Discovery

Diagnosis

Prognosis

Care

Deep Learning for Biomedical Discovery and Data Mining

truyen.tran@deakin.edu.au

truyentran.github.io

@truyenoz

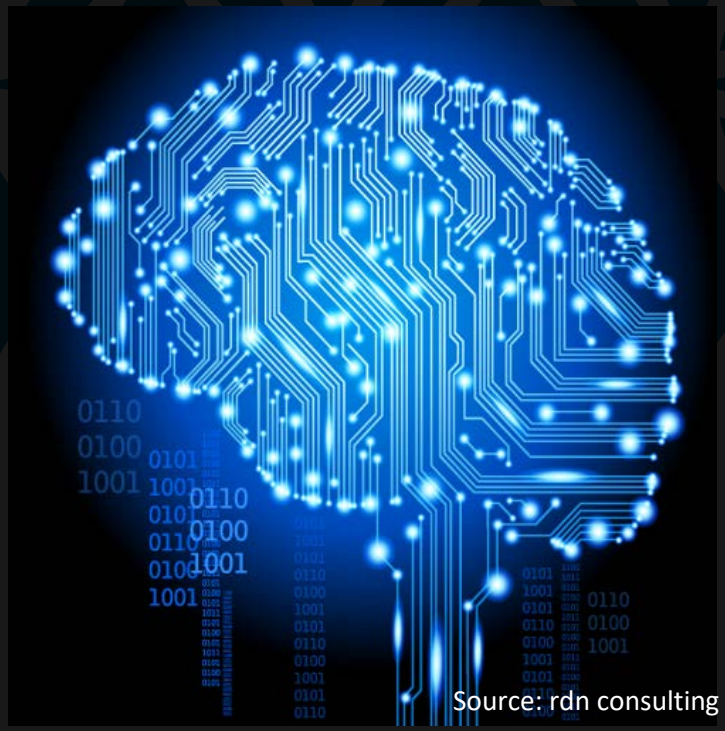
letdataspeak.blogspot.com

goo.gl/3jJ1O0

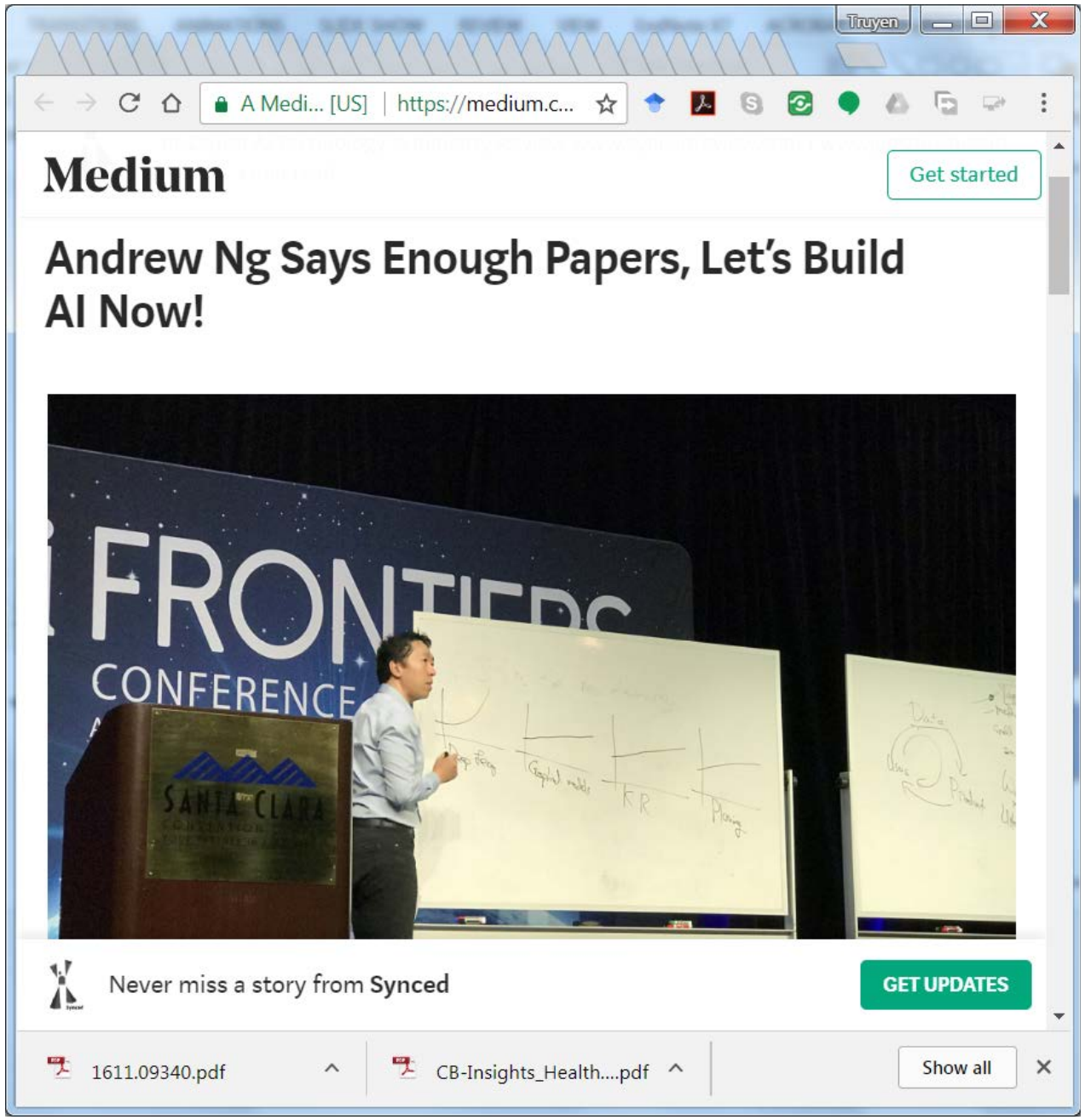
Truyen Tran

Deakin University

Melbourne, June 2017



Source: rdn consulting



"We have enough papers. Stop publishing, and start transforming people's lives with technology!"

We will quickly solve "easy" problems of the form:

A → B

BUT … Should we solve all problems of and for those Internet giants like Google, Facebook & Baidu?

Resources

Slides and references:

- https://truyentran.github.io/pakdd18-tute.html
- Shorten URL: goo.gl/UuZZJ9

Key survey paper (updated frequently):

- **Ching, Travers, et al. "Opportunities And Obstacles For Deep Learning In Biology And Medicine." *bioRxiv* (2018): 142760**

The Team

Agenda

Topic 1: Introduction (20 mins)

Topic 2: Brief review of deep learning (30 mins)

- Classic architectures
- Capsules & graphs
- Memory & attention

Topic 3: Genomics (30 mins)

- Nanopore sequencing
- Genomics modelling

QA (10 mins)

Break (30 mins)

Topic 4: Healthcare (40 mins)

- Time series (regular & irregular)
- EMR analysis: Trajectories prediction
- EMR analysis: Sequence generation

Topic 5: Data efficiency (40 mins)

- Few-shot learning
- Generative models
- Unsupervised learning of drugs

Topic 6: Future outlook

QA (10 mins)

Topics not covered

Privacy preserving

Medical imaging (2-4D)

Neuroscience

- Models for spike trains
- Deep learning for connectomics

Biomedical NLP

- Classical & social NLP
- Knowledge graphs

Wearables

- Tracking the state of physical and mental health
- Lifestyle management & monitoring

Health Insurance

- Future illness/spending prediction
- Proactive prevention programs
- **WARNING: Working for insurance companies does raise ethical concerns!**

Nutrition: Mobile phone vision → calories

Explainable AI

- Seeing through the black-box, e.g., visualization, motifs
- Explainable architectures that use biological mechanisms and medical ontologies
- Dual architecture: predictor & explainer

-omics: Gene expression & Proteomics

Why now?

High-impact & **data-intensive**.

- Andrew Ng's rule: impact on 100M+ people.
- Biomedicine is the only industry that will never shrink!

Ripe for innovations fuelled by deep learning techniques.

- Major recent advances and low hanging fruits are being picked.

Great **challenges**:

- High volume and high dimensional;
- Great privacy concerns;
- Need integrated approach to encompass great diversities.

It is the right time to join force with biomedical scientists!

Biomedicine is ripe for ML/KDD – (or other way around?)

ML/KDD that matters

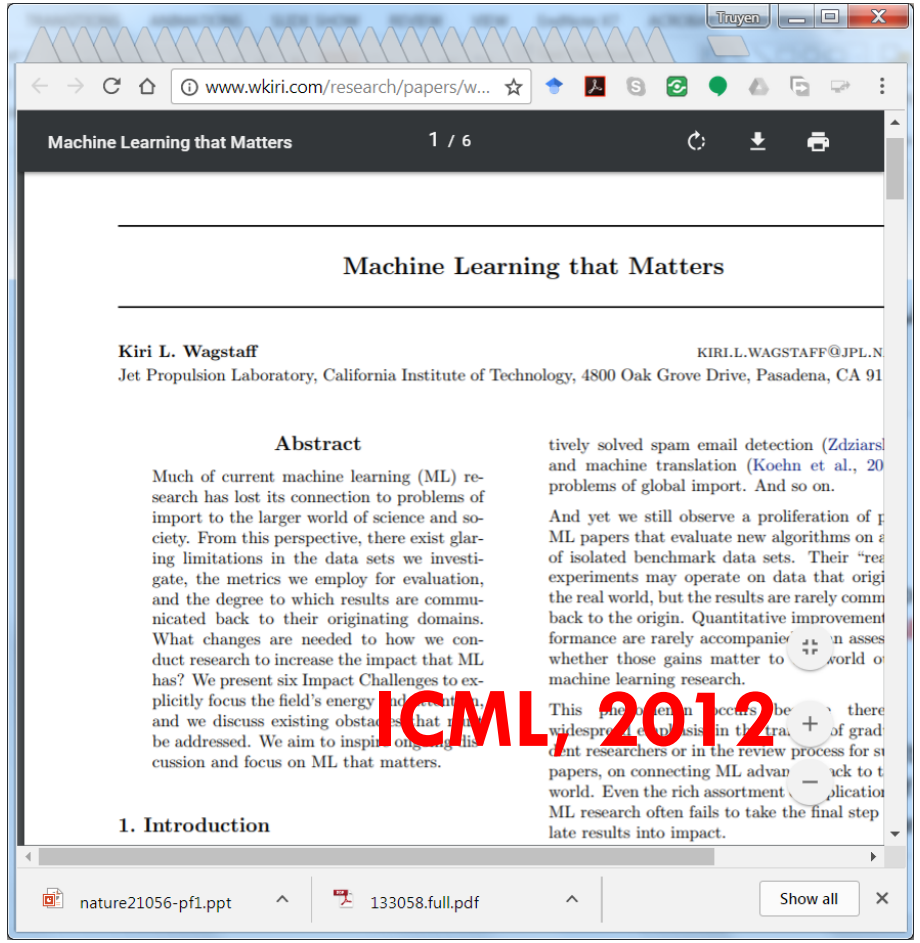
- E.g., huge successes in radiology with off-the-shell CNNs
- Big business opportunities, e.g., IBM Watson for Health

Excellent testbed for machine learning techniques

- Any modality: 2D-4D vision, time-series, 1D signals, sound, text, social network, graphs.
- **For DL, any neural architectures: CNN/CapsNet, RNN, Memory, DBN/VAE/GAN**
- An excellent escape from the UCI datasets!

Excellent sources of new problems

- Metric scale from nano-meter (atoms) to meters (human body and brain).
- Time scale from mini-seconds (ion channels) to 100 years.
- Complexity unimaginable (e.g., brain, DNA, cell networks).



Recent AI/ML/KDD activities

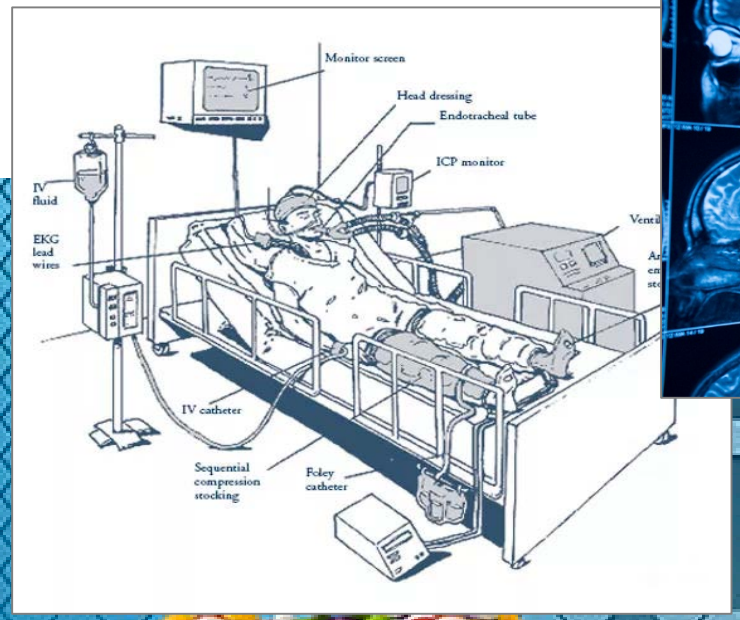
Conference on Machine Learning for Healthcare (MLHC), 2018

ICML/IJCAI/AAAI (2018)

- Joint Workshop on Artificial Intelligence in Health
- The 3rd International Workshop on Knowledge Discovery in Healthcare Data
- The 3rd International Workshop on Biomedical Informatics with Optimization and Machine Learning
- AI for synthetic biology
- Health Intelligence
- Workshop on Computational Biology

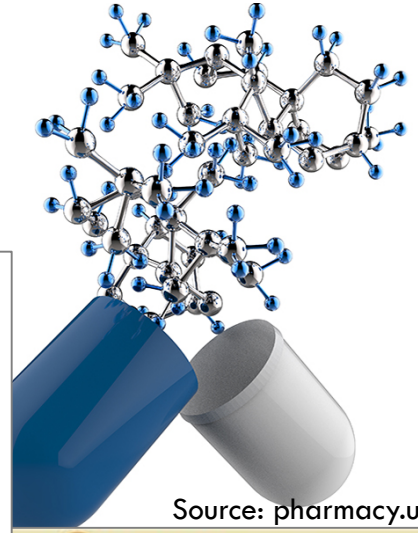
KDD/SDM/ICDM (2018)

- **Health Day at KDD'18**
- epiDAMIK: Epidemiology meets Data Mining and Knowledge discovery
- 2018 KDD Workshop on Machine Learning for Medicine and Healthcare
- 17th International Workshop on Data Mining in Bioinformatics
- Workshop on Data Mining in Bioinformatics (BIOKDD 2018)

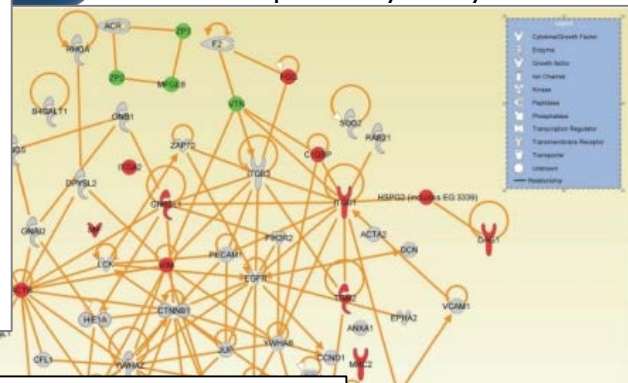


http://www.healthpages.org

http://engineering



Source: pharmacy.umaryland.edu



PubMed

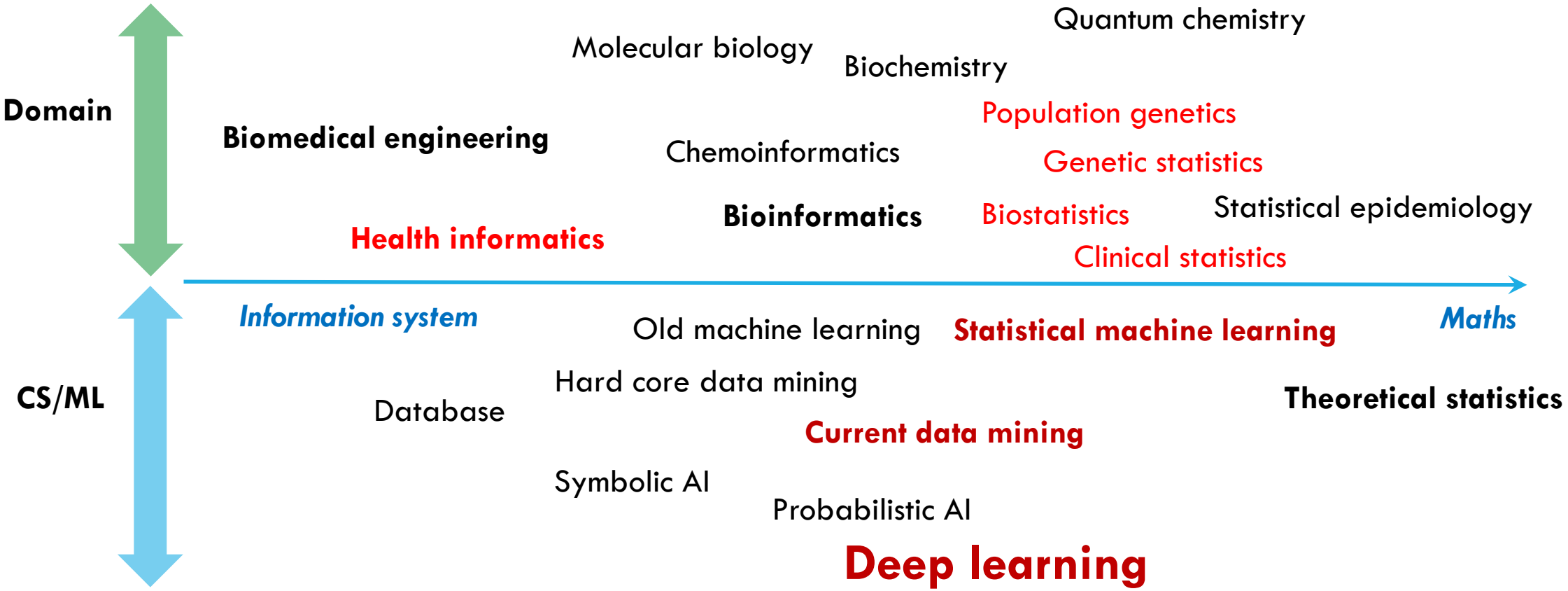
marketingland.com

s://openi.nlm.nih.g

AGATGGATC

Big Rooms in Biomedicine

First thing first: Speak their languages



Using intuition and domain knowledge

Intuition is important to reduce hypothesis space

- There are infinite number of hypotheses
- We need to search for some highly probably ones!

But it can be deadly wrong!

- A recently discharged patient can be readmitted right away (just like not treated).
- A good doctors can be associated with high rate of mortality and readmission.

Domain knowledge is critical

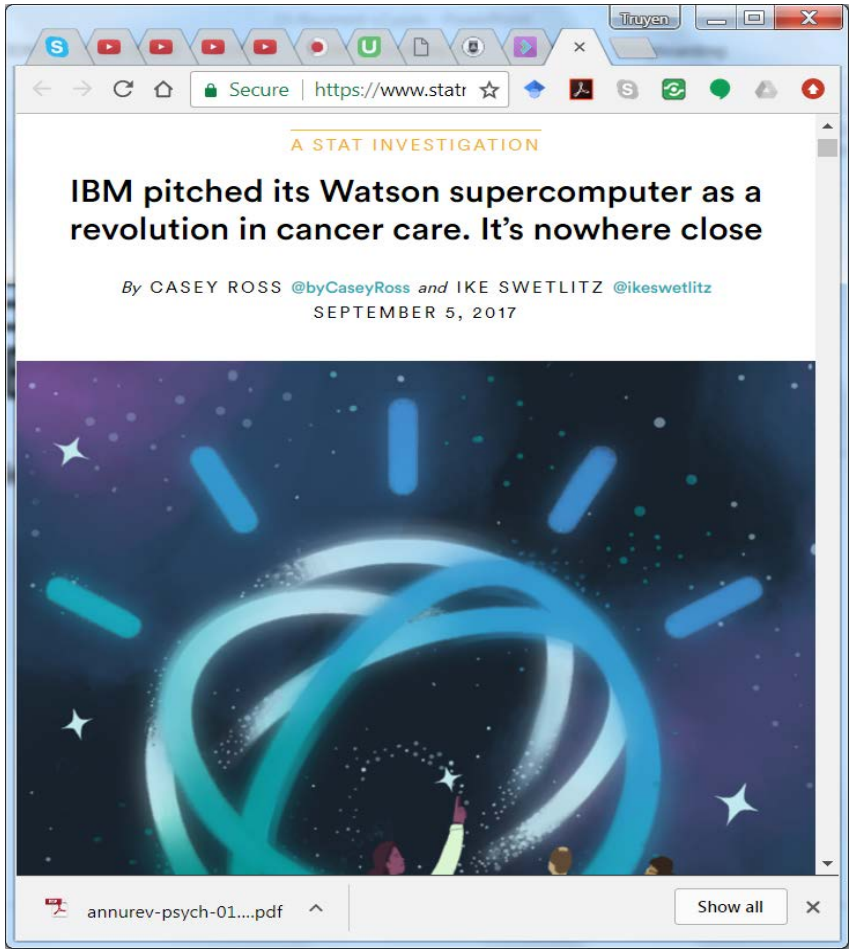
- Check the literature. Obey the laws. Follow protocols.
- Do the home work, e.g., pregnancy diabetes; women with prostate cancer; men with breast cancer.
- **Choose right neural architectures!**

But … a lot of data can support any dump tricks!

Let’s be warned!

2011

2017



“He said later that the background information Watson provided, including medical journal articles, was helpful, giving him more confidence that using a specific chemotherapy was a sound idea.

But the system did not directly help him make that decision, nor did it tell him anything he didn’t already know.”

What make biomedicine hard for deep learning?

Great diversity but may be small in size

High uncertainty, low-quality/missing data

Reusable models do not usually exist

Human doesn't know how to read biomedicine (Brendan Frey, U of Toronto)

Require deep thinking for a reasonable deep architecture

However, at the end of the day, we need only a few generic things:

- Vector → DNN (e.g., highway net) | Sequence → RNN (e.g., LSTM, GRU)
- Repeated motifs → CNN | Set → Attention
- Graphs → Conv graphs; Column Networks

Health data is exceptionally hard to model

Many diseases are not fully understood

- Cancers, mental health

Many treatments have little or unknown effects

Care processes are complex: Protocols, regulations, multiple stakeholders (patients, family, nurse, doctor, hospital, community, government, the public).

Data is (usually) small, biased, noisy, irregular/episodic, missing with external interventions

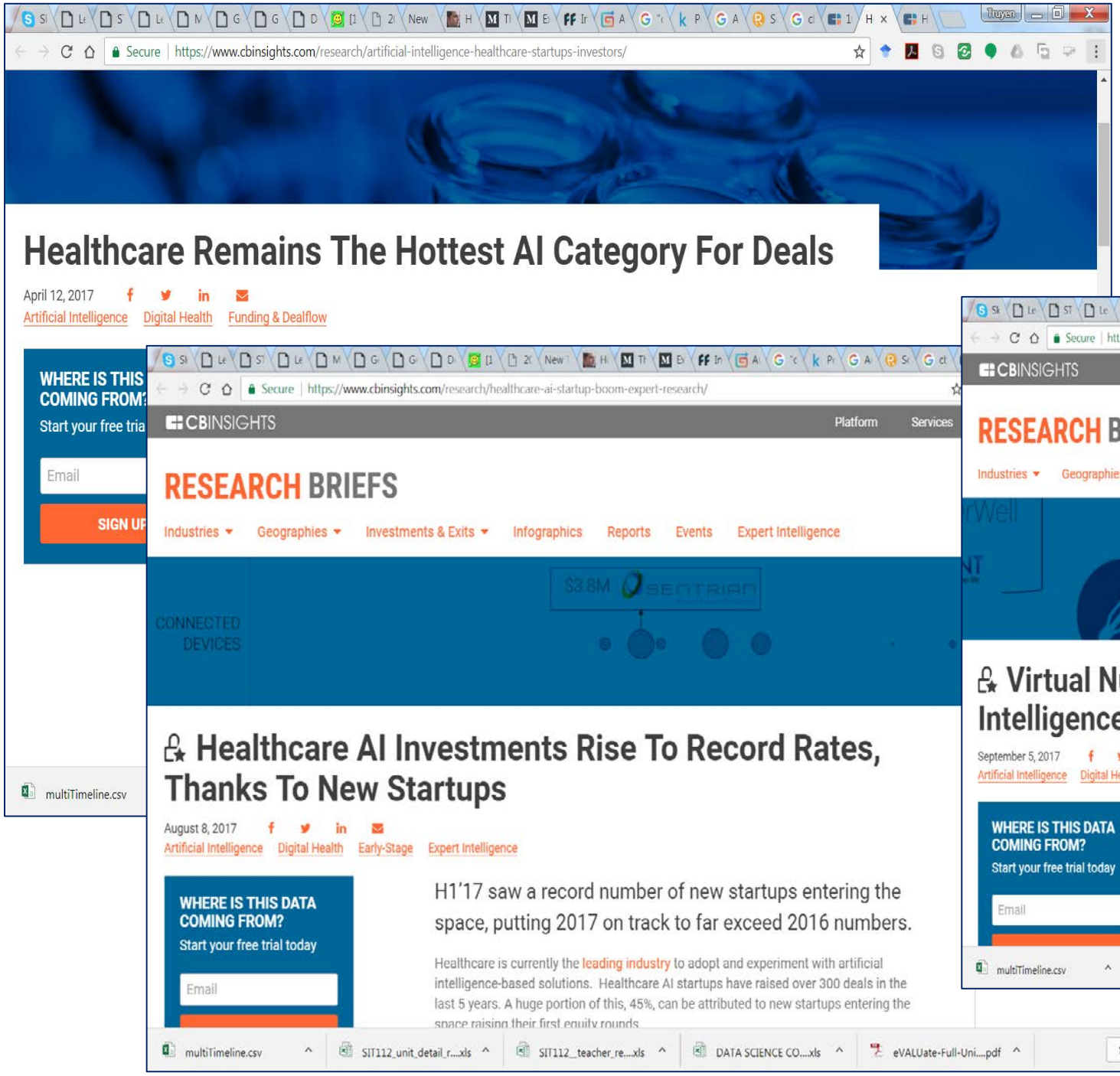
Data is not shared, due to privacy and ethical concerns

Predicting into future isn't like finding out what is there (e.g., classification)

Decision making is complex.

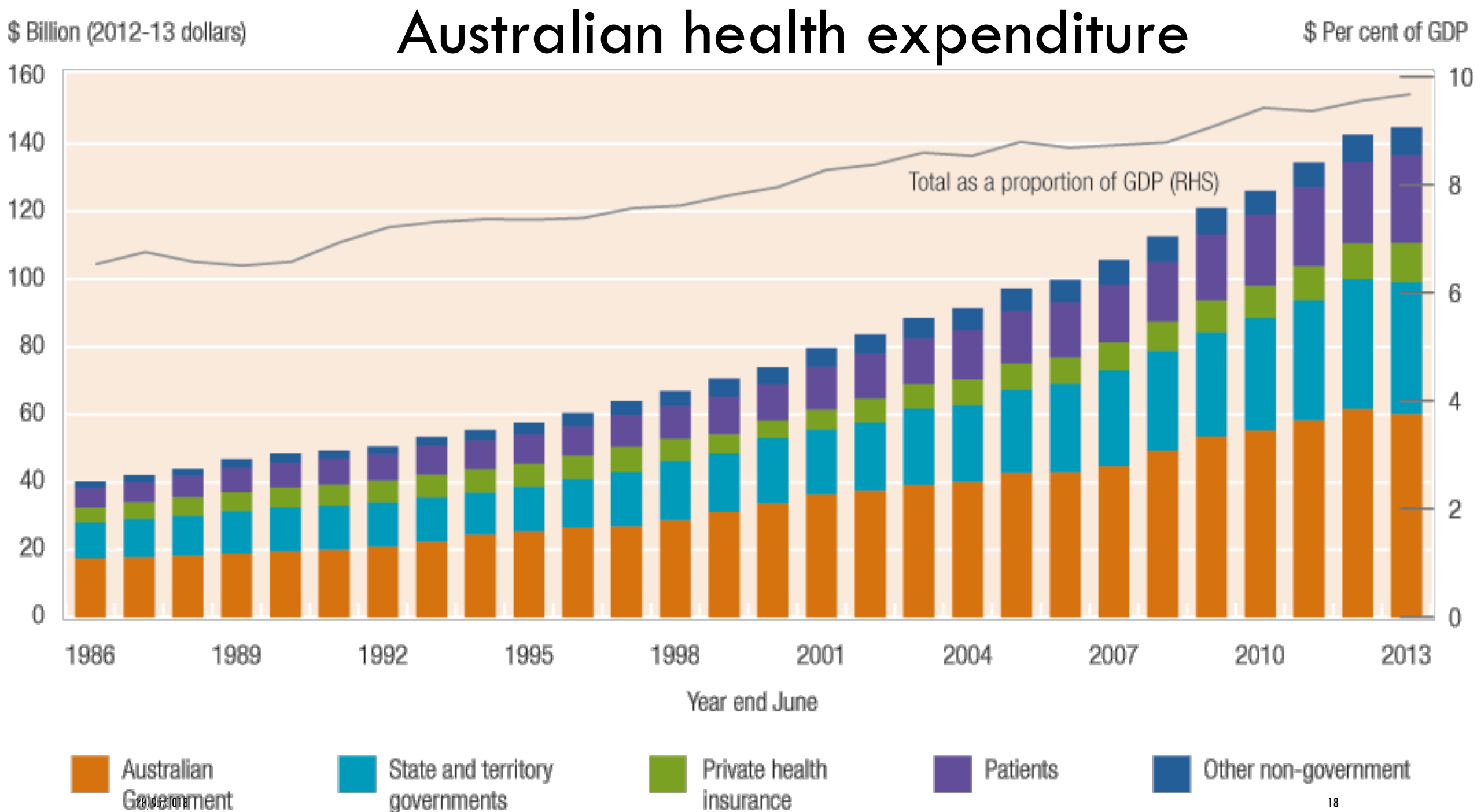
Doctors are rightfully sceptical of what is new.

Then the good news



Google Trends: "artificial intelligence" + healthcare

Australian health expenditure



http://www.pc.gov.au/news-media/pc-news/pc-news-may-2015/improving-australia-health-system

It has just started.

Supervised learning

(mostly machin

A → B

Will be quickly solved for "easy problems (Andrew Ng)

Anywhere in between: semi-supervised learning, reinforcement learning, lifelong learning, meta-learning, few-shot learning, knowledge-based ML

Unsupervised learning

han)

$$\mathbf{v} \sim P_{model}(\mathbf{v})$$

$$(\mathbf{v}) \approx P_{data}(\mathbf{v})$$

Biologist

Bioinformatician

Physician

Health informatician

AI/ML/DL

How does deep learning work for biomedicine?

Discovery

Diagnosis

Prognosis

Efficiency

http://hubpages.com/education/Top-Medical-Inventions-of-The-1950s
http://www.ctrr.net/journal
https://cdn1.iconfinder.com

Agenda

Topic 1: Introduction (20 mins)

Topic 2: Brief review of deep learning (30 mins)

- Classic architectures
- Capsules & graphs
- Memory & attention

Topic 3: Genomics (30 mins)

- Nanopore sequencing
- Genomics modelling

QA (10 mins)

Break (30 mins)

Topic 4: Healthcare (40 mins)

- Time series (regular & irregular)
- EMR analysis: Trajectories prediction
- EMR analysis: Sequence generation

Topic 5: Data efficiency (40 mins)

- Few-shot learning
- Generative models
- Unsupervised learning of drugs

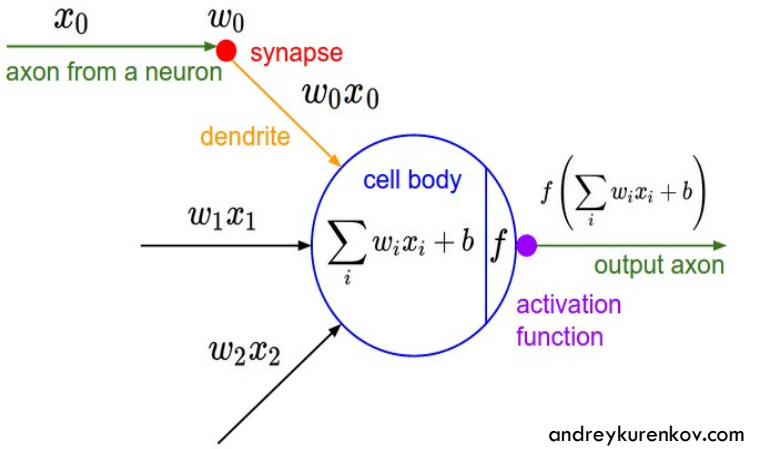
Topic 6: Future outlook

QA (10 mins)

What is deep learning?

Quick answer: multilayer perceptrons (aka deep neural networks) of the 1980s rebranded in 2006

- Same backprop trick, as of 2017.
- Has a lot more hidden layers (100-1000X).
- Much bigger labelled datasets.
- Lots of new arts (dropout, batch-norm, Adam/RMSProp, skip-connections, Capsnet, external memory, GPU/TPU, etc.).
- Lots more people looking at lots of (new) things (VAE, GAN, meta-learning, continual learning, fast weights, etc.)



Feature ~~engineering~~ learning

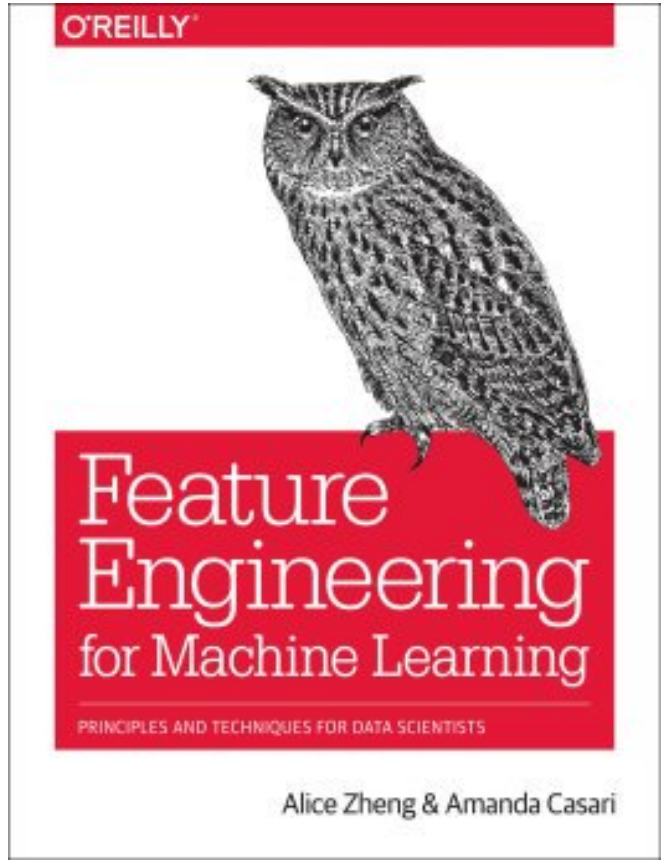
In typical machine learning projects, 80-90% effort is on feature engineering

- A right feature representation doesn't need fancy classifiers to work well.

Text : BOW, n-gram, POS, topics, stemming, tf-idf, etc.

Software: token, LOC, API calls, #loops, developer reputation, team complexity, report readability, discussion length, etc.

Try yourself on Kaggle.com!



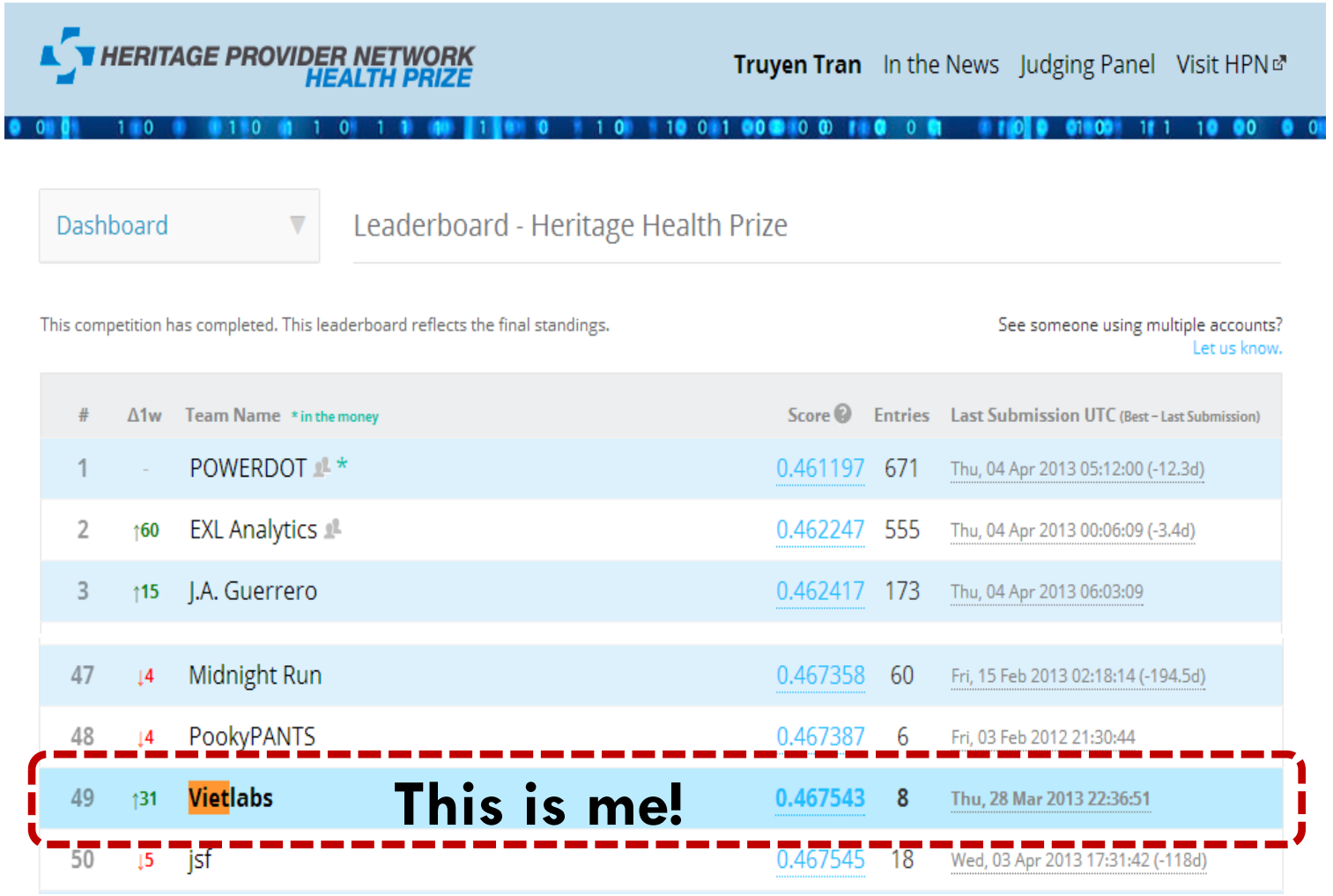
Feature engineering = $$$

$3M Prize, 3 years

170K patients, 4 years worth of data

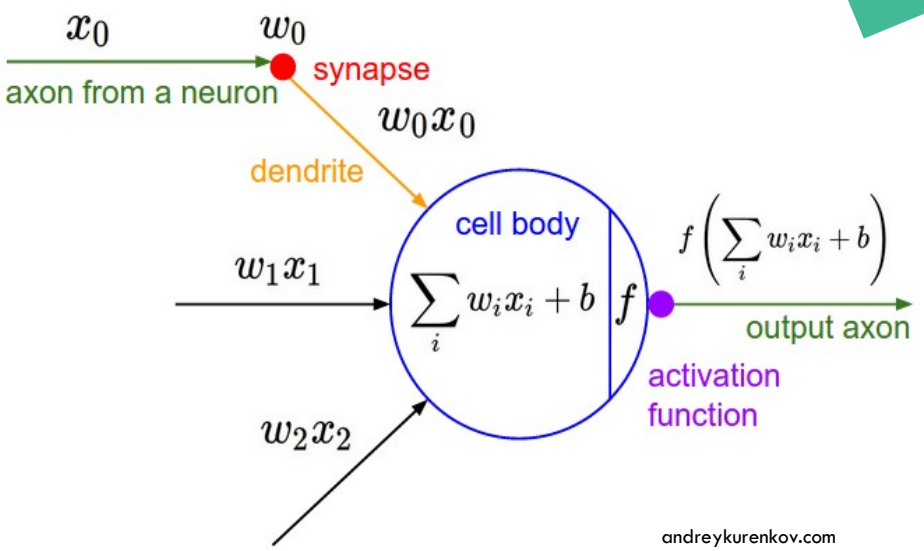
Predict length-of-stay next year

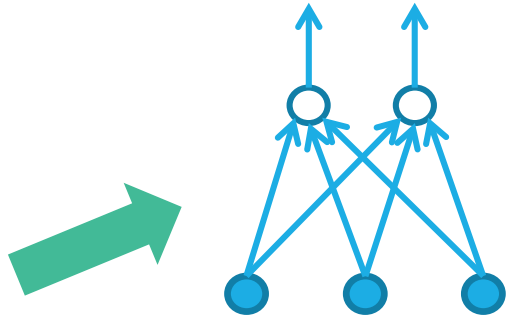
Not deep learning yet (early 2013), but strong ensemble needed → suggesting dropout/batch-norm



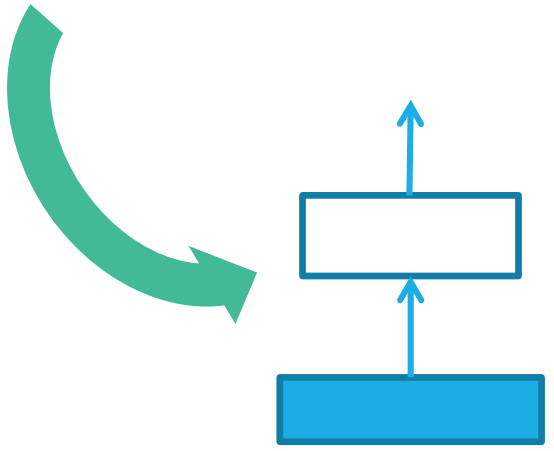
Deep learning as feature learning

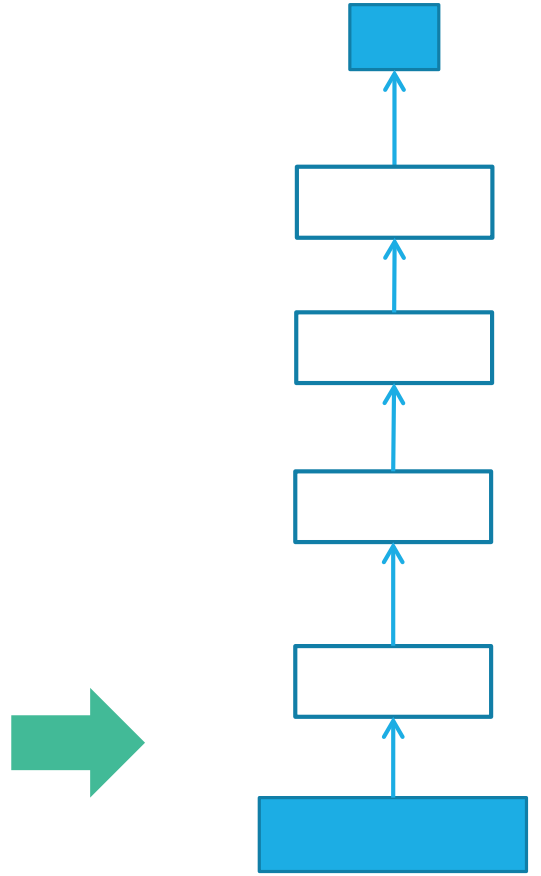
Integrate-and-fire neuron





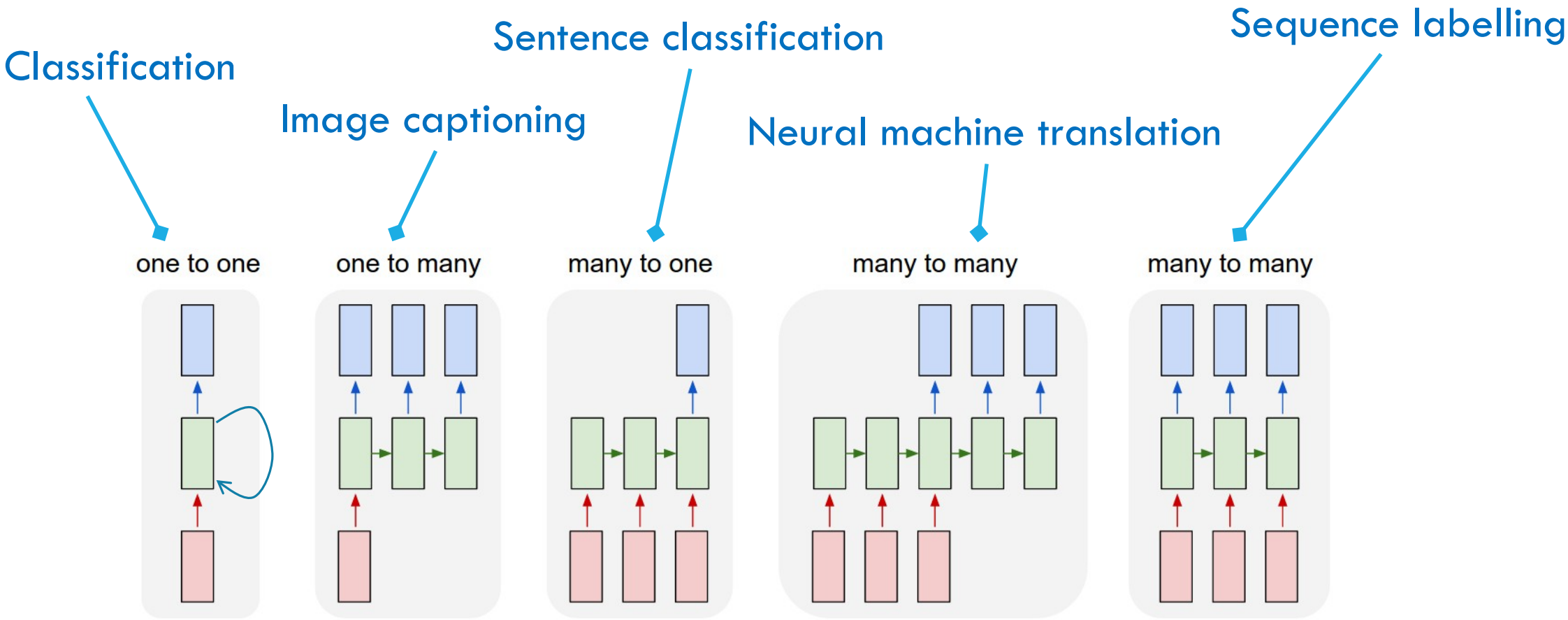
Feature detector





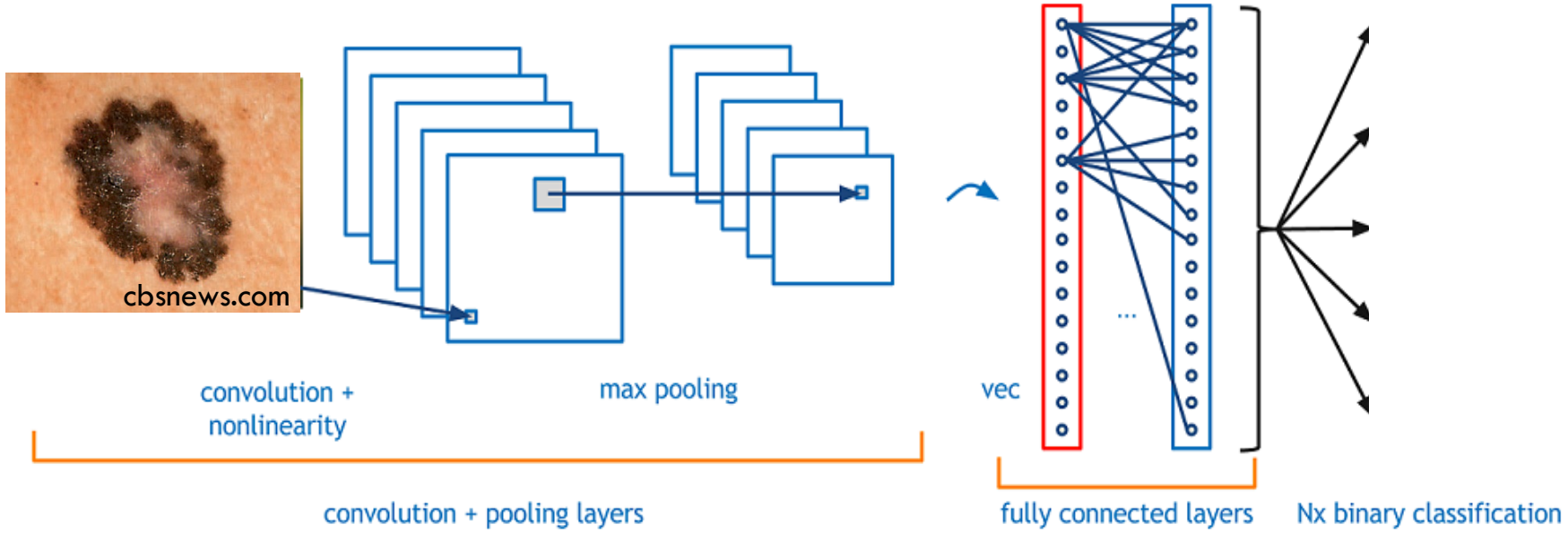
Block representation

Recurrent neural networks



Source: http://karpathy.github.io/assets/rnn/diags.jpeg

Convolutional nets

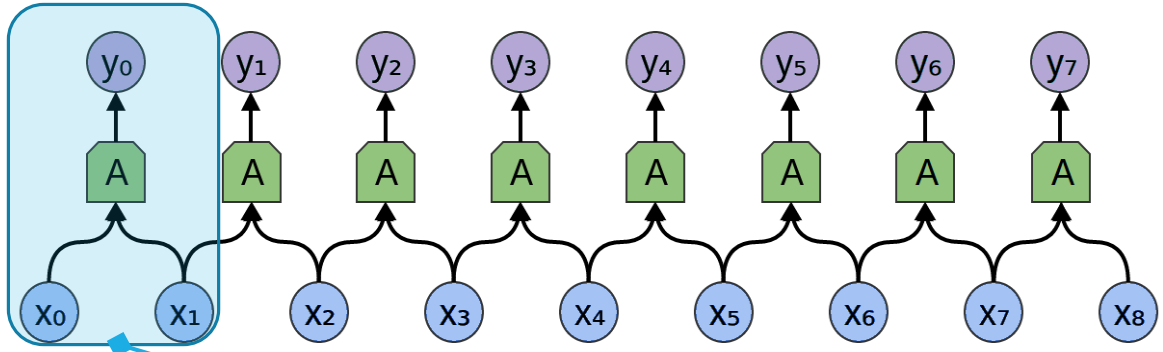


adeshpande3.github.io

Learnable convolution

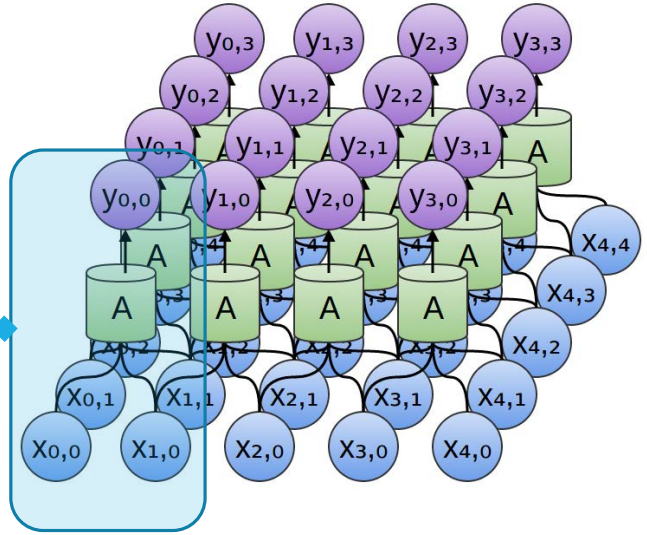
Learnable kernels

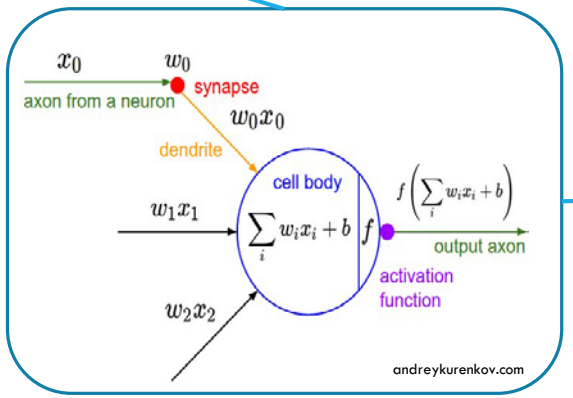
$$y_i = \sum_c K(c) x_{i+c}$$



http://colah.github.io/posts/2015-09-NN-Types-FP/

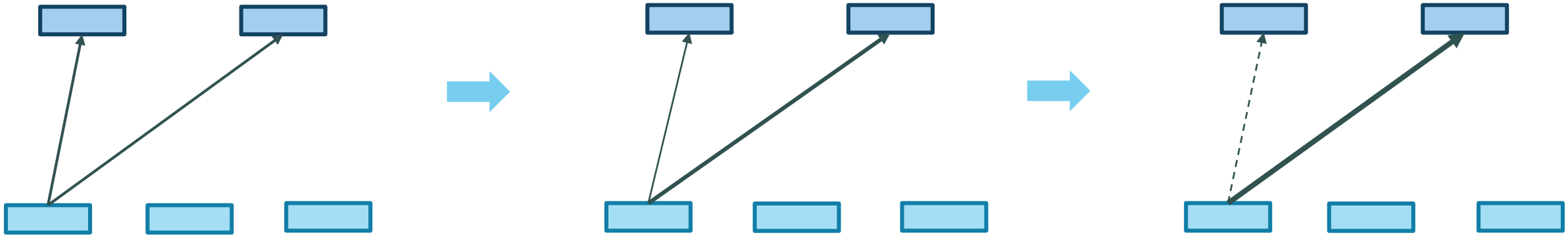
$$y_{ij} = \sum_{c,d} K(c,d) x_{i+c,j+d}$$

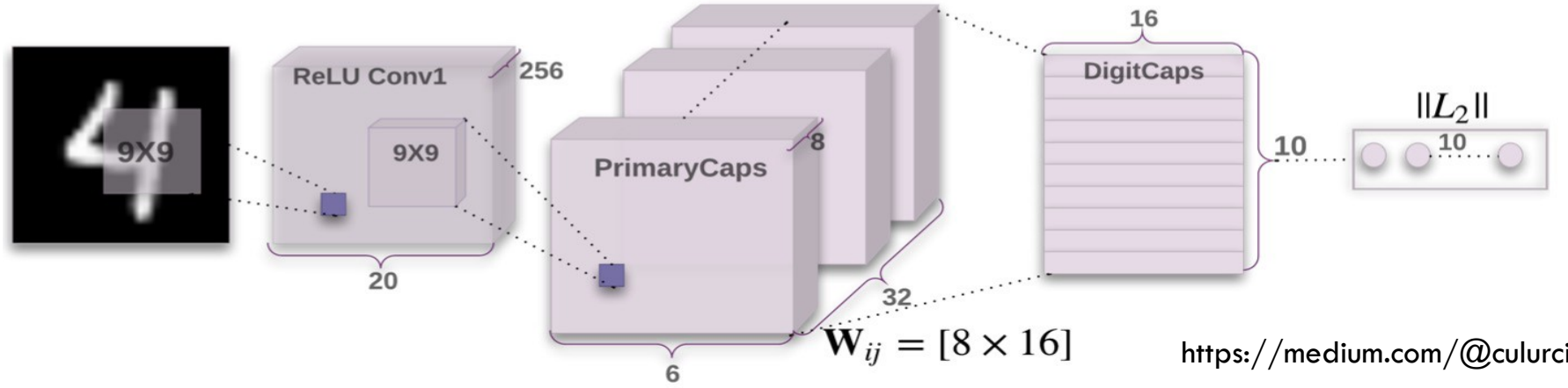




Feature detector, often many

CapsNet (Hinton's group)





https://medium.com/@culurciello

Graphs

Goal: representing a graph as a vector

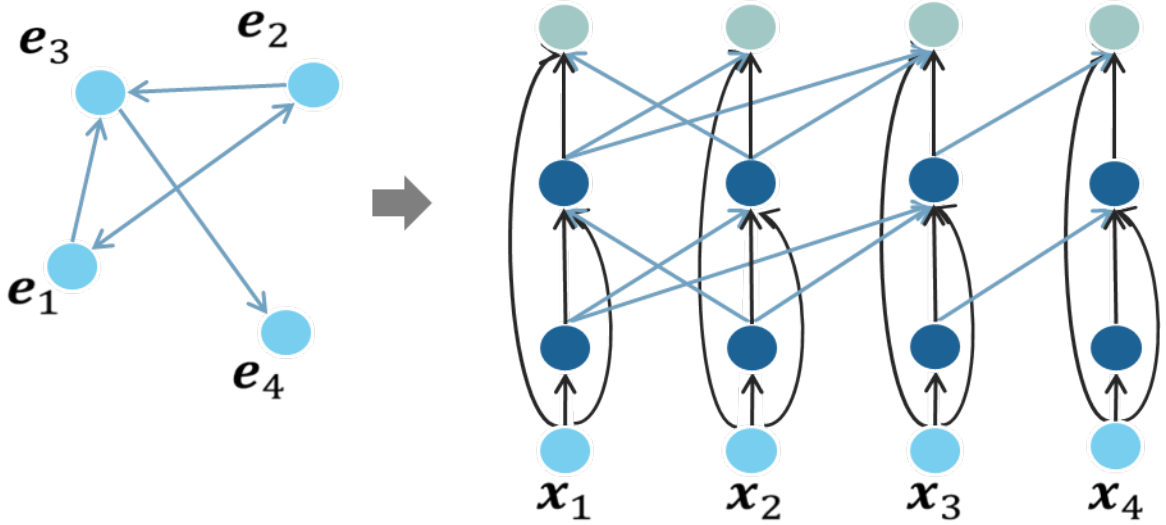
Many applications

- Drug molecules
- Object sub-graph in an image
- Dependency graph in software deliverable

Recent works:

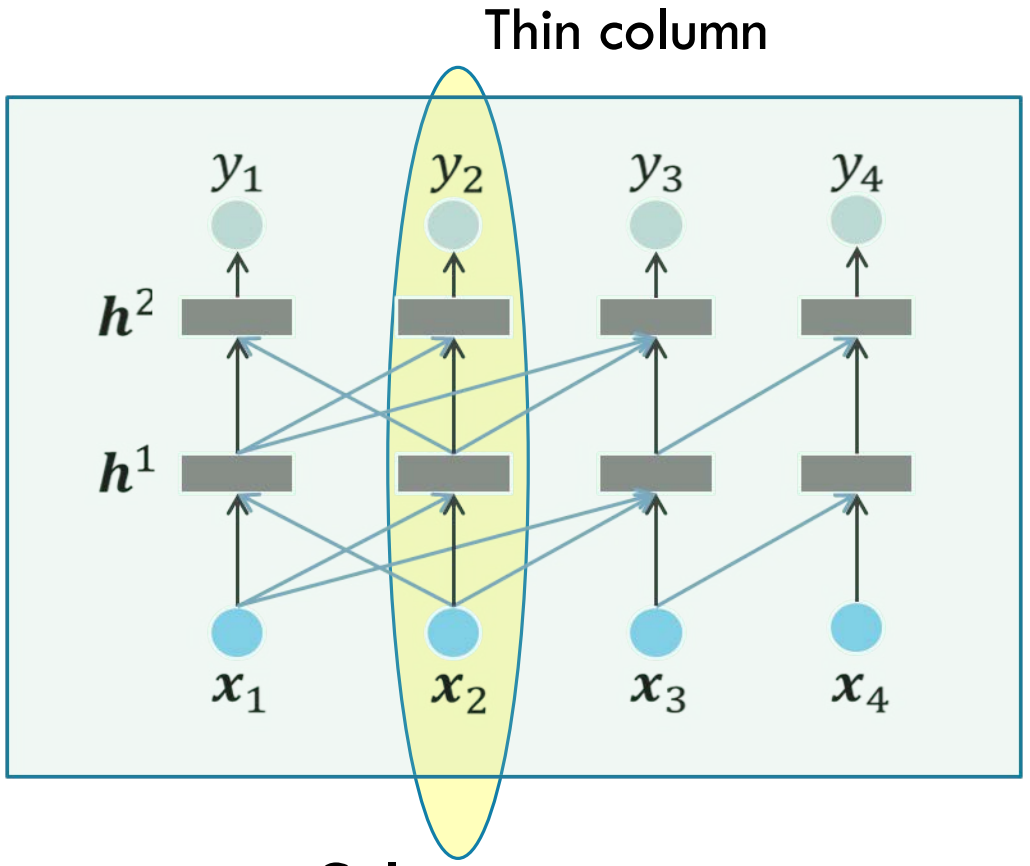
- Graph recurrent nets, column nets (Pham et al, 2017).
- Graph variational autoencoder (Kipf & Welling, 2016)
- Graph convolutional nets (LeCun, Welling and many others)

Column networks



Relation graph

Stacked learning

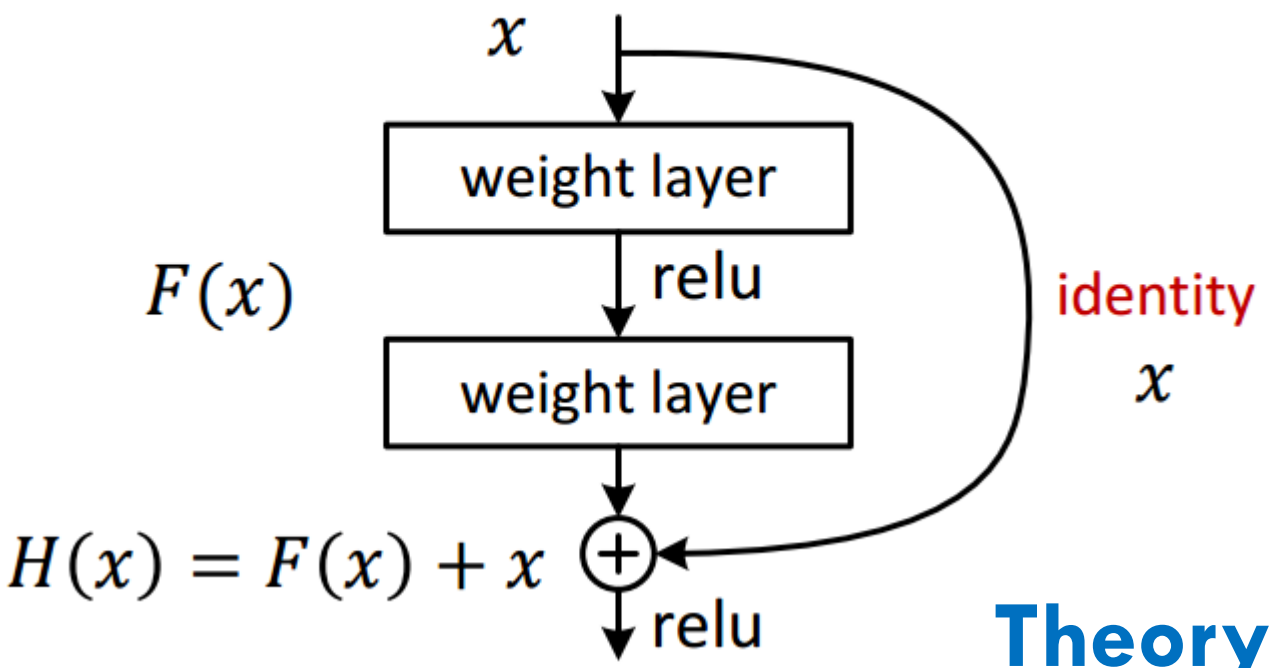


Column nets

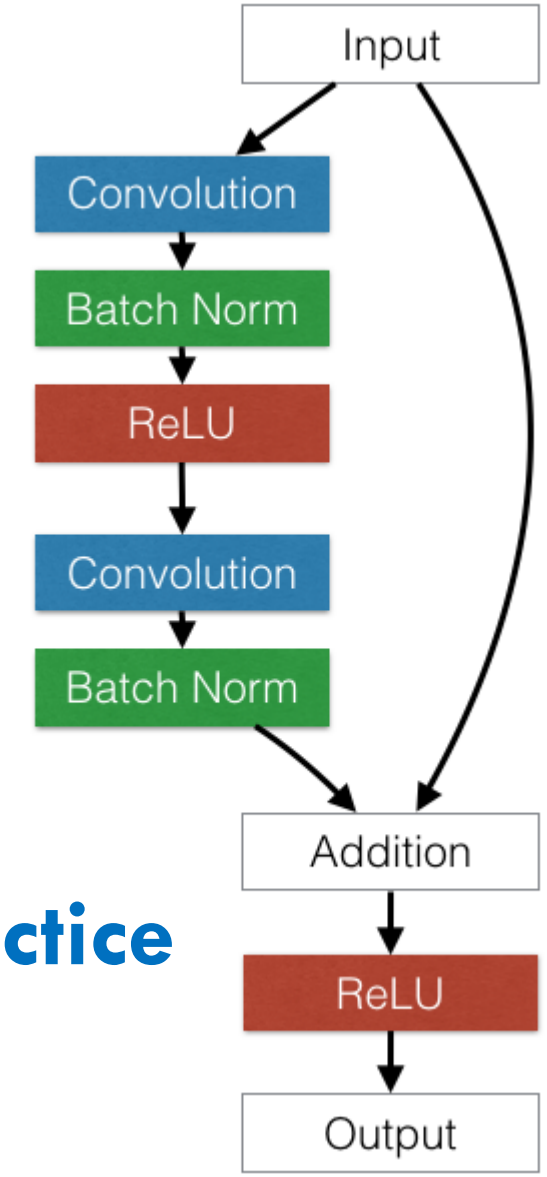
#REF: Pham, Trang, et al. "Column Networks for Collective Classification." *AAAI*. 2017.

Skip-connections

- Residual net



Theory



Practice

http://qiita.com/supersaiakujin/items/935bbc9610d0f87607e8

http://torch.ch/blog/2016/02/04/resnets.html

MANN: Memory-augmented neural networks

Long-term dependency

- E.g., outcome depends on the far past
- Memory is needed (e.g., as in LSTM)

Complex program requires multiple computational steps

- Each step can be selective (attentive) to certain memory cell

Operations: Encoding | Decoding | Retrieval

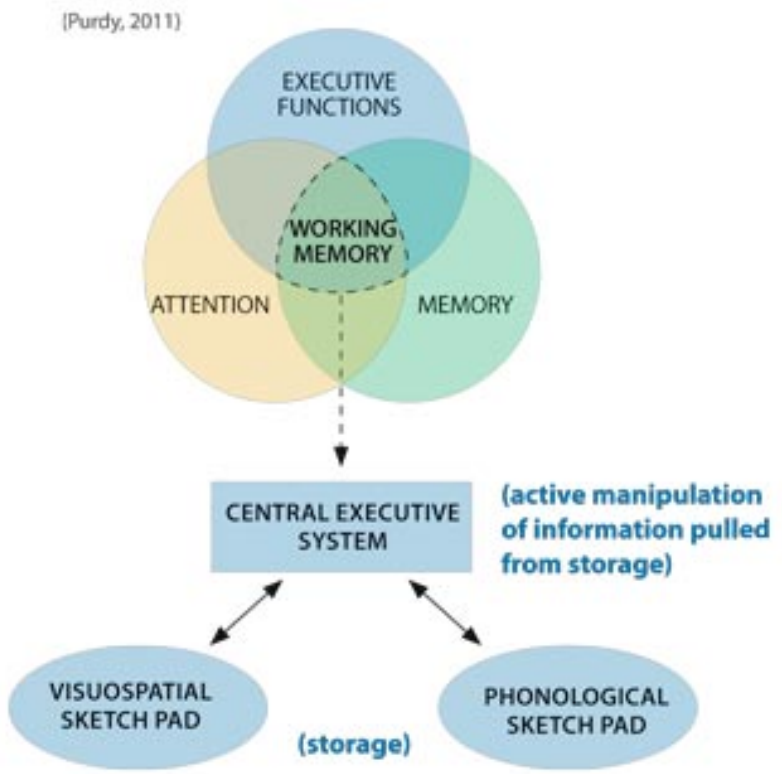
Memory types

Short-term/working (temporary storage)

Episodic (events happened at specific time)

Long-term/semantic (facts, objects, relations)

Procedural (sequence of actions)



http://www.rainbowrehab.com/executive-functioning/

Attention mechanisms

Need attention model to select or ignore certain inputs

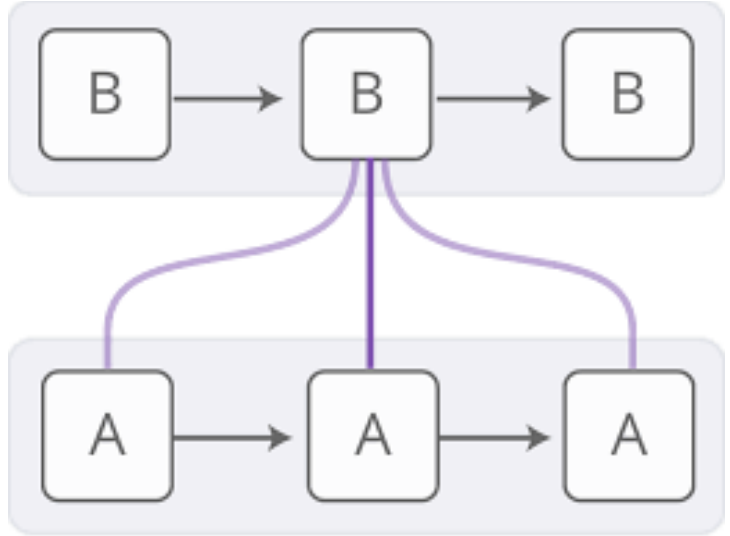
Human exercises great attention capability – the ability to filter out unimportant noises

- Foveating & saccadic eye movement

In life, events are not linear but interleaving.

Pooling (as in CNN) is also a kind of attention

Routing (as in CapsNet) is another example.



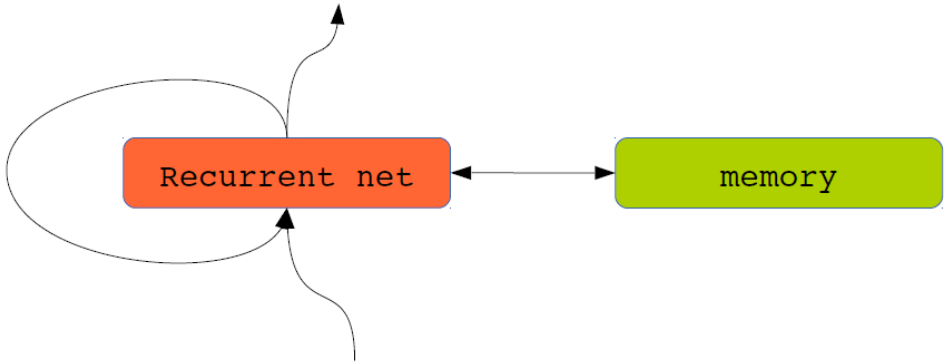
http://distill.pub/2016/augmented-rnns/

MANN: examples

Memory networks of Facebook: (Weston et al, Facebook, 2015); (Sukhbaatar et al, 2015) – associative memory

Dynamic memory networks of MetaMind: (Kumar et al, 2015) – episodic memory

Neural Turing machine and Differential Neural Computer of DeepMind (Graves et al. 2014, 2016) -- tape



(LeCun, 2015)

Supervised deep learning: steps

Step 0: Collect LOTS of high-quality data

- Corollary: Spend LOTS of time, $$ and compute power

Step 1: Specify the **computational graph** Y = F(X; W)

Step 2: Specify the loss L(W; D) for data D = {(X1,Y1), (X2,Y2),… }

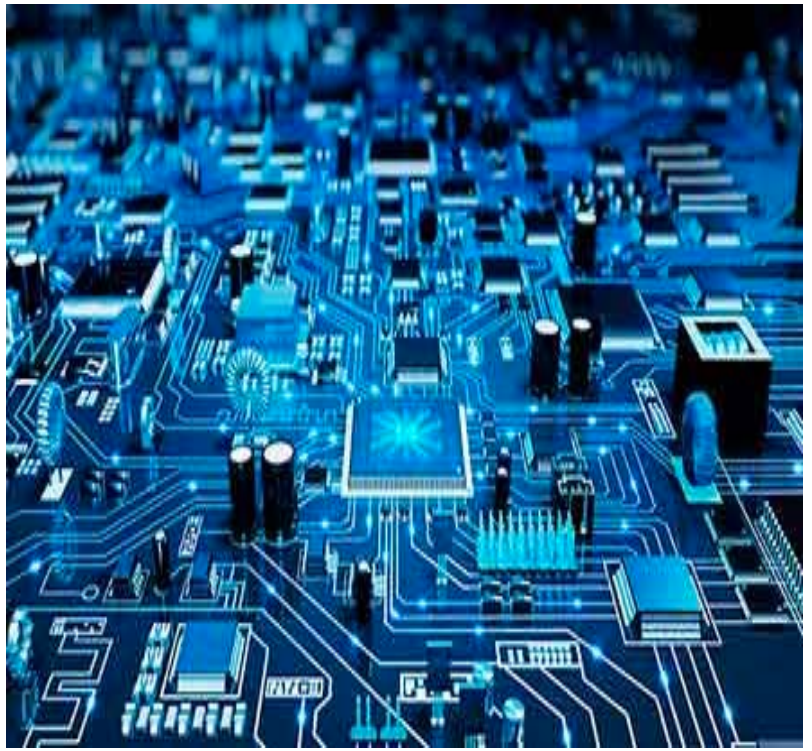
Step 3: Differentiate the loss w.r.t. W (now mostly automated)

Step 4: Optimize the loss (a lot of tools available)

Deep learning as new electronics (or LEGO?)

Analogies:

- Neuron as feature detector → SENSOR, FILTER
- Multiplicative gates → AND gate, Transistor, Resistor
- Attention mechanism → SWITCH gate
- Memory + forgetting → Capacitor + leakage
- Skip-connection → Short circuit
- Computational graph → Circuit
- Compositionality → Modular design



Relationships

- **Now**: Electronics redesigned to support tensors in deep learning
- Prediction: Deep learning helps to design faster electronics

NIPS18 submissions

4855 submissions

Agenda

Topic 1: Introduction (20 mins)

Topic 2: Brief review of deep learning (30 mins)

- Classic architectures
- Capsules & graphs
- Memory & attention

Topic 3: Genomics (30 mins)

- Nanopore sequencing
- Genomics modelling

QA (10 mins)

Break (30 mins)

Topic 4: Healthcare (40 mins)

- Time series (regular & irregular)
- EMR analysis: Trajectories prediction
- EMR analysis: Sequence generation

Topic 5: Data efficiency (40 mins)

- Few-shot learning
- Generative models
- Unsupervised learning of drugs

Topic 6: Future outlook

QA (10 mins)

Human genome

3 billion base-pairs (characters), 20K genes, 98% non-coding regions

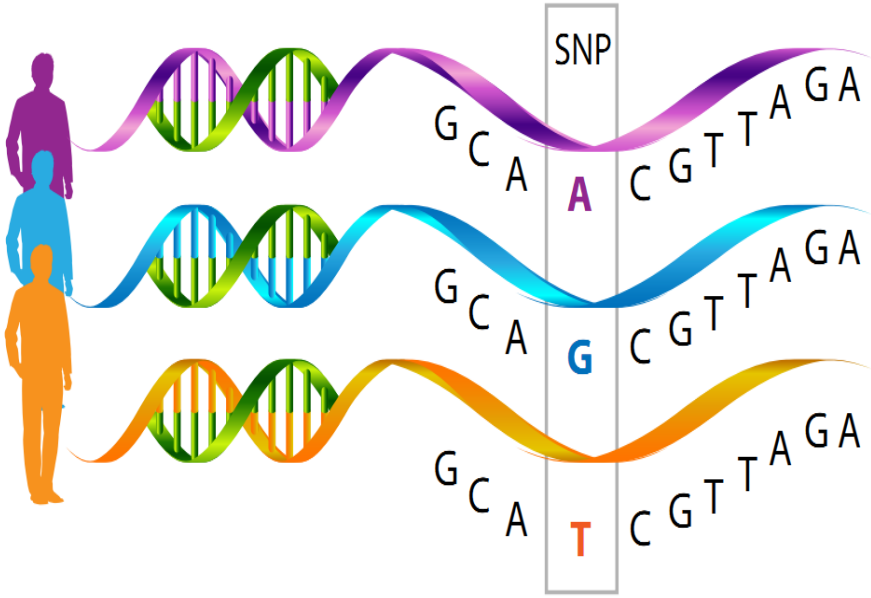
Any two random persons share 99.9% genome

The 0.1% difference is thought to account for all variations between us

- Appearance: Height (80% heritable), BMI, hair, skin colors
- IQ, education levels
- Genetic disorders such as cancers, bipolar, schizophrenia, autism, diabetes, etc.

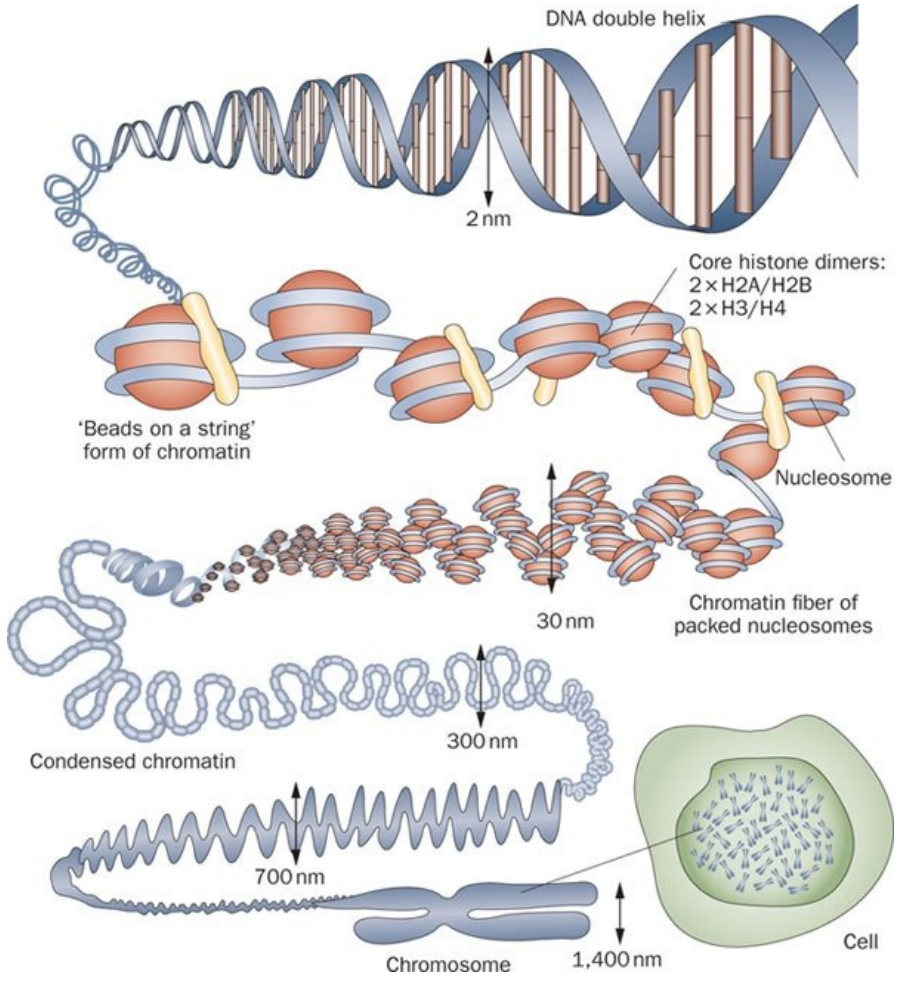
Any two random persons share about 60% variations (SNV/SNP)

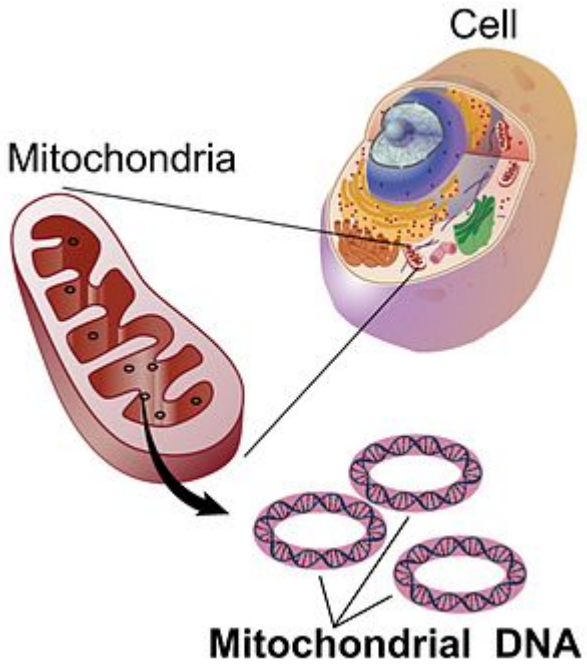
As we age, there are small mutations within our cells

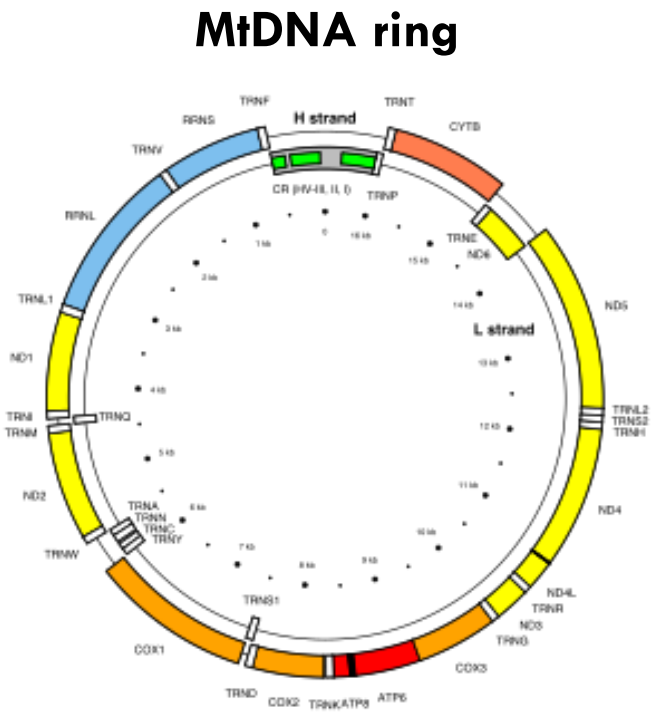


https://neuroendoimmune.files.wordpress.com

The cell, nuclear DNA & MtDNA







https://qph.ec.quoracdn.net/main-qimg-2c39fede406d71fb534bbae6cc9b8aad-c
https://en.wikipedia.org/wiki/Mitochondrial_DNA

Sequencing

The first step is to read (sequence) the DNA/MtDNA, and represent the information as string of characters (A,C,G,T) in computer.

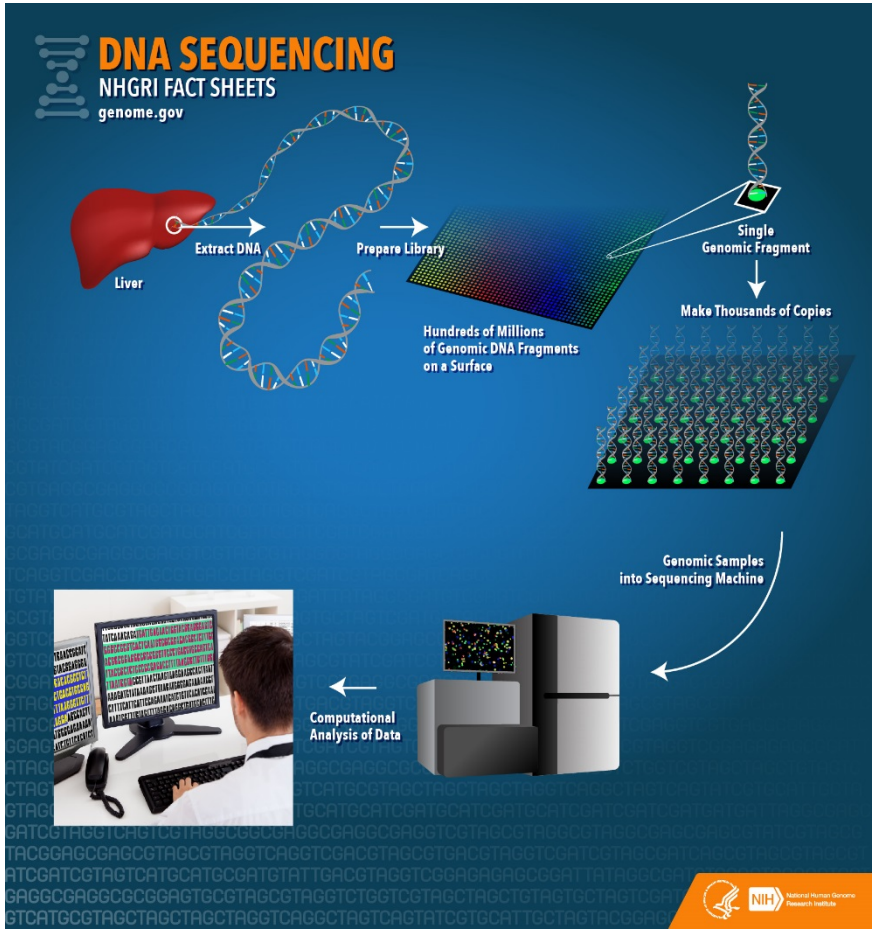
The most popular technique these days read short sequences (hundreds of characters), and align.

Each position is read typically at least 30 times to get enough confidence → Huge storage!!!

String alignment is then the key to final sequence → Need super-computer to do this fast.

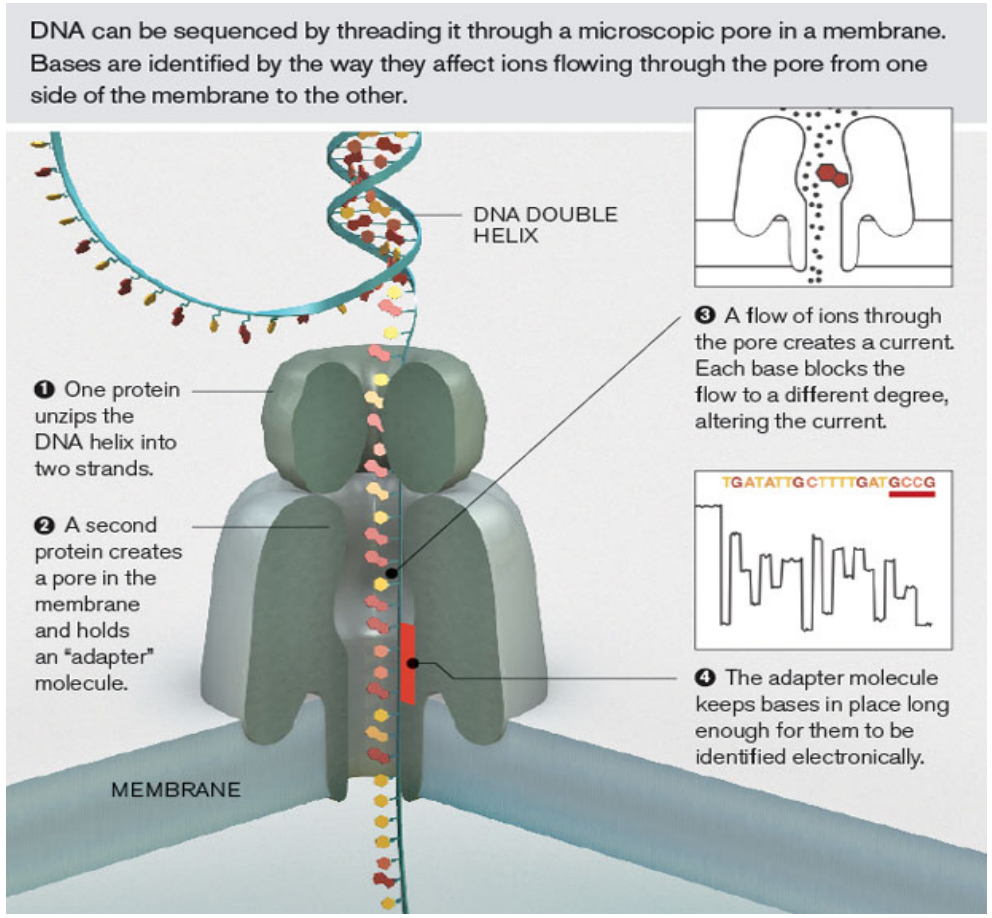
A DNA sequence is compared against the reference genome. Only the difference (0.1%) need to be stored.

- This does not usually apply for MtDNA, as each cell has as many as 500 MtDNAs, they are slightly different! More different as we age.



Source: https://www.genome.gov

The latest: nanopore sequencing (electrical signals → A|C|G|T)



Source: technologyreview.com

Source: ibtimes.co.uk

Continuous segmentation & labelling

Deep architectures for nanopore sequencing

Aimed at real time recognition

The setting is similar to speech recognition!

- → The early days used HMMs. Now LSTMs.

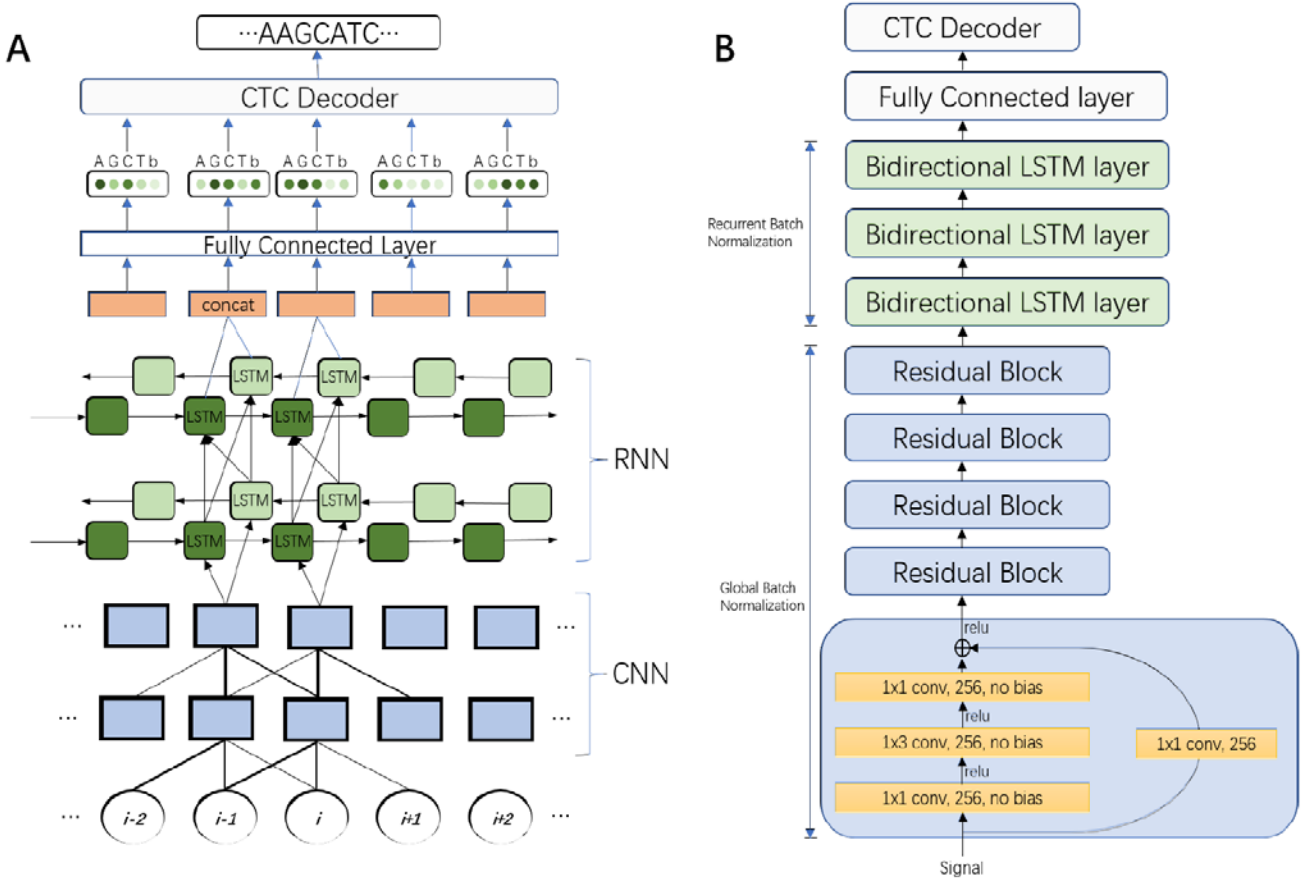
We will briefly review the latest:

- **Chiron** (Teng et al., August 2017, UQ, Australia)

Other GRU/LSTM variants

- Nanonet (Oxford Nanopore Technologies, 2016)
- BasecRAWller (Stoiber & Brown, May 2017)
- **DeepNano** (Boza et al., June 2017, Comenius University in Bratislava, Slovakia)

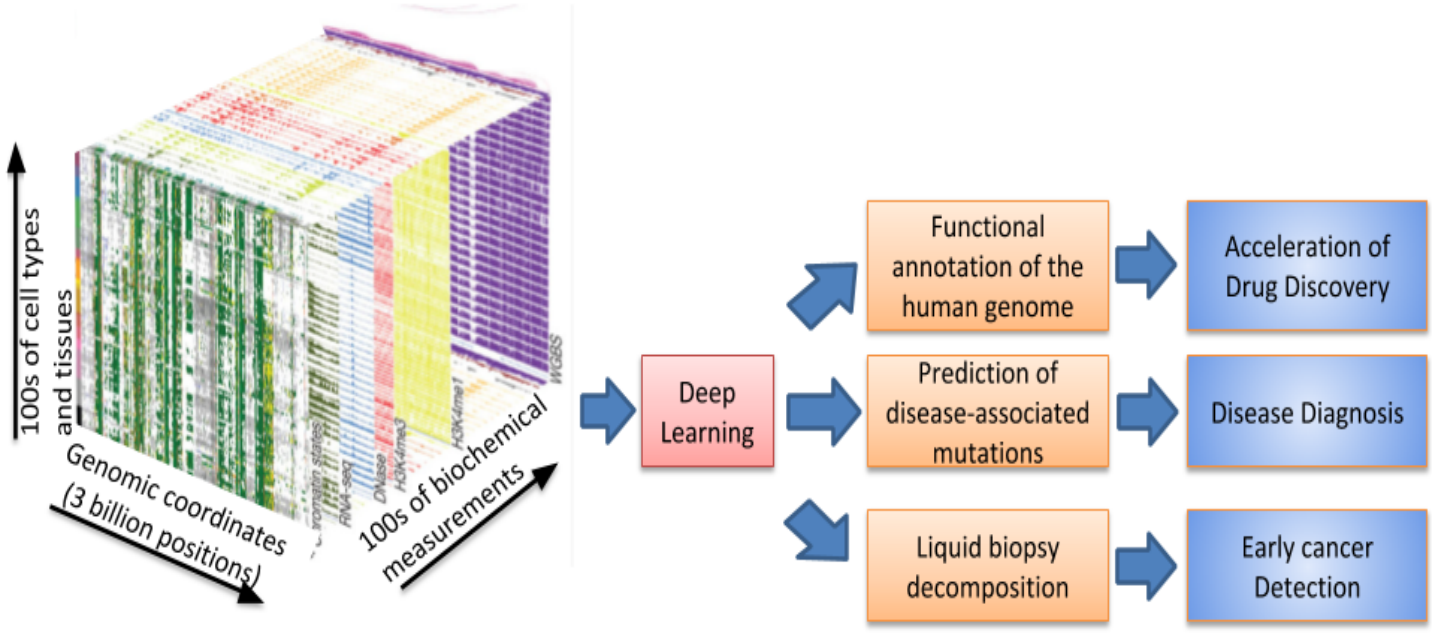
Chiron



Dataset	Basecaller	**Identity Rate**
Lambda	Metrichor	0.8650 (-0.0246)
	Albacore	**0.8896**
	BasecRAWller	0.8154 (-0.0742)
	Chiron	0.8776 (-0.012)
E. coli	Metrichor	0.8864 (-0.0193)
	Albacore	0.901 (-0.0047)
	BasecRAWller	0.8254 (-0.0803)
	Chiron	**0.9057**
M. tuberculosis	Metrichor	0.8802 (-0.0117)
	Albacore	**0.8919**
	BasecRAWller	0.8241 (-0.0678)
	Chiron	0.8851 (-0.0068)
Human	Metrichor	0.794 (-0.0446)
	Albacore	**0.8386**
	BasecRAWller	0.8149 (-0.0237)
	Chiron	0.8154 (-0.0232)

#REF: Teng, Haotien, et al. "Chiron: Translating nanopore raw signal directly into nucleotide sequence using deep learning", GigaScience, Volume 7, Issue 5, 1 May 2018, giy037.

Opportunities for Deep Learning in Genomics



- Genetic diagnostics
- Refining drug targets
- Pharmaceutical development
- Personalized medicine
- Better health insurance
- Synthetic biology

https://towardsdatascience.com/opportunities-and-obstacles-for-deep-learning-in-biology-and-medicine-6ec914fe18c2
https://www.oreilly.com/ideas/deep-learning-meets-genome-biology

Some AI problems

DNA is a book, easy to read (costs less than $1K to sequence), extreme difficult to comprehend.

- It has 3B characters (A,C,T,G), 46 volumes (chromosomes), 20K chapters.
- The longest book has less than 10M characters, 13 volumes (“A la recherche du temps perdu" (In Search of Lost Time), by Marcel Proust, 2012) – as recognized by Guinness World Records.

Short sequences (100 chars) are predictive of protein binding, also gene start/end.

Proteins are big 3D graphs interacting with the 1D-2D strings (DNA, RNA), and other proteins & drugs (which are graphs themselves).

Long chains of influence, from SNP to cell, tissue and organ functions.

Viruses can be generated/edited on computer, hence discrete sequence generation problem.

Filling the genotypes → phenotypes gap

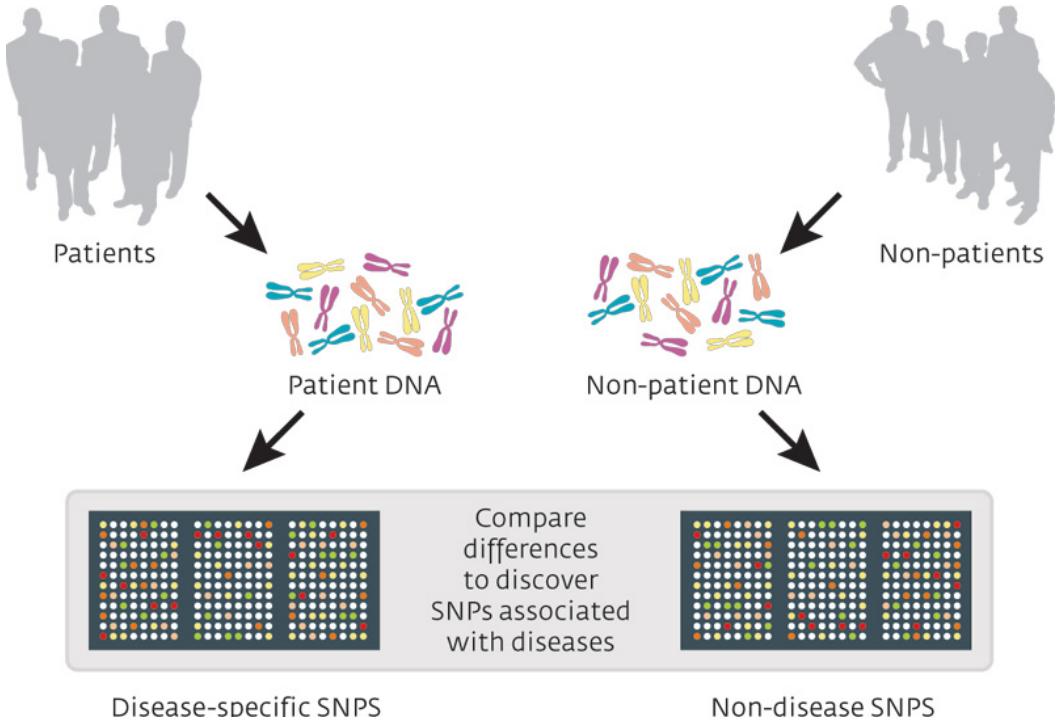
Ultimate goals:

- Estimating explained variance in inheritability
- Discover risk factors
- Predicting individual phenotypes: Height, Glucose, BMI, IQ, Edu, Mental, Cancers…

Some paths under investigation

- Predicting the bio of the cells, DNA + MtDNA, and more
- Statistical modeling of genetic architectures, e.g., Bayesian, mixed linear models, Gaussian Processes.
- Motif modeling with DNA/RNA/protein, e.g., predict binding sites
- Developing data-efficient techniques for genomics
- Integrating multimodalities

GWAS: Genome-Wide Association Study



Setting:

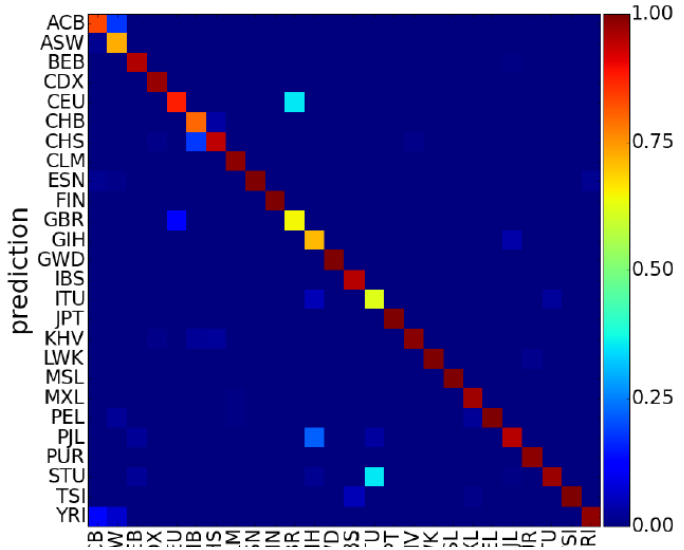
- For each DNA, only differences from a reference genome are recorded.
- The differences are SNPs, one per dimension.

Problems

- Very high dimensional (typically hundreds of thousands), low sample size (typically hundreds)
- Missing/unreliable data
- Typically very weak association
- Combating the False Discovery Rate (FDR) due to multiple parallel hypotheses: Individual *p*-value must be extremely small, e.g. 5×10e-8

Source: http://vignette4.wikia.nocookie.net

Diet networks for GWAS



#REF: Romero, Adriana, et al. "Diet Networks: Thin Parameters for Fat Genomic." *arXiv preprint arXiv:1611.09340* (2016).

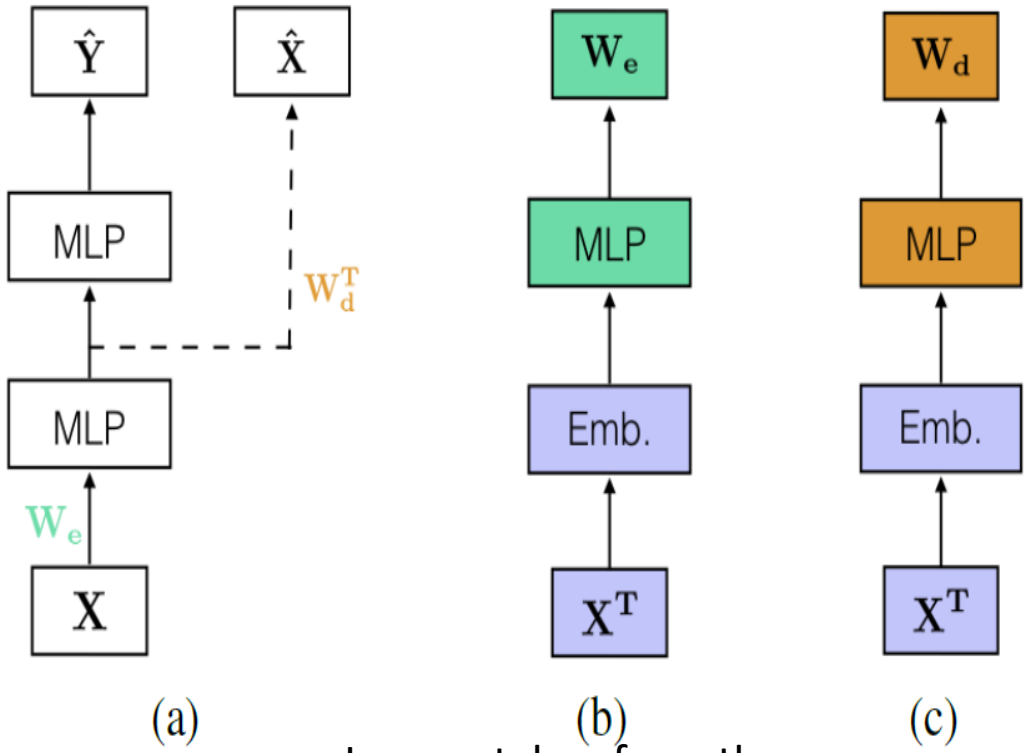
Use a “hypernet” to generate the main net.

Features are embedded (not data instance).

Unsupervised autoencoder as regularizer.

Works well on country prediction on the 1000 Genomes Project dataset.

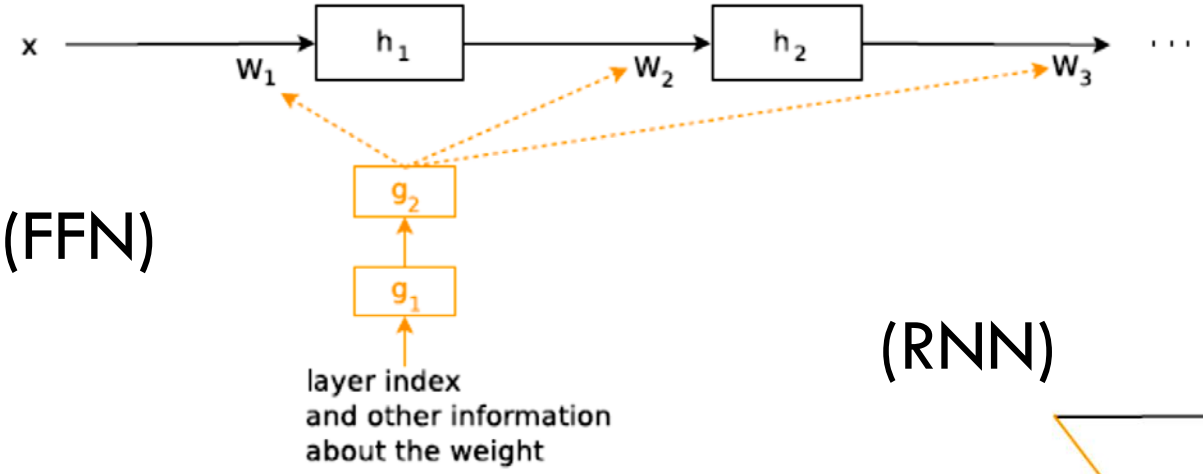
- But this is a relatively easy problem. PCA, even random subspace can do quite well!



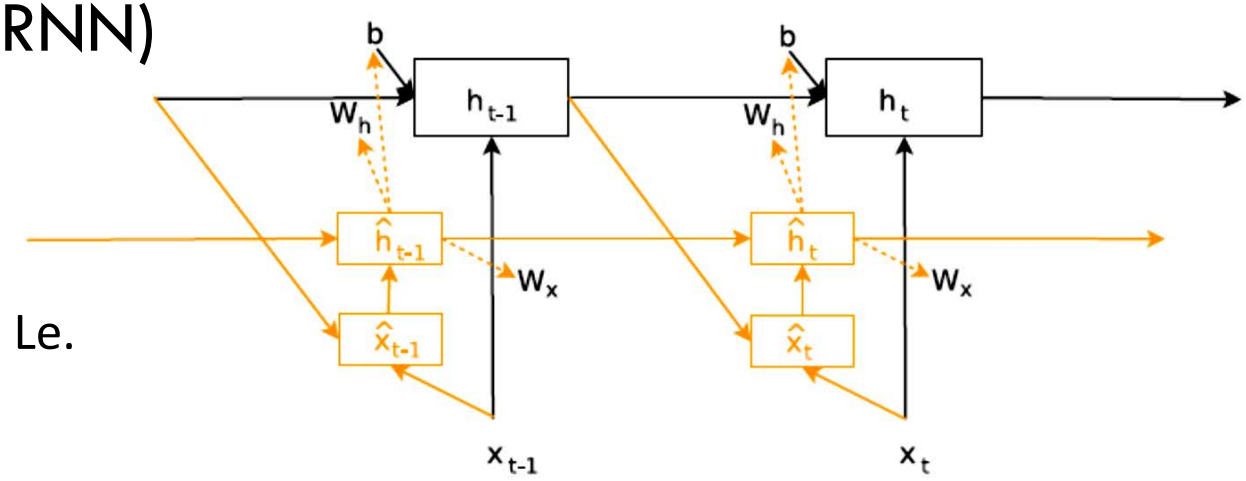
Images taken from the paper

HyperNetworks: Network to generate networks

(FFN)



(RNN)



#REF: Ha, David, Andrew Dai, and Quoc V. Le. "HyperNetworks." *arXiv preprint arXiv:1609.09106* (2016).

GWAS: Challenges

We are detecting rare events!!!

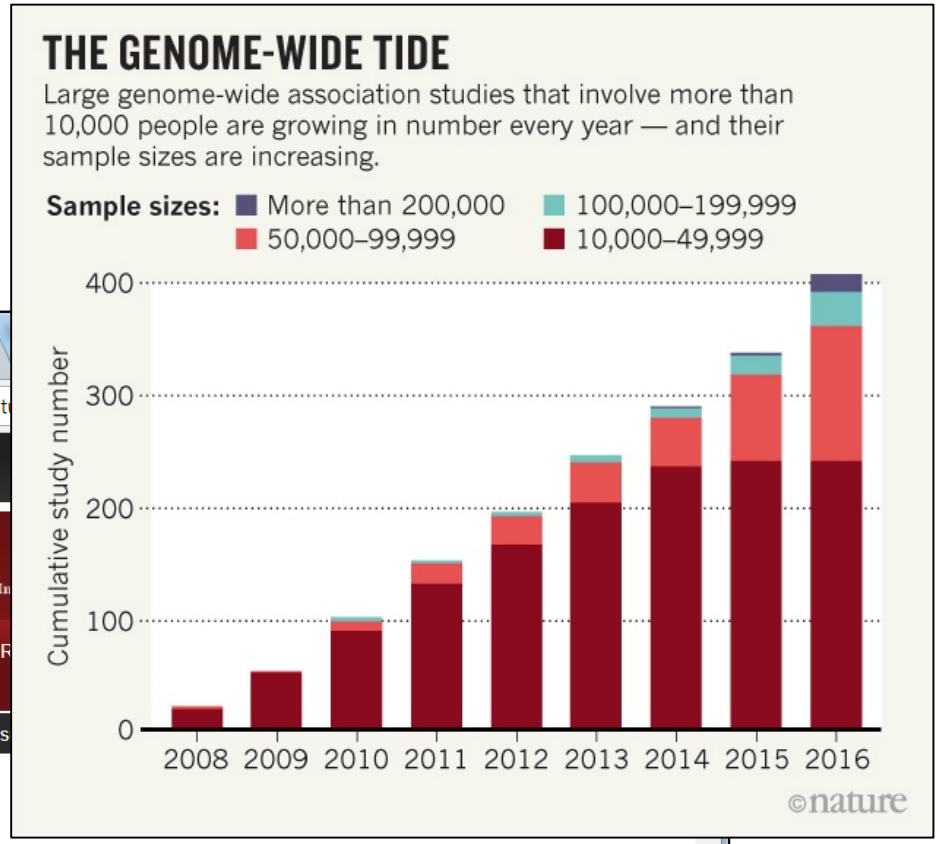
Results hard to replicate across studies.
- Model stability?

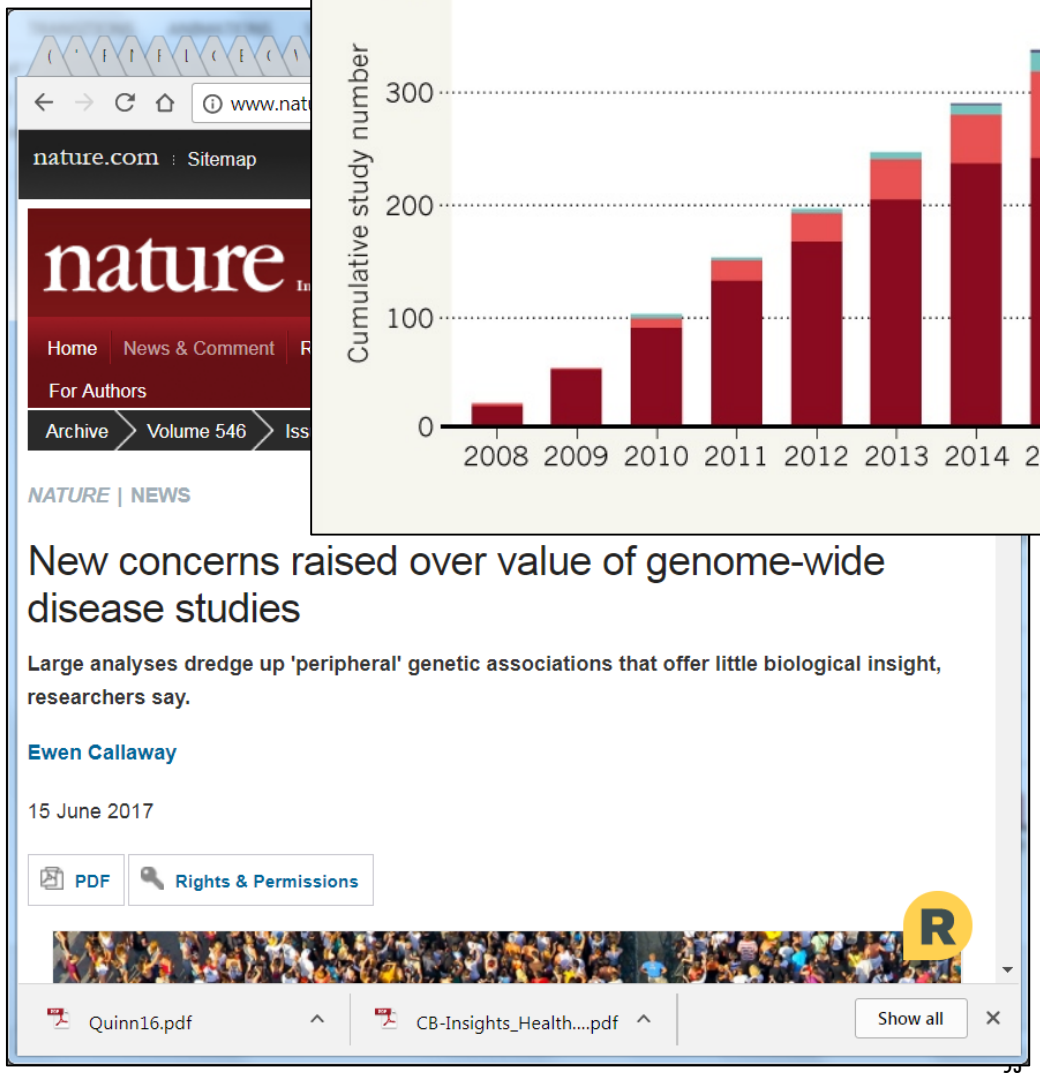
SNP → phenotypes seem impossible.

If it is (e.g., race prediction), little insights can be drawn upon.

The pathway is deep and complex
- Room for deep learning?

Room for structured models
- SNP annotations
- Spatial relationships
- Evolutionary trees





Rooms for deep learning

Bridge the genotype-phenotype gap

- Incorporating HUGE amount of data
- Modelling the multiple layers of complex biological processes in between.
- Starting from the DNA and its immediate functions, e.g., protein binding, gene start, alternative splicing, SNP annotations.

Deep learning has shown to work well in cognitive domains, where human can perform in less than a second.

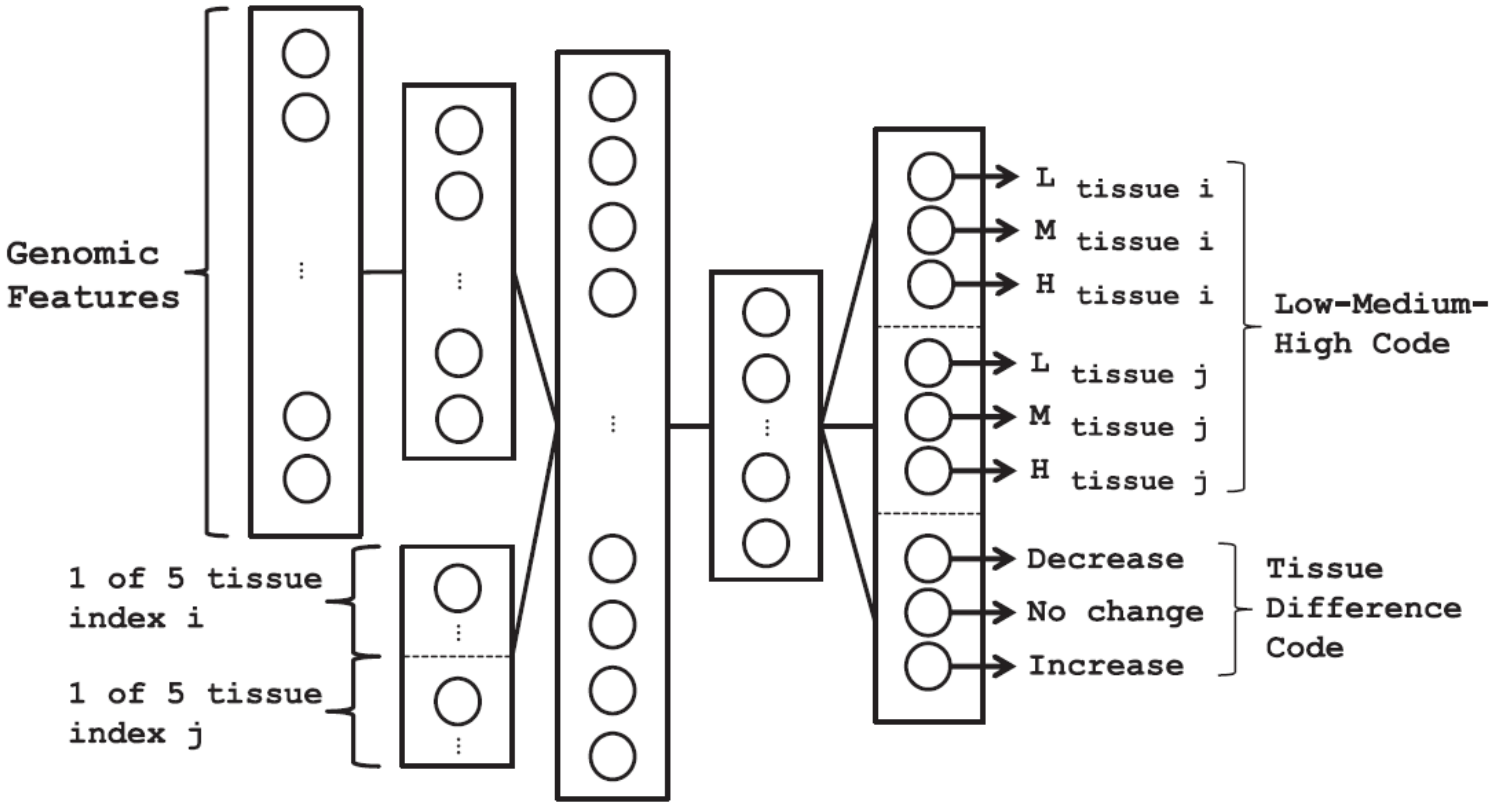
- We need to be super-human to bridge the gap.

New models for 2% of coding part, as well as 98% non-coding (probably having regulatory functions)

Incorporating biological understanding into model, not the black-box.

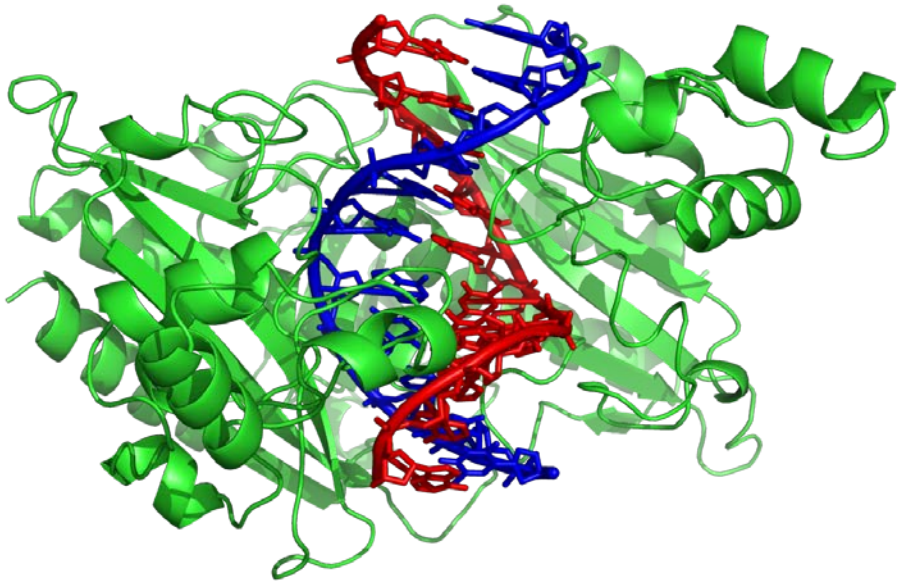
#Ref: https://www.oreilly.com/ideas/deep-learning-meets-genome-biology

Use of feedforward nets: Tissue-regulated splicing code



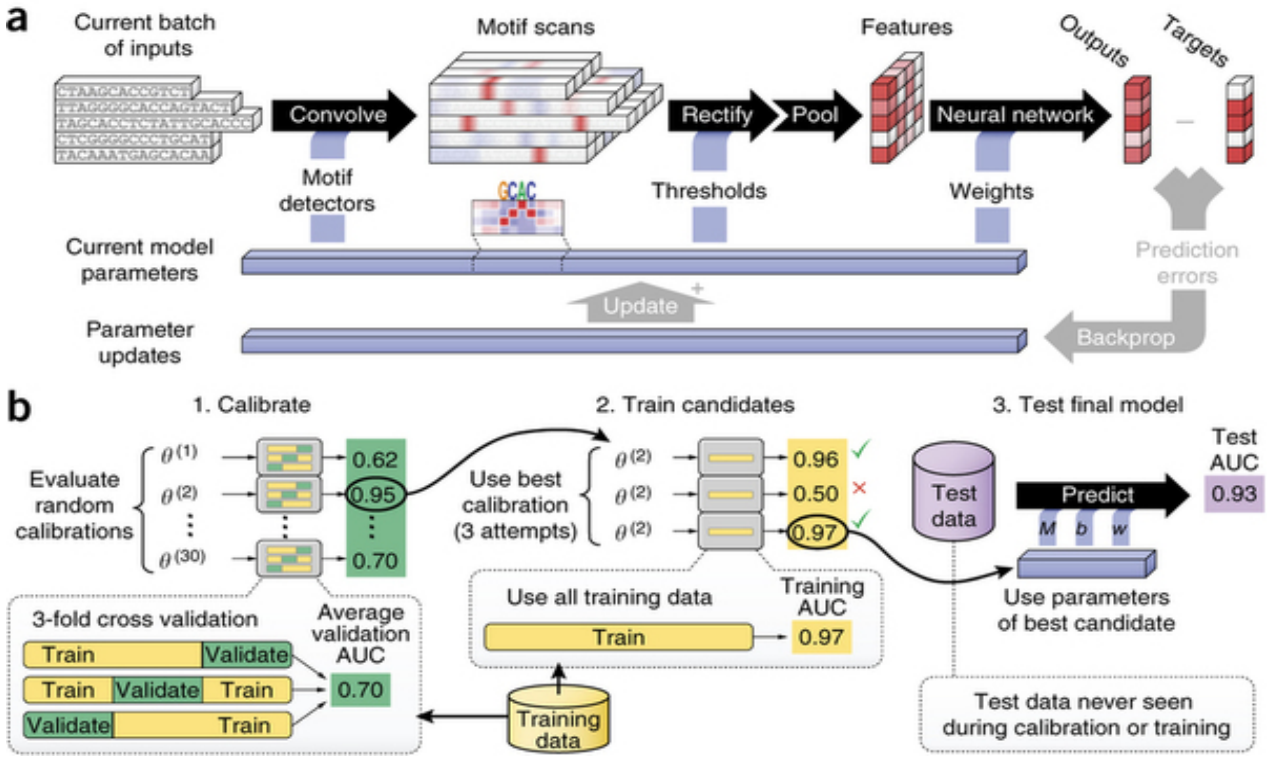
#REF: Leung, Michael KK, et al. "Deep learning of the tissue-regulated splicing code." *Bioinformatics* 30.12 (2014): i121-i129.

Use of CNNs: Discovery of DNA motifs



The restriction enzyme EcoRV (green)
Source: wikipedia.org/wiki/DNA-binding_protein

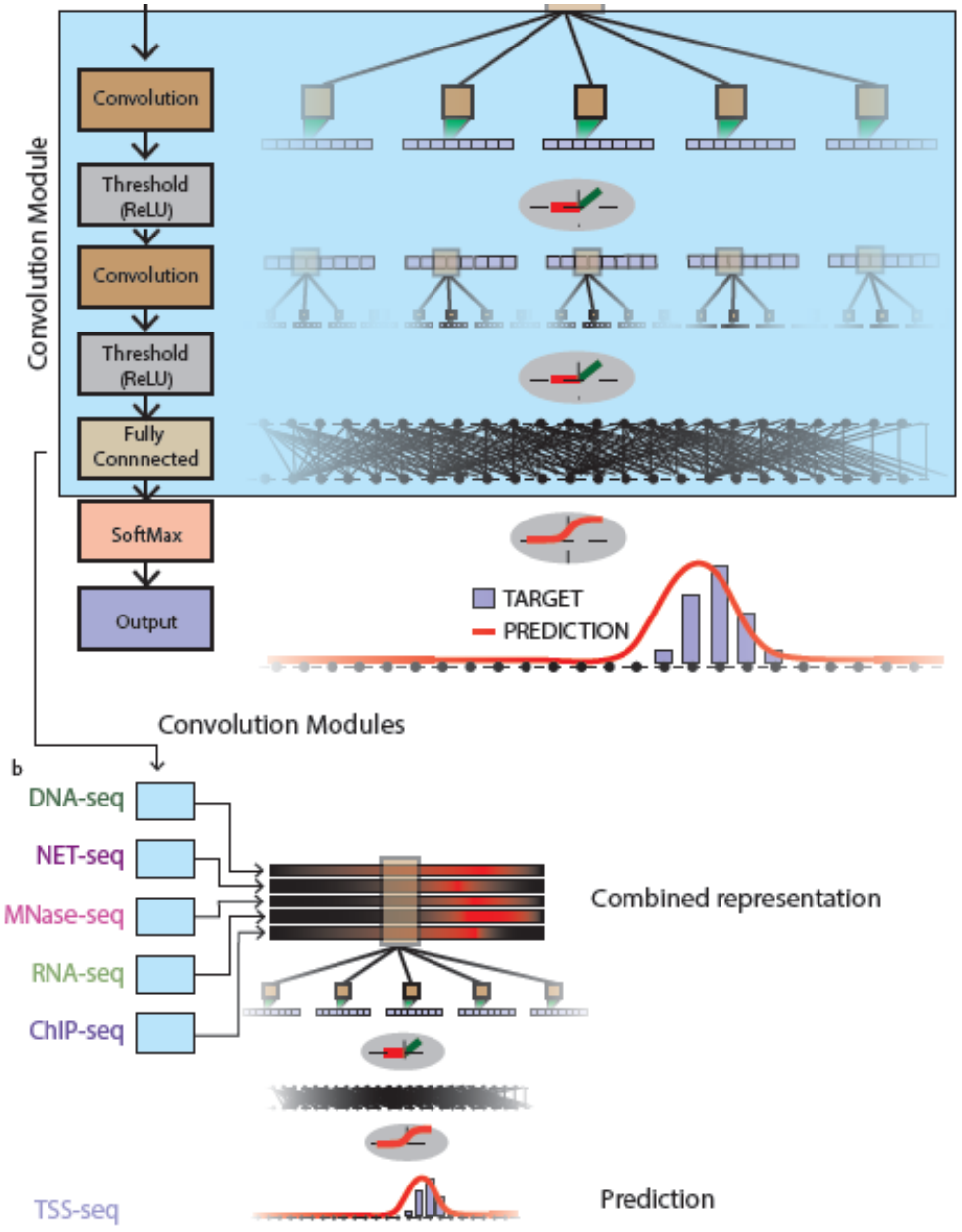
DeepBind (Alipanahi et al, Nature Biotech 2015)

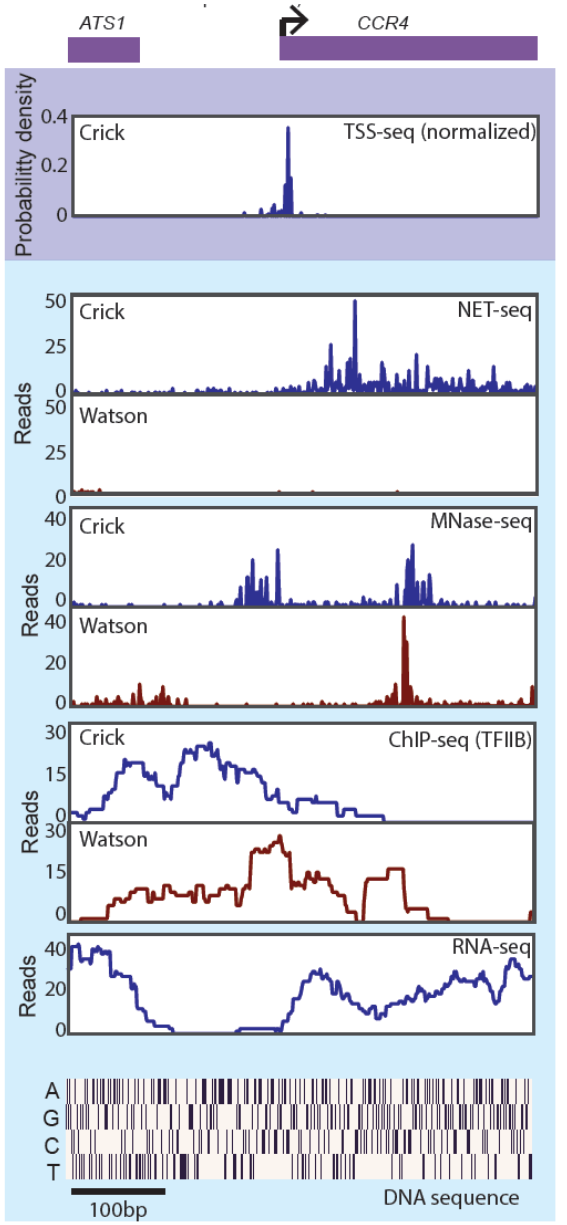


http://www.nature.com/nbt/journal/v33/n8/full/nbt.3300.html

Use of CNNs: FIDDLE

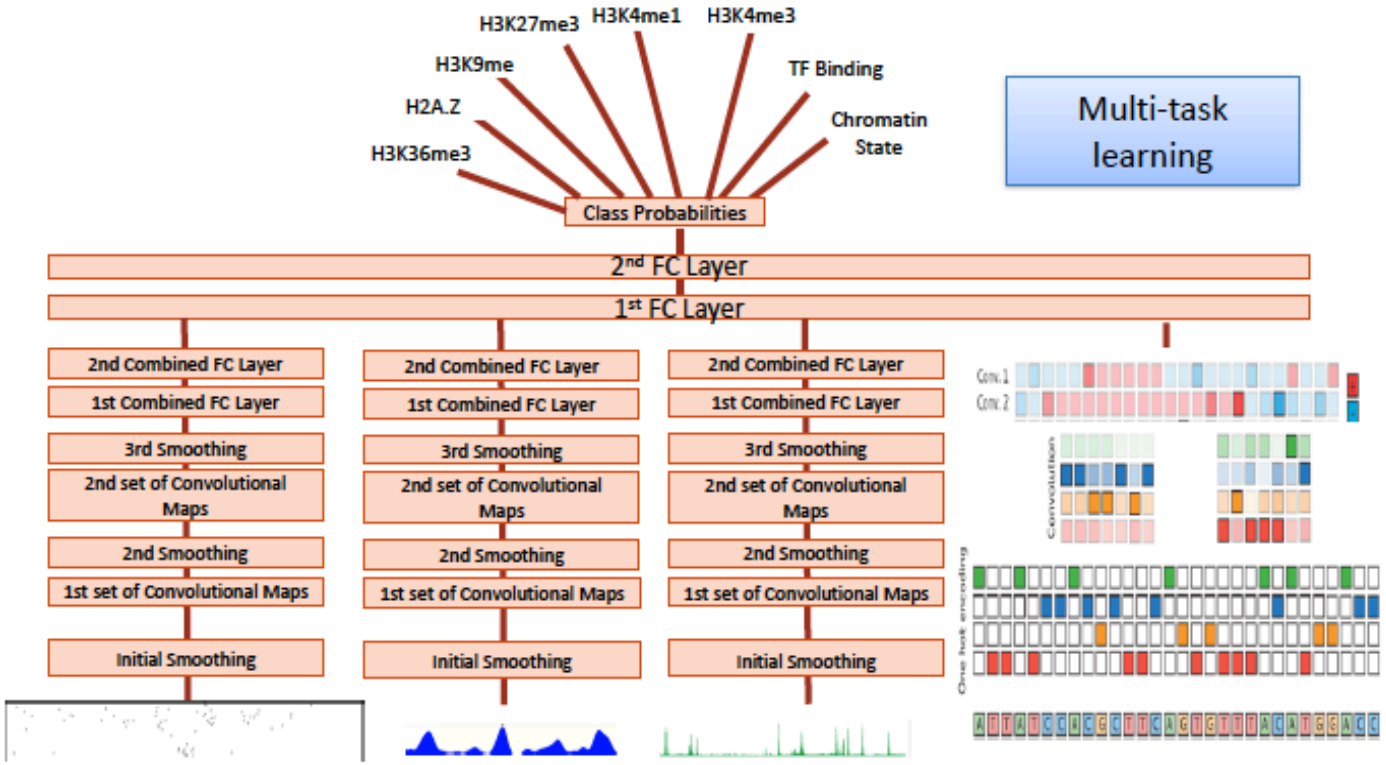
#REF: Eser, Umut, and L. Stirling Churchman. "FIDDLE: An integrative deep learning framework for functional genomic data inference." *bioRxiv* (2016): 081380.





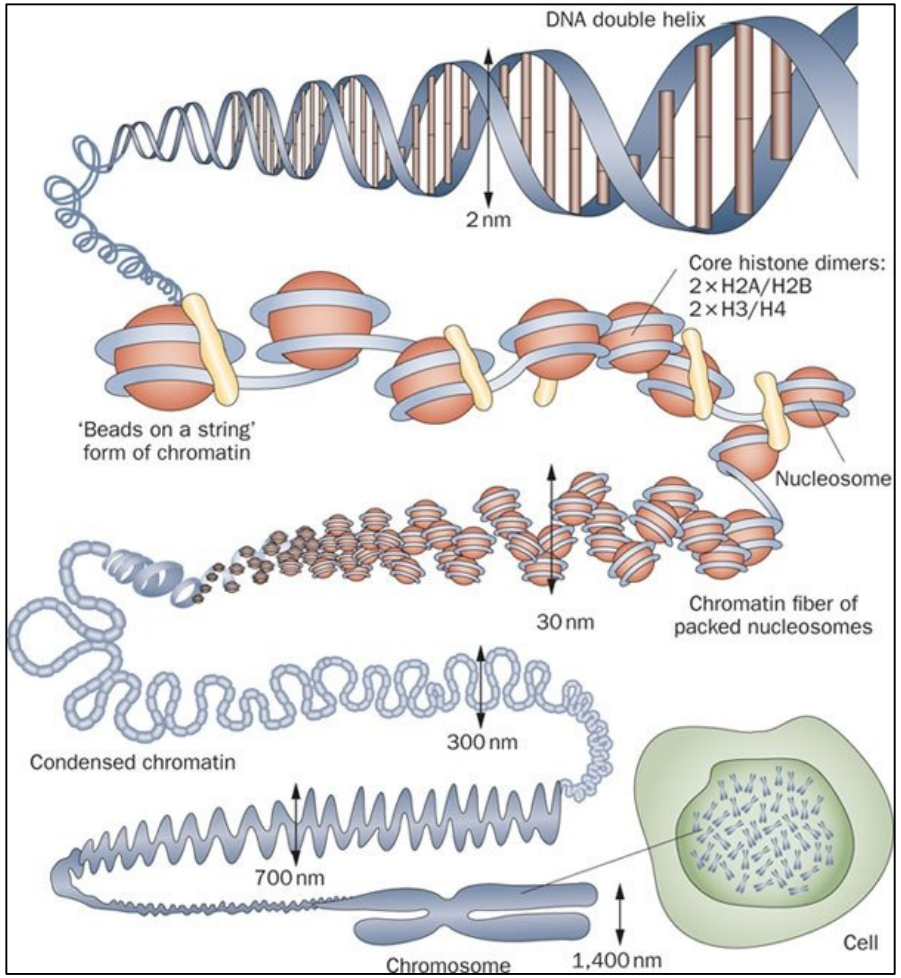
THE CHROMPUTER

Integrating multiple inputs (1D, 2D signals, sequence) to simulatenously **predict multiple outputs**



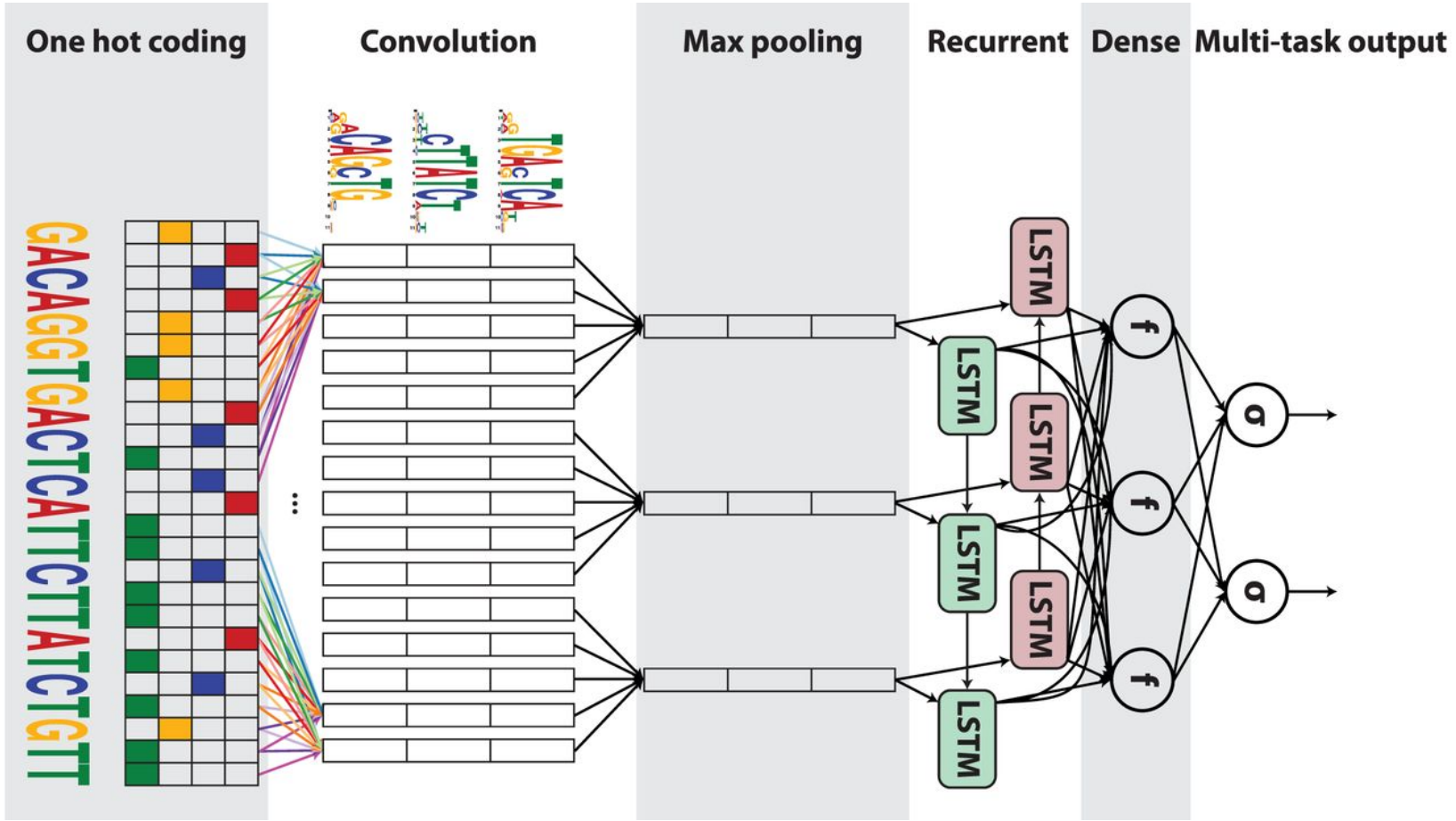
Source: https://simons.berkeley.edu/sites/default/files/docs/4575/2016-kundaje-simonsinstitute-deeplearning.pdf

Chromatins



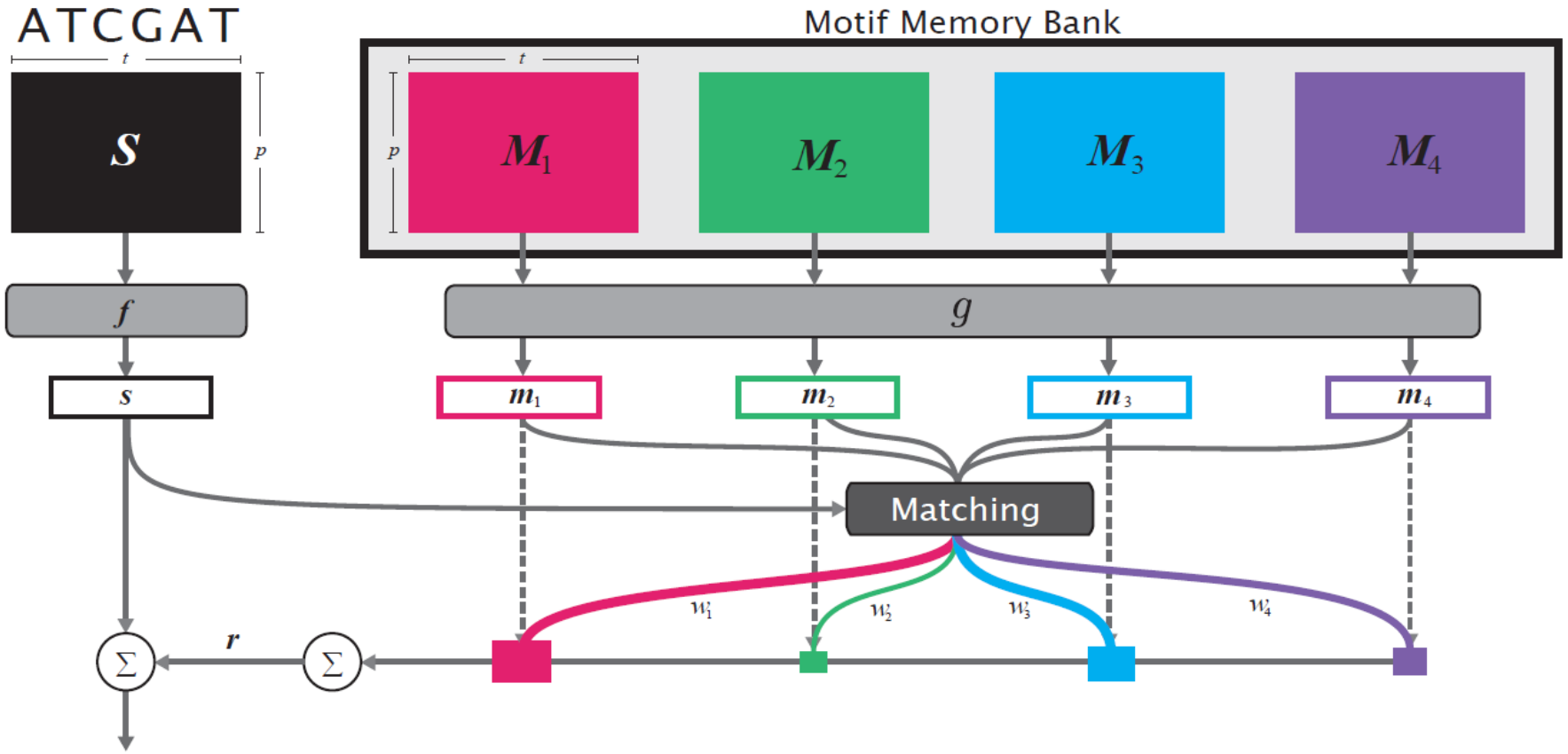
https://qph.ec.quoracdn.net

User of CNN+RNNs: DanQ



#REF: Quang, Daniel, and Xiaohui Xie. "DanQ: a hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences." *Nucleic acids research* 44.11 (2016): e107-e107.

Use of MANN: Matching nets



Output

#REF: Lanchantin, Jack, Ritambhara Singh, and Yanjun Qi. "Memory Matching Networks for Genomic Sequence Classification." *arXiv preprint arXiv:1702.06760* (2017).

More models/frameworks

DragoNN

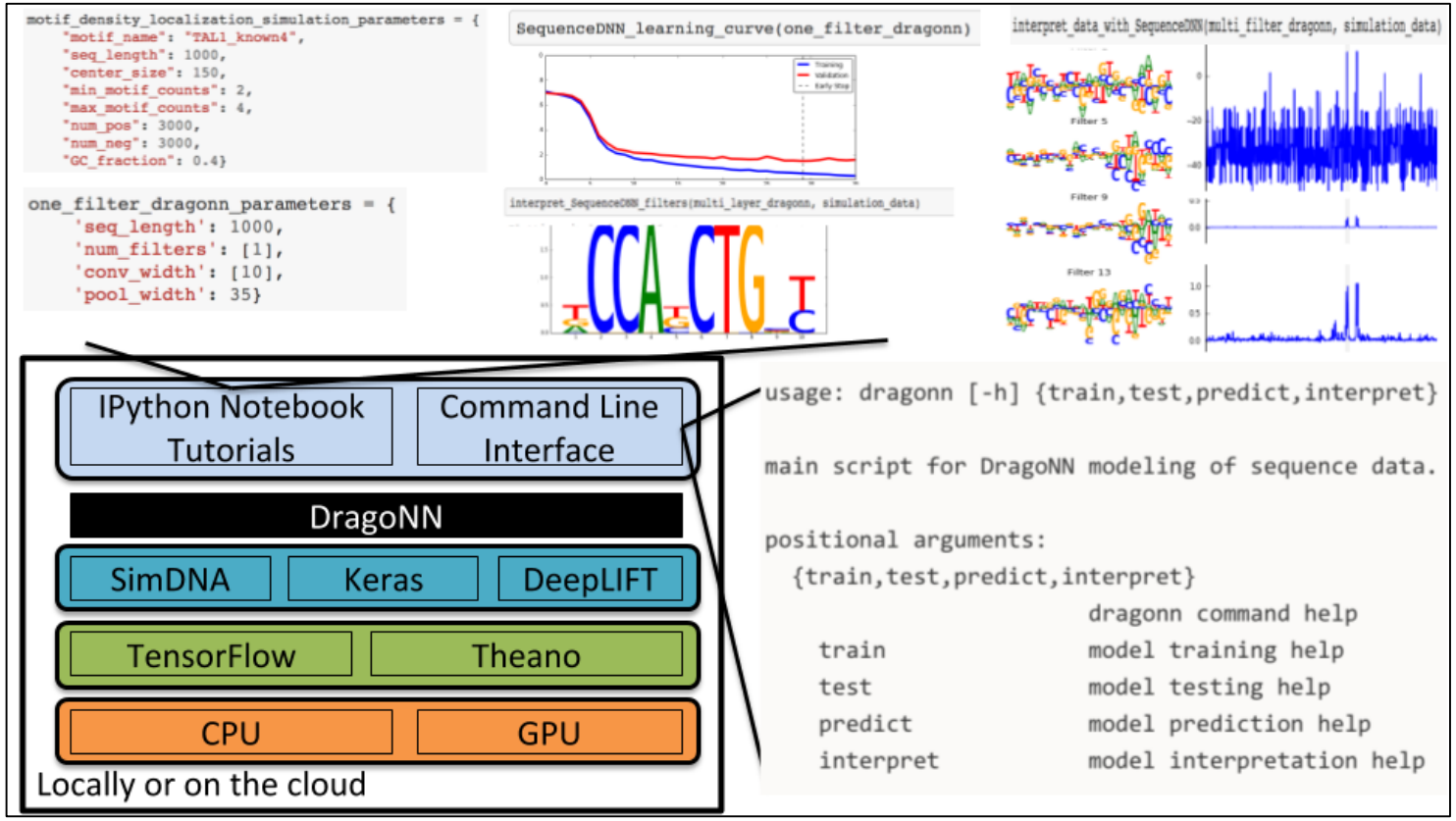
DeepChrome

DeepSEA

Basset

DeepBound

…



http://kundajelab.github.io/dragonn

The outlook

Read an extremely long book and answer any queries about it

- Memory-augmented neural networks (MANN), and
- Multiple hierarchical attentions and grammars

Instead of read, write (DNA/viruses/RNA/proteins)

Supper-rich genome SNP annotation

The society of things (DNA/RNA/protein)

Transfer learning between cell types, tissues and diseases

Biology-driven deep nets (e.g., knowledge as memory)

Handling rare events (e.g., the role of memory)

and break

We're hiring

PhD & Postdocs

truyen.tran@deakin.edu.au

http://ahsanqawl.com/2015/10/qa/

References

Ching, Travers, et al. "Opportunities And Obstacles For Deep Learning In Biology And Medicine." *bioRxiv* (2018): 142760

Eser, Umut, and L. Stirling Churchman. "FIDDLE: An integrative deep learning framework for functional genomic data inference." *bioRxiv* (2016): 081380.

Ha, David, Andrew Dai, and Quoc V. Le. "HyperNetworks." *arXiv preprint arXiv:1609.09106* (2016).

Leung, Michael KK, et al. "Deep learning of the tissue-regulated splicing code." *Bioinformatics* 30.12 (2014): i121-i129.

Lanchantin, Jack, Ritambhara Singh, and Yanjun Qi. "Memory Matching Networks for Genomic Sequence Classification." *arXiv preprint arXiv:1702.06760* (2017).

Pham, Trang, et al. "Column Networks for Collective Classification."*AAAI*. 2017

Quang, Daniel, and Xiaohui Xie. "DanQ: a hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences." *Nucleic acids research* 44.11 (2016): e107-e107.

Teng , Haotien, et al. "Chiron: Translating nanopore raw signal directly into nucleotide sequence using deep learning", *GigaScience*, Volume 7, Issue 5, 1 May 2018, giy037.

Wagstaff, K. L. (2012, June). Machine learning that matters. In *Proceedings of the 29th International Coference on International Conference on Machine Learning* (pp. 1851-1856). Omnipress.