Section 5: AI and Deep Learning

Vadim Sokolov

Suggested Reading

Google and Nvidia

Last Section

- Multiple Regression (Newfood study, Golf Analysis)
- Interactions (how advertisement change price elasticity?)
- Predictive analytics cases(Target, Walmart, Airbnb, Stitch Fix)
- Logistic regression (NBA predictions, Horse predictions, LinkedIn)

This Section

- AI Case studies (Nvidia, Google)
- Reinforcement Learning
- Uber analysis of food delivery
- Deep Learning

Three Main Problems

1. Image Processing (photos and videos)
   - GPU (nvidia) and TPU (google)
2. Natural Language Processing (speech recognition, text analysis and machine translation)
   - Word2Vec and BERT
3. Robotics (automation)
   - AWS, Q-learning and bandit problems

Deep learning solves all of this problems

AIQ Chapters

- Chapter 1: Bayes and Personalization
- Chapter 2: Image Classification
- Chapters 3: Robotics
- Chapter 4: Natural Language Processing
- Chapter 5: Real time monitoring and decision making
- Chapter 6: Health analytics
- Chapter 7: Sports Analytics and Big Data

AI: Introduction

What Does “AI" Really Mean?

Think of an algorithm.

Two distinguishing features of AI algorithms:

1. Algorithms typically deal with probabilities rather than certainties.
2. There’s the question of how these algorithms “know" what instructions to follow.

AI Failures

AI: Pattern Recognition

Predictive Analytics uses Pattern Recognition

Good prediction rule to map input to output

Two key ideas

1. In AI, a “pattern" is a prediction rule mapping an input to expected output.
2. “Learning a pattern" means fitting a good prediction rule to a data set.

In AI, prediction rules are often referred to as “models”. The process of using data to find a good prediction rule is called “training the model”.

Ethics of Automating your job

Turing’s warning

Pattern Matching: Mammograms

Pattern matching is a powerful empirical tool

- Original findings were done by looking at data!!
- Side-by-side images with and without breast cancer.
- To the human eye what's different? White blobs.

Nowadays done by deep learning ...

NPR Article on Training A Computer To Read Mammograms As Well As A Doctor

Pattern Matching: Chess

Shannon number 10^{152}

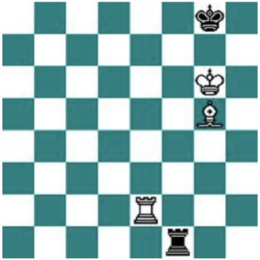
- Hand-coded Rules
- Chess.com

  Dataset on all games. Openings, Middle game, End game
- 23 million human games. Logistic regression, predict next move of Grandmaster given board position. 57% accuracy
- Deep and Reinforcement Learning!

  Maximize probability of winning. Build Value and Policy functions.
- alphaGoZero. Play itself billions of times!!

  No need for humans . . . Spatial not lines of play

Pattern Matching: Chess



$$v(s, \mathbf{w}) = \mathbf{x}(s) \cdot \mathbf{w} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix} \cdot \begin{array}{l} \begin{bmatrix} +5 \\ +3 \\ +1 \\ -5 \\ -3 \\ -1 \\ \vdots \end{bmatrix} \end{array} \begin{array}{c} \text{♖} \\ \text{♗} \\ \text{♙} \\ \text{♜} \\ \text{♝} \\ \text{♟} \\ \end{array}$$

$$v(s, \mathbf{w}) = 5 + 3 - 5 = 3$$

Karpov Unbeaten (Queen's Gambit Declined)

chessgames.com
online chess database and community

Search | Kibitz | Profile | Help

Members • Prefs • Laboratory • Collections • Openings • Endgames • Sacrifices • History • Search Kibitzing • Kibitzer's Café • Chessforums • Tournament Index • Players • Kibitzing

Karpov playing Queen's Gambit Declined (D58) as White

Player: **Anatoly Karpov**
Opening: **Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst** (D58)

Talking:

page 1 of 1; 22 games

PGN Download

Game	Result	Moves	Year	Event/Locale	Opening
1. Karpov vs Spassky	1-0	35	1974	Karpov - Spassky Candidates Semifinal	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
2. Karpov vs Unzicker	1-0	104	1980	Horten Tournament	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
3. Karpov vs Geller	1-0	33	1981	Moscow	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
4. Karpov vs Kasparov	½-½	23	1984	Karpov - Kasparov World Championship Match 1984/85	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
5. Karpov vs Kasparov	½-½	22	1984	Karpov - Kasparov World Championship Match 1984/85	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
6. Karpov vs Kasparov	½-½	35	1984	Karpov - Kasparov World Championship Match 1984/85	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
7. Karpov vs Kasparov	½-½	48	1985	Karpov - Kasparov World Championship Match 1984/85	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
8. Karpov vs Kasparov	½-½	49	1985	Karpov - Kasparov World Championship Match	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
9. Karpov vs Kasparov	½-½	62	1987	Kasparov - Karpov World Championship Match	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
10. Karpov vs Portisch	½-½	56	1988	Brussels World Cup	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
11. Karpov vs Short	½-½	47	1988	Belfort World Cup	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
12. Karpov vs Short	½-½	84	1989	Rotterdam World Cup	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
13. Karpov vs Portisch	½-½	56	1989	Skelleftea World Cup	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
14. Karpov vs Short	½-½	30	1991	Amsterdam VSB	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
15. Karpov vs Beliavsky	1-0	41	1991	Reggio Emilia 9192	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
16. Karpov vs Ljubojevic	1-0	52	1992	Melody Amber Rapid	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
17. Karpov vs Short	1-0	45	1992	Candidates semi-final	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
18. Karpov vs U Boensch	1-0	41	1992	Baden-Baden Group A	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
19. Karpov vs Short	½-½	41	1993	Melody Amber Rapid 2nd	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
20. Karpov vs Kiril D Georgiev	1-0	34	1994	Tilburg	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
21. Karpov vs Beliavsky	1-0	62	1996	YUG-chT	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst
22. Karpov vs Topalov	½-½	31	2002	FIDE GP	D58 Queen's Gambit Declined, Tartakower (Makagonov-Bondarevsky) Syst

The Netflix show: Harmon vs Borgov

Deep Learning: Introduction

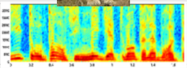
Deep Learning is the most widely used machine learning tool for high dimensional input-output problems

- Image Recognition
- Google Translate
- Driverless Cars

The applications are endless ....

> *Sir, you complain a machine can't do something. If you can define the task, I can build a machine to do it*
>
> *John von Neumann*

Deep Learning Recovers Patterns

Input	Output
Pixels: 	"lion"
Audio: 	"see at tuhl res taur aun ts"
<query, doc>	P(click on doc)
"Hello, how are you?"	"Bonjour, comment allez-vous?"
Pixels: 	"A close up of a small child holding a stuffed animal"

Why do we care about DL?

Input space (X) includes numerical, text (word2vec), images, videos

Vectors, matrices and tensors, ...

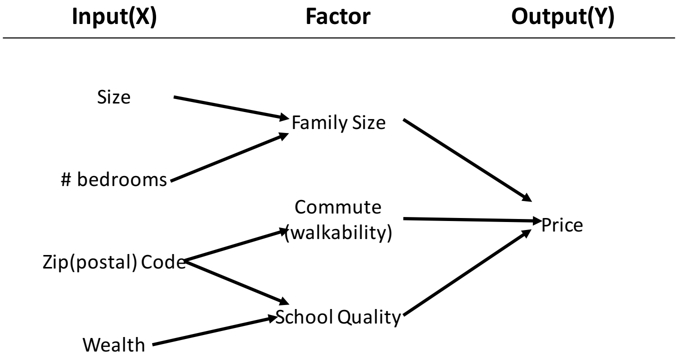
- Google's translation algorithm $\sim$ 1-2 billion parameters
- Alexa's speech recognition: 100 million parameters

  Networks will get larger and more efficient
- Google Waymo
- ImageNet $n = 14{,}197{,}122$, $p = 256 \times 256 = 65536$

Multi-Layer Deep Models

- NN models one layer!! Key is to use multi “deep” layers
- Learn weight and connections in hidden layers

Predicting House Prices ...



Kolmogorov-Arnold

There are no multivariate functions just superpositions of univariate ones

Let $f_1, \ldots, f_L$ be given univariate activation functions. We set

$$f_l^{W,b}(z) = f_l\left(\sum_{j=1}^{N_l} w_{lj} z_j + b_l\right) = f_l(W_l z + b_l)\,, \quad 1 \leq l \leq L,$$

Deep predictor has hidden units N_l and depth L.

$$\hat{Y}(X) = F(X) = \left(f_1^{W_1,b_1} \circ \ldots \circ f_L^{W_L,b_L}\right)(X)$$

Composite map Unrolled

The final output is the response y, which can be numeric or categorical. A deep prediction rule is then

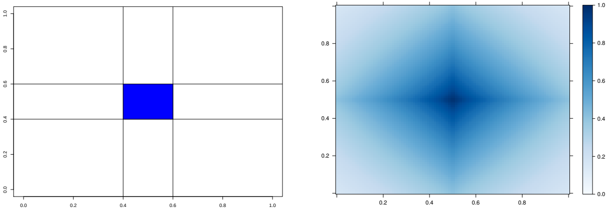
$$
\begin{aligned}
z^{(1)} &= f^{(1)}\left(w^{(0)}x + b^{(0)}\right), \\
z^{(2)} &= f^{(2)}\left(w^{(1)}z^{(1)} + b^{(1)}\right), \\
&\dots \\
z^{(L)} &= f^{(L)}\left(w^{(L-1)}z^{(L-1)} + b^{(L-1)}\right), \\
\hat{y}(x) &= w^{(L)}z^{(L)} + b^{(L)}.
\end{aligned}
$$

Dimension reduction!

Deep Learning Predictors

Smart conditional averaging

The competitors: Trees and RF.

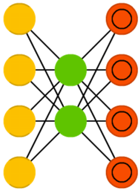


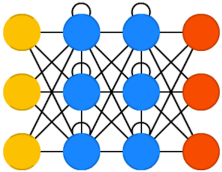
(a) Tree Kernel (b) Random Forest Kernel

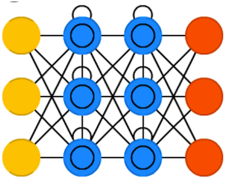
Few points will be neighbors in a high dimensional input space.

Deep Architectures: `TensorFlow`

Many possible superpositions of univariate semi-affine functions



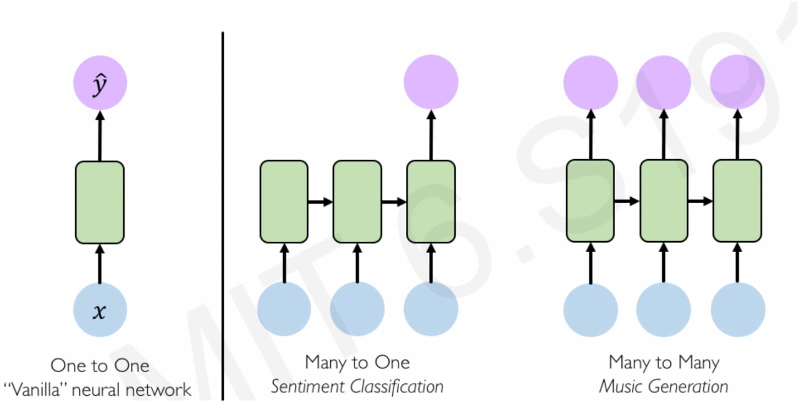




Auto-encoder Recurrent Long Short Term Memory

http://www.asimovinstitute.org/neural-network-zoo/

LSTM



$\hat{y}$
x
One to One
"Vanilla" neural network
Many to One
Sentiment Classification
Many to Many
Music Generation

Dropout

Dropout is a model selection technique designed to avoid over-fitting in the training process.

It does so by removing input dimensions in X randomly with a given probability p.

The dropout architecture becomes

$$
\begin{aligned}
D_i^{(l)} &\sim \mathrm{Ber}(p), \\
\tilde{Y}_i^{(l)} &= D^{(l)} \star X^{(l)}, \\
Y_i^{(l)} &= f(Z_i^{(l)}), \\
Z_i^{(l)} &= W_i^{(l)} X^{(l)} + b_i^{(l)}.
\end{aligned}
$$

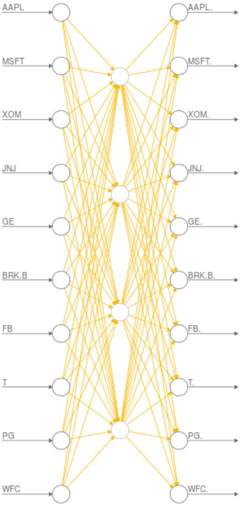
Auto-Encoder

An auto-encoder trains the architecture to approximate X by itself (i.e., $X = Y$) via a bottleneck structure.

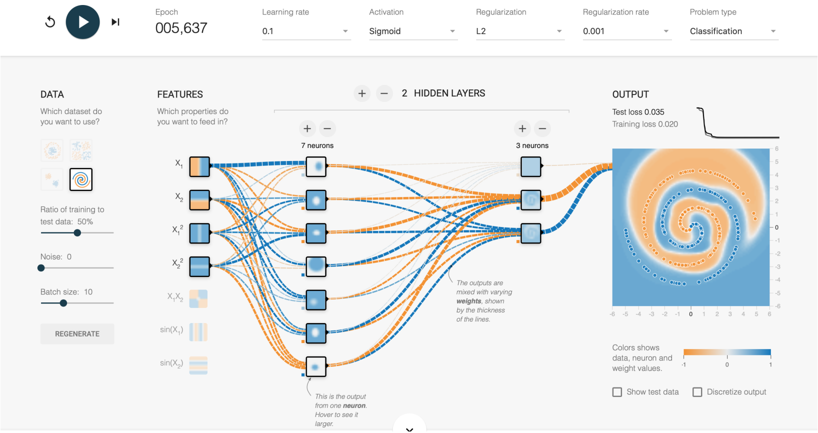
- An auto-encoder creates a more cost effective representation of X.
- Under an L^2-loss function, we wish to find

  $$\arg\min_{W,B} \|F^{W,b}(X) - X\|_2^2$$

  subject to a regulariziation penalty on the weights and offsets.
- In an auto-encoder, for a training data set $\{X_1, X_2, \ldots\}$, we set the target values as $Y_i = X_i$.


AAPL
MSFT
XOM
JNJ
GE
BRK.B
FB
T
PG
WFC
AAPL.
MSFT.
XOM.
JNJ.
GE.
BRK.B.
FB.
T.
PG.
WFC.

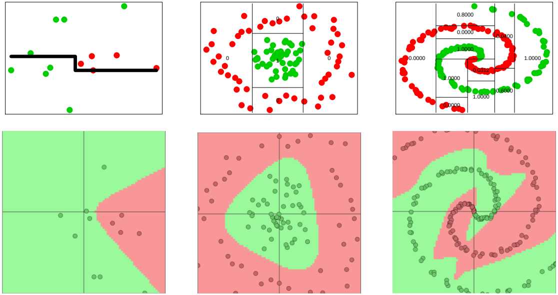
Neural Networks: Tensorflow and PyTorch



Tree vs DL example

$$Y = \mathrm{softmax}(w^0 Z^2 + b^0)$$

$$Z^2 = \tanh(w^2 Z^1 + b^2) \quad Z^1 = \tanh(w^1 X + b^1).$$



An advantage of deep architectures is that the number of hyper-planes grow exponentially with the number of layers

DL Solves the Problem

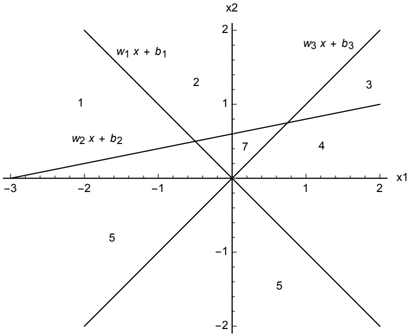


Figure: Hyperplanes defined by three neurons with ReLU activation functions.

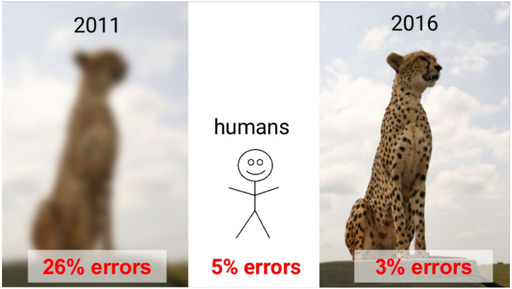
$$\hat{Y}(X) = \sum_{k \in K} w_k(X)\,\hat{Y}_k(X),$$

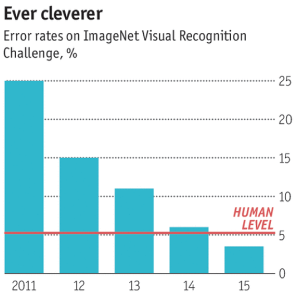
Image Processing

Image Processing

- Algorithms
  - deep learning on GPU and TPU
- Data sets (benchmark problems)
-
  - MNIST, CIFRA
- Business applications
  - L'Oréal, Advertising

Image recognition has improved





Machines are becoming better than humans

Image Processing: MNIST

Hand-written digits.

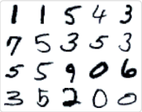
Multi-layer fully-connected neural network.

Convolution neural network

Classification Dataset Results

MNIST

who is the best in MNIST ?



MNIST 50 results collected

Units: error %

> Classify handwriten digits. Some additional results are available on the original dataset page.

Result	Method	Venue	Details
0.21%	Regularization of Neural Networks using DropConnect	ICML 2013	
0.23%	Multi-column Deep Neural Networks for Image Classification	CVPR 2012	
0.23%	APAC: Augmented PAttern Classification with Neural Networks	arXiv 2015	
0.24%	Batch-normalized Maxout Network in Network	arXiv 2015	Details
0.29%	Generalizing Pooling Functions in Convolutional Neural Networks: Mixed, Gated, and Tree	AISTATS 2016	Details
0.31%	Recurrent Convolutional Neural Network for Object Recognition	CVPR 2015	
0.31%	On the Importance of Normalisation Layers in Deep Learning with Piecewise Linear Activation Units	arXiv 2015	

DL Model

Dataset contains 60k training observations and 10k samples out-of-sample performance (validation). 2-hidden layers with ReLU activation function

$$
\begin{aligned}
Y =& \operatorname{softmax}(W^3 \tilde{Y}_i^{(3)} + b^3) \\
\tilde{Y}_i^{(3)} =& D^{(3)} \star Z^2 \\
D_i^{(3)} \sim& \operatorname{Ber}(0.5) \\
Z^2 =& \max(W^2 \tilde{Y}_i^{(2)} + b^1, 0) \\
\tilde{Y}_i^{(2)} =& D^{(2)} \star Z^1 \\
D_i^{(2)} \sim& \operatorname{Ber}(0.5) \\
Z^1 =& \max(W^1 \tilde{Y}_i^{(1)} + b^1, 0) \\
\tilde{Y}_i^{(1)} =& D^{(l)} \star X, \quad X \in R^{1024} \\
D_i^{(1)} \sim& \operatorname{Ber}(0.5)
\end{aligned}
$$

Cross-entropy–negative log-likelihood–as loss function to be optimized.

Mnist Example

First load the keral library

```
library(keras)
```

and load the data

```
mnist <- dataset_mnist()
x_train <- mnist$train$x
y_train <- mnist$train$y
x_test <- mnist$test$x
y_test <- mnist$test$y
```

we need to convert outputs to categorical varibale

```
y_train <- to_categorical(y_train, 10)
y_test <- to_categorical(y_test, 10)
```

Data preparation

Reshape the data and vectorize the images

```
dim(x_train) <- c(nrow(x_train), 784)
dim(x_test) <- c(nrow(x_test), 784)
```

Rescale so the inputs are between 0 and 1

```
x_train <- x_train / 255
x_test <- x_test / 255
```

Build the model

```
model <- keras_model_sequential()
model %>%
  layer_dense(units = 256, activation = 'relu', input_shape = c(784)) %>%
  layer_dropout(rate = 0.4) %>%
  layer_dense(units = 128, activation = 'relu') %>%
  layer_dropout(rate = 0.3) %>%
  layer_dense(units = 10, activation = 'softmax')
```

You can print your model

```
summary(model)
```

```
## Model: "sequential"
## ________________________________________________________________________________
## Layer (type)                        Output Shape                    Param #
## ================================================================================
## dense (Dense)                       (None, 256)                     200960
## ________________________________________________________________________________
## dropout (Dropout)                   (None, 256)                     0
## ________________________________________________________________________________
## dense_1 (Dense)                     (None, 128)                     32896
## ________________________________________________________________________________
## dropout_1 (Dropout)                 (None, 128)                     0
## ________________________________________________________________________________
## dense_2 (Dense)                     (None, 10)                      1290
## ================================================================================
## Total params: 235,146
## Trainable params: 235,146
## Non-trainable params: 0
## ________________________________________________________________________________
```

Train it

It takes a few minutes!

```
model %>% compile(
loss = 'categorical_crossentropy',
optimizer = optimizer_rmsprop(),
metrics = c('accuracy')
)

history <- model %>% fit(
  x_train, y_train,
epochs = 15, batch_size = 128,
validation_split = 0.2
)
```

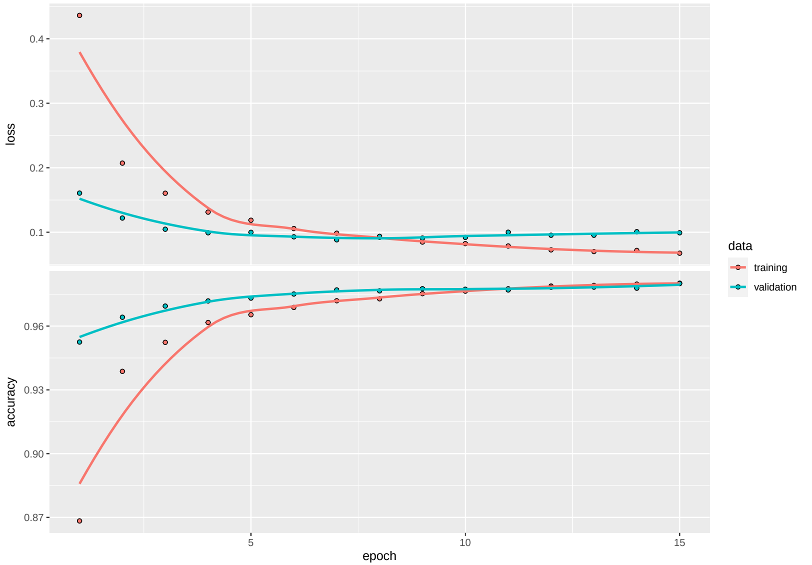
Now we can predict!

```
yhat = model %>% predict_classes(x_test)
print(yhat[1:5])
```

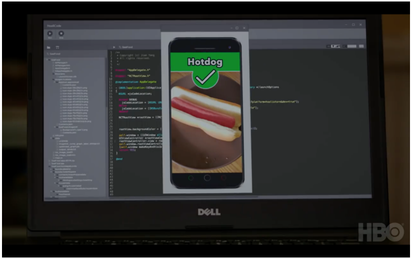
```
## [1] 6 5 8 6 4
```

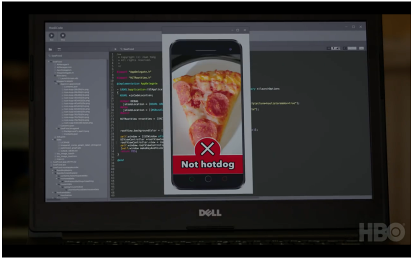
Plot the loss function values

```
plot(history)
```



Not HotDog





Silicon Valley: Season 4 Episode 4: https://youtu.be/ACmydtFDTGs

Examples: Image Classification

Pattern: relationship between the visual features and its class.

- Cucumber sorter of images to sort them into nine different classes.
- Toilet paper assignment at the Temple of Heaven in Beijing
- Identification of untagged friends on Facebook
- Detection collisions between subatomic particles at CERN

Ultimately, computers are agnostic about the type of input you give them, because to a computer, it's all just numbers.

Japanese Farming: Cucumbers

TensorFlow

Makoto Koike sorts cucumbers at his parents farm. 9 classes of cucumbers.

His mother spent up to eight hours per day at peak harvesting times.

Learning to sort cucumbers can take months.

Size, thickness, color, texture, scratches, crooked or have prickles.

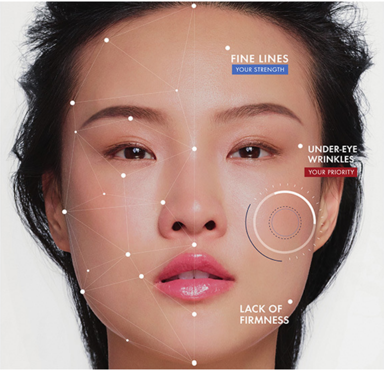
- Data 7000 pictures of cucumbers sorted by his mother. What are the important “features”?
- Recognition accuracy exceeded 95% for test images, out-of-sample the accuracy drops to about 70%.

Farming and TensorFlow

How AI Helps us Understand & Protect Snow Leopards

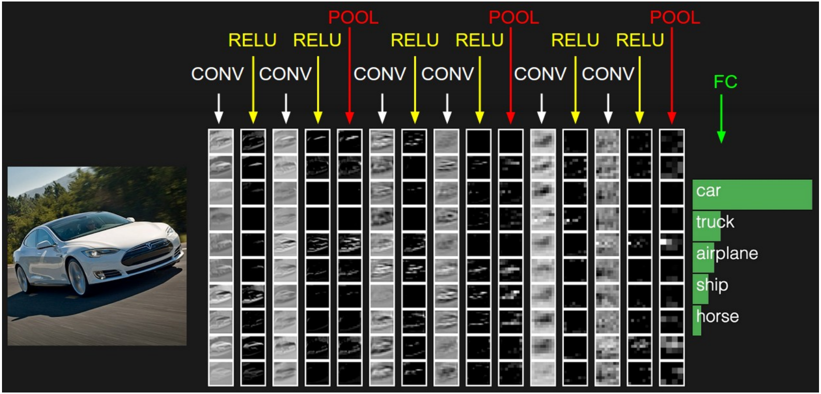
L’Oréal AI-powered skin diagnostic

Skin Atlases allow to evaluate or predict the general aging of the face and are used today for clinical evaluations of cosmetic or dermatological treatments



loreal article

CNN



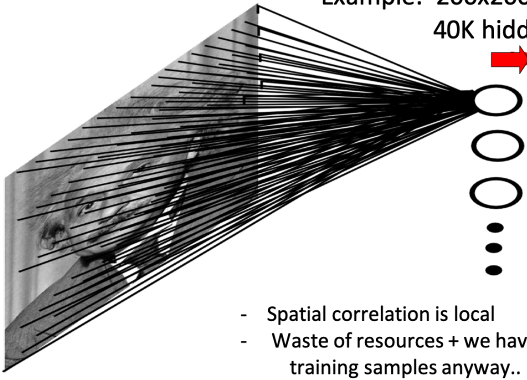
Source: http://cs231n.github.io/convolutional-networks/

Fully Connected Layer

Example: 200x200 image

40K hidden units

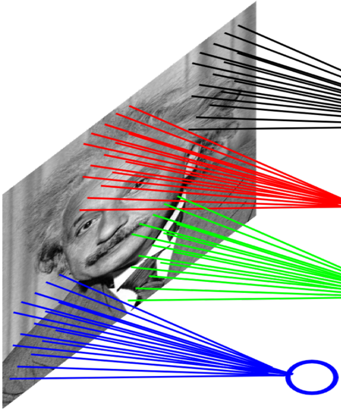
parameters!!!



- Spatial correlation is local
- Waste of resources + we have not enough training samples anyway..

Ranzato @ Facebook

Locally Connected Layer

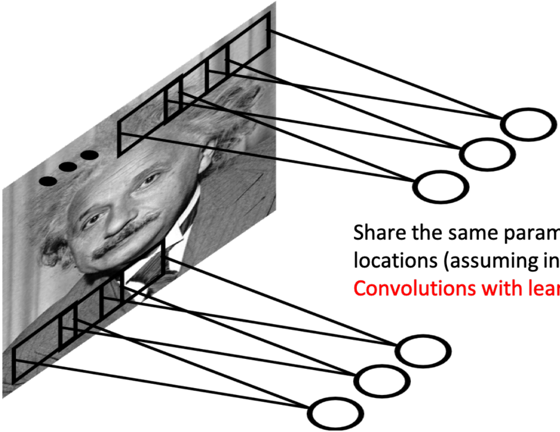


Example: 200x200 image
40K hidden units
Filter size: 10x10
4M parameters

Note: This parameterization is good when input image is registered (e.g., face recognition).

Ranzato @ Facebook

Convolutional Layer

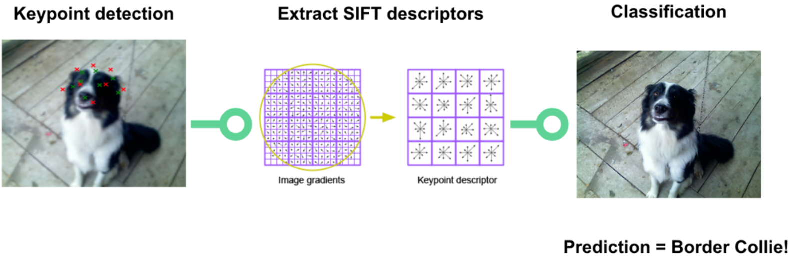


Share the same parameters across different locations (assuming input is stationary):
Convolutions with learned kernels

Ranzato @ Facebook

TPU Advantages for real-time data analysis

- Matrix analysis programmed into chips
- XLA: Accelerated Linear Algebra
- AD: Automatic Differentiation



CPU is a neural net processor

AI in China

Xinhua is a China's state news agency uses virtual anchor

- Sogou, the anchor was designed to simulate human voice, facial expressions and gestures
- AI to measure level of focus and engagement among students

WSJ Video

SenseFace: AI for Profiling Uighurs in China

Shoppers lined up for identification checks outside the Kashgar Bazaar last fall. Members of the largely Muslim Uighur minority have been under Chinese surveillance and persecution for years.Credit. Paul Mozur

Source: “One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority” by NYT

https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html

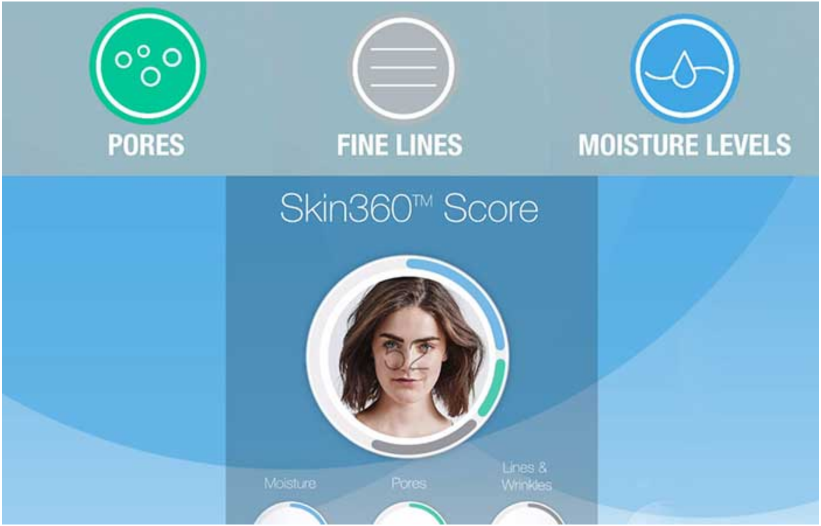
SenseFace: AI for Profiling Uighurs in China

- Law enforcement in the central Chinese city of Sanmenxia screened whether residents were Uighurs 500,000 times in 1 month
- Law enforcement from the central province of Shaanxi, requested a smart camera system last year that “should support facial recognition to identify Uighur/non-Uighur attributes”.

Facial Recognition is is becoming a commodity. Can buy from IBM:

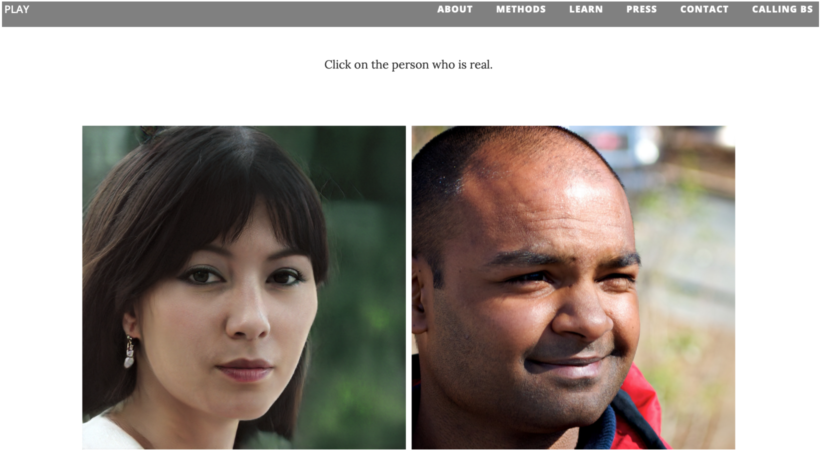
Attribute detection with Body Camera Analytics

Facial Recognition For Personalized Customer Experience



The Verge Article

Which Person is Real?

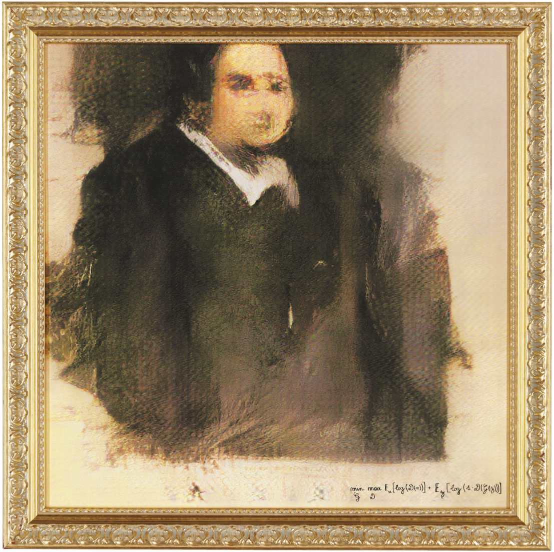


One of these models doesn't exist

Virtual model Imma

engadget article

AI artwork sells for $432,500



$$\min_G \max_D \mathbb{E}_x[\log(D(x))] + \mathbb{E}_z[\log(1 - D(G(z)))]$$

Application: Training A New Rembrandt

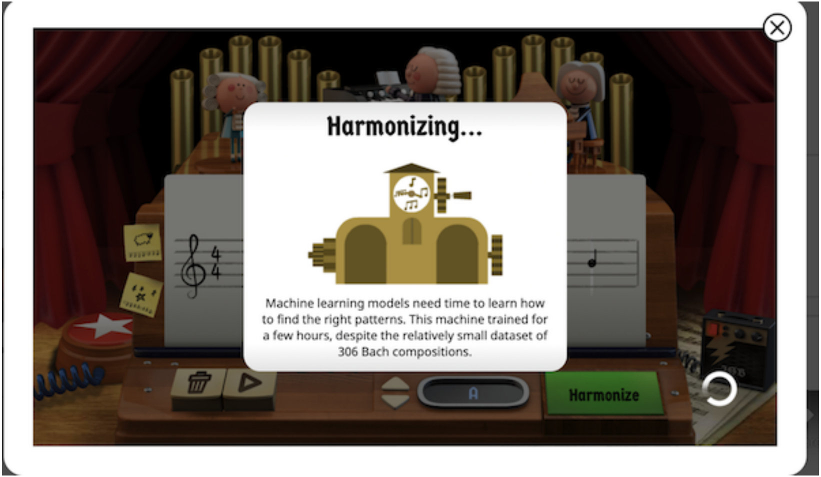
Analyze all 346 of Rembrandt's paintings

- Identify all geometric patterns used by Rembrandt.
- Reassemble into a fully formed face and bust

Nvidia Faces

Google's AI: Music like Bach

Try Yourself!



Google's AI: Music like Bach

Google has launched its first ever AI-powered Doodle. All you have to do is plunk down a few notes and the Doodle will add harmony to them in the signature style of Bach.

In this case, 306 chorale harmonizations composed by Bach were fed into a model. His chorales make for great training data because their structure is pretty consistent and concise – they all contain four voices, which take on a pleasing depth when layered on top of one another. After the model "learned" Bach's style by picking out the patterns, the machine learning system was refitted to run within the confines of your humble web browser.

Google Verily: Identifying Skin Cancer

- Dataset: 130,000 images of skin lesions/2,000 different diseases
- Test data: 370 high-quality, biopsy-confirmed images
- Better performance than 23 Stanford dermatologists
- 10,000 hours no match for deep learning and large datasets

Google's AI: Heart disease from eye scan

Google's Verily scans of the back of a patient's eye able to accurately deduce individual's age, blood pressure, and whether or not they smoke, etc.

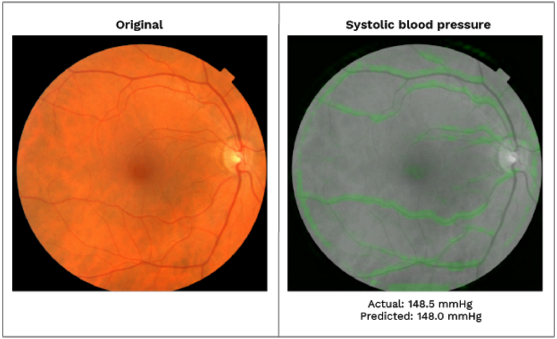
Predict their risk of suffering a major cardiac event–such as a heart attack–with roughly the same accuracy as current leading methods.

Quicker to analyze a patient's cardiovascular risk—doesn't require a blood test.

Training $300,000$ patients. Eye scans plus general medical data. Deep learning to mine for patterns

Google's Verily: Diabetic Retinopathy

Diabetic Retinopathy is the fastest growing cause of preventable blindness!



Two images of the fundus–interior rear of your eye.

The left is a regular image; the right shows how Google's algorithm picks out blood vessels (in green) to predict blood pressure.

The DL algorithm outperforms ophthalmologists 82 vs 84

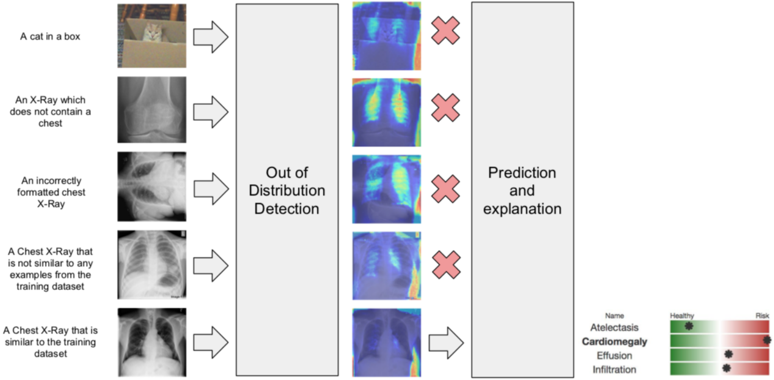
Intelligent Scanning Using Deep Learning for MRI

The scan operator first acquires a set of low-resolution “localizer" images from which approximate location and orientation of desired landmarks can be identified. These anatomical references are then used to manually plan the exact locations, orientation, and required coverage for images that will be used for the high-resolution scans that are used for diagnosis.

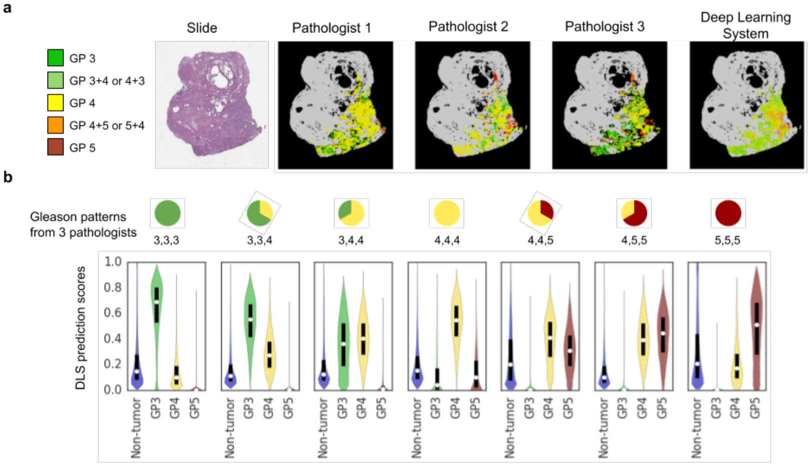
To aid the scan operator we developed a deep-learning (DL) based framework for intelligent MRI slice placement (ISP) for several commonly used brain landmarks. TensorFlow library with the Keras interface is used to implement the DL based framework for ISP.

As compared to the classical approaches, a DL-based approach is less affected by factors that affect MRI image quality or appearance. And it can be easily extended across other anatomies.

Chester: Chest X-Ray Disease Prediction System



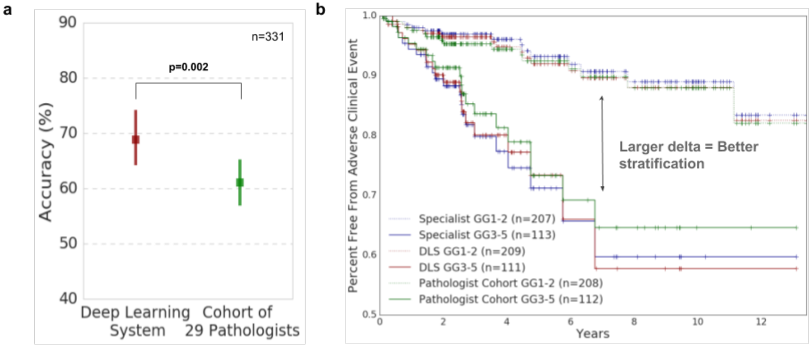
Google Verily: Gleason grading of prostate cancer



Source: https://ai.googleblog.com/2018/11/improved-grading-of-prostate-cancer.html

Image Analysis: Gleason grading of prostate cancer

Deep learning outperforms pathologists

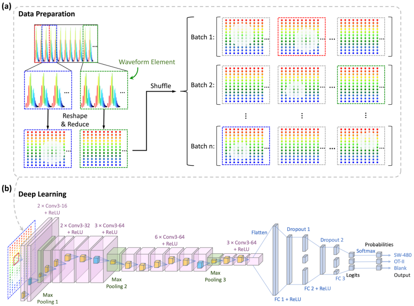


Source: https://ai.googleblog.com/2018/11/improved-grading-of-prostate-cancer.html

AI Diagnostics

Artificial intelligence-powered device that detects cancer cells

- Hundreds of times faster than previous methods.
- Can extract cancer cells from blood immediately after they are detected
- Larry Ellison claims it is the future (video)



Source: Deep learning enables scientists to identify cancer cells in blood in milliseconds

Natural Language Processing

Natural Language Processing

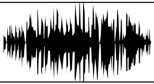
- Algorithms
  - Word2Vec, deep learning
- Business applications
  - Classification
  - Semantic Search
  - Question & Answer
  - Content Generation
  - Language Translation
  - Summarization

NLP: Watson and Jeopardy

Jeopardy

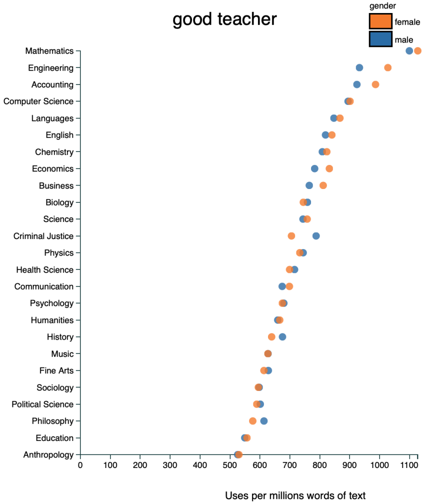
- IBM super computer Watson beat two former champions of TV game jeopardy and took home one million dollars prize. Watson is a significant leap a machine's ability to understand context in human language.
- IBM believes the technology behind Watson can be applied to a variety of fields, most notably medicine.

Examples: Google Assistant

Input	Output
	Speech to text: “Chi-ca-go hot-dog."
68F/20C, 70% humidity, mostly sunny	Numerical prediction: “Power consumption in London will be 25,500 megawatt-hours."

Google Assistant

Text Analytics



http://benschmidt.org/profGender

Natural Language Processing (NLP)

Use word2vec combined with classification model (inputs = word vectors) to solve different problems

- Sequence tagging: automated analysis of documents (extract item names, shipping address,...)
- Text classification: sentiment analysis (is this article negative about my product?)
- Seq2seq: machine translation, chat bots, summarization (law)

Example: conversational AI by NVIDIA

Example: Can a Machine Learn to Write for The New Yorker?

Train the model by asking it to fill-in the blaks

Original Text

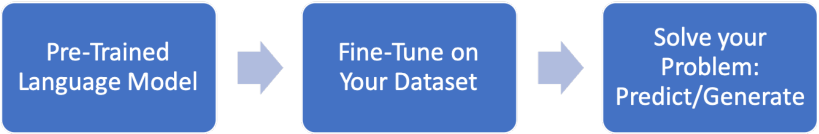
Former President Barack Obama returned to his academic roots on Tuesday, visiting the University of Chicago campus where he spent 12 years, and the neighborhood where his life last held some semblance of normalcy.

Masked Words

Former ________ Barack Obama returned to his academic _______ on Tuesday, visiting the __________ of Chicago campus where he spent 12 years, and the neighborhood where his _____ last held some semblance of normalcy.

Use pre-trained model for your task

- Train a model by asking it to fill-in the blanks on large corpuses of text (all digitized books, text from web pages, twitts and FB posts, Google queries,...)
- Fine tune this model on the individual task with small amounts of data
- No need to have large datasets!
- Some existing pre-trained models: ULMFiT, Transformer, Google's BERT, Transformer-XL, OpenAI's GPT-2, word2vec



Human and Machines, Talking Together

A few likely trends stand out:

- Language models will become personalized.

  The machines around you will adapt to the way you speak, just as they adapt to your movie-watching preferences
- Good policy and thoughtful regulations will be hugely important.

  An algorithm that can write an episode of Friends seems cute, if a bit useless. That same algorithm will seem a lot more pernicious when someone can program it to flood the internet with fake news around election time.

ChatBots and DeepFakes

AI Can create speech and video content that humans cannot You Won't Believe What Obama Says In This Video!

You cannot distinguish human voice from the synthesized one: Google WaveNet

China's AI news anchor

Digital Einstein

AI can talk to you and you won't even know it: Silicon Valley: Gilfoyle Made A Bot (Season 6 Episode 1)

Natural Language Processing (NLP): David Bowie and Verbasizer

David Bowie used machine learning (ML) to write his songs. Bowie comments that the questionable quality of the lyrics from his old band Tin Machine should be “blamed” on the computer.

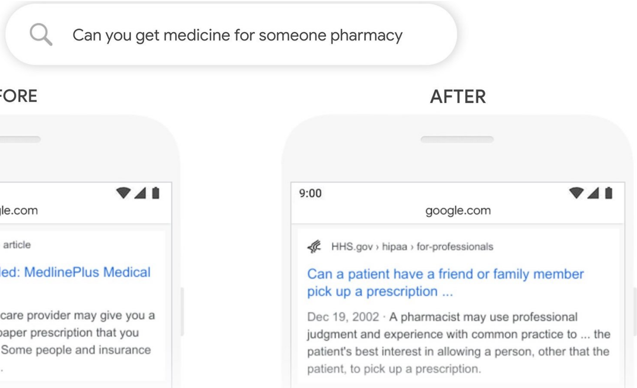
Verbasizer automates a technique used by Bowie to write songs

Form random phrases based on shuffled word cut-outs from newspapers and other sources.

Example: Google Search: BERT

Google is using deep learning to pre-process your search query.

Bidirectional Encoder Representations from Transformers



Source: Understanding searches better than ever before

GPT3

- Powering Next Generation Apps: https://openai.com/blog/openai-api/
- Transformer architecture

OpenAI API

We're releasing an API for accessing new AI models developed by OpenAI. Unlike most AI systems which are designed for one use-case, the API today provides a general-purpose "text in, text out" interface, allowing users to try it on virtually any English language task. You can now request access in order to integrate the API into your product, develop an entirely new application, or help us explore the strengths and limits of this technology.

Musk criticises OpenAI

From Grace to Alexa: The Natural Language Revolution

Three basic problems to overcome:

- Rule Bloat

  It's hard to write down all the rules for natural languages. Exceptions create trouble.

- Robustness

  Top-down rules usually break upon contact with the real world. People's understanding of language is highly robust to background noise, “mistakes" and pronunciation.

- Ambiguity

  Language is full of ambiguity: homophones and syntactic ambiguity.

1980–2010:The Growth of Statistical Natural Language Processing

A new approach is necessary:

- flexible rather than rigid
- probabilistic rather than deterministic
- bottom-up, based on real-world data, rather than top-down, based on a profusion of rules

From the 1980s onward, NLP shifted its focus from understanding to mimicry–from knowing how, to knowing that. Language became a prediction-rule problem based on input/output pairs.

- speech recognition:“brekfustahkoz"=“breakfast tacos"
- English to Russian:“reset"=“perezagruzka"

The Math of 20 Questions

Numerical representation of words could be like:

	Animal	Agreeable	Growls or grunts	Talks	Lives in London	Is a bear
Scrooge	1	0	1	1	1	0
Rafael Nadal	1	1	1	1	0	0
Tiny Tim	1	1	0	1	1	0
Paddington Bear	1	1	0	1	1	1
Trafalgar Square Christmas tree	0	1	0	0	1	0

How AI Plays 20 Questions

The rules are changed in three ways:

1. Instead of just a yes or no, each answer is a number between 0 (completely no) and 1 (completely yes).
2. Scores are decided by “semantic closeness," or how close you are to the word’s underlying meaning.

   For JFK saying “Ich bin ein Berliner", “I am a cronut" is a lot closer than “I am a German".
3. The same questions must be asked in every game.

HOW AN ALGORITHM PLAYS 20 QUESTIONS

	Question 1: "computers"	Question 2: "universities"	Question 3: "cooking"	Question 4: "law"
nvidia	1	0.045	0.156	0.083
servers	0.999	0.944	0.214	0.184
username	0.999	0.468	0.842	0.963
ethernet	0.999	0.587	0.617	0.072
interface	0.999	0.355	0.831	0.032
router	0.998	0.697	0.986	0.911
displays	0.998	0.693	0.111	0.174
port	0.997	0.646	0.583	0.184
pixels	0.997	0.253	0.017	0.21
firewall	0.995	0.729	0.957	0.636
undergraduate	0.089	0.999	0.107	0.627
faculty	0.365	0.999	0.114	0.944
scholarships	0.063	0.999	0.291	0.398
applicants	0.153	0.999	0.22	0.77
colleges	0.206	0.997	0.132	0.514
fellowship	0.216	0.997	0.035	0.688
committee	0.32	0.996	0.912	0.824
departments	0.42	0.994	0.502	0.77
residential	0.145	0.993	0.569	0.801
publications	0.173	0.993	0.524	0.938

Google word2vec: Putting Word vectors to Work

In effect, word2vec has learned to take the SAT Verbal test using only skills from the SAT Math test.

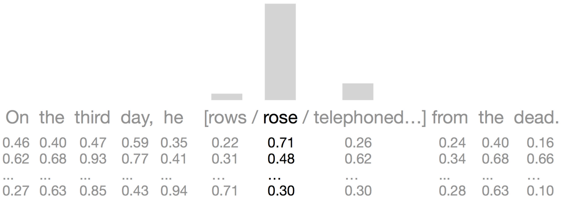
- World capitals: London-England + Italy. Word2vec's answer: “Rome."
- Word tenses: captured - capture + go. Word2vec's answer: “went."
- Which hockey teams play in which cities: Canadiens - Montreal + Toronto. Word2vec's answer: “Maple Leafs."

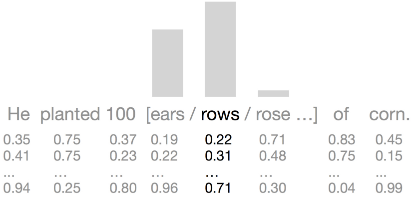
Word2Vec

Word vectors provide a clear-cut mathematical description of something that to any human listener seems simple: sometimes one word fits better, and sometimes the other.

How	lovely,	it	smells	like	a	[skunk /	rows /	rose /	goat /	sewer …]
0.45	0.61	0.83	0.39	0.45	0.43	0.66	0.22	0.71	0.32	0.20
0.37	0.18	0.51	0.39	0.71	0.98	0.22	0.31	0.48	0.87	0.26
…	…	…	…	…	…	…	…	…	…	…
0.99	0.33	0.39	0.24	0.29	0.19	0.68	0.71	0.30	0.26	0.06

Word2Vec


On the third day, he [rows / rose / telephoned…] from the dead.
0.46 0.40 0.47 0.59 0.35 0.22 0.71 0.26 0.24 0.40 0.16
0.62 0.68 0.93 0.77 0.41 0.31 0.48 0.62 0.34 0.68 0.66
… … … … … … … … … … …
0.27 0.63 0.85 0.43 0.94 0.71 0.30 0.30 0.28 0.63 0.10


He planted 100 [ears / rows / rose …] of corn.
0.35 0.75 0.37 0.19 0.22 0.71 0.83 0.45
0.41 0.75 0.23 0.22 0.31 0.48 0.75 0.15
… … … … … … … …
0.94 0.25 0.80 0.96 0.71 0.30 0.04 0.99

Natural Language Revolution

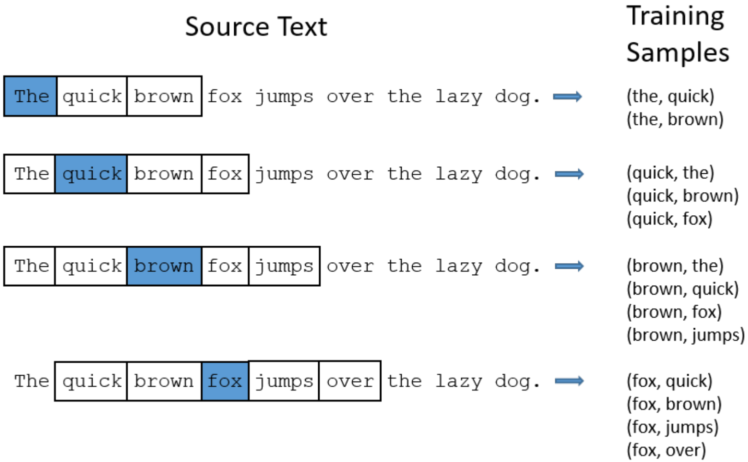
How words become numbers:

1. numerical representation for words:

	Animal	Agreeable	Growls or grunts	Talks	Lives in London	Is a bear
Scrooge	1	0	1	1	1	0

2. word co-location statistics: “ketchup", “fries", “bun"
3. vector operation(word2vec): queen=king-man+woman
4. speech recognition: mathematical language provides crucial tie-breaking information for homophones

How do we train word2vec?



Robotics

The Robotics Revolution

- In the 1950s, the state of the art was Theseus, a life-size autonomous mouse built by Claude Shannon at Bell Labs, and powered by a bank of telephone relays. He would navigate by trial and error until he found the cheese.
- In the 1960s and '70s, there was the Stanford Cart: a wagon-sized with four small bicycle wheels, an electric motor, and a single TV camera. The Cart could steer itself across a chair-filled room in five hours, without human intervention
- Today? Self-driving cars are just the start.

Robot Shoot Hoops

This basketball-loving robot by Toyota played against pro ballers in a shooting competition. According to its creators, the bizarre-looking humanoid robot called ‘Cue’ learned how to shoot thanks to artificial intelligence.

Examples: Robotics

Two classic examples

- Rory and the Robot (Rory vs The Robot)
- Robot Back-flip (Be very afraid . . . robots can now do backflips)

Key idea: train and fail many-many times until you "learn"

Deep Learning

Stanford Cart

Cart

- Hans Moravec and Stanford Cart
- KL10 processor, at about 2.5 MIPS, Moravec was eventually able to use multi-ocular vision to navigate slowly around obstacles in a controlled environment. The cart moved in one meter spurts punctuated by ten to fifteen minute pauses for image processing and route planning. In 1979, the cart successfully crossed a chair-filled room without human intervention in about five hours.

Google DeepMind’s Deep Q-learning playing Atari Breakout

Google DeepMind

- Google DeepMind created an AI using deep reinforcement learning that plays Atari games and improves itself to a superhuman level.
- Capable of playing many Atari games and uses a combination of deep artificial neural networks and reinforcement learning.
- This was the beginning for Google DeepMind.

OpenAI and Dota2 is the current state-of-the-art

The question is: Where am I?

In AI, this is called the SLAM problem, for “simultaneous localization and mapping." The word “simultaneous" is key here.

Whether you’re a person or a robot, knowing where you are means doing two things at once:

1. constructing a mental map of an unknown environment
2. inferring your own unknown location within that environment

The solution to both problems involved Bayes’s rule.

Smart cities. Copenhagen bicycles.

Copenhagen wants to cut bus travel times by 5 to 20 percent and cycling travel times by 10 percent. Reduce the number of times cyclists have to stop by 10 percent.

To better manage the traffic, there's an AI that identifies cars/bicyclists and gives priority to cyclists at morning peak hours.

Installing 380 "intelligent traffic signals" that will spot, and prioritize, buses and bikes.

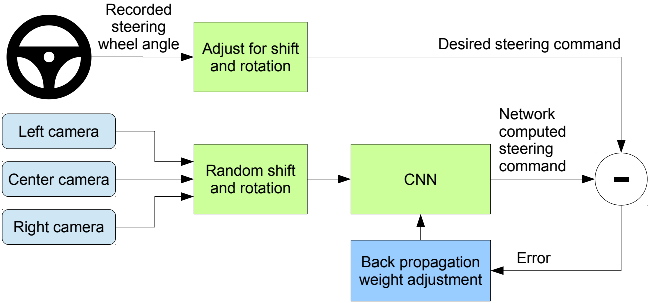
Self-driving Cars: perception

Tesla presentation

Imitation Learning

Machines can learn from billions of cases ....

- Learn form the best. Take all the moves of a chess grand master: $\{s_i, a_i\}_{i=1}^{N}$ state-action pairs
- Learn conditional distribution over actions $\pi_\theta(a_t \mid s_t)$
- Use deep neural network



Source: Mariusz 2016

Reward Function

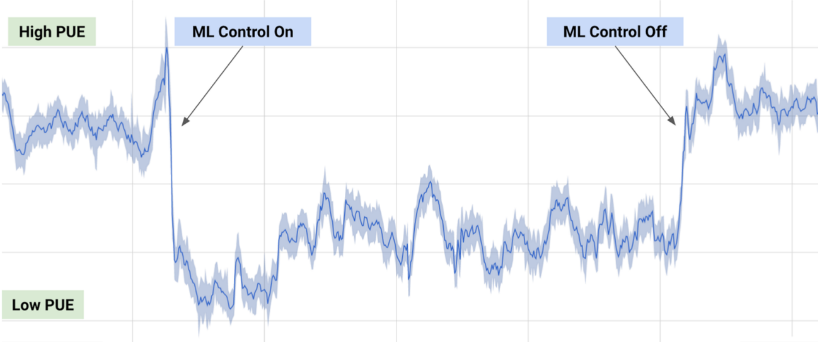
- What if some observed actions are bad, e.g in a self-play setting
- How do we distinguish a good action from a bad action?
- Introduce the reward function $r(s, a)$

High Reward

Low Reward

Google Energy: Data Center Cooling Costs Reduced by 40%

Monitoring real-time conditions and adjusting data center climate control based on past experience



DL Predicting Energy Demand

Cornell University

W
the Simon

arXiv.org > stat > arXiv:1808.05527

Search...

Help | Advanced

Statistics > Machine Learning

[Submitted on 16 Aug 2018 (v1), last revised 10 Apr 2019 (this version, v3)]

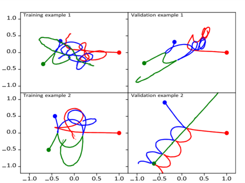
Deep Learning for Energy Markets

Michael Polson, Vadim Sokolov

Deep Learning is applied to energy markets to predict extreme loads observed in energy grids. Forecasting energy loads and prices is challenging due to sharp peaks and troughs that arise due to supply and demand fluctuations from intraday system constraints. We propose deep spatio-temporal models and extreme value theory (EVT) to capture theses effects and in particular the tail behavior of load spikes. Deep LSTM architectures with ReLU and $\tanh$ activation functions can model trends and temporal dependencies while EVT captures highly volatile load spikes above a pre-specified threshold. To illustrate our methodology, we use hourly price and demand data from 4719 nodes of the PJM interconnection, and we construct a deep predictor. We show that DL-EVT outperforms traditional Fourier time series methods, both in-and out-of-sample, by capturing the observed nonlinearities in prices. Finally, we conclude with directions for future research.

Three-Body Problem

- Calculation of a position of three interacting bodies are very sensitive
- Used to be used for navigation now is used for determining the structure of globular star clusters and galactic nuclei
- Interactions of black hole binaries interact with single black holes.
- Neural network provides an accurate solutions and is up to 100 million times faster than a state-of-the-art conventional solver.



Run state of the art Brutus solver for 10,000 random starting positions. x: starting position, y: trajectories

MIT Technology Review

Q-learning

There's a matrix of Q-values that solves the problem.

- Let s denote the current state of the system and a an action.
- The Q-value, $Q_t(s, a)$, is the value of using action a today and then proceeding optimally in the future. We use $a = 1$ to mean no deal and $a = 0$ means deal.
- The Bellman equation for Q-values becomes

$$Q_t(s, a) = u(s, a) + \sum_{s^\star} P(s^\star | s, a) \max_a Q_{t+1}(s^\star, a)$$

where P denotes the transition matrix of states

The value function and optimal action are given by

$$V(s) = \max_a Q(s, a) \quad \text{and} \quad a^\star = \text{argmax} Q(s, a)$$

Reinforcement Learning

- Generate their training data via optimally designed experiments

  Maximize a reward to find an optimal policy or action function $d(x_t \mid \theta)$
- Action function predicts response of the system to event t with characteristics x_t. e.g. t - customer arriving on the website and d is the list of ads to be shown Customer generates response y_t, e.g. click/no click
- Calculate the reward $r(d(x_t \mid \theta), y_t)$

  Sequentially learn parameters $\theta, \ldots, \theta, \ldots$

Goal: find an optimal configuration θ^* that minimizes the objective

$$\sum_{t=1}^{T} \left[r(d(x_t \mid \theta), y_t) - r(d(x_t \mid \theta), y_t) \right]$$

Policy Gradient

- Specify parametric π and r (e.g. deep learning)
- Generate state-action samples s_t^i, a_t^i and associated reward r_t^i, $i = 1, \ldots, N$

$$E_\theta \left[\sum_t r_t \right] \approx \frac{1}{N} \sum_{i=1}^{N} \sum_t r_t^i$$

- Run a step of batch SGD to update θ: policy update via back-propagation, after seeing the reward
- Only works for deterministic dynamics!
- Naive algorithm

Deep Learning: Chess

Shannon number 10^{152}

- Deep and Reinforcement Learning!
- Maximize probability of winning.
- Build Value and Policy functions. (ReLU networks)
- alphaGoZero. Play itself billions of times!!

  No need for humans . . . Spatial not lines of play

Magnus Carlsen VS Bill Gates, Alpha Zero's "Immortal Zugzwang Game" against Stockfish

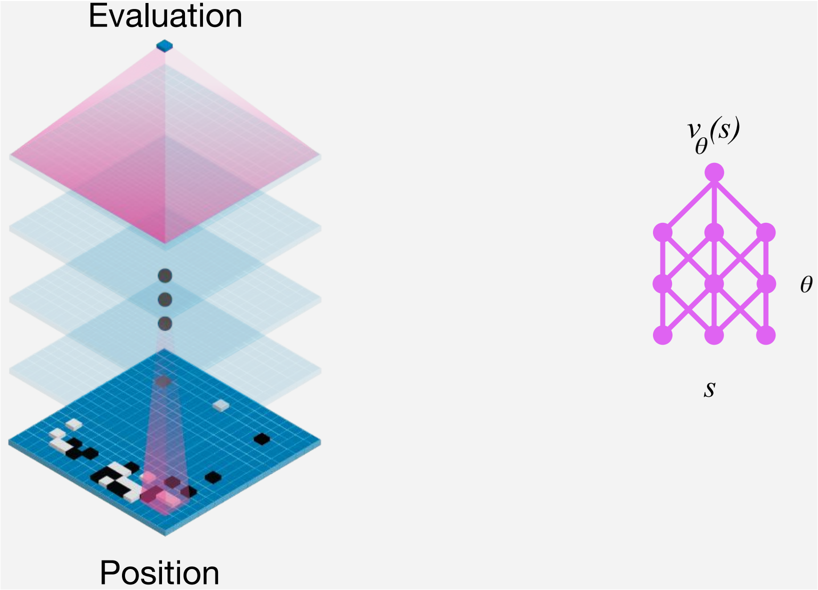
AlphaGo and AlphaGo Zero

Hand-crafted heuristic rules with deep learners

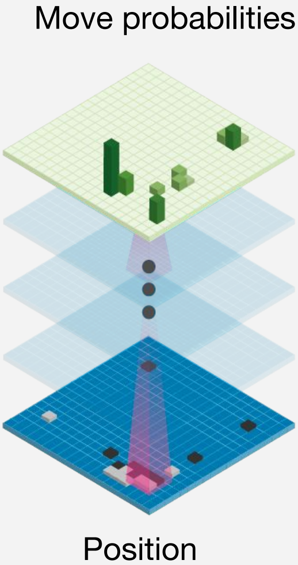
- Maximize probability of winning (Value function)
- Use SGD to update network weights based on self-play samples
- 4 hours to train grand-master level algorithm with no human inputs
- Same idea can be applied to many other settings: replace models of the world with neural nets
- Humans do the same. Tennis players do not use Newton's laws to predict trajectory of a ball

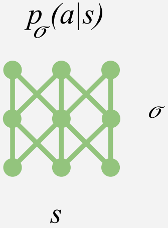
AlphaGo Movie Trailer

Value Function

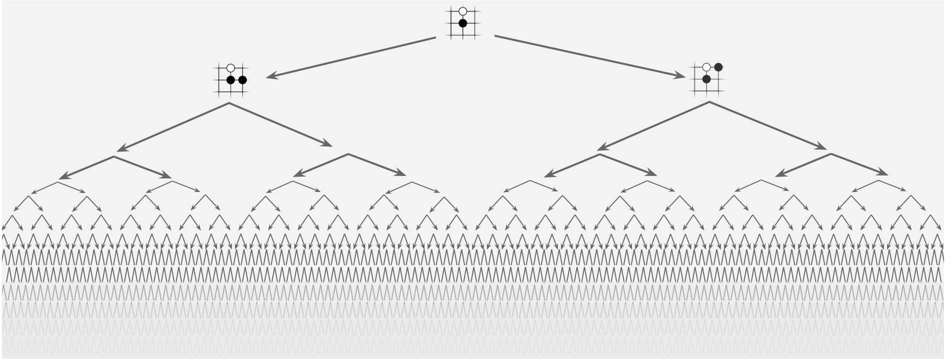


Policy Function

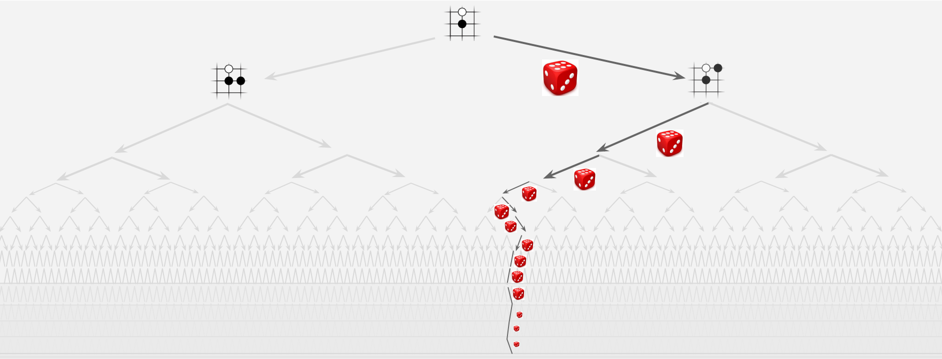




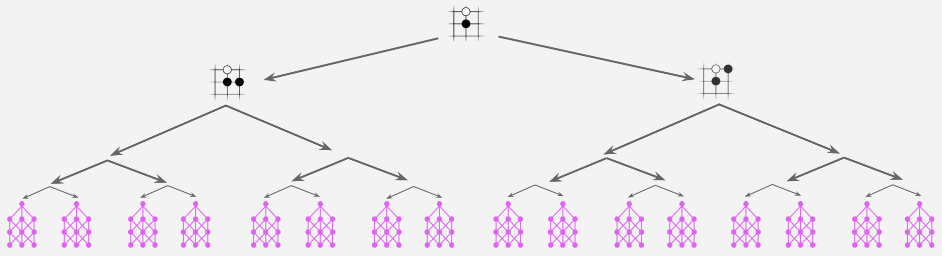
Full Tree



Monte-Carlo rollouts

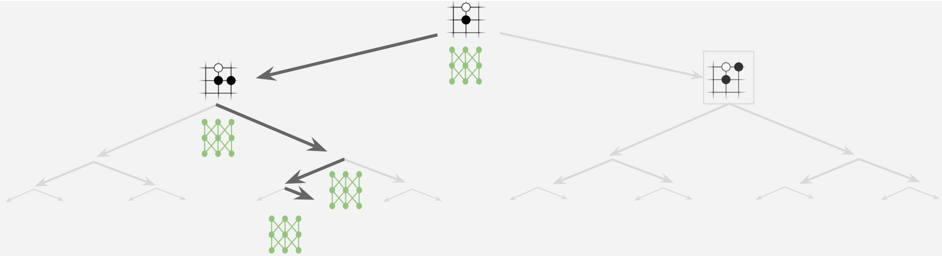


Reducing depth with value network



- Value function approximates probability of winning.
- Pick the path with highest approximated chance to win the game
- No need to explore the tree till the end

Reducing breadth with policy network



- Policy function gives a histogram over possible moves
- Pick a few with highest probabilities
- No need to explore low probability moves, reduce breadth of the search

Summary: αGo

alphaGo Movie

Supervised and Reinforcement Learning, Value Function and Tree Search

Convenient

- Fully observed
- Discrete action space
- Perfect simulator
- Relatively short game
- Trial-and error experience
- Large human datasets

Inconvenient

- Actions executed awkwardly
- Incomplete information
- Imperfect simulator
- Longer tasks, hard to assess value
- Hard to practice millions of times
- Small human data sources

AlphaGoZero's Strategies

You do not need to see all possible games. Predictive model interpolates!

StarCraft 2

You Need Both Exploration and Exploitation to Achieve the Grand Master Level

Source: AlphaStar: Grandmaster level in StarCraft II using multi-agent reinforcement learning by DeepMind

Nature Video

Multi-Armed Bandit: Thompson Sampling

- $r(d(x_t \mid \theta), y_t)$ pick one of the J possible ads
- Reward $y_t = 1$ if the user clicks on the ad and $y_t = 0$ otherwise
- Parameters are $\theta_j = p(y_t = 1 \mid d_t = j)$

$$p(d_{t+1} = j) = p\left(\theta_j = \max_k \theta_k \mid y^t\right)$$

Replace it with a Monte-Carlo simulation draw

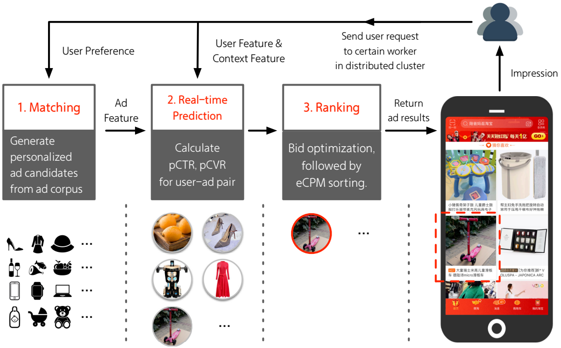
$$\theta_{t+1} \sim p(\theta \mid y^t)$$

then the best decision is

$$d_{t+1} = \arg\max_j \theta_{t+1,j}$$

Optimizing Advertising Budgets

Matching, RTP and Ranking modules sequentially pro- cess user requests, and finally return specified quantity of ads. These ads are shown in Guess What You Like of Taobao App, tagged by Hot (as shown in red dashed box) and sur- rounded with recommendation results



Jin 2018, Breen 2019

Waymo: 9 billion hours of autonomous driving simulations

Alphabet-owned Waymo uses AI to run simulations in gaming-like environments. These simulations are being used to inform their autonomous vehicle efforts and the company has amassed an impressive 5 billion hours in edge case scenarios.

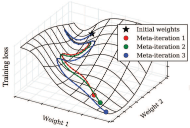
These edge cases are the really tricky 1% or 0.1% or even 0.001% of possible interactions that could occur when you vehicles moving through complex environments.

While Waymo have some actual cars, it's clear their emphasis on safe simulations digitally, is the complete opposite to Tesla's strategy.

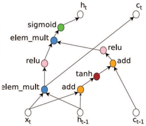
The big difference is that Tesla is deploying Autopilot into real customer vehicles and taking the data from events where the humans have taken over, therefore identifying an edge case that wasn't already accommodated in the system.

Leanring how to Learn

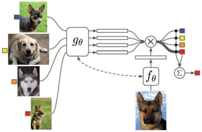
Can machines replicate human's ability to learn new concepts: General Artificial Intelligence or Learning to Learn



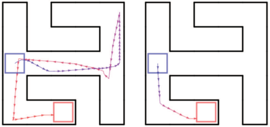
hyperparameter optimization
Maclaurin et al. '15



learned recurrent cell
Zoph & Le '17



few-shot image classifier
Vinyals et al. '16


learning to quickly navigate new mazes
Duan et al. '16

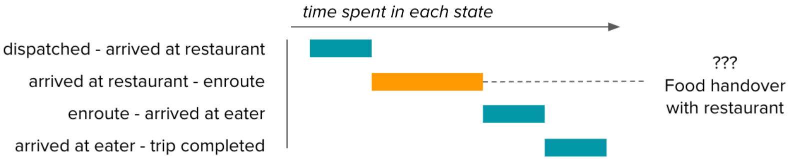
https://bair.berkeley.edu/blog/2017/07/18/learning-to-learn/

AI at Deutsche Bahn

- Deutsche Bahn is looking at replacing traditional algorithms for optimal scheduling and dispatching with RL techniques (InstaDeep, Laterre 2018)
- They also use Vision systems to manage service providers (clean snow), damage detection (schedule maintenance). object detection (left suitcases and people counting)

Source: https://www.dbsystel.de/dbsystel-en/digitalisation/ventures/zero-one-data/Video-analysis-3714252

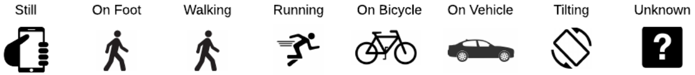
Uber

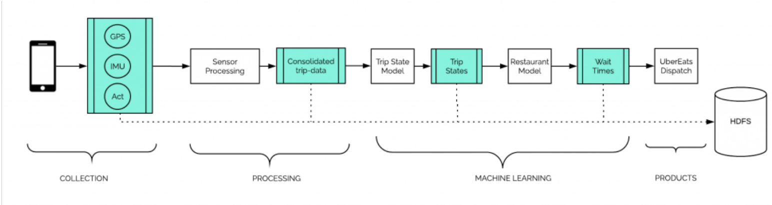


- Which restaurants have the longest parking times and why?
- How long does it take for the delivery-partner to walk to the restaurant?
- Does a restaurant have a difficult pickup process for delivery-partners?
- When should we dispatch the delivery-partner? Does the delivery-partner have to wait for the food? Is the food waiting for the delivery-partner?
- Did the delivery-partner have trouble with a delivery?

Uber

Android activity recognition

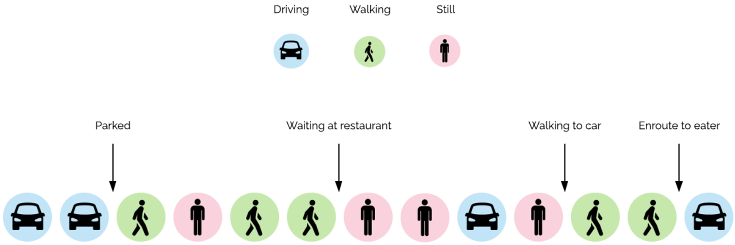




Raw data collection on phone to processing it as part of a batch pipeline.

Machine learning (ML) aimed at improving the product experience for the delivery-partner and eater by accurately calculating delivery times

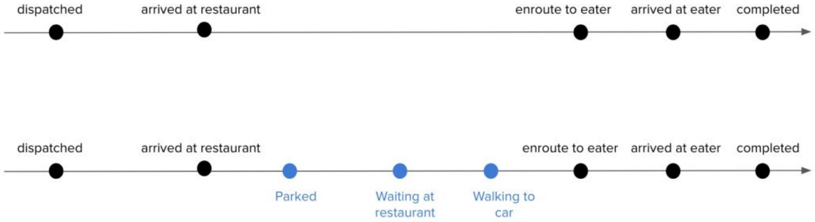
Uber



Sequence models are capable of finding the delivery-partner change-points amongst a sequence of state observations. These observations consist of activities or activities fused with other modalities, such as GPS and motion sensors

Uber

The additional detail achieved in our Uber Eats Trip State Model



The ML model gives an in-depth look at how an Uber Eats trip proceeds, allows to control dispatch time

Business Statistics: 41000

Enterprise AI

Vadim Sokolov

The University of Chicago Booth School of Business

http://vsokolov.org/courses/41000/

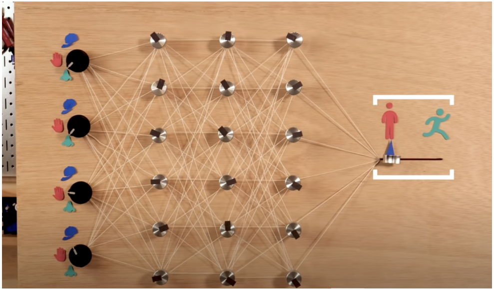
Sports Analytics

EPL

- Data experts are becoming football's best signings
- Analysis of Premier League Data
- How data analytics killed the Premier League's long ball game
- The Real Reason Behind Leicester City's 2015-16 Premier League Success

Stealing Baseball Signs

Machine Vision for sign-stealing



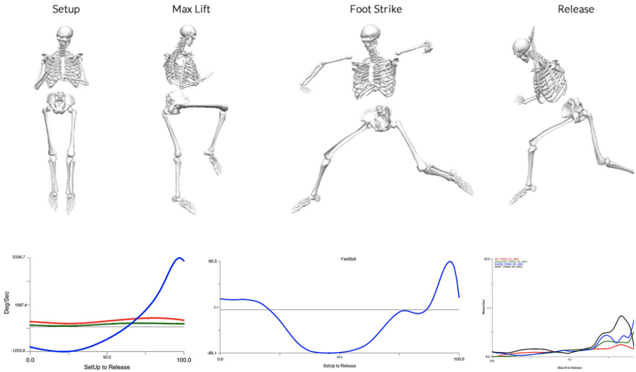
Like in the Astros sign-stealing scandal!

Machine learning is definitely the next competitive edge in MLB.

Teams that are able to classify limb movements to find the optimal swing or pitch mechanics will be able to more accurately identify high-potential talent.

YouTube Video

Kinatrax

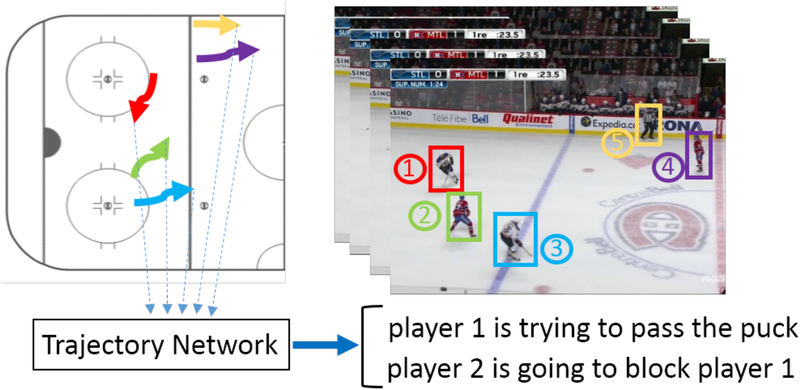


Neural network extracts data for biometric reports from images

Sporttechie, Chicago Cubs, SimiSystems, Automating socuts, The Unlikeliest No-Hitter

Sports Analytics: Learning Player Trajectories: NHL, NBA and EPL

Characteristics of group dynamics from their trajectories alone.



Homecourt Can Help You Improve Your Jump Shot

HomeCourt is a basketball training app that uses deep learning to record, track, and chart shots for basketball players in real time. Using NVIDIA Tesla P100 GPUs on the Google Cloud, with the cuDNN-accelerated TensorFlow and Keras deep learning framework, the team trained their neural network on hours of basketball footage they filmed in local Bay Area high schools.

Not only can you review their workout videos with instant stat analysis but also workouts from other players and engage, and interact with the broader basketball community through basketball on a device they have with them all the time.

Homecourt

- No sensors.

  HomeCourt is ready to use with iPhone for real-time shot tracking and analysis.

- Interactive drills.

  Your phone is transformed into your own virtual on-demand skills coach, with mobile AI-powered assessment tools.

- Shot Science.

  HomeCourt's proprietary Shot Science technology provides meaningful insights for every shot you take.

- Analyze every shot.

With data tracked from $>$10M shots, $>$9M dribbles from $>$150 countries.

Racehorse Big Data Unlocks the Formula for Human Superathletes

Jeff Seder had spent decades and millions of dollars collecting a huge database of physiological and biological data in an effort to discover which traits corresponded most closely with greatness. He pioneered portable ultrasound device that allowed him to examine horses on the inside and scanned tens of thousands of animals. It was only through this that he came to the conclusion that one of the most important data points in selecting a horse is the size of its heart – and American Pharoah had a huge one.

The wisdom of big data is increasingly being applied to human competition, although the field most certainly has room to grow. It will likely change not only how athletes perform, but how they look, as the tools and technologies of sports medicine become more and more sophisticated.

Artificial Intelligence in Formula 1

Formula One

- Strategy teams at the race track and at the team's HQ are constantly trying to predict the next best optimal move to improve their drivers' positions.
- Teams are limited to 60 data scientists AI (a.k.a machine learning/deep learning) provides better predictions of when best to stop, when to change tyres, overtake, ...
- Best strategies can vary quickly from moment to moment.

Formula One: Features

Data collected from 200 mile race, 70 laps, 300 sensors per car.

Millions of data points

What is the fastest way from start to finish. Use this variables to predict

- Weight effect
- Fuel loaf
- Fuel load time
- When to pit-stop
- Pit lane
- Tyre degradation

Used to be pre-race, now in-race (AWS)

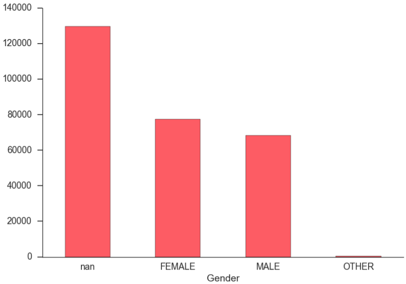
Monte-Carlo simulations of all cars and all traffic situations

Enterprise Applications

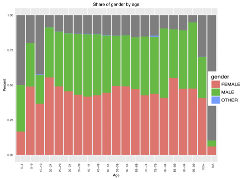
Airbnb Data

- Dataset from Airbnb Kaggle competition
- Predict in which country a new user will make his or her first booking
- 10 countries where users make frequent bookings: 12 classes (10 major countries + other + NDF)
- Data: users table + sessions of each user
- 213,451 users + 1,056,7737 individual sessions.

A lot of Missing Data

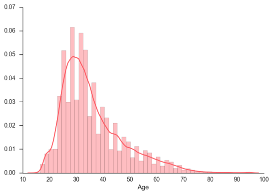


(a) Number of observations for each gender

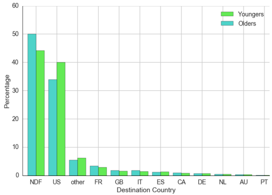


(c) Relationship between age and gender

Age is Important



(a) Distribution of user's age



(b) Destination by age category

(c) Destination by age group

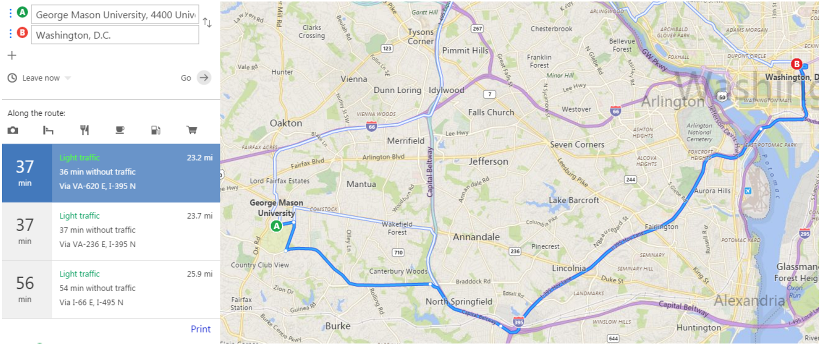
DL Model

- Two hidden dense layers and ReLU activation function $f(x) = \max(0, x)$.
- ADAGRAD optimization
- Evaluation metric
$$\mathrm{NDCG}_k = \frac{1}{n}\sum_{i=1}^{n} \mathrm{DCG}_5^i,$$
- 20 epochs and batch size of 256.
- 10% of the sample for evaluating the model

Traffic Prediction

Google maps real-time travel predictions ...

- Path search algorithms to calculate fastest route



When Iowa's snow piles up, TensorFlow can keep roads safe

To improve road safety and efficiency, the Iowa Department of Transportation has teamed up with researchers at Iowa State University to use machine learning, including our TensorFlow framework, to provide insights into traffic behavior. Iowa State's technology helps analyze the visual data gathered from stationary cameras and cameras mounted on snow plows.

They also capture traffic information using radar detectors. Machine learning transforms that data into conclusions about road conditions, like identifying congestion and getting first responders to the scenes of accidents faster..

In California, snow may not be an issue, but traffic certainly is, and college students there used TensorFlow to identify pot holes and dangerous road cracks in Los Angeles.

Tesla deploys massive new Autopilot neural net in v9

Based on the new capabilities of Autopilot under version 9, the new computer vision neural net had to be significantly updated.

It can now track vehicles and other objects all around the car – meaning that it makes better use of the 8 cameras around the car and not just the front-facing ones.

Scaling computational power, training data, and industrial resources plays to Tesla's strengths and involves less uncertainty than potentially more powerful but less mature techniques.

Automated Rotterdam Port

Port Rotterdam Port is one of the most automated ports and one of the largest ports in the world.

Automated container carriers are completely computer controlled, carrying containers to cranes. Meanwhile, the cranes are human controlled and move the containers to the ship.

With the fully automated cranes, the terminal can be run by a team of no more than 10 to 15 people on a day-to-day basis.

Automated Port: Port of Qingdao

Qingdao

- Port of Qingdao is the first automated container terminal in Asia. The terminal is called a "ghost port" since it is all controlled by AI and no workers found in sight.
- Through laser scanning and positioning, the program is able to locate the four corners of each container. It accurately grabs them and puts them onto the driverless trucks. And it is capable to work in complete darkness. The smart autopilot trucks, driven by electricity, have their routes and tasks under digital control. They even know when it's time to go for a recharge.

Rio Tinto Mining Automation to Boost Efficiency

Rio Tinto

- 73 self-driving trucks that reportedly haul payloads at a cost 15 percent less than those operated by human drivers. In addition to the trucks, they also have robotic, rock-drilling rigs plugging away at the topography. In the near future. Rio Tinto is looking to upgrade the trains that haul the ore to port to not only drive themselves but also have the ability to load and unload automatically.
- 15% reduction in the cost of operating the automated trucks compared to those driven by humans, as hauling is among the largest costs to a mining operation.

AI to Eliminate Bias from Hiring

Traditional hiring is biased.

- The deepest-rooted source of bias in AI is the human behavior it is simulating
- Current hiring process leads to significant unconscious bias against women, minorities and older workers.
- Often only applicants to be considered are those coming from Ivy League campuses, passive candidates from competitors (HBR article)

Can AI eliminate unconscious human bias and accelerate the hiring process?

Amazon's Hiring Tool and Discrimination

- After a year of use, Amazon decided to abandon AI-based hiring process
- "In 2015, the team realized that its creation was biased in favor of men when it came to hiring technical talent"
- The training sample was biased to begin with!

MONEYBOX

Amazon Created a Hiring Tool Using A.I. It Immediately Started Discriminating Against Women.

By JORDAN WEISSMANN — OCT 10, 2018 • 4:52 PM

Slate article

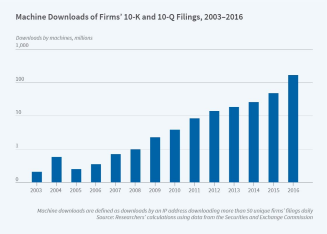
Goldman's Apple Card Algo Discriminates

- David Heinemeier Hansson the Ruby on Rails developer
- Railed on twitter against the Apple Card for giving him 20 times the credit limit that his wife got
- They filed joint tax returns and that his wife has a better credit score than he does
- Can we interpret the decision of the model?
- Is there are hidden features that explain the decision of the model?
- Regulator opens probe after sexism was alleged in viral tweets
- Steve Wozniak urges tougher regulation on credit algorithms

Bloomberg article

Corporate Reporting and AI

- Mechanical downloads of 10-K and 10-Q from 360K (2003) to 165M (2016)
- Companies adjust their language and reporting in order to achieve maximum impact with algorithms
- Companies expecting higher levels of machine readership prepare their disclosures in ways that are more readable by this audience
- Companies avoid words that are listed as negative in the directions given to algorithms



Original Article

Amazon Go Stores

- Fully automated. All of your actions are tracked using CNNs
- Items in your cart are charged to your credit card when you step out.

  In Chicago since 2019

Amazon Go Video

Beyond FAANG: a public option for A.I

Public program that provides universal access to goods and services, with a private opt-out

- pathway for start-ups and public-sector organizations to develop abilities and products that would compete with those of the tech giants
- public data pool that would make data accessible to registered users
- sensitive data would be highly regulated

Health care, transportation, energy and other areas could also benefit significantly from A.I. NYT Article

Beyond FAANG: From Agriculture to Art

- Medicine: diagnostics (image analysis), doctor's notes and chat bots (natural language processing); genomics and drug discovery
- Agriculture: remote sensing from satellites (image analysis), plant disease monitoring (image analysis)
- Heavy equipment: health monitoring, quality control
- Insurance: more accurate risk estimates
- Finance: Capital One has a deep learning group
- Art, fashion, digital advertisement, skin care, self driving cars, ...

NYT article

“Arms race" with Moscow’s intelligence agencies

REPORT

Zuckerberg: We’re in an ‘Arms Race’ With Russia, but AI Will Save Us

Buckle up — the technology won’t be ready for another decade.

BY ELIAS GROLL | APRIL 10, 2018, 7:57 PM

- Bad actors seeking are always changing their tactics
- Ahead of the Alabama election detected fake accounts in Macedonia
- Took down 30,000 fake accounts before France election in 2017

Foreign Policy article

AI is used on the other side of misinformation

How A.I. Could Be Weaponized to Spread Disinformation

- Campaign spread against the “White Helmets” in 2017: “an arm of Western governments to sow unrest in Syria"
- Russians posted the same text many times and gave itself away
- “As artificial intelligence becomes more powerful, experts worry that disinformation generated by A.I. could make an already complex problem bigger and even more difficult to solve.”

One of the statements below is an example from the disinformation campaign. A.I. technology created the other. Guess which one is A.I.:

The White Helmets alleged involvement in organ, child trafficking and staged events in Syria.	The White Helmets secretly videotaped the execution of a man and his 3 year old daughter in Aleppo, Syria.

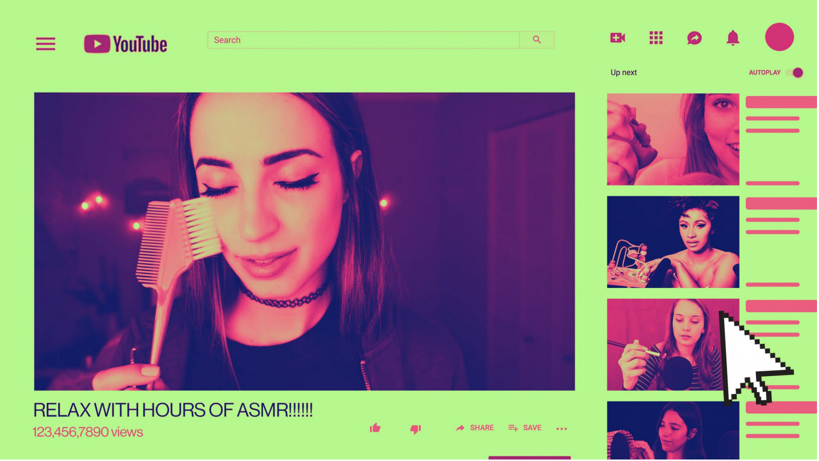
Pinterest: deep learning

One of the most popular ways people find ideas on Pinterest is through Related Pins, an item-to-item recommendations system that uses collaborative filtering.

Previously, candidates were generated using board co-occurrence, signals from all the boards a Pin is saved to. Now, for the first time, Pinterest is applying deep learning to make Related Pins even more relevant. Ultimately, they developed a scalable system that evolves with their product and people's interests, the most relevant recommendations can surface through Related Pins.

Pin2Vec is built to embed all the Pins in a 128-dimension space. Pin tuples are used in supervised training to train the embedding matrix for each of the tens of millions of Pins of the vocabulary. TensorFlow is used as the trainer. At serving time, a set of nearest neighbors are found as Related Pins in the space for each of the Pins.

Recommendation Systems: YouTube



YouTube is experimenting with ways to make its algorithm even more addictive

Recommendation Systems: YouTube

“Recommendation algorithms are some of the most powerful machine-learning systems today because of their ability to shape the information we consume”

Goal is to maximize amount of time spent watching

Confirmation bias: can create an addictive experience that shuts out other views.

Recommendation Systems: YouTube Algorithm

1. Compiles a shortlist of several hundred videos by finding ones that match the topic and other features of the one you are watching
2. Ranks the list according to the user's preferences, which it learns by feeding all your clicks, likes, and other interactions into a machine-learning algorithm

The effect is that over time, the system can push users further and further away from the videos they actually want to watch.

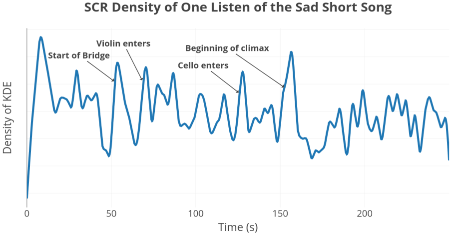
YouTube's algorithms created an isolated far-right community, pushed users toward videos of children, and promoted misinformation (online extremism)

AI for Personalized Music

Inputs are auditory features extracted from a composition (dynamics, timbre, harmony, rhythm). Total of 74 inputs

Outputs are brain activity + subjective description of the listener.

Used unpopular songs for training. To avoid confounding. Used equal number of "happy" and "sad songs"



Multimodal View into Music's Effect on Human Neural, Physiological, and Emotional Experience

MIT Technology Review Article

Data Team vs Business Experts

At Netflix, Who Wins When It's Hollywood vs. the Algorithm?

As the company plunges deeper into originals, its L.A. wing is doing the once-unthinkable: overriding the metrics

By Shalini Ramachandran and Joe Flint
Nov. 10, 2018 12:00 am ET

- The Netflix War: How to promote "Grace and Frankie"?
- Product team only included Ms. Fonda's co-star, Lily Tomlin.
- Tests showed that more users clicked on the show when the photo did not include Ms. Fonda.
- Could violate the contract!

Vulture's Inside the Binge Factory

Cindy Holland and Ted Sarandos, who are in charge of content at Netflix: It's 70 percent gut and 30 percent data

AI can predict if you'll die soon – but we've no idea how it works

- 1.77 million electrocardiograms (ECG) from 400,000 people
- DL extracts features from the time series. Better features that doctors came up with.
- No matter what, the voltage-based model was always better than any model you could build out of things that we already measure from an ECG
- The AI accurately predicted risk of death even in people deemed by cardiologists to have a normal ECG

Still unclear what patterns the AI is picking up which makes some physicians reluctant to use such algorithms

NewScientist

Explainable AI

Open.ai

In policy applications interpretability is important.

Cost of a mistake is high.

Explainable AI “A good deep learning approach could give us more comfort that we know what’s happening in the system than having 1000 of these human-created rules, created over decades.”

AGI = Artificial General Intelligence

SpaceX Landing

SpaceX lands rocket at sea, makes history

More Booth Stats!

Lots more to do:

41201: Big Data

41204: Machine Learning

41913: Bayes, AI and Deep Learning

HAVE FUN!!!