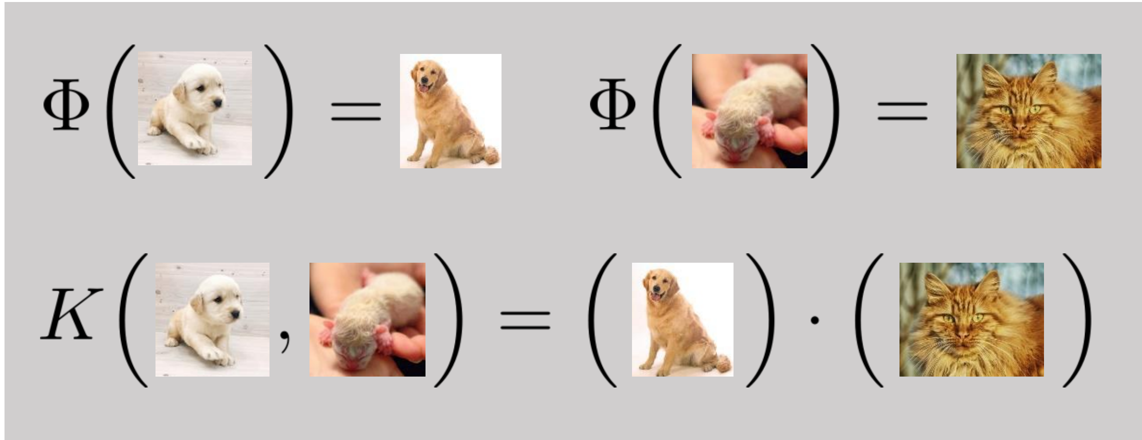
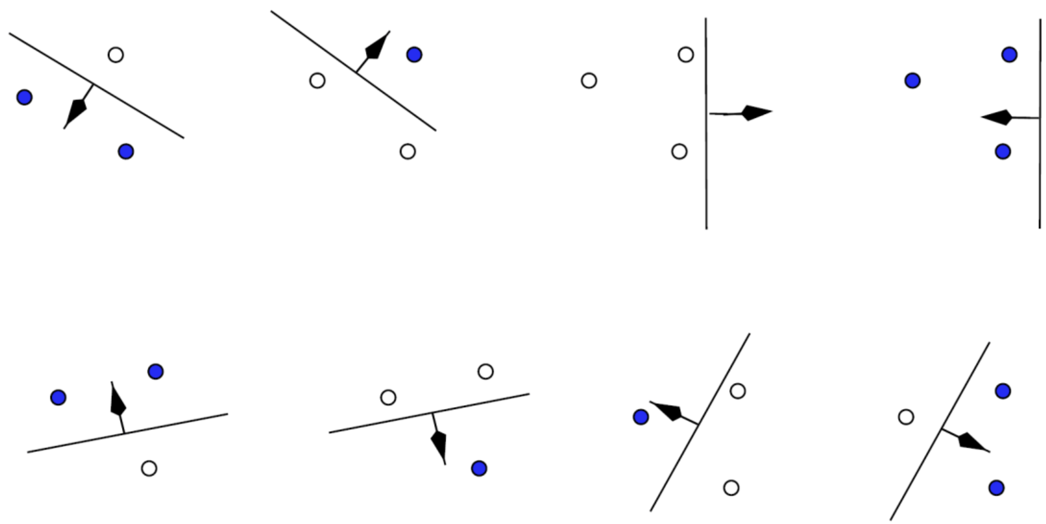


Kernel-based Learning: Theoretical Foundation

Pilsung Kang
School of Industrial Management Engineering
Korea University

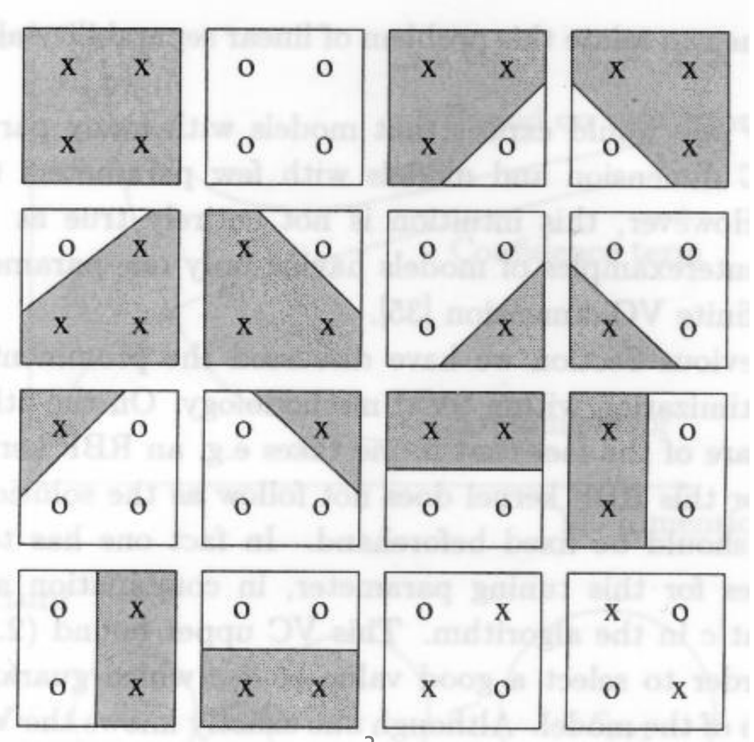
The Concept of Shatter

- Shatter
  - ✓ A set of points is said to be shattered by a class of functions if, no matter how we assign a binary label to each points, a member of the class can perfectly separate them.
    - A linear classifier can shatter (n+1) instances in n-dimensional space
  - ✓ 함수 F는 n개의 points를 Shatter할 수 있다 → 함수 F에 의해 n개의 point는 임의의 +1 또는 -1을 Target value로 하는 분류 경계면의 생성이 가능하다.



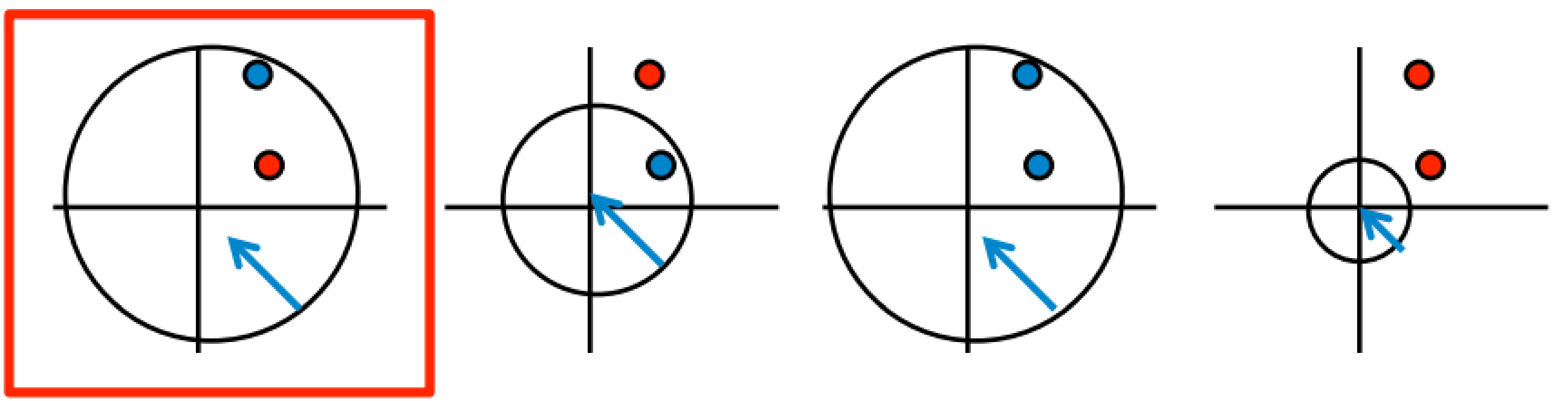
The Concept of Shatter

- Shatter
  - ✓ A set of points is said to be shattered by a class of functions if, no matter how we assign a binary label to each points, a member of the class can perfectly separate them.
    - Can a linear classifier shatter (n+2) instances in n-dimensional space?



The Concept of Shatter

- Shatter
  - ✓ A set of points is said to be shattered by a class of functions if, no matter how we assign a binary label to each points, a member of the class can perfectly separate them.
    - A circle cannot shatter 2 instances in 2-dimensional space

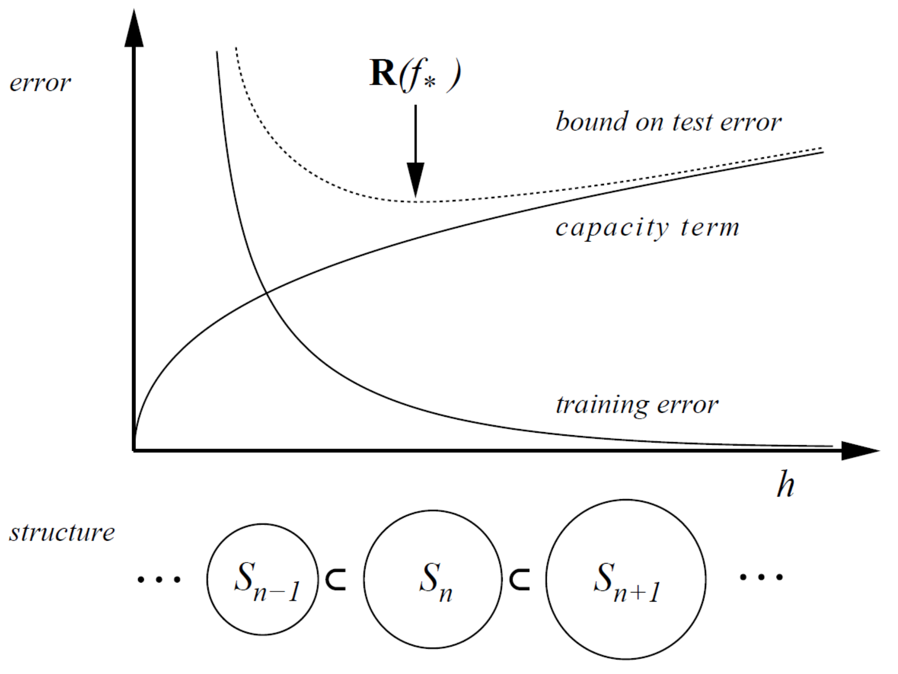


Vapnik-Chervonekis (VC) Dimension

- VC dimension
  - ✓ Measures the capacity of a hypothesis space
  - ✓ Capacity is a measure of complexity and measures the expressive power, richness or flexibility of a set of functions by assessing how wiggly its members can be
  - ✓ The maximum number of points that can be shattered by H is called VC dimension

ERM vs. SRM

- Structural Risk Minimization (SRM)



✓ An inductive principle for model selection used for learning from finite training data

✓ Describe a general model of capacity control and provides a trade-off between **hypothesis space complexity (VC dim.)** and the **quality of fitting the training data (empirical error)**

ERM vs. SRM

- Structural Risk Minimization

Let h denote the VC dimension of the function class F and let R_{emp} be defined as follows using the 0/1 loss.

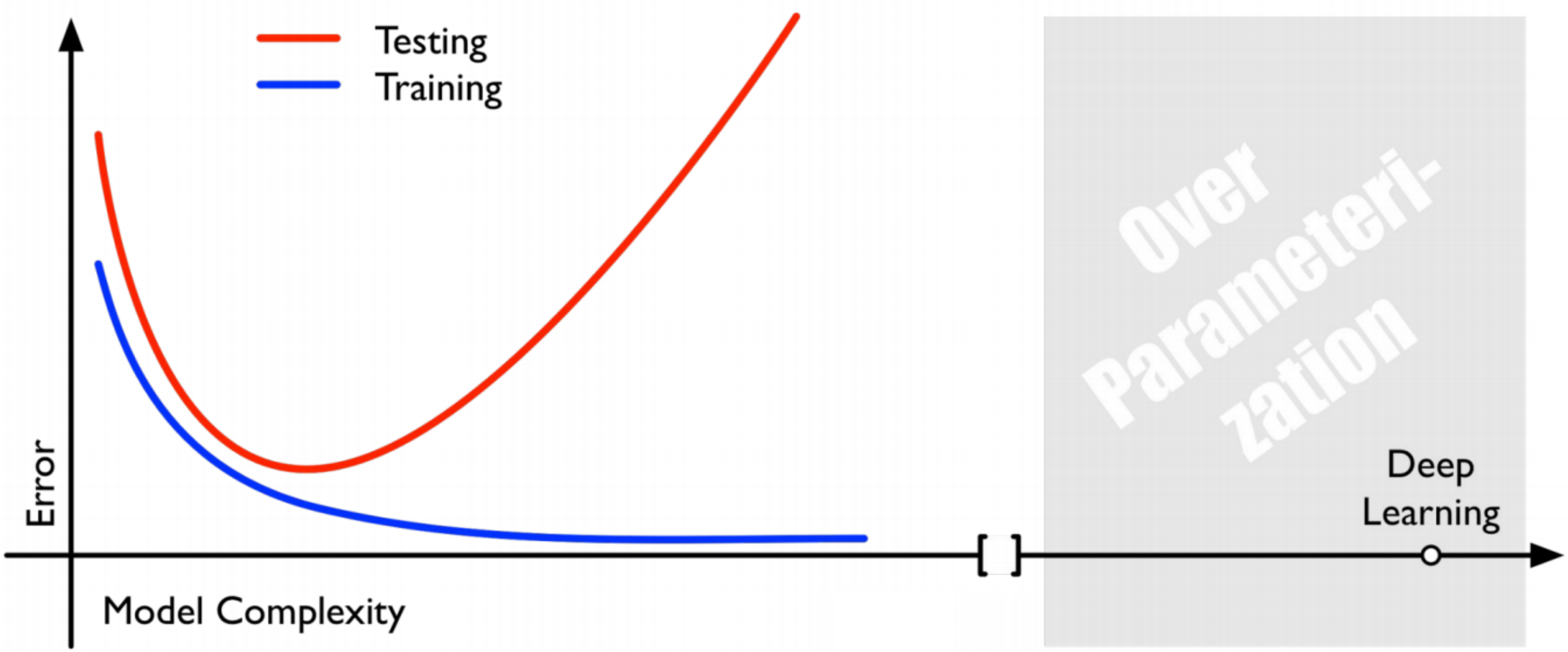
$$R_{emp}[f] = \frac{1}{n}\sum_{i=1}^{n} L(f(x_i), y_i)$$

For all $\delta > 0$ and $f \in F$ the inequality bounding the risk below holds with probability of at least $1 - \delta$ for $n > h$.

$$R[f] \leq R_{emp}[f] + \sqrt{\frac{h\left(ln\frac{2n}{h} + 1\right) - \ln\left(\frac{\delta}{4}\right)}{n}}$$

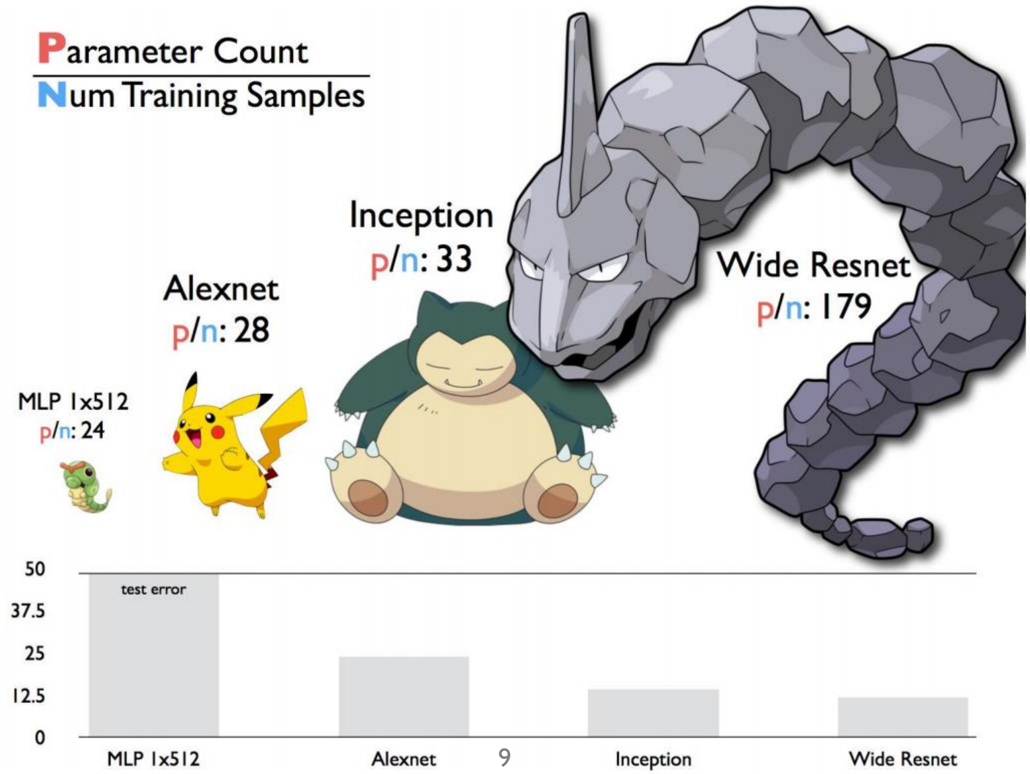
ERM vs. SRM

- Structural Risk Minimization (SRM)
  - ✓ Trade-off between **hypothesis space complexity (VC dim.)** and the **quality of fitting the training data (empirical error)**
  - ✓ Understanding Deep Learning Requires Rethinking Generalization (Zhang et al 2017a, 2017b, 2017c, Park 2017)



ERM vs. SRM

- Structural Risk Minimization (SRM)
  - ✓ Trade-off between **hypothesis space complexity (VC dim.)** and the **quality of fitting the training data (empirical error)**
  - ✓ Understanding Deep Learning Requires Rethinking Generalization (Zhang et al 2017a, 2017b, 2017c, Park 2017)



고려대학교 KOREA UNIVERSITY

DSBA Data Science & Business Analytics

Kernel Machine Kids vs. Deep Learning Kids

- Fantastic Four in Deep Learning

Yann Lecun	Geoffrey Hinton	Yoshua Bengio	Andrew Ng
Facebook AI Director Professor @NYU	Fellow @ Google Professor @U of Toronto	Head of MILA	Chief Scientist @Baidu Co-Founder of Coursera Professor @Stanford Univ.

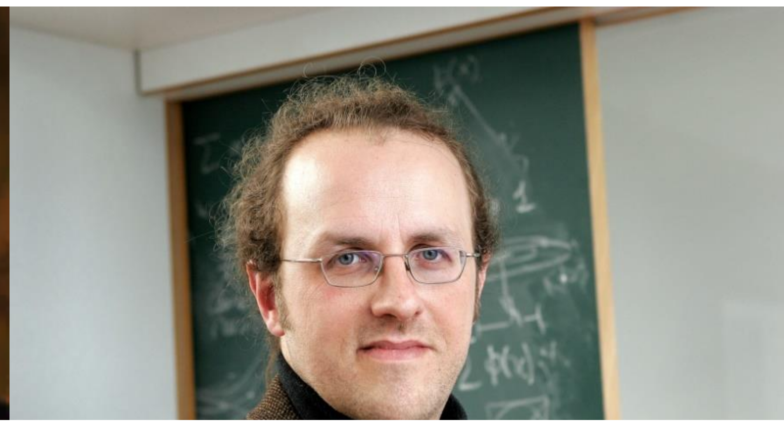
Kernel Machine Kids vs. Deep Learning Kids

- The Three Musketeers in Kernel Machine

Vladimir Vapnik
Professor @RHUL
Book: Statistical Learning Theory

Bernhard Scholkopf
Director of Intelligent Systems
@Max Planck Institute
Book: Learning with Kernels

Alexander J. Smola
Professor @CMU
Book: Advances in Kernel Methods

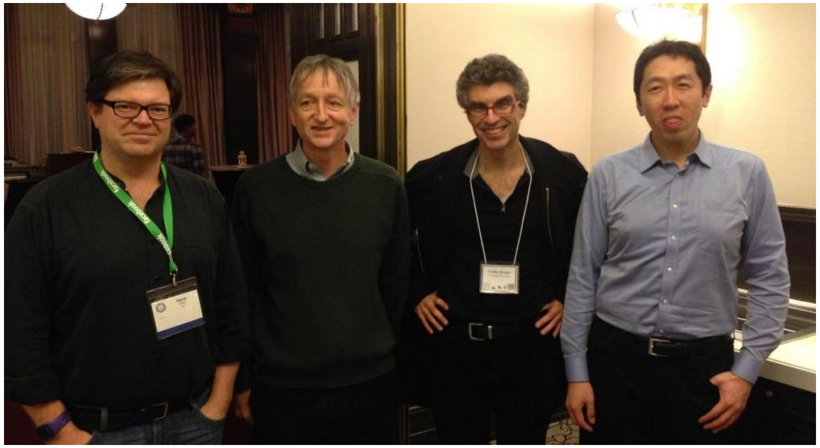




BIG DATA

Facebook's AI team hires Vladimir Vapnik, father of the popular support vector machine algorithm

JORDAN NOVET @JORDANNOVET NOVEMBER 25, 2014 1:23 PM

HOT Kids vs. BTS Kids



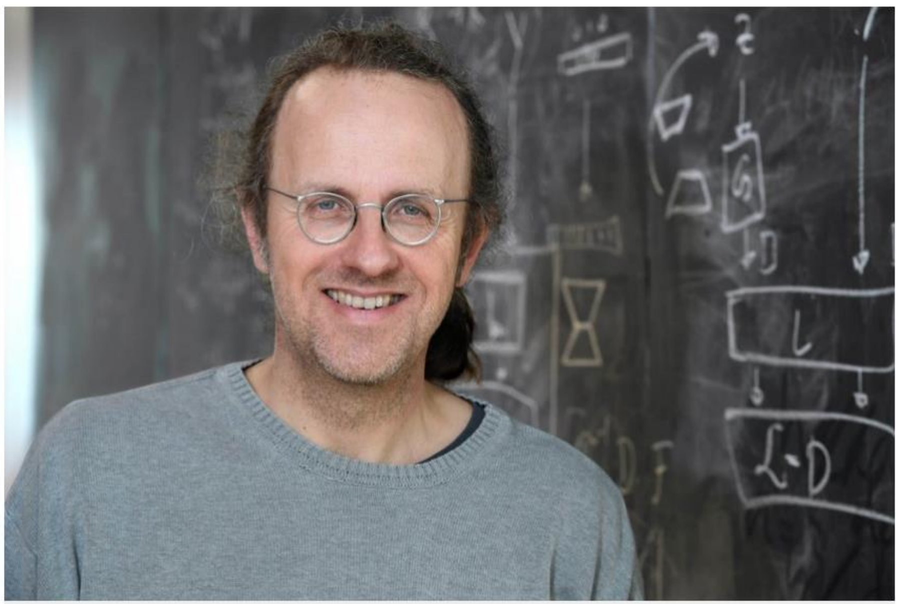
Kernel Machine → Deep Learning

Körber Prize 2019 for Bernhard Schölkopf

This year, the scientific distinction with the highest prize money in Germany goes to the pioneer of artificial intelligence

SEPTEMBER 13, 2019

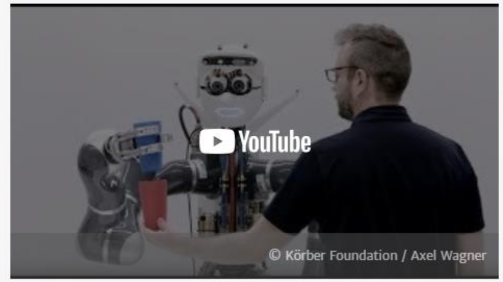
Bernhard Schölkopf, Director at the Max Planck Institute for Intelligent Systems in Tübingen, is honoured with the Körber Prize for European Science 2019. The Körber Foundation awards the prize to honour the computer scientist's contributions to machine learning, which today supplies one of the most important methods of Artificial Intelligence (AI). The Körber Prize includes prize money of one million Euros.



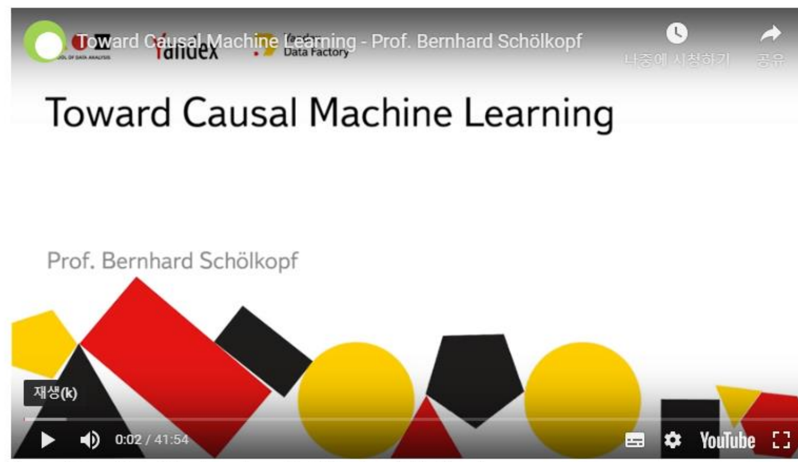
A pioneer of Artificial Intelligence: Bernhard Schölkopf receives the 2019 Körber Prize for European Science.
© David Ausserhofer

https://www.mpg.de/13645470/schoelkopf-koerber-prize

Causal relationships in data



Teaching robots how to learn



Toward Causal Machine Learning – Prof. Bernhard Schölkopf

https://www.youtube.com/watch?time_continue=2&v=ooeRlw3U2zU

Kernel Machine → Deep Learning

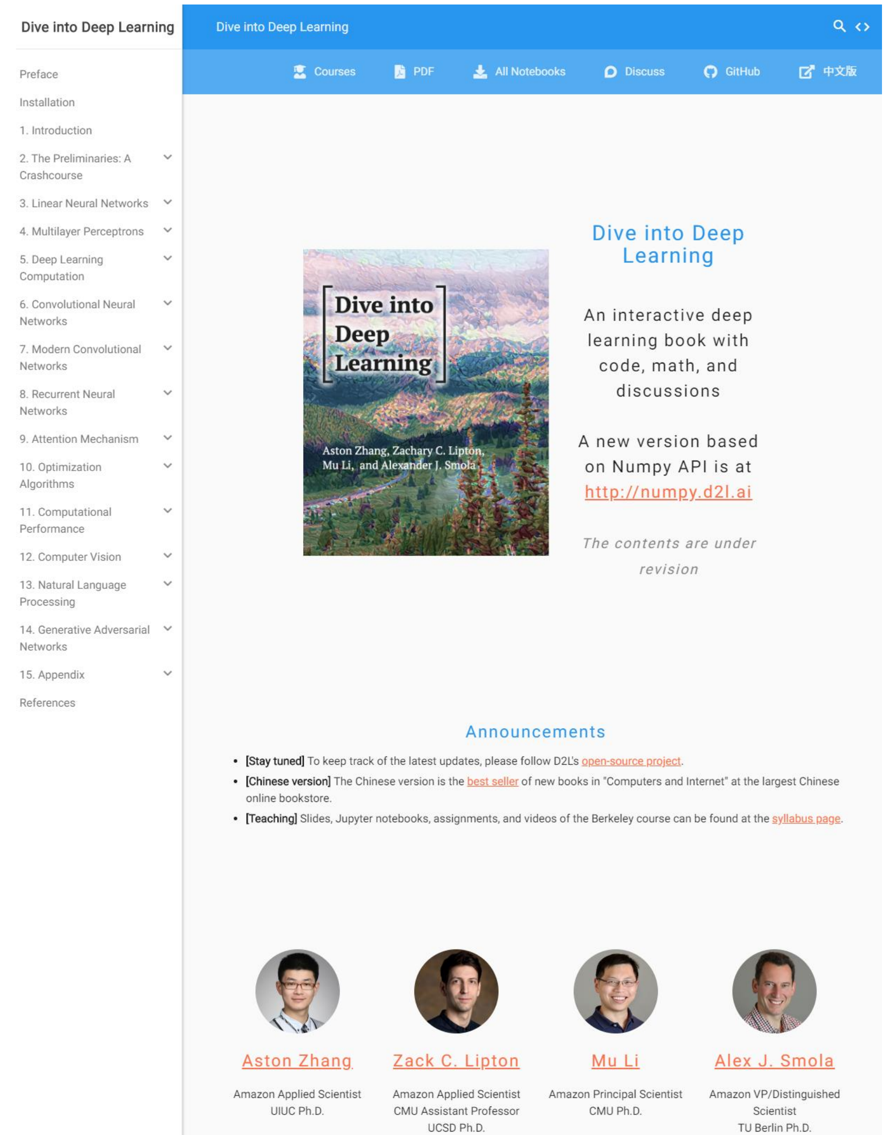
AWS Startups Blog

Alex Smola Showcases the Breadth of AWS's Machine Learning Capabilities at Collision 2018

by Michelle Kung | on 07 MAY 2018 | in Events | Permalink | Share

https://aws.amazon.com/ko/blogs/startups/alex-smola-aws-machine-learning-collision-conference-new-orleans/

훗날 이 말을 한 당사자인 이진영이 은퇴 후 인터뷰에서 밝히길 야잘잘의 진짜 속뜻은 **'야구를 잘하는 사람이 잘하지만 그걸 이겨내기 위해선 자신과 같은 길을 가는 경쟁자들을 어떻게 해야 이길 수 있는지 고민하고 노력하는 것'**라고 밝혔다. 그러니까 재능 말고도 노력도 뒷받침해야 한다는 이야기



https://www.d2l.ai/


Q & A

References

Research Papers

- Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2016). Understanding deep learning requires rethinking generalization. *arXiv preprint arXiv:1611.03530*.

Other materials

- http://www.cs.nyu.edu/~mohri/icml2011-tutorial/
- Suykens, J. (2003). Least Squares Support Vector Machines. IJCNN 2003 Tutorial.
- Abu-Mostafa, Y. (2012). Lecture 14: Support Vector Machines. Learning From Data. Caltech.
- http://www.ci.tuwien.ac.at/~meyer/svm/final.pdf
- http://pubs.rsc.org/en/content/articlehtml/2010/an/b918972f
- Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2017b). Understanding deep learning requires rethinking generalization. International Conference on Learning Representations. [Slides] http://pluskid.org/slides/ICLR2017.key
- Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2017c). Understanding deep learning requires rethinking generalization. International Conference on Learning Representations. [Poster] http://pluskid.org/slides/ICLR2017-Poster.pdf
- Park, J.S. (2017). Deep learning and data. DSBA Lab Seminar. http://dsba.korea.ac.kr/wp/wp-content/seminar/Paper%20Review/Deep%20Learning%20and%20Data%20-%20%EB%B0%95%EC%9E%AC%EC%84%A0.pdf