Course Policy: Fundamentals of Machine Learning

Data Intelligence and Learning (DIAL) Lab

Prof. Jongwuk Lee

Course Information

- **Fundamentals of Machine Learning (SWE3050-41)**
  - Time: 13:30 – 14:45 Monday, 12:00 – 13:15 Wednesday
  - Language: English
    - All lecture materials are written in English.

- **Required course for third-year and fourth-year students**

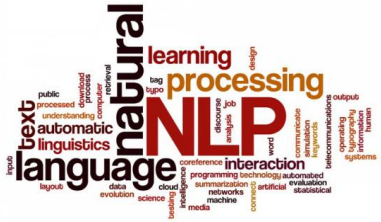
- **Prerequisite**
  - Programming basics, data structures, algorithms
  - Statistics and Probability, Linear algebra, Artificial Intelligence

Brief Biography

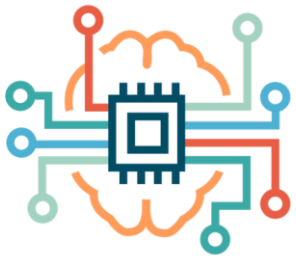
Contact information

- Homepage: http://jongwuklee.weebly.com
- Email: jongwuklee@skku.edu
- Phone: 031-299-4329
- Office: Corporate Collaboration Center 85464

Research interests



- Recommender systems
- Information retrieval
- Natural language processing
- Machine learning and deep learning optimization



Brief Biography

➢ Data Intelligence and Learning (DIAL) lab

- https://diallab.github.io/

DIAL Lab

Data Intelligence and Learning Lab

lab is primarily interested in (1) recommendation systems, (2) information retrieval, (3) data mining, (4) nachine learning algorithms. Specifically, our research aims at improving user personalization by developing ıpplications such as recommendation systems and search engines. Also, we study developing more s that effectively learn noisy and biased data. Over the past three years, we published several papers in top- /W, ICDM, CIKM, WSDM, NAACL, and CVPR) in data mining, natural language processing, and machine

Details >

Recent News

02 Aug 2022
최은성 석박통합과정, 이선경 석사과정, 최민진 석박통합과정 CIKM 2022 국제 학술대회 논문 채택

01 Aug 2022
한상우 석사 졸업생, 최은성 석박통합과정, 임찬 석사 졸업생 CIKM 2022 국제 학술대회 논문 채택

16 May 2022
2022년 하계 여름 인턴 모집 (DIAL Lab Summer Internship)

14 Apr 2022
이종욱 교수, 2022년 4월 전자정보연구정보센터(EIRIC) "라이징스타" 선정

04 Apr 2022
이재웅 박사과정, 박성민 석박통합과정 SIGIR 2022 국제 학술대회 논문 채택

more >

Information Explosion

- **The explosion of data is approximately 40 zettabytes in 2020.**
  - 1.7MB of data is created every second for every person.

WHAT'S A ZETTABYTE?	
1 kilobyte	1,000
1 megabyte	1,000,000
1 gigabyte	1,000,000,000
1 terabyte	1,000,000,000,000
1 petabyte	1,000,000,000,000,000
1 exabyte	1,000,000,000,000,000,000
1 zettabyte	1,000,000,000,000,000,000,000

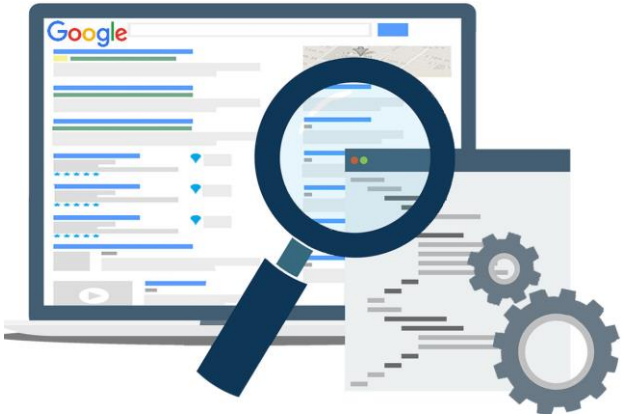
- **Google gets over 3.5 billion searches daily.**
  - 1.2 trillion searches yearly and more than 40,000 queries per second.

https://www.ibm.com/blogs/journey-to-ai/2020/06/netezza-and-ibm-cloud-pak-a-knockout-combo-for-tough-data/
https://techjury.net/blog/big-data-statistics/#gref

How to Access to Relevant Items?

- **How can we help users get access to exciting items?**

- **Pull mode (search engines)**
  - **Users take the initiative.**
  - Ad-hoc information needs



- **Push mode (recommender systems)**
  - **Systems take the initiative.**
  - Systems infer users' potential needs.

Information Retrieval

- **This is a field of searching for data with unstructured natures that satisfies information needs from huge data collection.**

- **Retrieve for what?**
  - Usually **documents**, and it can be extend to **images**, **video**, and **audio**.

- **Web search engines are the most visible IR application, but there are many other cases.**
  - E-mail search
  - Searching your laptop
  - Legal information retrieval
  - Medical information retrieval

Recommender Systems

- Netflix: 2/3 of the movies watched are recommended
- Amazon: 35% sales from recommendation
- Google News: generate 38% more clickthrough rates.

Course Logistics

Goals of this Course

- **Understanding basic models of machine learning**
  - **Supervised learning**
    - Linear regression, logistic regression
    - Decision tree, K-nearest neighbor, Naïve Bayes
    - Support vector machine (SVM)
    - Neural networks, deep neural network (DNN)
  - **Unsupervised learning**
    - Clustering, dimensionality reduction

- **Understanding fundamental principles of machine learning**
  - Overfitting problem, generalization, regularization
  - Evaluation metrics, cross-validation

Grading Policy

Attendance	10%
Midterm exam	20%
Final exam	20%
Homework	50%

Attendance (10%)

- **Attendance: 100 pts (10%)**
  - **5 per absence, 2.5 per lateness**
  - **> 7 absences → grade F unconditionally**

- **Exceptions (학교 기준 참고)**
  - 가족이 사망하여 상중인 경우
  - 학교 공식행사, 교육실습, 현장수업, 예비군 행사의 결석 사유
  - 학기 중 취업이 확정되어 출석하지 못하는 경우
  - 기타 학장이 부득이하다고 인정되는 사유가 있는 경우

Exam (40%)

- Midterm exam (200 pts)
  - Expected: Oct 17 (Mon)
- Final exam (200 pts)
  - Expected: Dec 5 (Mon)
- No make-up exam: unconditionally, grading is F!

Homework (50%)

- **6 homework (500 pts)**
  - Late submission
    - **75%**: 1 day late, **50%**: 2 days late, **25%**: 3 days late
    - **0%**: 4 days and more

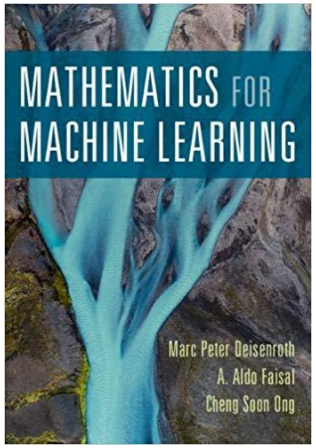
- **Cheating is not to be tolerated.**
  - Copy other students' work or allow them to copy your work.
  - **Unconditionally, grading is F!**

Classroom Participation

- Drive the lecture/discussion in this class!
- Remind me if you are still confusing!
- Help me quickly remember your names!

Useful Resources for ML/DL

Math for Machine Learning

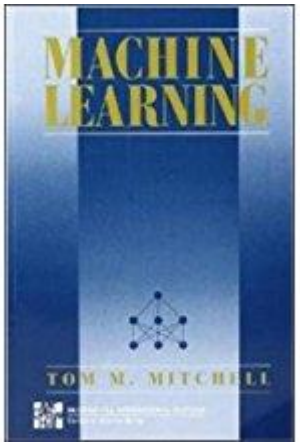


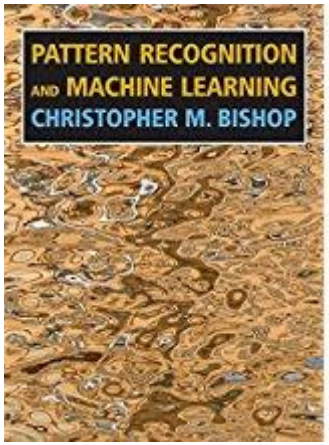
- ***Mathematics for Machine Learning,*** **Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong, 2020**
  - https://mml-book.github.io/

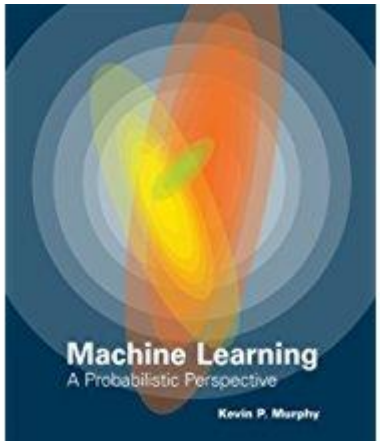
- **Review of probability**
  - http://web.stanford.edu/class/cs224n/readings/cs229-prob.pdf
- **Review of linear algebra**
  - http://web.stanford.edu/class/cs224n/readings/cs229-linalg.pdf
- **Convex optimization overview**
  - http://web.stanford.edu/class/cs224n/readings/cs229-cvxopt.pdf
- **Information theory**
  - https://arxiv.org/pdf/1802.05968.pdf

Popular ML Books

- ***Machine Learning*, Tom Mitchell, 1997**
  - http://www.cs.cmu.edu/~tom/mlbook.html
- ***Pattern Recognition and Machine Learning*, Christopher M. Bishop, 2006**
  - https://www.microsoft.com/en-us/research/people/cmbishop/
  - http://norman3.github.io/prml/
- ***Machine Learning*, Kevin Murphy, 2012**
  - https://www.cs.ubc.ca/~murphyk/MLbook/

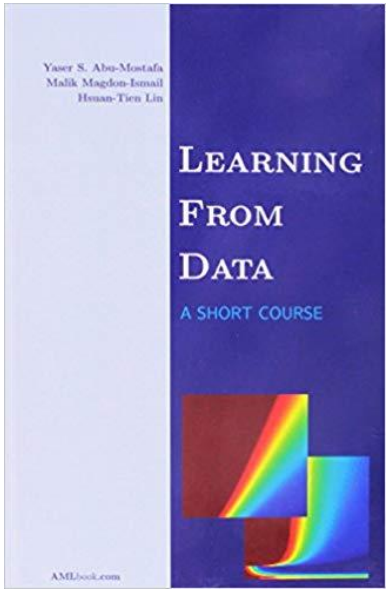


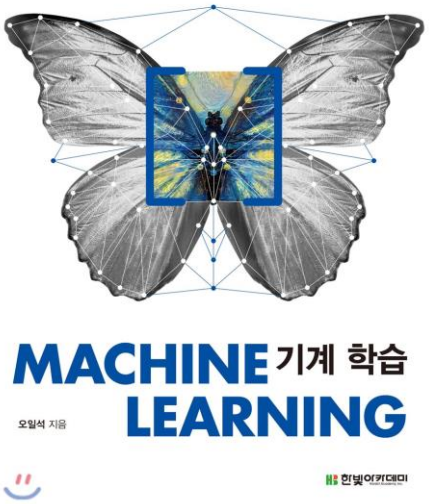




Popular ML Books

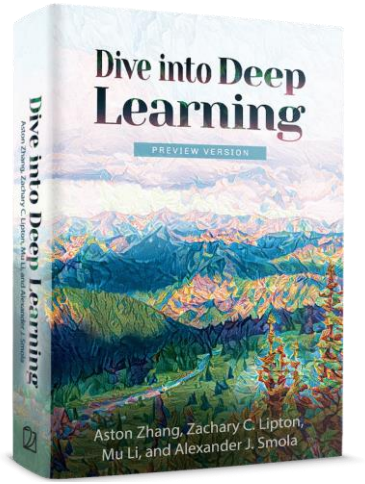
- **Learning from Data, Yaser S. Abu-Mostafa, Malik Magdon-Ismail, Hsuan-Tien Lin (2012)**
  - http://amlbook.com/
- **Machine Learning, 오일석 (2017)**
  - https://book.naver.com/bookdb/book_detail.nhn?bid=12873234





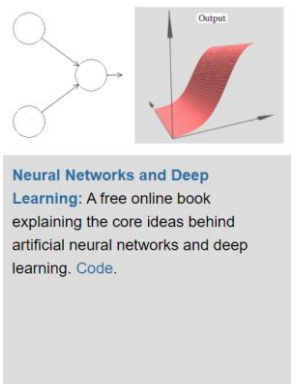
Popular DL Books

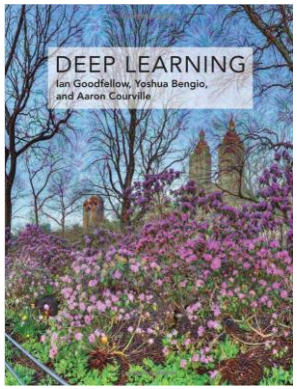
- ***Dive intro deep learning*, Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola (Online book)**
  - https://d2l.ai/index.html
  - https://arxiv.org/ftp/arxiv/papers/2106/2106.11342.pdf
  - http://courses.d2l.ai/berkeley-stat-157/syllabus.html

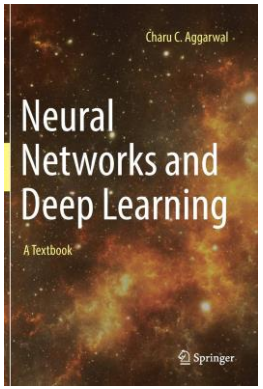


Popular DL Books

- ***Neural Networks and Deep Learning***, **Michael Nielsen, 2019 (Online book)**
  - http://neuralnetworksanddeeplearning.com/
- ***Deep Learning***, **Ian Goodfellow, Yoshua Bengio, Aaron Courville, 2016**
  - https://www.deeplearningbook.org/
- ***Neural Networks and Deep Learning: A Textbook***, **Charu C. Aggarwal, 2018**
  - http://www.charuaggarwal.net/neural.htm

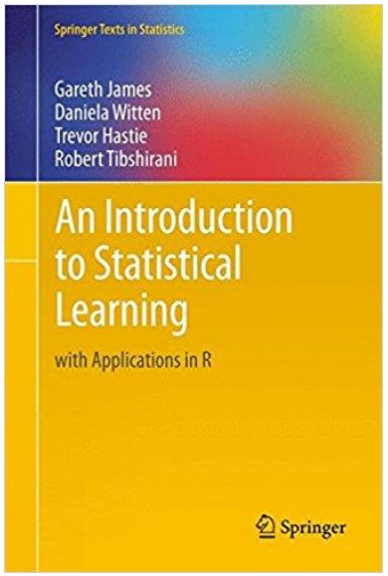


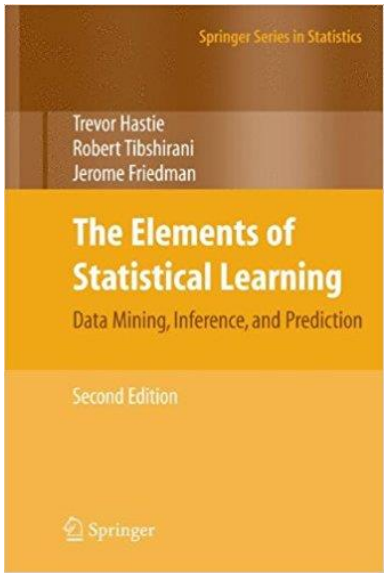




Popular Statistical Learning Books

- **An Introduction to Statistical Learning**
  - Hareth James, Daniela Witten, Trevor Hasite, Robert Tibshirani (2014)
  - PDF: http://www-bcf.usc.edu/~gareth/ISL/
  - https://www.youtube.com/watch?v=UMeDN_Z7Ne0
- **Lectures**
  - https://lagunita.stanford.edu/courses/HumanitiesSciences/StatLearning/Winter2016/about





Online Lecture

- **Machine Learning (Prof. Andrew Ng, Stanford)**
  - http://cs229.stanford.edu/syllabus.html
  - https://www.coursera.org/learn/machine-learning
- **Machine Learning (Prof. Tom Mitchell)**
  - http://www.cs.cmu.edu/~ninamf/courses/601sp15/lectures.shtml
- **Introduction to Machine Learning (Prof. Il-chul Moon)**
  - https://www.youtube.com/watch?v=t6S7ekXz3aY&list=PLbhbGI_ppZISMV4tAWHlytBqNq1-lb8bz
- **Deep Learning (Prof. Andrew Ng, Stanford)**
  - http://cs230.stanford.edu/syllabus/

Online Lecture

- **CS231n: Convolutional Neural Networks for Visual Recognition**
  - http://cs231n.stanford.edu/
- **CS224d: Deep Learning for Natural Language Processing**
  - http://cs224d.stanford.edu/
- **Deep Learning (Oxford)**
  - https://www.cs.ox.ac.uk/people/nando.defreitas/machinelearning/
- **Deep Learning (CMU)**
  - https://www.cs.cmu.edu/~rsalakhu/10707/lectures.html

Finding Resources for ML & DL

- Google Scholar: https://scholar.google.co.kr/
- Cornell University Archive: https://arxiv.org/
- Deep Learning Archive: http://www.arxiv-sanity.com/
- Papers with Code: https://paperswithcode.com/

Q&A