

Big Data & Deep Learning: A Powerful Mix

Max Welling



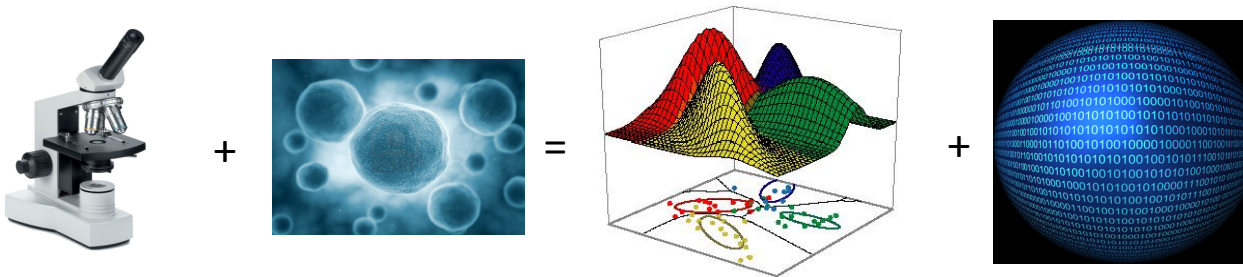
Universiteit van Amsterdam



University of California, Irvine

Overview

- Three exponential growth laws
- Machine learning 101
 - Unsupervised learning
 - Supervised (deep) learning
 - Generative vs. discriminative models

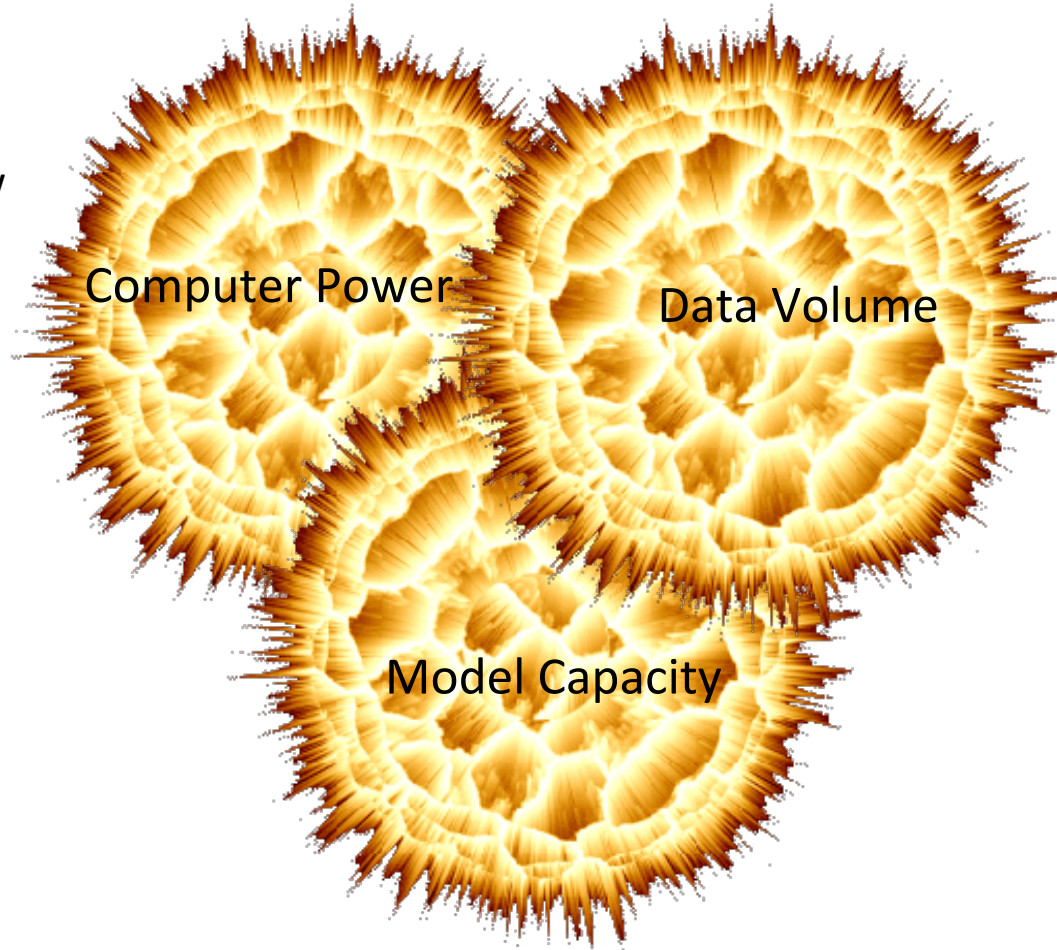


(ML as a "datascope")

Three Exponential Growth Laws

Exponential Growth: Big Data

Moore's law



big data

deep learning

Big Data



How much Data = Big Data?

- Total amount generated by humanity so far:
4 zettabyte = 4,000,000,000,000,000,000,000 byte.
- That's 8 billion hard disks or a pile as high as the moon if every hard disc is 5 cm thick.
- *Amount of data doubles about every two years.*



Big Data in Astronomy

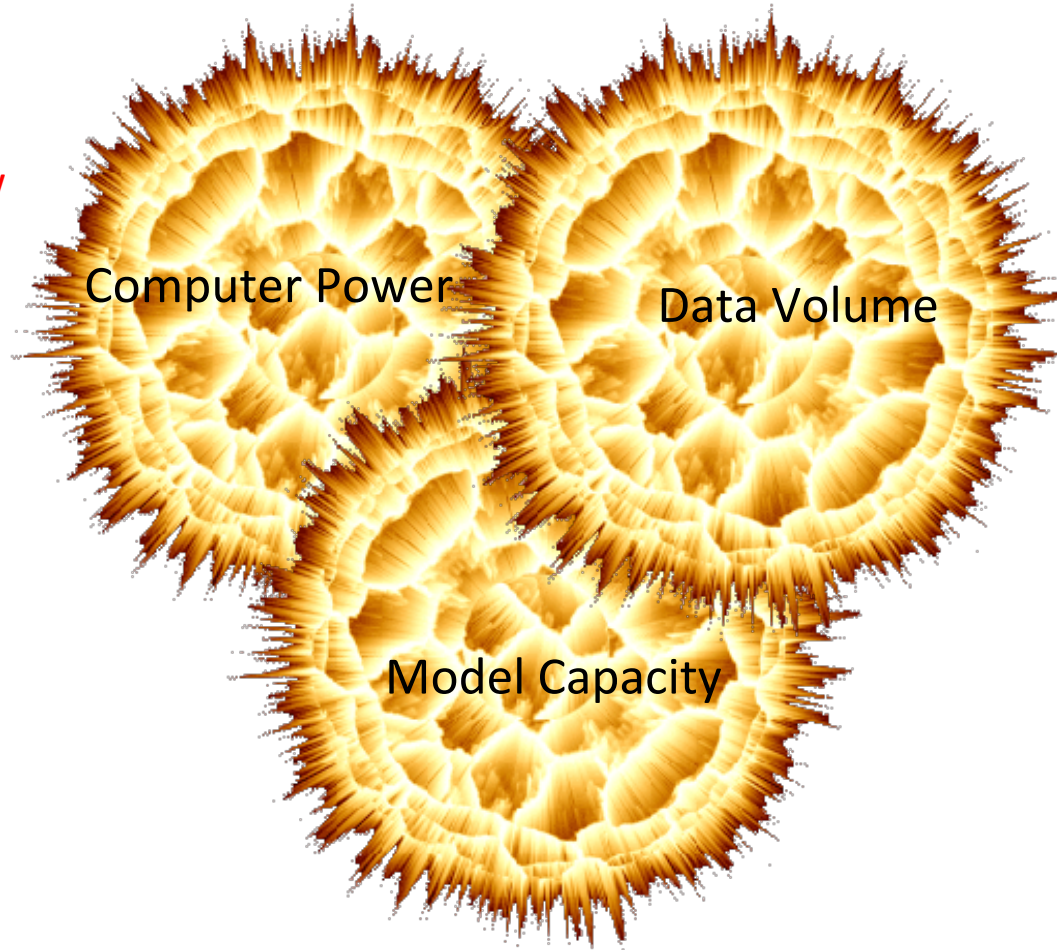
- "Square Kilometer Array": 1 exabyte per day in 2024.

(1 exabyte = 1000 petabytes = 1 million terabytes = 1 million hard discs *per day*)



Exponential Growth: Compute Power

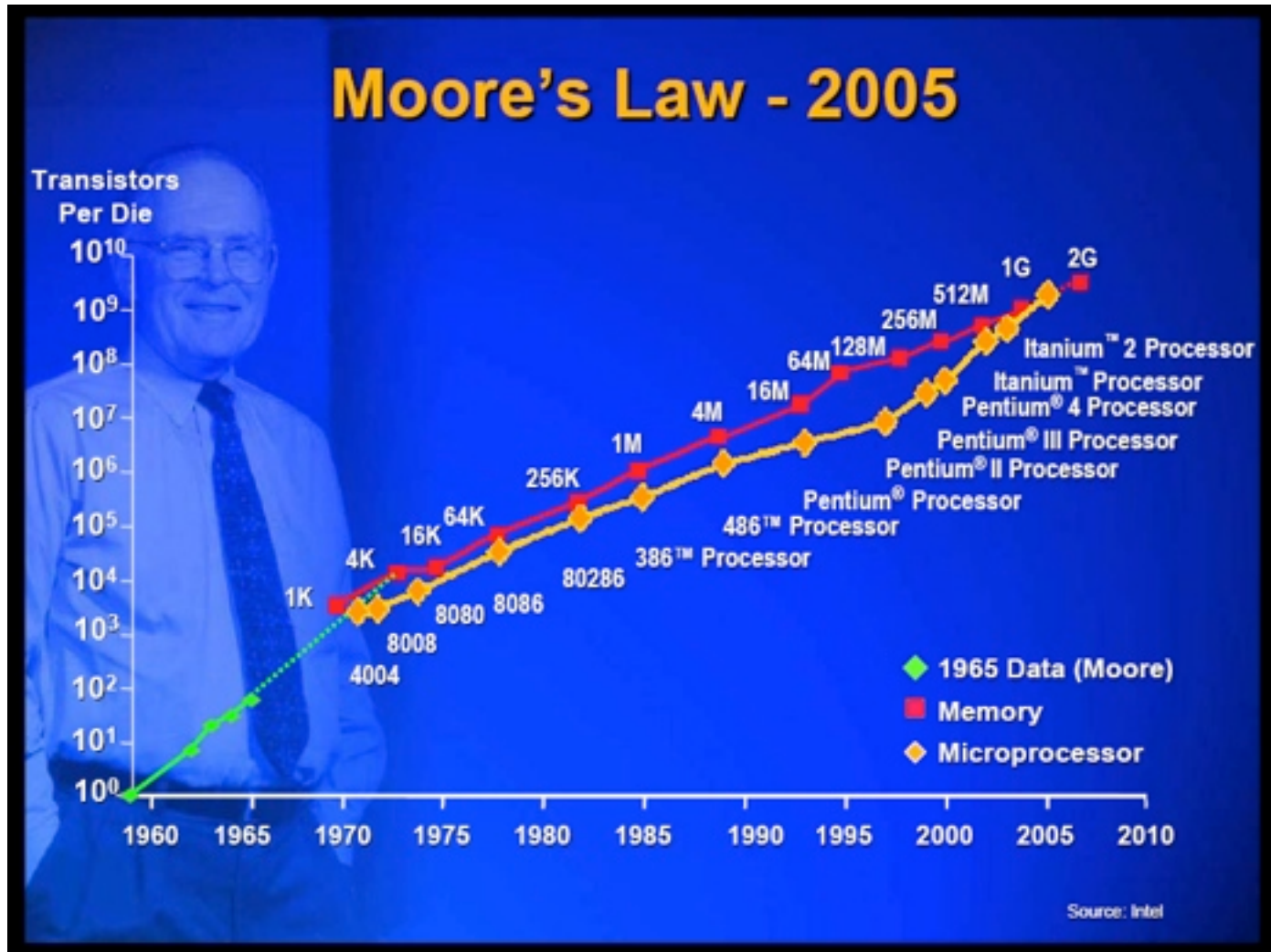
Moore's law



big data

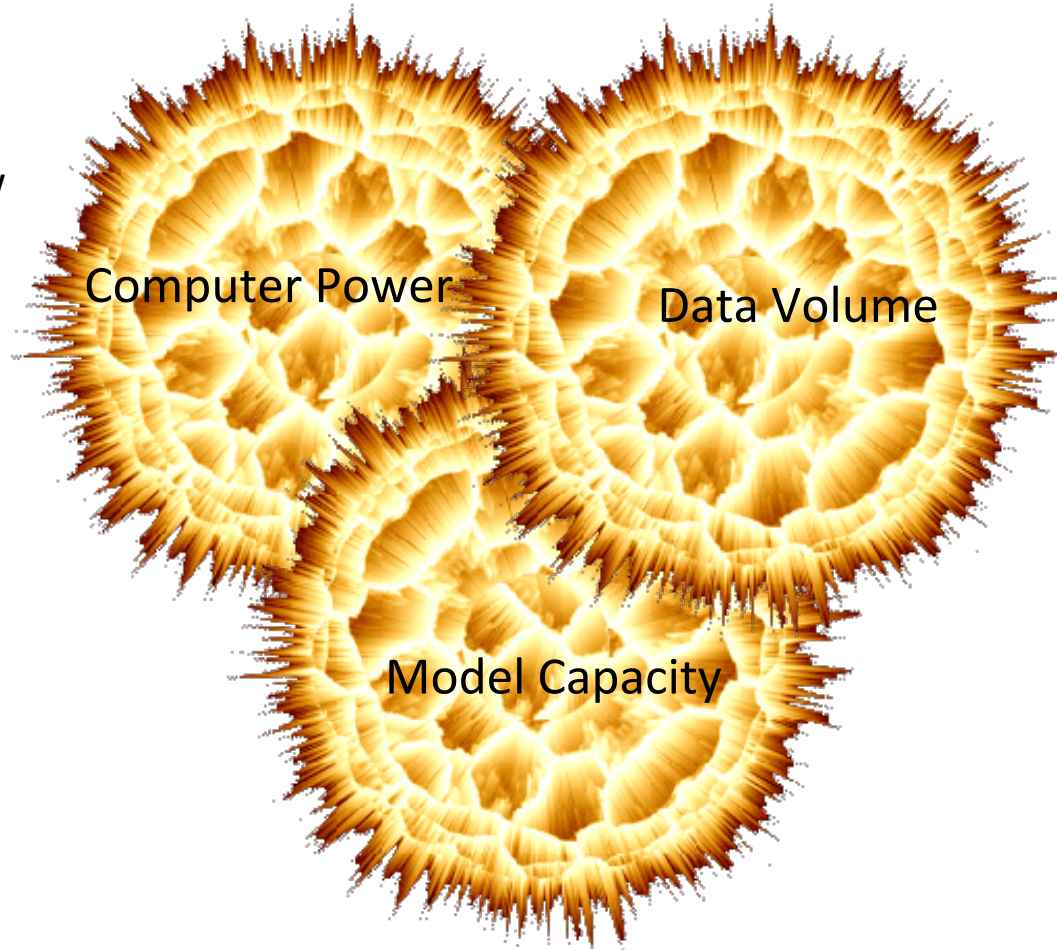
deep learning

Moore's Law Powers Big Data



Exponential Growth: Model Capacity

Moore's law



Computer Power

Data Volume

Model Capacity

big data

deep learning

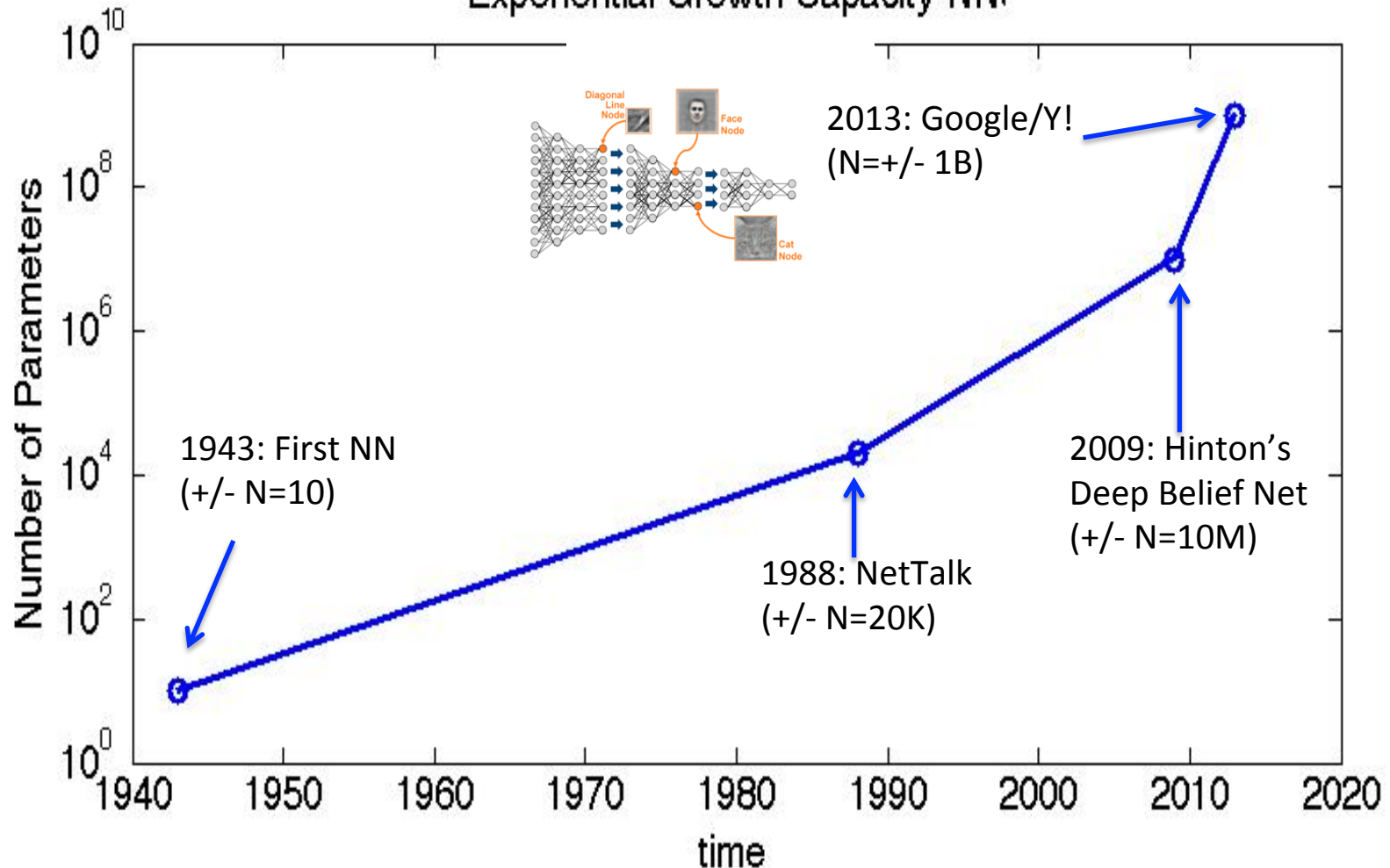
10^{14}
N=100T



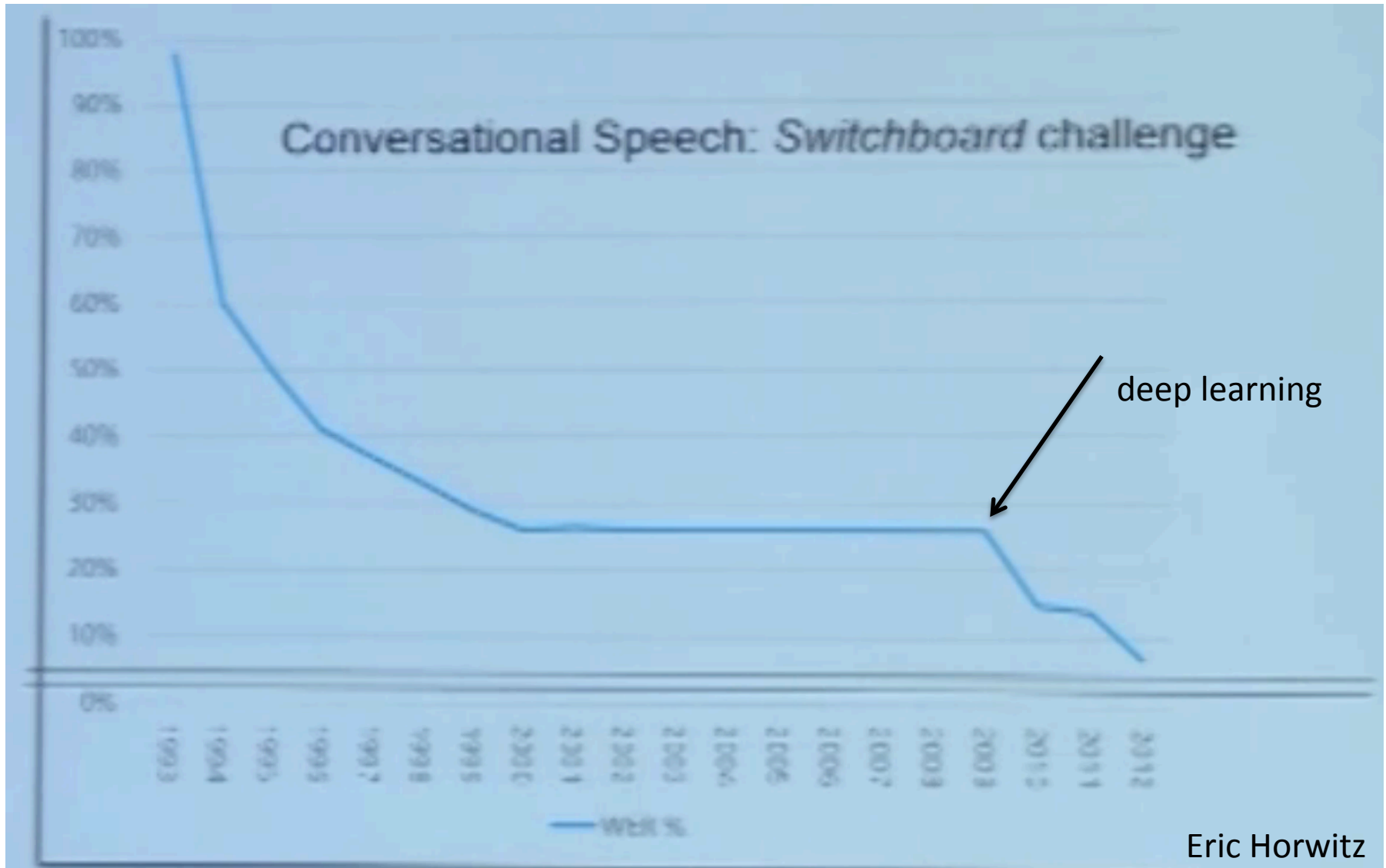
2030-2050 capacity
human brain is
reached

Deep Neural Networks are Big!

Exponential Growth Capacity NN:



The Power of Deep Learning



Machine Learning 101

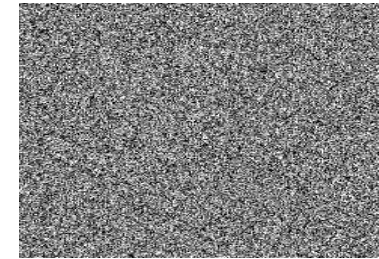
Data-mining: Digging for Information



data = ore
informatie = gold
machine learning = pickaxe

Useful Information

Useful Information is information with which you can make predictions

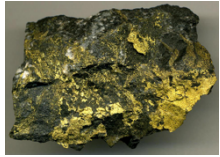


No information

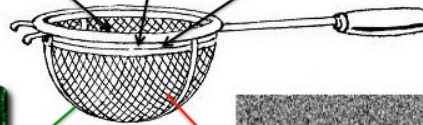
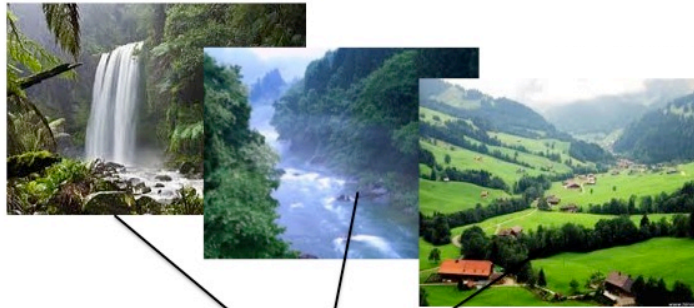
Lots of **useful** information

Lots of information but no
useful information

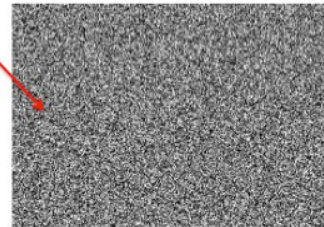
The Information Sieve



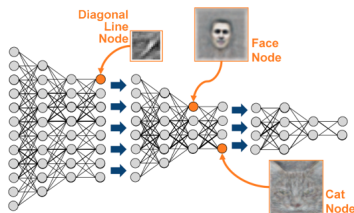
data = ore



useful
information
= gold



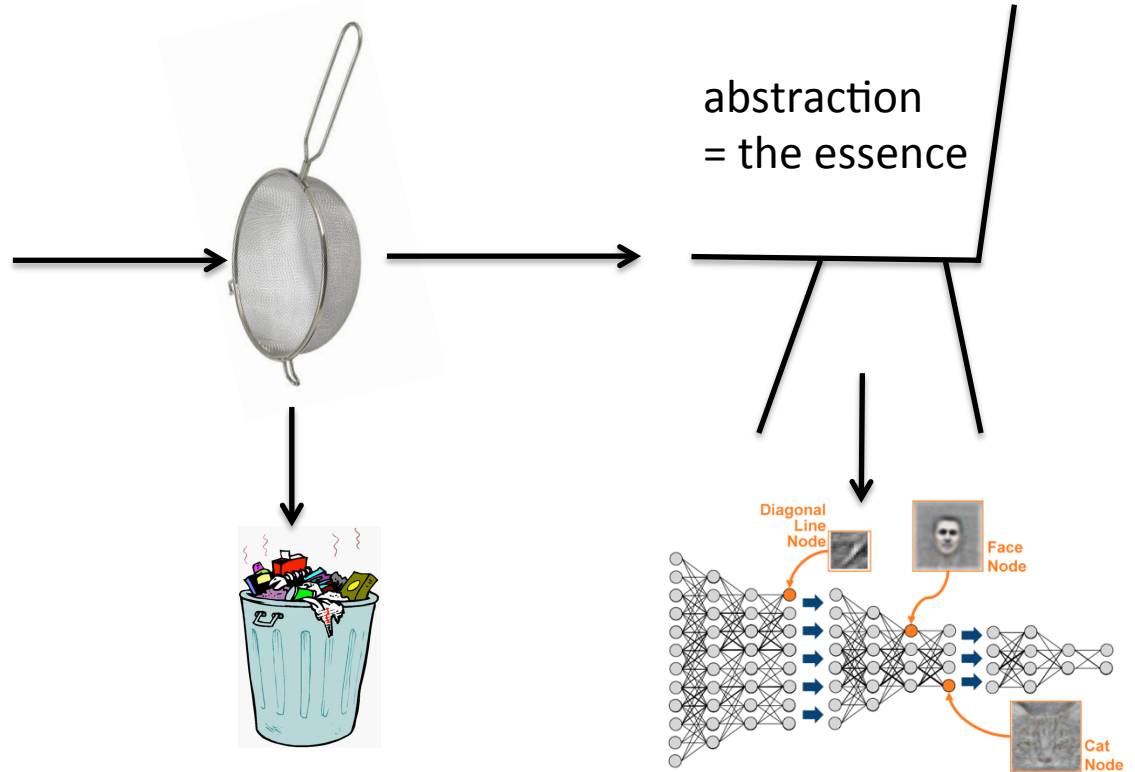
noise = grit



What is this?

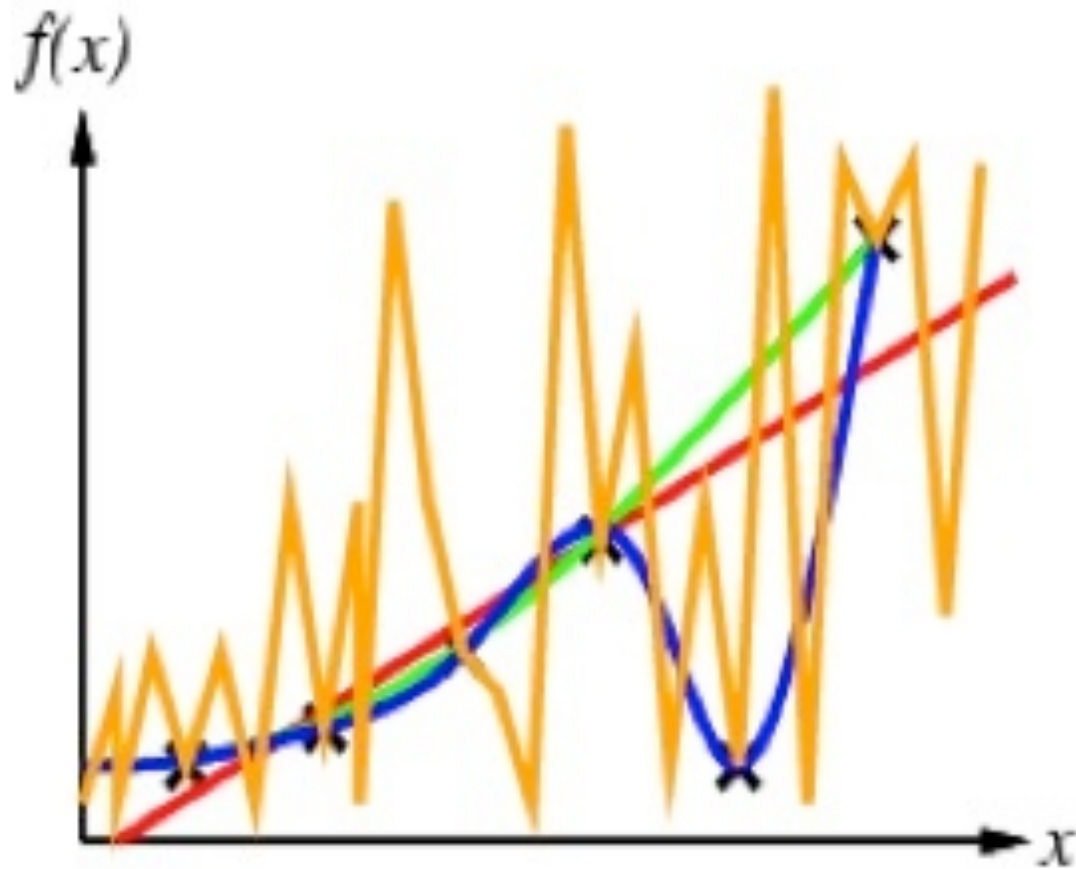


“To Know is to Forget”



To generalize one needs to forget the details and remember the essence.

Overfitting: An Experiment



Wisdom of the Crowd



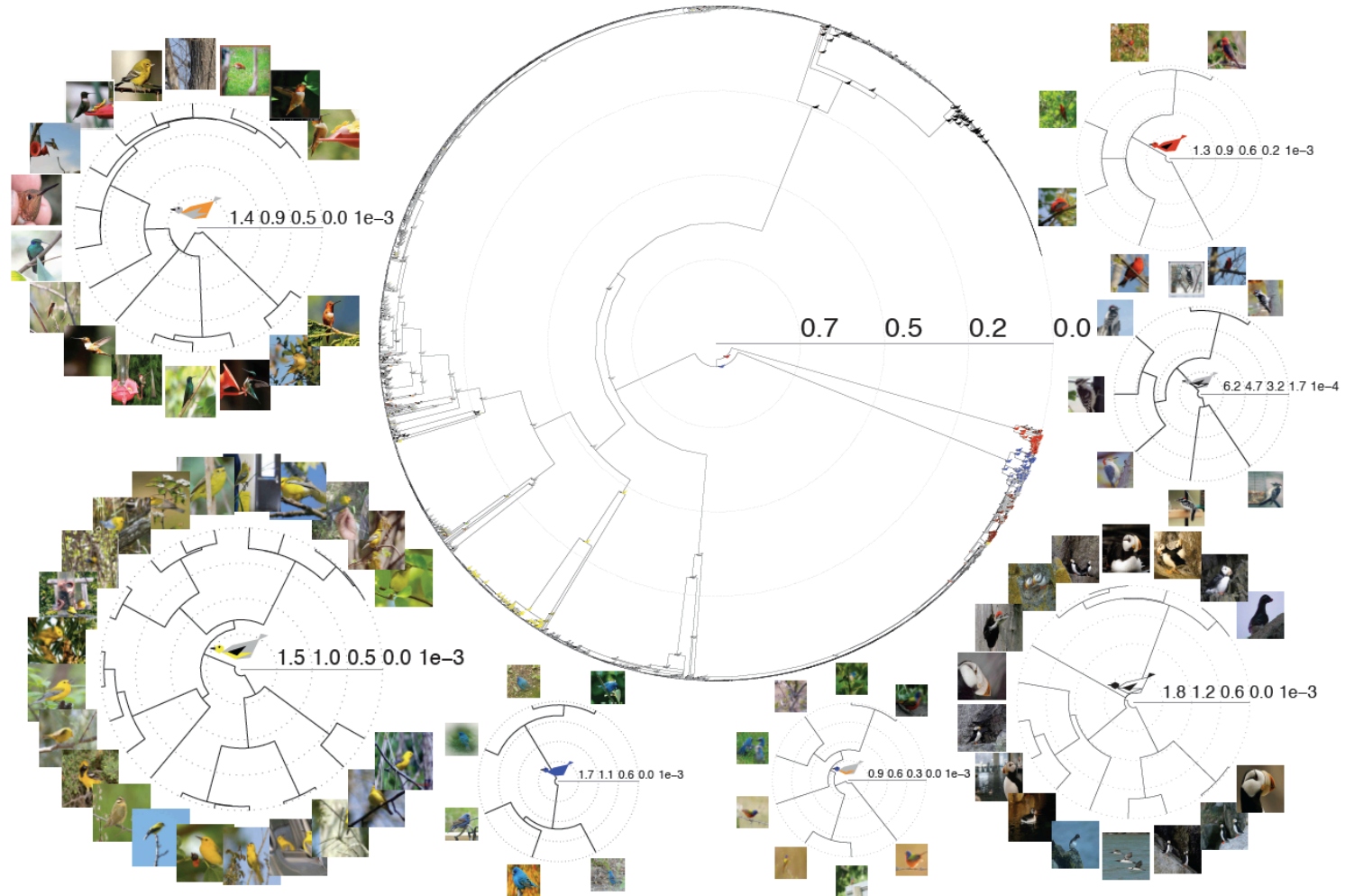
- Everyone guesses the weight of this cow.
- Order all estimates.
- Take the middle guess (1,2,3,4,5)

Answer: **600 kg**

Unsupervised Learning (no labels)

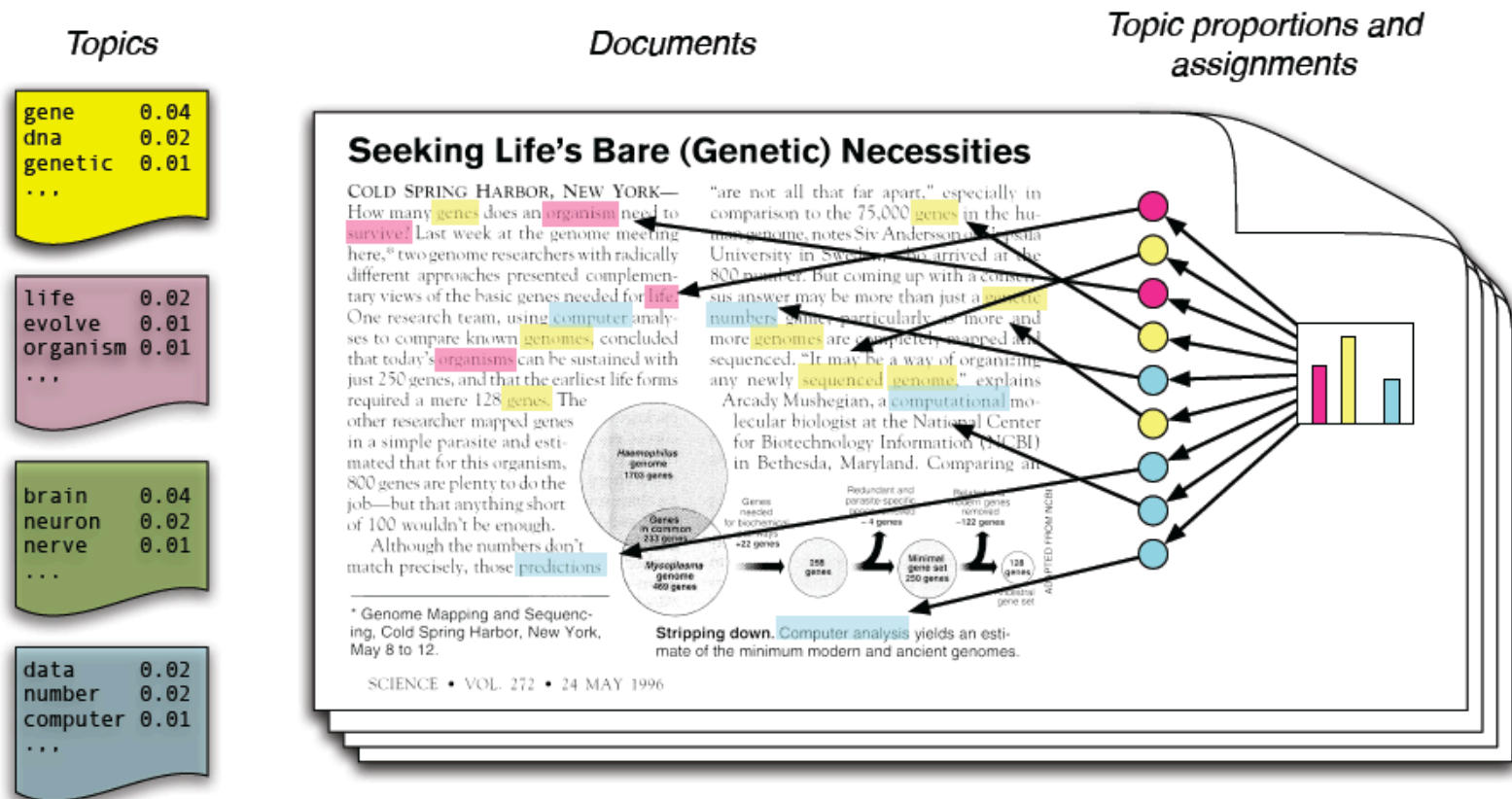
Hierarchical Clustering of Birds

(with Dilan Görür)



Topic Models

- Every document consists of a small number of topics.
- The algorithm learns the topics distribution per document as well as the words in a topic.

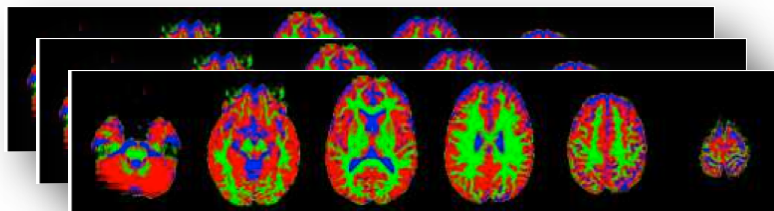


Supervised Learning (with labels)

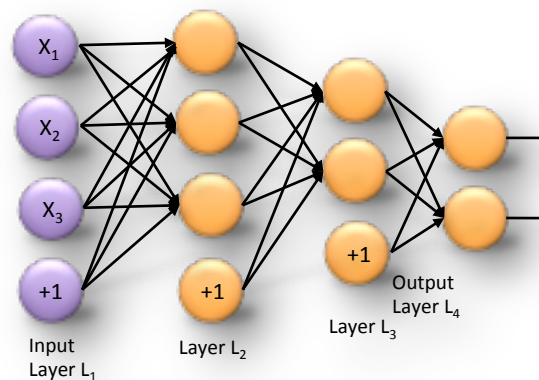
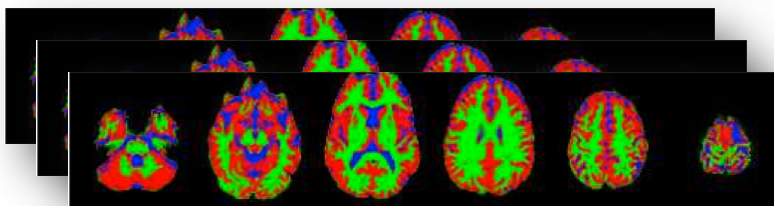
Training a Classifier

Step 1: Train model

Patiënts: 3D scan + profile



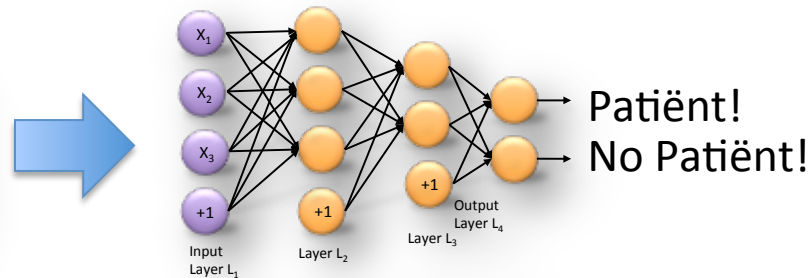
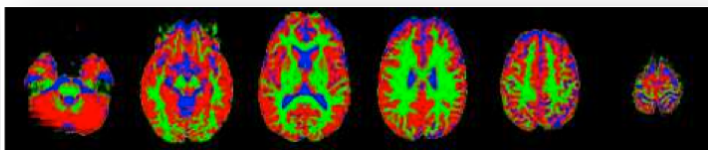
Controls: 3D scan + profile



Deep learning, multi layer network

Step 2: Model Toepassen

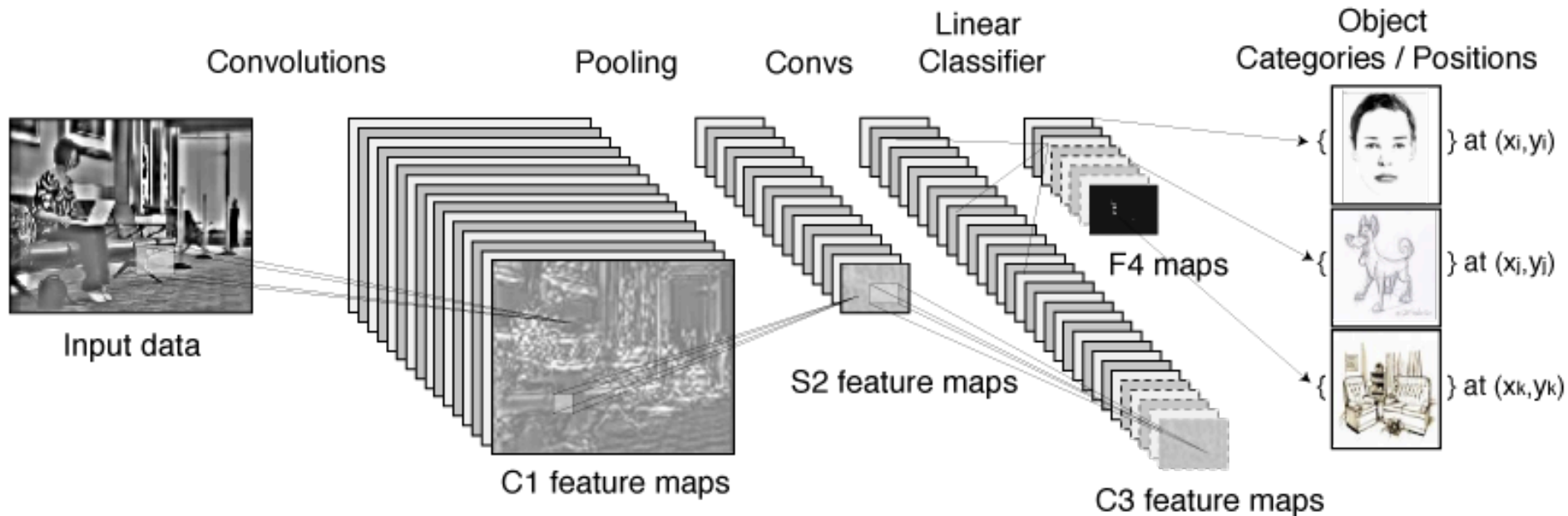
New Subjects: 3D scan + profile





Deep learning, multi layer network

Deep Learning

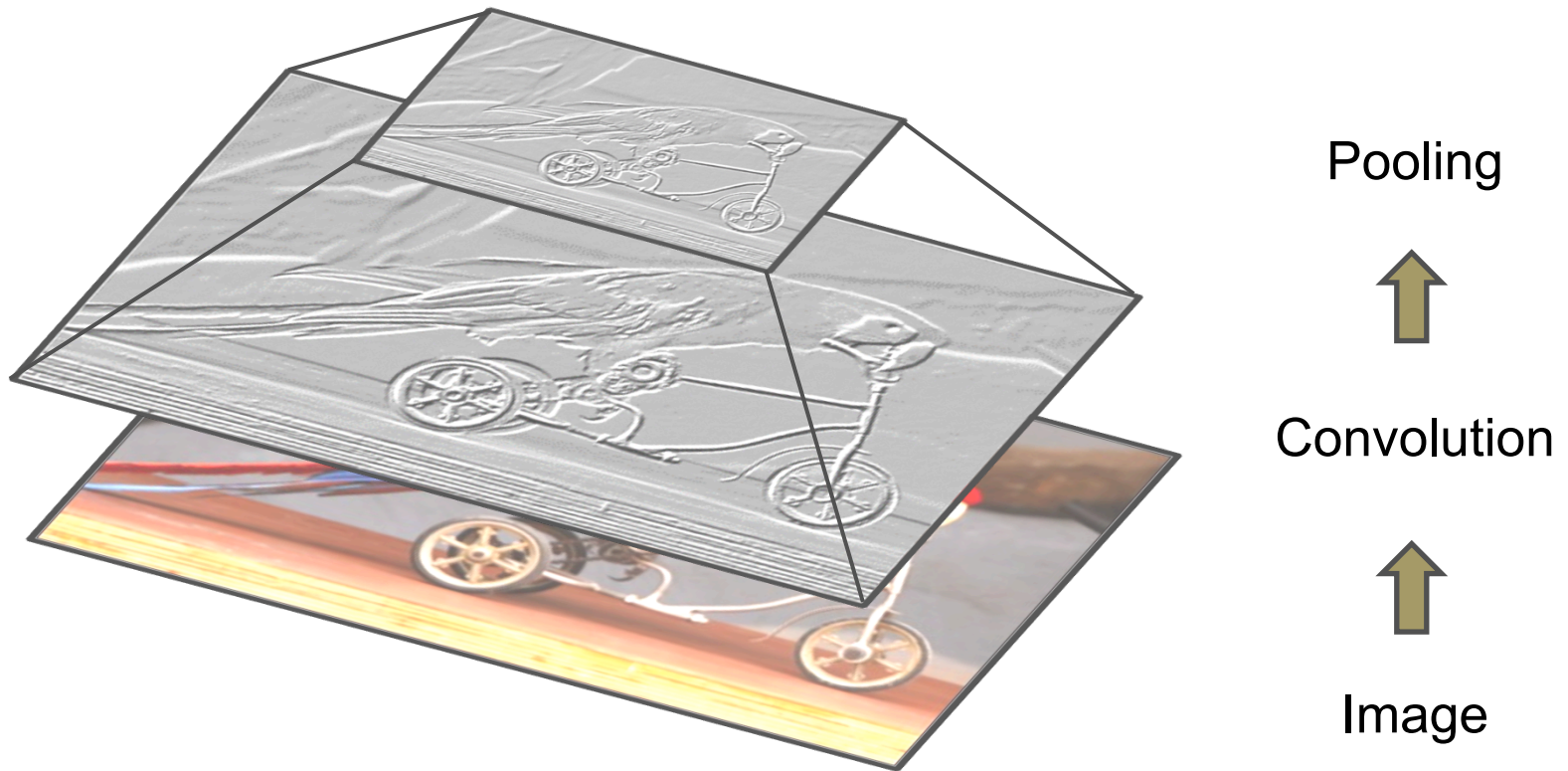
- Neural networks with many layers of artificial neurons (10B parameters)
- Trained on GPUs (supercomputing at home)



 Forward: Filter, subsample, filter, subsample, ..., classify
 Backward: backpropagation

Basic Convolutional Network Operations

(slide borrowed from Li Deng)



Increasingly Abstract Features

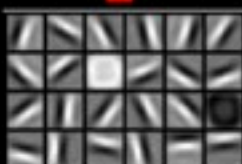
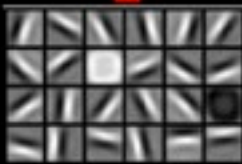
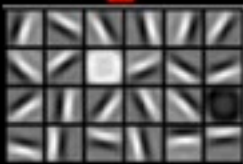
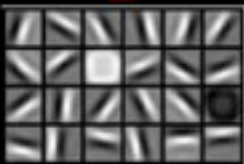
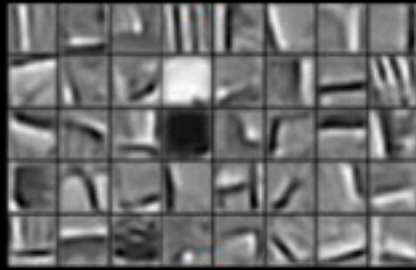
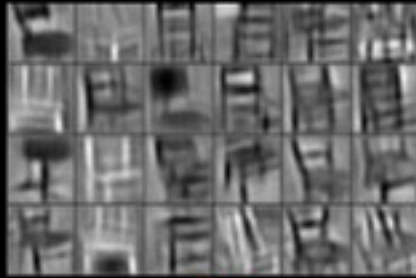
Features learned from training on different object classes.

Faces

Cars

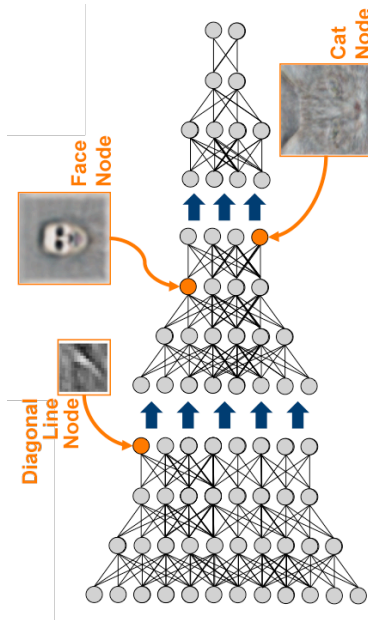
Elephants

Chairs

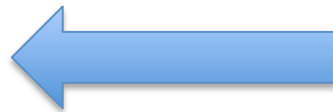


Generative vs Discriminative?

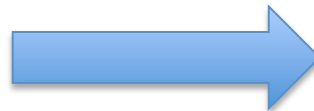
Deep neural network



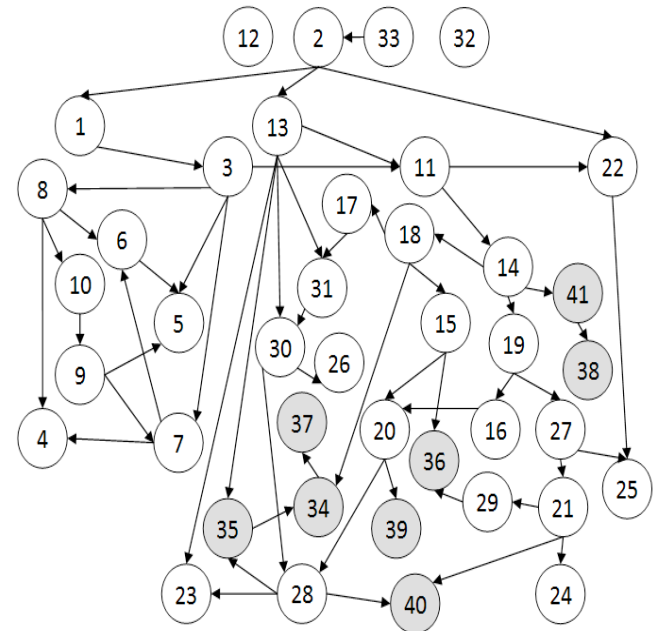
Regularize



Optimize

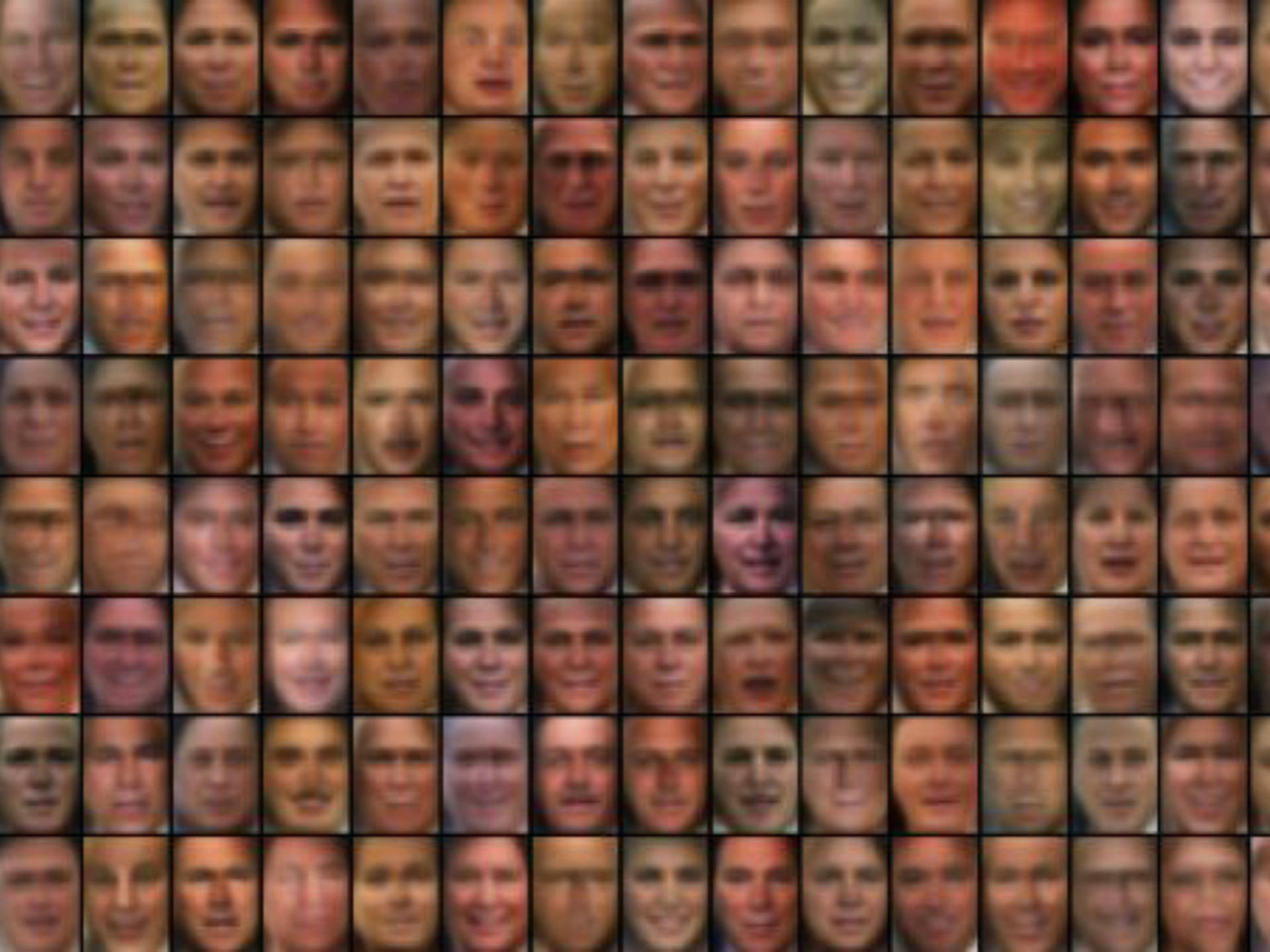


Bayesian network



- Advantages discriminative models:
 - low bias / flexible map from input to target
 - efficient training algorithms available
 - solve the problem you are evaluating on.
 - don't need Bayes rule to classify

- Advantages generative models:
 - inject expert knowledge
 - model causal relations
 - Interpretable
 - probabilities
 - unlabeled data (semi-supervised learning)



0 1 2 3 4 5 6 7 8 9

0 1 2 3 4 5 6 7 8 9

0 1 2 3 4 5 6 7 8 9

0 1 2 3 4 5 6 7 8 9

0 1 2 3 4 5 6 7 8 9

0 1 2 3 4 5 6 7 8 9

0 1 2 3 4 5 6 7 8 9

Conclusions

- Machine Learning tries to make predictions on future data by learning models from historical data.
- Big data, Moore's law and deep learning are revolutionizing AI.
- So, much, that people are getting scared....(so jump on the bandwagon!)

Don't Let Artificial Intelligence Take Over, Top Scientists Warn



Questions

