quote

Machine learning basics

Grégory Châtel

Disaitek
Intel Software Innovator

@rodgzilla
github.com/rodgzilla

2019/02/12

Machine learning

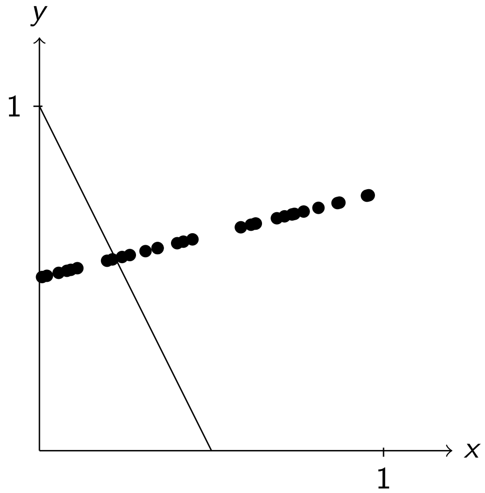
Machine learning (ML) is a subfield of artificial intelligence.

Intuitively We want to *learn from* and *make predictions on* data.

Technically We want to update the parameters of a model to make it describe our training data as well as possible ("well" being defined by a *loss function*).

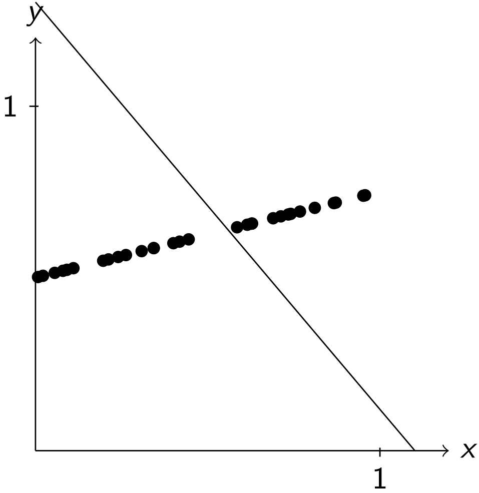
Model example

Linear regression



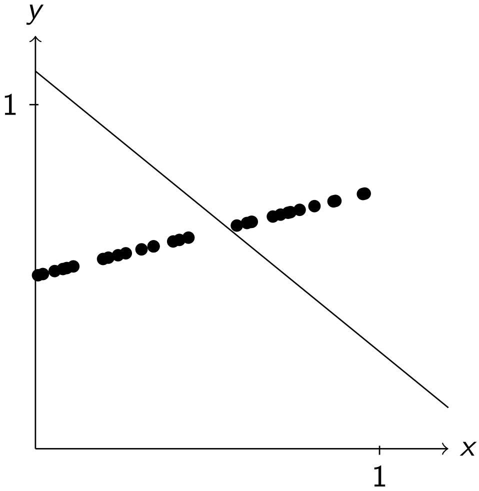
Model example

Linear regression



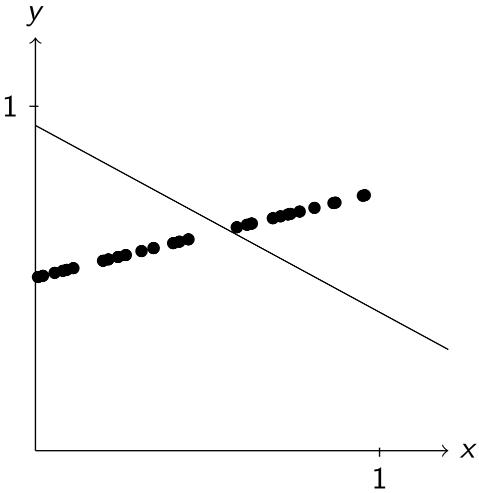
Model example

Linear regression



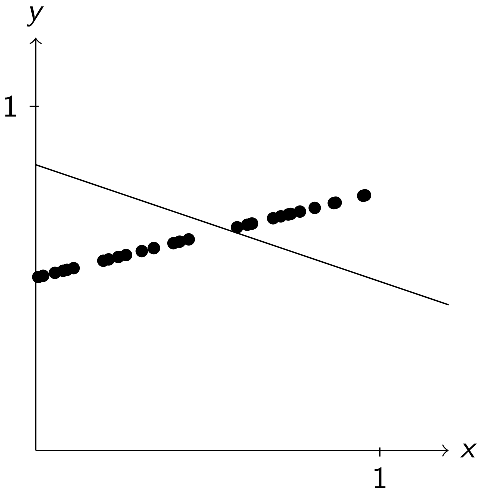
Model example

Linear regression



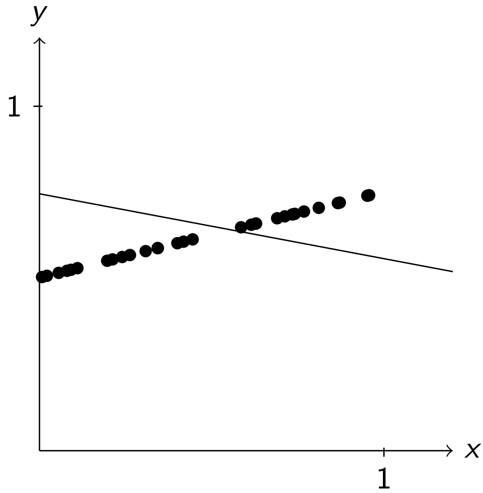
Model example

Linear regression



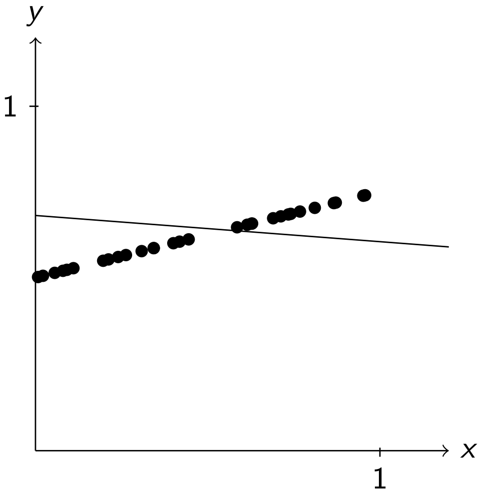
Model example

Linear regression



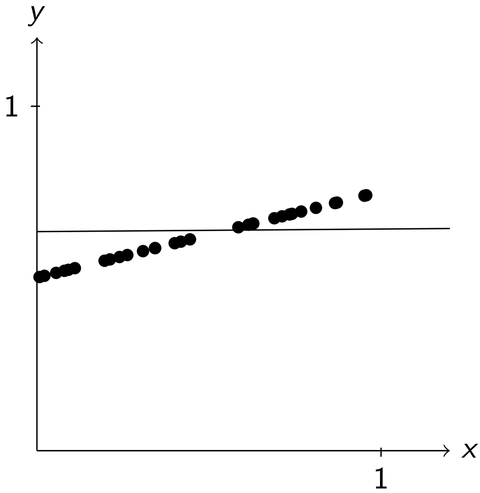
Model example

Linear regression



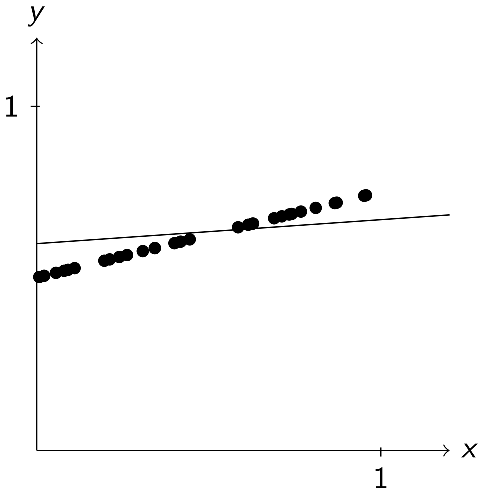
Model example

Linear regression



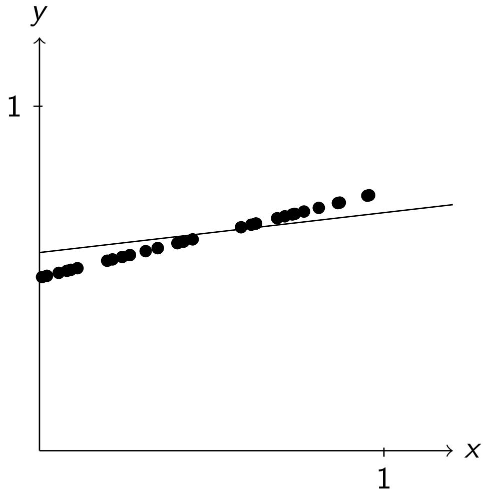
Model example

Linear regression



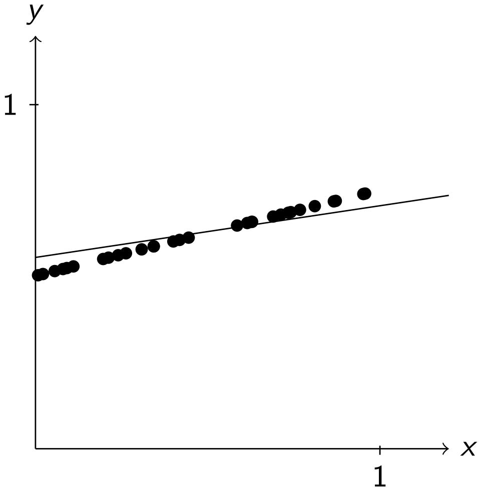
Model example

Linear regression



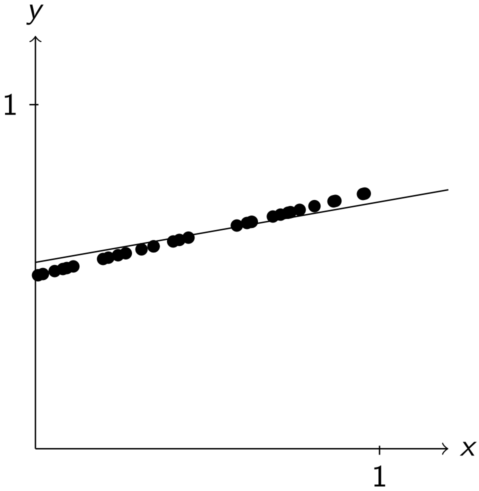
Model example

Linear regression



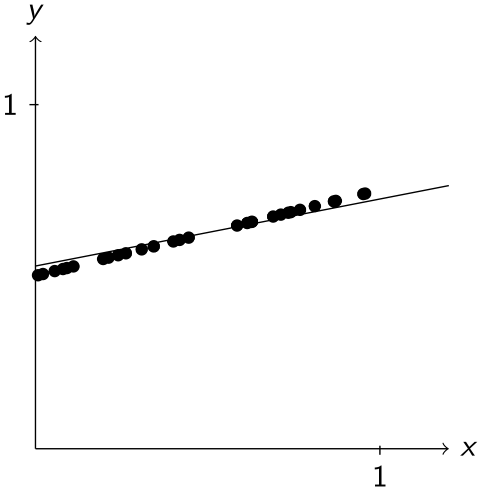
Model example

Linear regression



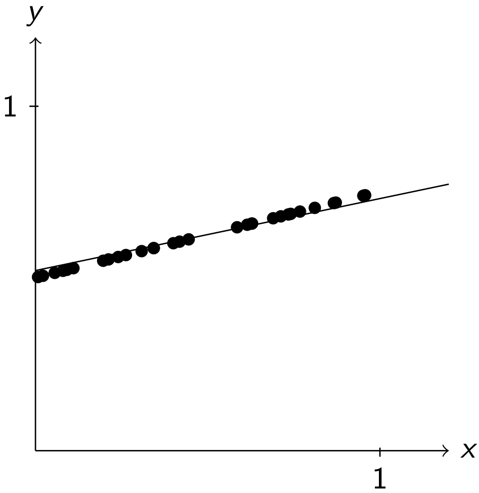
Model example

Linear regression



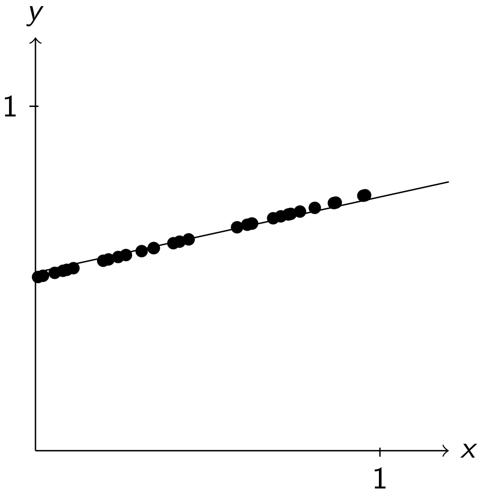
Model example

Linear regression



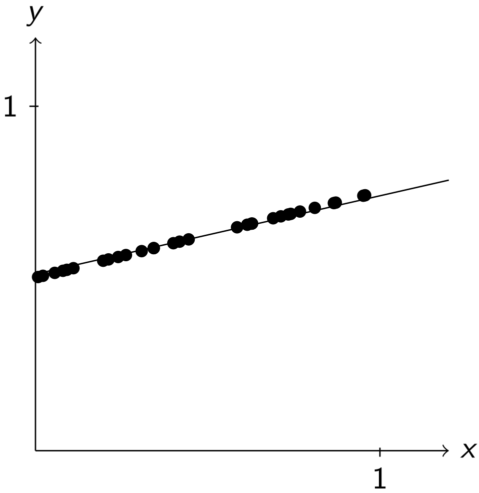
Model example

Linear regression



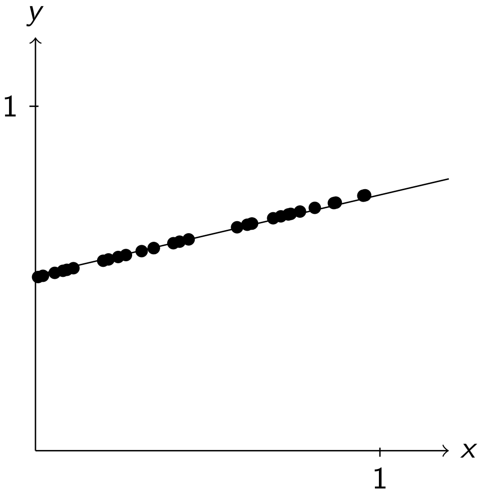
Model example

Linear regression



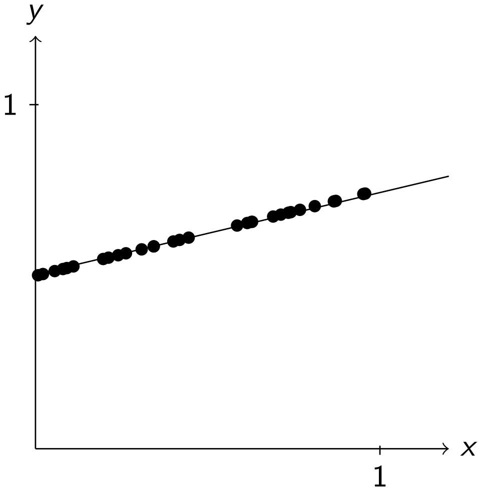
Model example

Linear regression



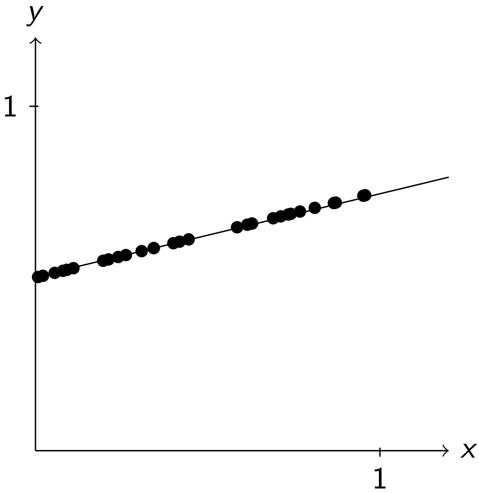
Model example

Linear regression



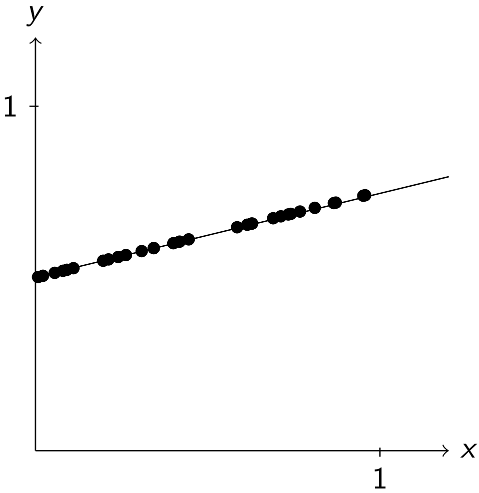
Model example

Linear regression



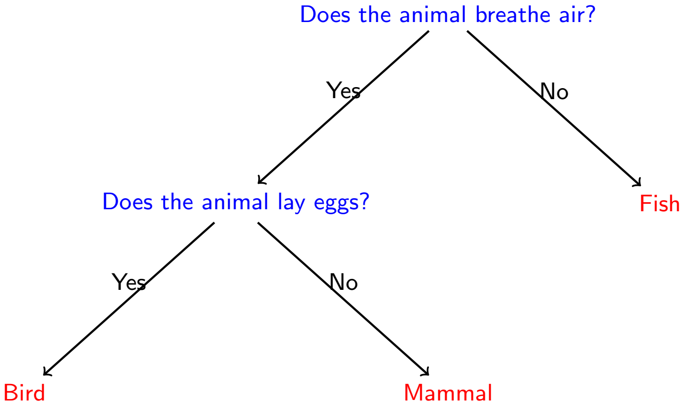
Model example

Linear regression



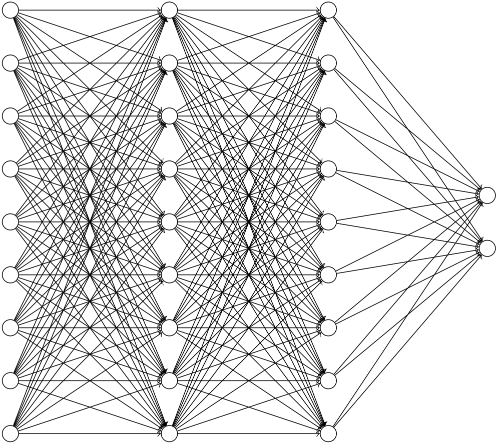
Model example

Decision tree



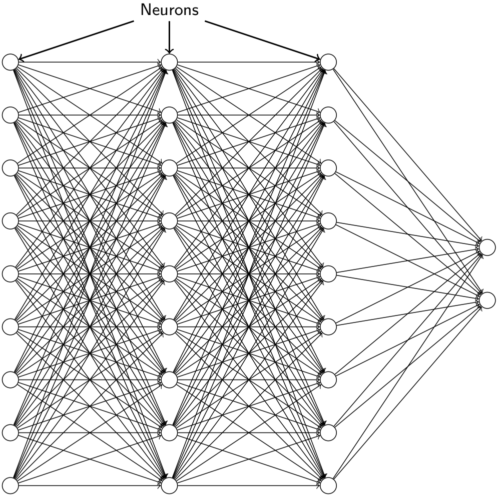
Model example

Neural network (deep learning)



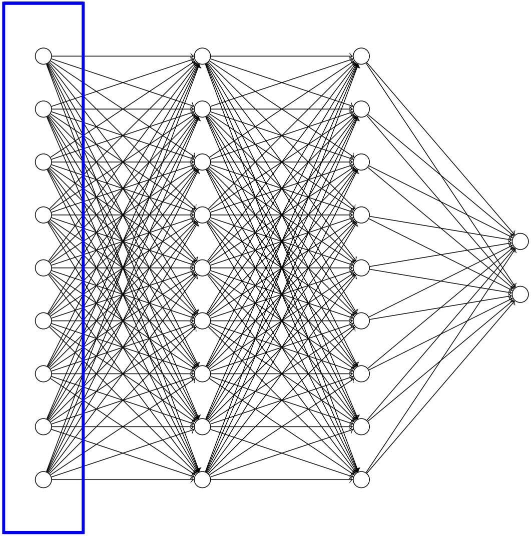
Model example

Neural network (deep learning)



Model example

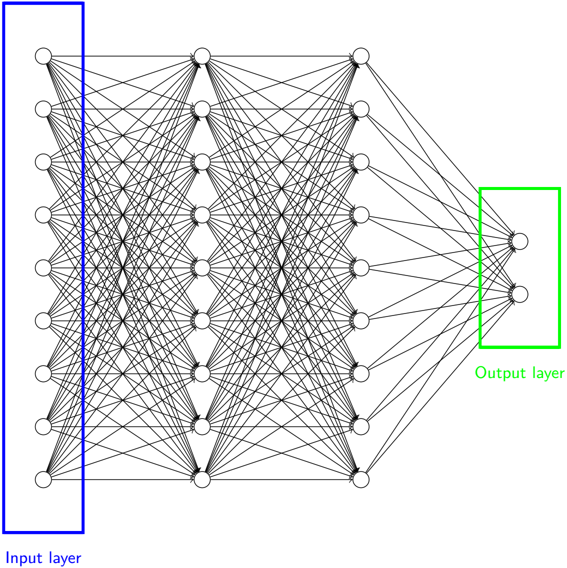
Neural network (deep learning)



Input layer

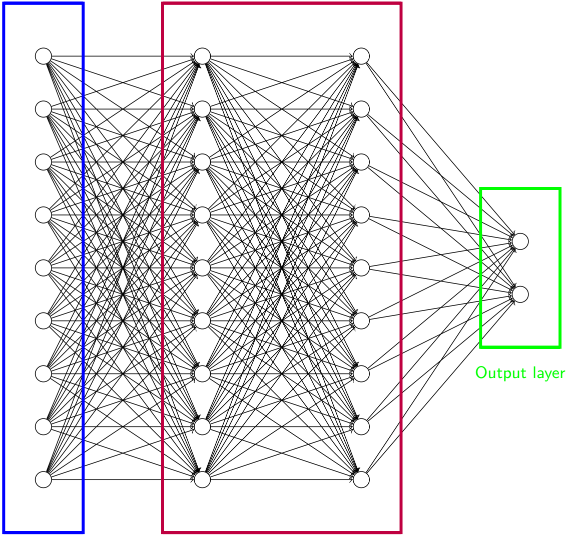
Model example

Neural network (deep learning)



Model example

Neural network (deep learning)



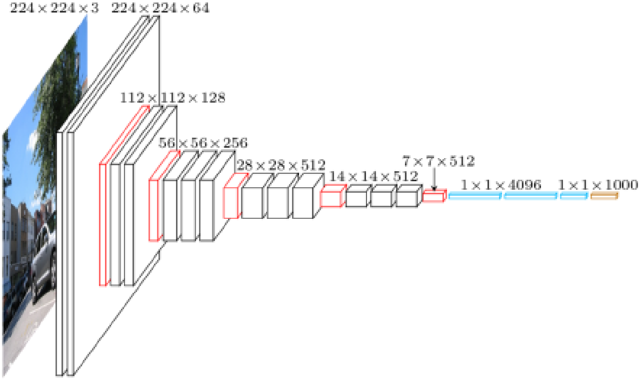
Output layer

Input layer

Hidden layers

Deep learning architecture

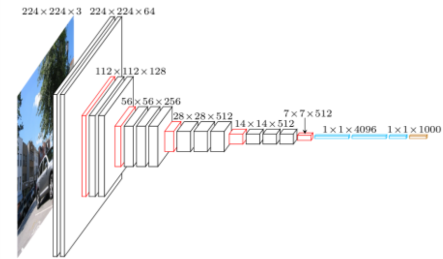
Image recognition (VGG 16)

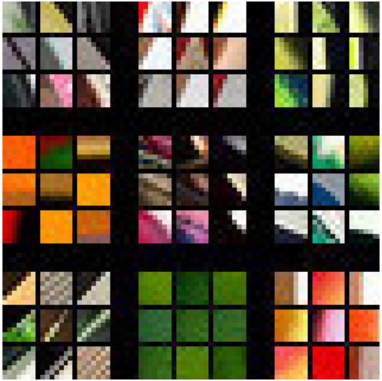

$224 \times 224 \times 3$
$224 \times 224 \times 64$
$112 \times 112 \times 128$
$56 \times 56 \times 256$
$28 \times 28 \times 512$
$14 \times 14 \times 512$
$7 \times 7 \times 512$
$1 \times 1 \times 4096$
$1 \times 1 \times 1000$

Deep learning architecture

Hierarchized pattern recognition

Layer 1

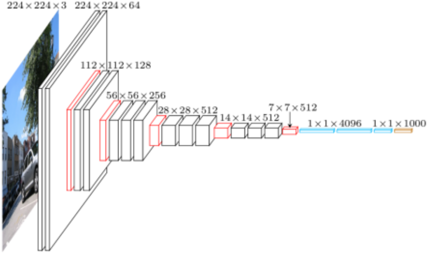


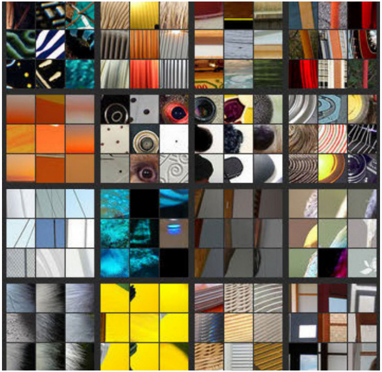


Deep learning architecture

Hierarchized pattern recognition

Layer 2

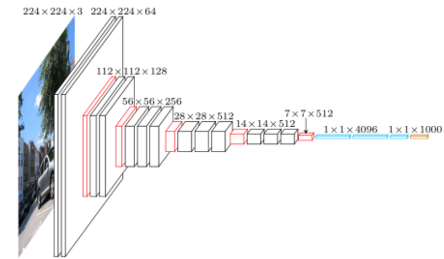


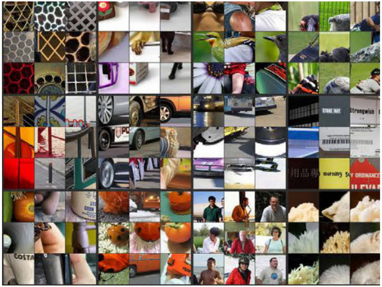


Deep learning architecture

Hierarchized pattern recognition

Layer 3

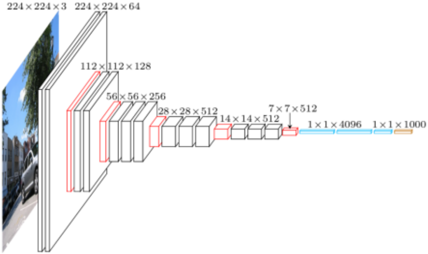


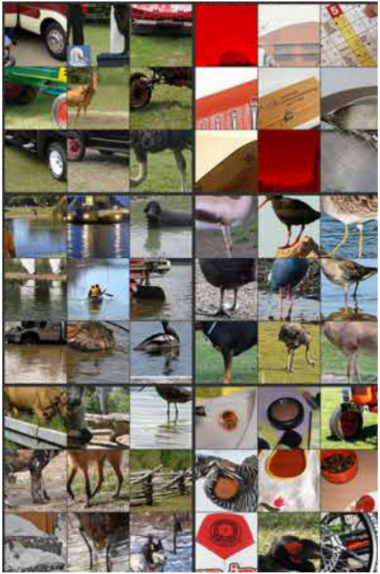


Deep learning architecture

Hierarchized pattern recognition

Layer 4

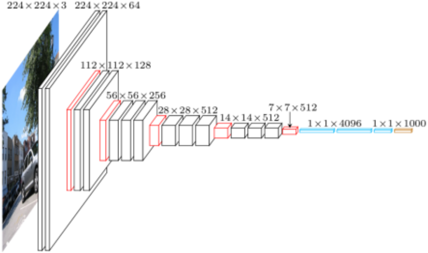


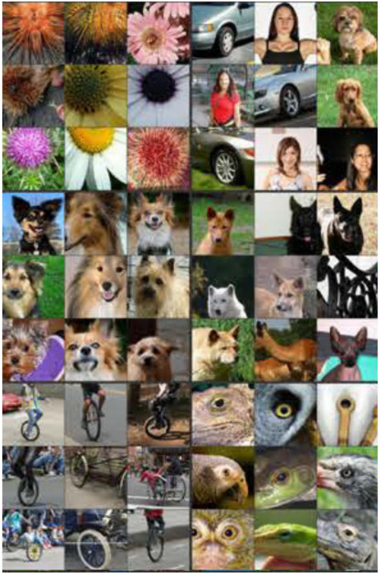


Deep learning architecture

Hierarchized pattern recognition

Layer 5





Application examples

Supervised learning

- Supervised tasks
  - Regression

Recommender system	(user, book) → rating
House price	(surface, nb rooms, city) → price

  - Classification

Image classification	pixel values → cat or dog
Text classification	list of words → spam or valid email

- Unsupervised tasks
  - Clustering

    Group clients by interests

  - Anomaly detection

    Credit card fraud detection

Deep Natural Language Processing (NLP)

Main ideas

- Learning the semantic meaning of words,

Deep Natural Language Processing (NLP)

Main ideas

- Learning the semantic meaning of words,
- Understanding the information hierarchy related to the task at hand,

Deep Natural Language Processing (NLP)

Main ideas

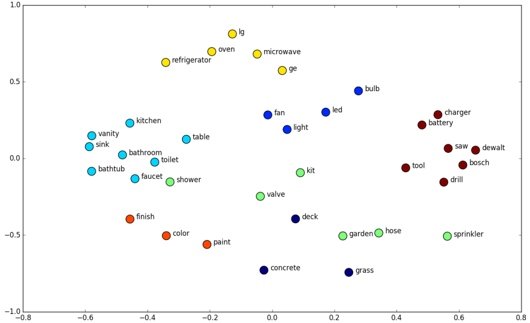
- Learning the semantic meaning of words,
- Understanding the information hierarchy related to the task at hand,
- Ability to make use of huge amounts of data.

Word embeddings

Semantic vectors

We associate to each word of the vocabulary a vector which represents its meaning.

$$\begin{aligned} \text{Oven} &\ [-0.2, 0.6] \\ \text{Microwave} &\ [-0.05, 0.57] \\ \text{Garden} &\ [0.22, -0.5] \end{aligned}$$

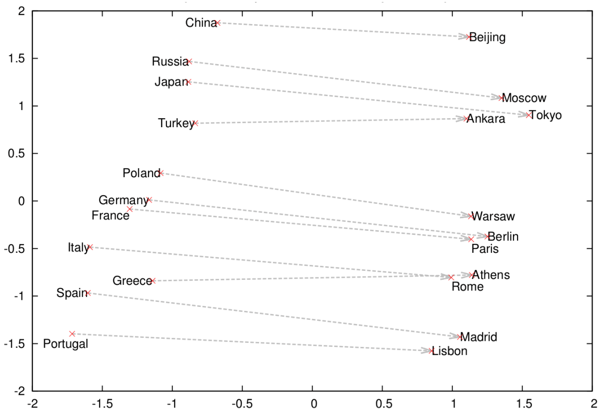


In real applications word embedding have 100 to 300 values, encoding all kind of characteristics about words.

Word embeddings

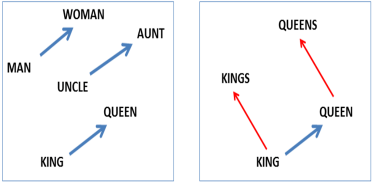
Links between concepts

When word embeddings are created using a large enough dataset, a lot of information is encoded in differences between vectors.

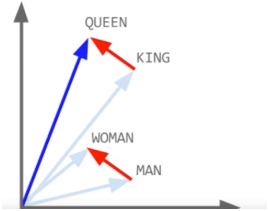


Word embeddings

Vector geometry

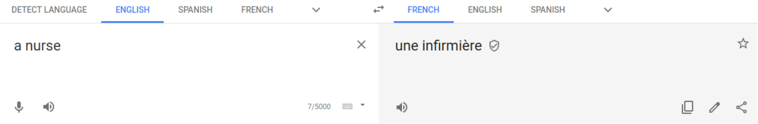


$$king - man + woman = queen$$



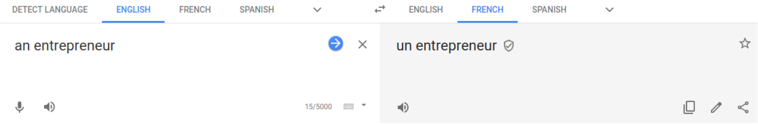
Word embeddings

Bias in representations



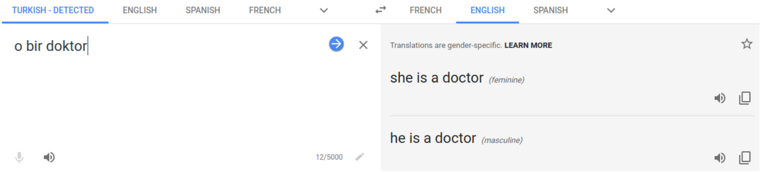
a nurse

une infirmière



an entrepreneur

un entrepreneur



o bir doktor

Translations are gender-specific. **LEARN MORE**

she is a doctor *(feminine)*

he is a doctor *(masculine)*

NLP tasks

Sentiment analysis

Automatized analysis of an item public perception:

- Negative
  - Even fans of Ismail Merchant's work, I suspect, would have a hard time sitting through this one.
  - Every conceivable mistake a director could make in filming opera has been perpetrated here.
  - Cheap, vulgar dialogue and a plot that crawls along at a snail's pace.
  - The material and the production itself are little more than routine.
- Positive
  - A rare and lightly entertaining look behind the curtain that separates comics from the people laughing in the crowd.
  - Rarely, indeed almost never, is such high-wattage brainpower coupled with pitch-perfect acting and an exquisite, unfakable sense of cinema.
  - Easily the most thoughtful fictional examination of the root causes of anti-Semitism ever seen on screen.

NLP tasks

Document tagging

Automatic tagging of documents, articles or books.

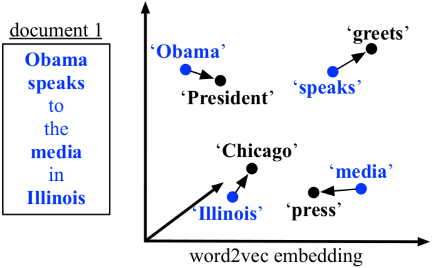
- Supervised way using classification (using past labels):
  - Harry Potter: Child book, Fantasy, Aventure, . . .
  - Lord Of The Rings: Fantasy, Aventure, . . .
  - Algorithms To Live By: Computer science, Textbook, . . .
- Unsupervised way using clustering (grouping books that looks the same):
  - Cluster 1: Harry potter, Lord Of The Rings, . . .
  - Cluster 2: Algorithms To Live By, The Art of Computer Programming, . . .
  - Cluster 3: Tofu from Scratch, Okinawa Diet

NLP tasks

Search engine

By using the natural language understanding capabilities of deep learning models, we can create more robust and performant search engines.

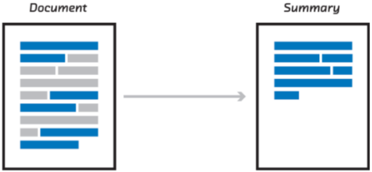
The matching performed by these search engines is semantic (*meaning* of the query) instead of lexical (finding *exactly* the word of the query in documents).



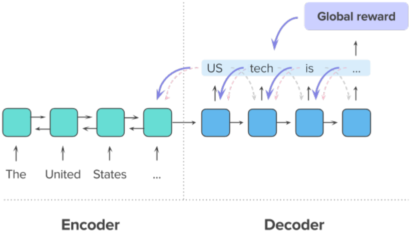
NLP tasks

Automatic summarization

Extractive summarization (copy-paste most important sentences)



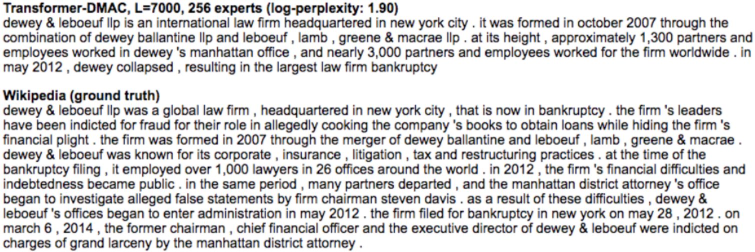
Abstractive summarization (generate new sentences that synthesize information)



NLP tasks

Abstractive summarization using wikipedia

The goal is to generate the abstract (first few paragraphs) of Wikipedia articles from their source documents.

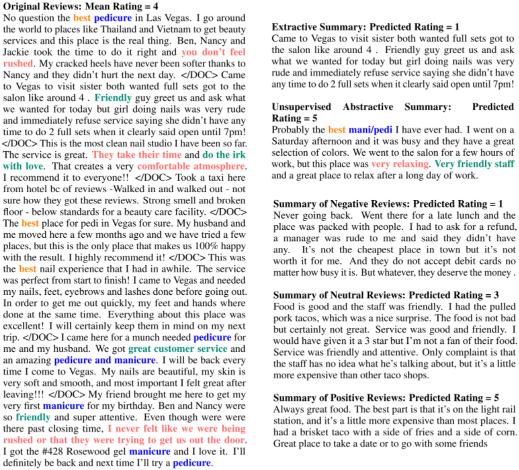


Transformer-DMAC, L=7000, 256 experts (log-perplexity: 1.90)
dewey & leboeuf llp is an international law firm headquartered in new york city . it was formed in october 2007 through the combination of dewey ballantine llp and leboeuf , lamb , greene & macrae llp . at its height , approximately 1,300 partners and employees worked in dewey 's manhattan office , and nearly 3,000 partners and employees worked for the firm worldwide . in may 2012 , dewey collapsed , resulting in the largest law firm bankruptcy

Wikipedia (ground truth)
dewey & leboeuf llp was a global law firm , headquartered in new york city , that is now in bankruptcy . the firm 's leaders have been indicted for fraud for their role in allegedly cooking the company 's books to obtain loans while hiding the firm 's financial plight . the firm was formed in 2007 through the merger of dewey ballantine and leboeuf , lamb , greene & macrae . dewey & leboeuf was known for its corporate , insurance , litigation , tax and restructuring practices . at the time of the bankruptcy filing , it employed over 1,000 lawyers in 26 offices around the world . in 2012 , the firm 's financial difficulties and indebtedness became public . in the same period , many partners departed , and the manhattan district attorney 's office began to investigate alleged false statements by firm chairman steven davis . as a result of these difficulties , dewey & leboeuf 's offices began to enter administration in may 2012 . the firm filed for bankruptcy in new york on may 28 , 2012 . on march 6 , 2014 , the former chairman , chief financial officer and the executive director of dewey & leboeuf were indicted on charges of grand larceny by the manhattan district attorney .

NLP tasks

Unsupervised abstractive summarization



Original Reviews: Mean Rating = 4
No question the **best** **pedicure** in Las Vegas. I go around the world to places like Thailand and Vietnam to get beauty services and this place is the real thing. Ben, Nancy and Jackie took the time to do it right and **you don't feel rushed**. My cracked heels have never been softer thanks to Nancy and they didn't hurt the next day. </DOC> Came to Vegas to visit sister both wanted full sets got to the salon like around 4 . **Friendly** guy greet us and ask what we wanted for today but girl doing nails was very rude and immediately refuse service saying she didn't have any time to do 2 full sets when it clearly said open until 7pm! </DOC> This is the most clean nail studio I have been so far. The service is great. **They take their time** and **do the irk with love**. That creates a very **comfortable atmosphere**. I recommend it to everyone!! </DOC> Took a taxi here from hotel bc of reviews -Walked in and walked out - not sure how they got these reviews. Strong smell and broken floor - below standards for a beauty care facility. </DOC> The **best** place for pedi in Vegas for sure. My husband and me moved here a few months ago and we have tried a few places, but this is the only place that makes us 100% happy with the result. I highly recommend it! </DOC> This was the **best** nail experience that I had in awhile. The service was perfect from start to finish! I came to Vegas and needed my nails, feet, eyebrows and lashes done before going out. In order to get me out quickly, my feet and hands where done at the same time. Everything about this place was excellent! I will certainly keep them in mind on my next trip. </DOC> I came here for a munch needed **pedicure** for me and my husband. We got **great customer service** and an amazing **pedicure and manicure**. I will be back every time I come to Vegas. My nails are beautiful, my skin is very soft and smooth, and most important I felt great after leaving!!! </DOC> My friend brought me here to get my very first **manicure** for my birthday. Ben and Nancy were so **friendly** and super attentive. Even though were were there past closing time, **I never felt like we were being rushed or that they were trying to get us out the door**. I got the #428 Rosewood gel **manicure** and I love it. I'll definitely be back and next time I'll try a **pedicure**.

Extractive Summary: Predicted Rating = 1
Came to Vegas to visit sister both wanted full sets got to the salon like around 4 . Friendly guy greet us and ask what we wanted for today but girl doing nails was very rude and immediately refuse service saying she didn't have any time to do 2 full sets when it clearly said open until 7pm!

Unsupervised Abstractive Summary: Predicted Rating = 5
Probably the **best** **mani/pedi** I have ever had. I went on a Saturday afternoon and it was busy and they have a great selection of colors. We went to the salon for a few hours of work, but this place was **very relaxing**. **Very friendly staff** and a great place to relax after a long day of work.

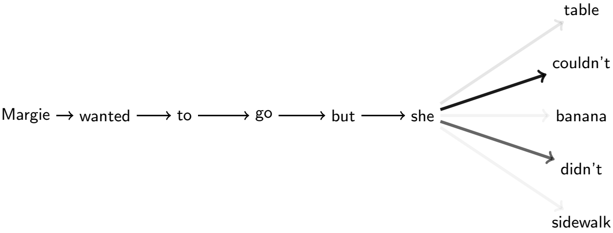
Summary of Negative Reviews: Predicted Rating = 1
Never going back. Went there for a late lunch and the place was packed with people. I had to ask for a refund, a manager was rude to me and said they didn't have any. It's not the cheapest place in town but it's not worth it for me. And they do not accept debit cards no matter how busy it is. But whatever, they deserve the money .

Summary of Neutral Reviews: Predicted Rating = 3
Food is good and the staff was friendly. I had the pulled pork tacos, which was a nice surprise. The food is not bad but certainly not great. Service was good and friendly. I would have given it a 3 star but I'm not a fan of their food. Service was friendly and attentive. Only complaint is that the staff has no idea what he's talking about, but it's a little more expensive than other taco shops.

Summary of Positive Reviews: Predicted Rating = 5
Always great food. The best part is that it's on the light rail station, and it's a little more expensive than most places. I had a brisket taco with a side of fries and a side of corn. Great place to take a date or to go with some friends

Language modeling

The goal of language modeling is to predict the word that is most likely to appear after a given sequence of words.



- Allows to learn the syntax of a language.
- Allows to learn the semantic of the words.
- Unlimited amount of data.
- Can be trained in multilingual setting.

The idea is to pretrain a model using this task and then use what it has learned to perform some other tasks.

Google BERT

The **B**idirectional **E**ncoder **R**epresentations for **T**ransformers model is a language model that has been trained on a massive corpus (7000 books + all Wikipedia pages for 102 languages). It is used as a base to perform many other linguistic tasks.

It allows user to work easily in a wide variety of languages.

It produces new state of the art results on 11 NLP tasks.

It is very fast and inexpensive to transfer on a new task.

A multilingual BERT model trained on a monolingual dataset will work (although with lower performances than with a proper training) on another language (zero-shot learning).

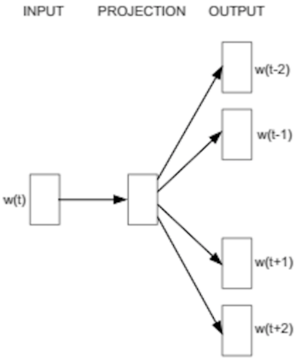
References

- IMDB movie review classification: Github repo for IMDB sentiment analysis with GPT
- Convolution layer visualization: matthewzeiler.com
- Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." Advances in neural information processing systems. 2013.
- Google blog post about gender bias in Google Translate: Reducing gender bias in Google Translate (2018)
- Chu, Eric, and Peter J. Liu. "Unsupervised Neural Multi-document Abstractive Summarization." arXiv preprint arXiv:1810.05739 (2018).
- Liu, Peter J., et al. "Generating wikipedia by summarizing long sequences." arXiv preprint arXiv:1801.10198 (2018).
- Devlin, Jacob, et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv preprint arXiv:1810.04805 (2018).

- Github repo for these slides.

Word2vec algorithm

The goal of the word2vec task is to predict the context of word (the words surrounding it) based on its vector representation.



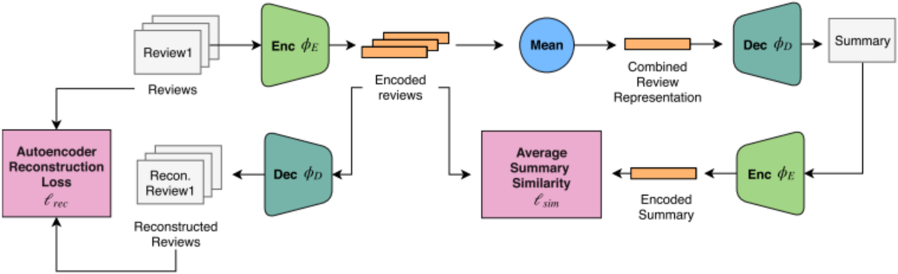
GDPR on model bias and discrimination

The following paragraph is an excerpt of the GDPR (from this article):

> *In order to ensure fair and transparent processing in respect of the data subject, taking into account the specific circumstances and context in which the personal data are processed, the controller should use appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures appropriate to ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimised, secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subject, and prevent, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or processing that results in measures having such an effect.*

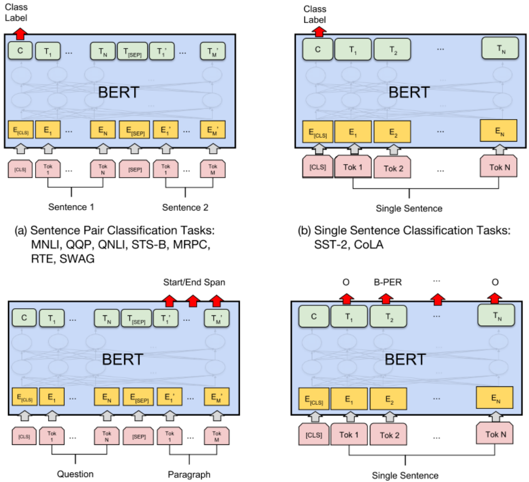
Unsupervised summarization model

The abstractive summary is generated by decoding the mean representation of the input documents.



Google BERT for NLP tasks

To perform other tasks with BERT, you need to format your input and objective according to the way described in the article.



(a) Sentence Pair Classification Tasks: MNLI, QQP, QNLI, STS-B, MRPC, RTE, SWAG

(b) Single Sentence Classification Tasks: SST-2, CoLA

(c) Question Answering Tasks: SQuAD v1.1

(d) Single Sentence Tagging Tasks: CoNLL-2003 NER