CS 4491/CS 7990
SPECIAL TOPICS IN BIOINFORMATICS

Mingon Kang, PhD
Computer Science, Kennesaw State University
Quotes

- “A breakthrough in machine learning would be worth 10 Microsofts” – Bill Gates, Microsoft
- “Machine Learning is the next Internet” – Tony Tether, former director, DARPA
- “Web rankings today are mostly a matter of machine learning” – Prabhakar Raghavan, Yahoo Director of Research
- “Machine learning is going to result in a real revolution” – Greg Papadopoulos, Sun CTO
New college graduates often ask me for career advice. … If I were starting out today and looking for the same kind of opportunity to make a big impact in the world, I would consider three fields.

One is artificial intelligence. We have only begun to tap into all the ways it will make people’s lives more productive and creative.

The second is energy, because making it clean, affordable, and reliable will be essential for fighting poverty and climate change.

The third is the biosciences, which are ripe with opportunities to help people live longer, healthier lives.

Ref: http://b-gat.es/2pHLHGa
What is Machine Learning?

- Algorithms that train data and improve the performance by using the knowledge

- Why?
  - It is often too difficult to design a set of rules “by hand”
  - Machine learning is about automatically extracting relevant information from data and applying it to analyze new data

- Examples
  - Face Recognition
  - Speech recognition
  - Stock prediction
Types of Learning

- Supervised learning (Classification and Regression)
  - Given labeled data, classifying or predicting unlabeled new data

- Unsupervised learning (Clustering)
  - Given unlabeled data, inferring a function to describe hidden patterns

- Feature Selection/Feature Reduction
  - Selecting a subset of relevant features

- Semi-supervised learning
  - Given both labeled/unlabeled data, classifying or predicting unlabeled new data

- And many topics…
What’s “Learning”?

Using past experiences (data) to improve future performance.

What does it mean to improve performance?

- “Minimize a loss” or “Maximize a gain”
- Minimize discrepancies between predictions and real results
- Maximize accuracy
What is Machine Learning?

Data → Training → Model:

\[ f(x) \]
What is Machine Learning?

New Data \( f(x) \) Make a decision

- yes
- no
What is Machine Learning?

- Cat vs Dog from images
What is Machine Learning?

- Vehicle Types from images

- Sedan
- SUV
- Pickup
- Coupe
- Minivan
- Wagon
- Hatchback
- Convertible
- Van
Classification

- Handwritten Digit Recognition

0, 1, ..., 9
Regression

- Stock Market
Clustering

- Grouping data sets
Journal/Conference

- Bioinformatics
  - Journals
    - Bioinformatics, PLoS Computational biology, BMC Bioinformatics, BMC Systems Biology, TCBB...
  - Conferences
    - ISMB, ACM RECOMB, PSB, ECCB, ACM BCB, WABI, BIBM, BIBE

- See “Google Scholar Metrics”
Database

- Medical Literature on line:
  - Online database of published literature since 1966
  - PubMED resource (Medline is its major component)
  - 4,600 journals
  - 12,000,000+ articles (most with abstracts)

Journal/Conference

- **Machine Learning**
  - **Conferences**
    - ICML, NIPS, CVPR, ICCV, AAAI, IJCAI, ECML, ECCV, KDD, UAI, COLT
  - **Journals**
    - Journal of Machine Learning Research, Machine Learning, PAMI, TKDE
More details

- More details of Journals/Conferences
  - Scopes
  - Paper types
    - Original Papers, Discovery Notes, Application Notes, and Reviews
  - Open-access journals
    - Otherwise use KSU Online Library

1https://academic.oup.com/bioinformatics/pages/General_Instructions
Machine Learning courses at KSU

- CS 4242 - Artificial Intelligence
- CS 4412 - Data Mining
- CS 4991/CS 7990 – Special Topics in Bioinformatics
- CS 4265/CS 7265 – Big Data Analytics
- CS 4267/CS 6278 – Machine Learning