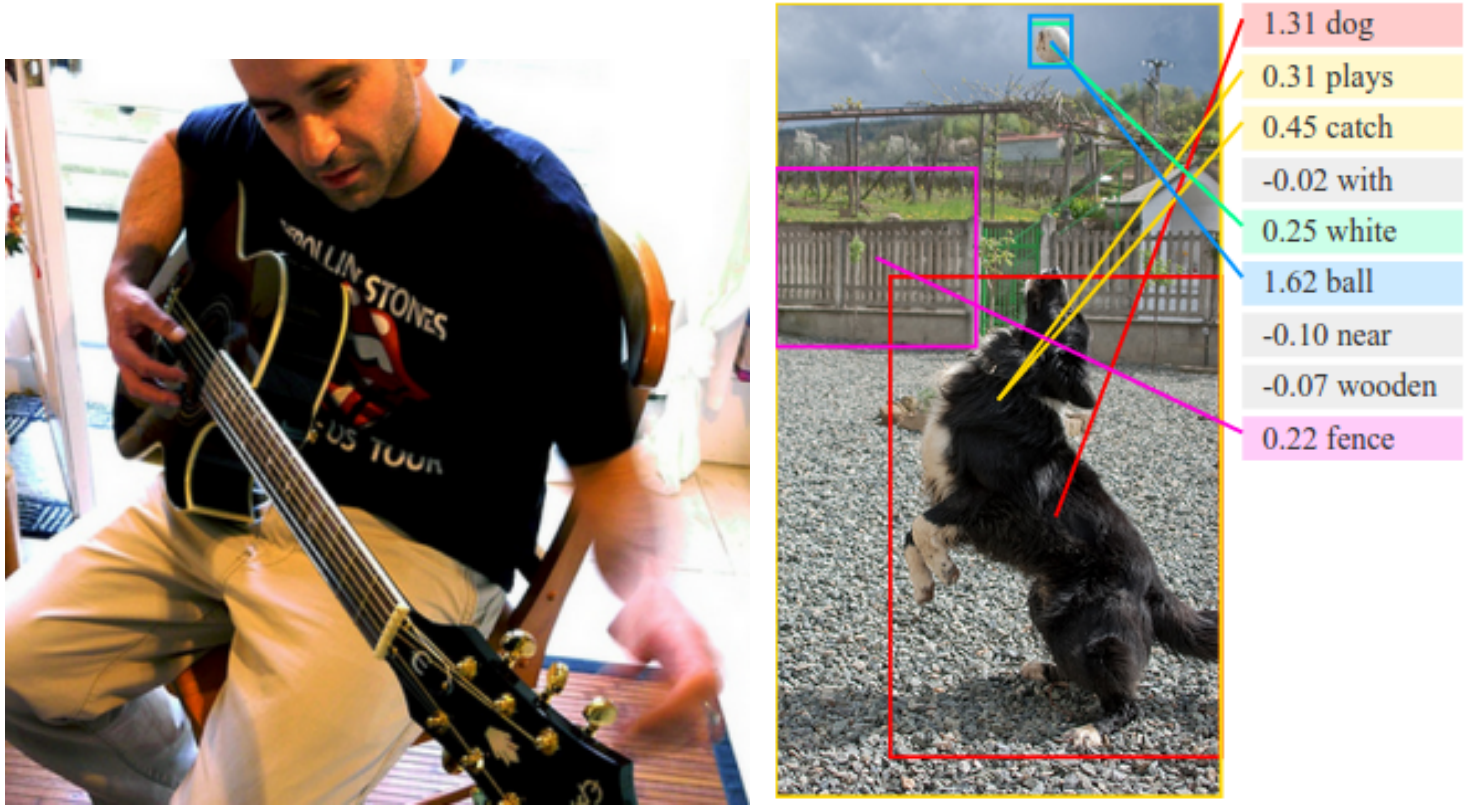



Figure: Predicted heat map of 1.6M candidate ternary compositions' stability rankings. Brighter colors imply greater stability. Figure: Saal et al. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD).

# Applications of statistics

Machine Learning: Multi-modal deep neural networks




(a) Man in black shirt is playing guitar.

(b) Association between words and the image

Figure: A. Karpathy and F. Li, Deep Visual-Semantic Alignments for Generating Image Descriptions

# Applications of statistics

## Inceptionism by deep neural networks


Figure: The original image influences what kind of objects form in the processed image. Image and text from: http://googleresearch.blogspot.co.uk/2015/06/inceptionism-going-deeper-into-neural.html

# Topics in this class

- Probability and statistics (discrete and continuous random variables)
- Statistical tests
- Confidence interval, uncertainty analysis
- Principal component analysis
- Data visualization
- Regression (OLS, DOE, Kriging, Neural nets)

# Class rules

- Homework due at the beginning of the class (submit in classroom or online)
- Two midterms (Sept. 29, Nov. 19); Do you need a final?
- Office hours: 2:30pm-4:30pm Thursdays; Location: GWC464
- Make good use of your time (You are *not* required to come to every class)
- Grading: 10% class participation, 30% homework, 30% midterm 1, 30% midterm 2
- We will use *Python*, but you are free to choose any other language.

# Python

- Why Python:
  - Widely used for scientific computing
  - Extensive built-in functions
  - Open source and *free*
  - Easy to learn, easy to write
- Who use Python: Google, Yahoo, IBM, CIA, National Labs (Los Alamos, Argonne, Lawrence Livermore), NASA, and many others
- Alternatives: Matlab, R
- Installation: Open DataJoy...and you are all set.

# Summary of the class

- Statistics is awesome, and important for your career with high probability
- Get familiar with Python
- **Homework 1 is due next Thursday**
- Please turn in the sign-off form no later than next Thursday