

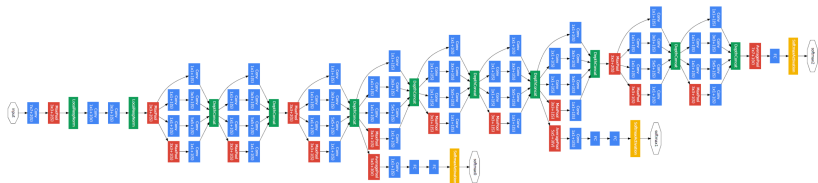
What's next?

Machine learning & neural networks

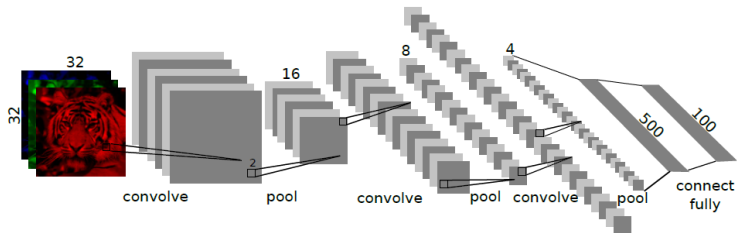
Anne Helby Petersen

Tuning and stealing

Main strategies for building more complicated NNs: Tuning *meta parameters* and stealing *architectures*.



Convolutional neural networks



- ▶ Mostly applied in image analysis and text/speech data analysis
- ▶ A strategy for autogenerating more informative input nodes when the data has a spatial/temporal structure
 - ▶ Variables (e.g. pixels) close to each other are "analyzed" jointly
 - ▶ Ends with "flattening" these new features and performing "classical" NN learning

What we did today

What we did today

1. Built machines from scratch

What we did today

1. Built machines from scratch
2. Built machines using logistic regression
 - 2.1 Tried out different performance metrics (accuracy, AUC)

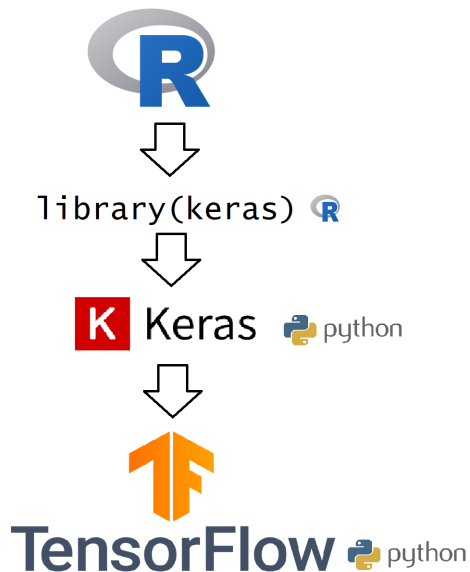
What we did today

1. Built machines from scratch
2. Built machines using logistic regression
 - 2.1 Tried out different performance metrics (accuracy, AUC)
3. Built neural networks
 - 3.1 Wide ones
 - 3.2 Deep ones
 - 3.3 Some using dropout
 - 3.4 Some using dropout tuned on data

What we did today

1. Built machines from scratch
2. Built machines using logistic regression
 - 2.1 Tried out different performance metrics (accuracy, AUC)
3. Built neural networks
 - 3.1 Wide ones
 - 3.2 Deep ones
 - 3.3 Some using dropout
 - 3.4 Some using dropout tuned on data
4. ... and, hopefully, understood the general ideas of what we were doing along the way

How did we do it?



What we didn't do today

- ▶ **Identify risk factors:** We can't tell the oncologists what to measure on their patients to best predict survival.

What we didn't do today

- ▶ **Identify risk factors:** We can't tell the oncologists what to measure on their patients to best predict survival.
- ▶ **Compute p -values/confidence intervals:** We don't know how certain our results are.

What we didn't do today

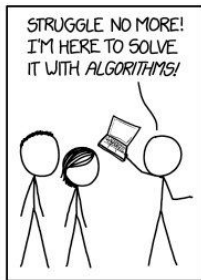
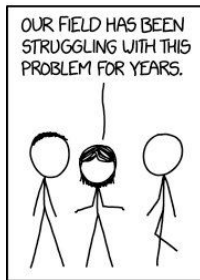
- ▶ **Identify risk factors:** We can't tell the oncologists what to measure on their patients to best predict survival.
- ▶ **Compute p -values/confidence intervals:** We don't know how certain our results are.
- ▶ **Investigate causal relationships:** We don't know what causes some patients to die sooner than others.

What we didn't do today

- ▶ **Identify risk factors:** We can't tell the oncologists what to measure on their patients to best predict survival.
- ▶ **Compute p -values/confidence intervals:** We don't know how certain our results are.
- ▶ **Investigate causal relationships:** We don't know what causes some patients to die sooner than others.

⇒ Deep learning is great for classification. But consider whether this is what you need.

Machine learning limitations: Classification is “easy”, but...




Examples in the wild:

- ▶ Apple's virtual assistant, Siri, uses DL for speech recognition
- ▶ The voice of Amazon's virtual assistant, Alexa, is generated by use of DL
- ▶ Google's virtual assistant uses DL for spoken language identification
- ▶ Google Image Search uses DL for labeling images
- ▶ Automatically adding color to black/white photos

A clever way to get labels for images

Select all images with
taxi



⏪ ⏮ ⓘ

VERIFY

Deep learning use cases: Let there be color!

Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification

[Satoshi Iizuka](#)* [Edgar Simo-Serra](#)* [Hiroshi Ishikawa](#) (*equal contribution)

SIGGRAPH 2016



Colorado National Park, 1941



Textile Mill, June 1937



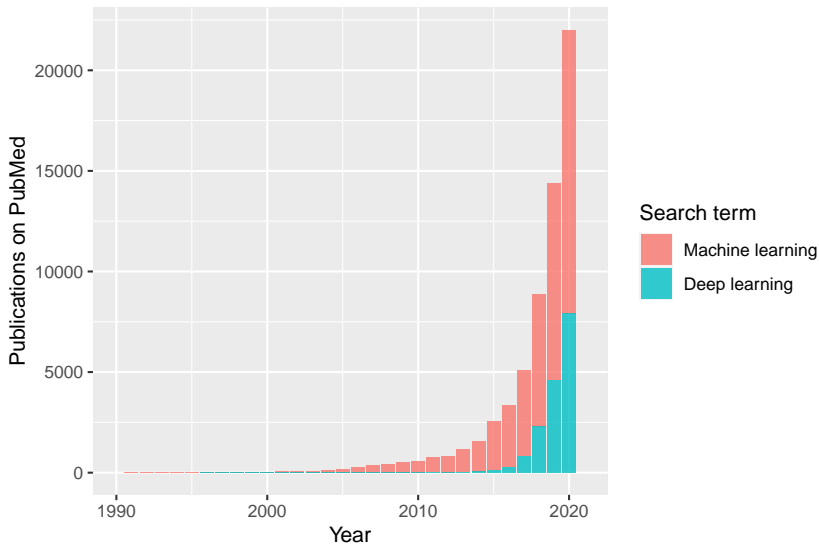
Berry Field, June 1909



Hamilton, 1936



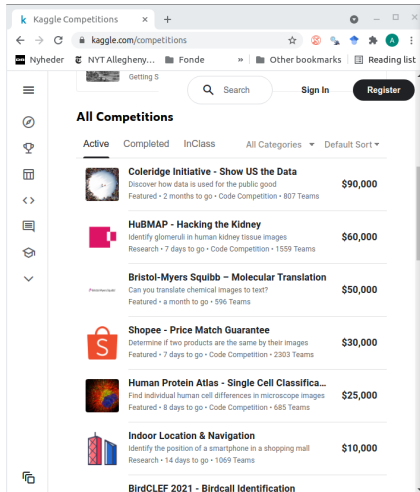
Deep learning use cases on PubMed



Machine learning limitations: Garbage in, garbage out



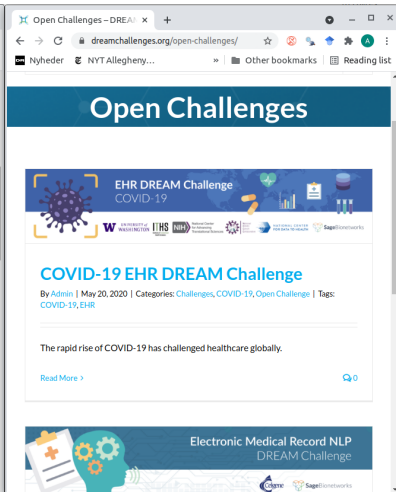
Want to use ML for more crowd sourced research?



The screenshot shows the Kaggle Competitions page. The URL is kaggle.com/competitions. The page features a search bar, a 'Sign In' button, and a 'Register' button. Below the navigation, there are tabs for 'Active', 'Completed', and 'InClass' competitions. A list of active competitions is displayed, each with a thumbnail, title, description, and prize amount.

Competition	Prize Amount
Coleridge Initiative - Show US the Data	\$90,000
HuBMAP - Hacking the Kidney	\$60,000
Bristol-Myers Squibb - Molecular Translation	\$50,000
Shopee - Price Match Guarantee	\$30,000
Human Protein Atlas - Single Cell Classifica...	\$25,000
Indoor Location & Navigation	\$10,000

At the bottom of the list, there is a link to 'BirdCLEF 2021 - Birdcall Identification'.



The screenshot shows the DREAM Challenges website. The URL is dreamchallenges.org/open-challenges/. The page features a large blue header with the text 'Open Challenges'. Below the header, there is a banner for the 'EHR DREAM Challenge COVID-19'. The main content area is titled 'COVID-19 EHR DREAM Challenge' and includes the following text:

By Admin | May 20, 2020 | Categories: Challenges, COVID-19, Open Challenge | Tags: COVID-19, EHR

The rapid rise of COVID-19 has challenged healthcare globally.

Read More >

At the bottom of the page, there is a banner for the 'Electronic Medical Record NLP DREAM Challenge'.

Kaggle competitions: kaggle.com, DREAM Challenges: dreamchallenges.org, Project Data Sphere: projectdatasphere.org

That's it!

Thank you for participating!

Further questions or feedback welcome now or at ahpe@sund.ku.dk.