Deep learning algorithms for automatic classification of electronic medical records starting from free text

Corrado Lanera*[1], Elisa Barbieri[2], Gianluca Piras[1], Arjun Maggie[3], Davy Weissenbacher[3], Daniele Dona[2], Luigi Cantarutti[4], Graciela Gonzalez[3], Carlo Giaquinto[2,4], and Dario Gregori[1]

2019/10/22 - @University of Torino - Department of Clinical and Biological Sciences

*corrado.lanera@unipd.it

Biomedicine Data Mining and Artificial Intelligence Lab

[1] Unit of Biostatistics, Epidemiology, and Public Health of the Department of Cardiac-Thoracic-Vascular Sciences and Public Health – University of Padova

[2] Division of Pediatric Infectious Diseases, Department for Woman and Child Health, University of Padua, Padua, Italy

[3] Health Language Processing Center, Institute for Biomedical Informatics of the Perelman School of Medicine – University of Pennsylvania

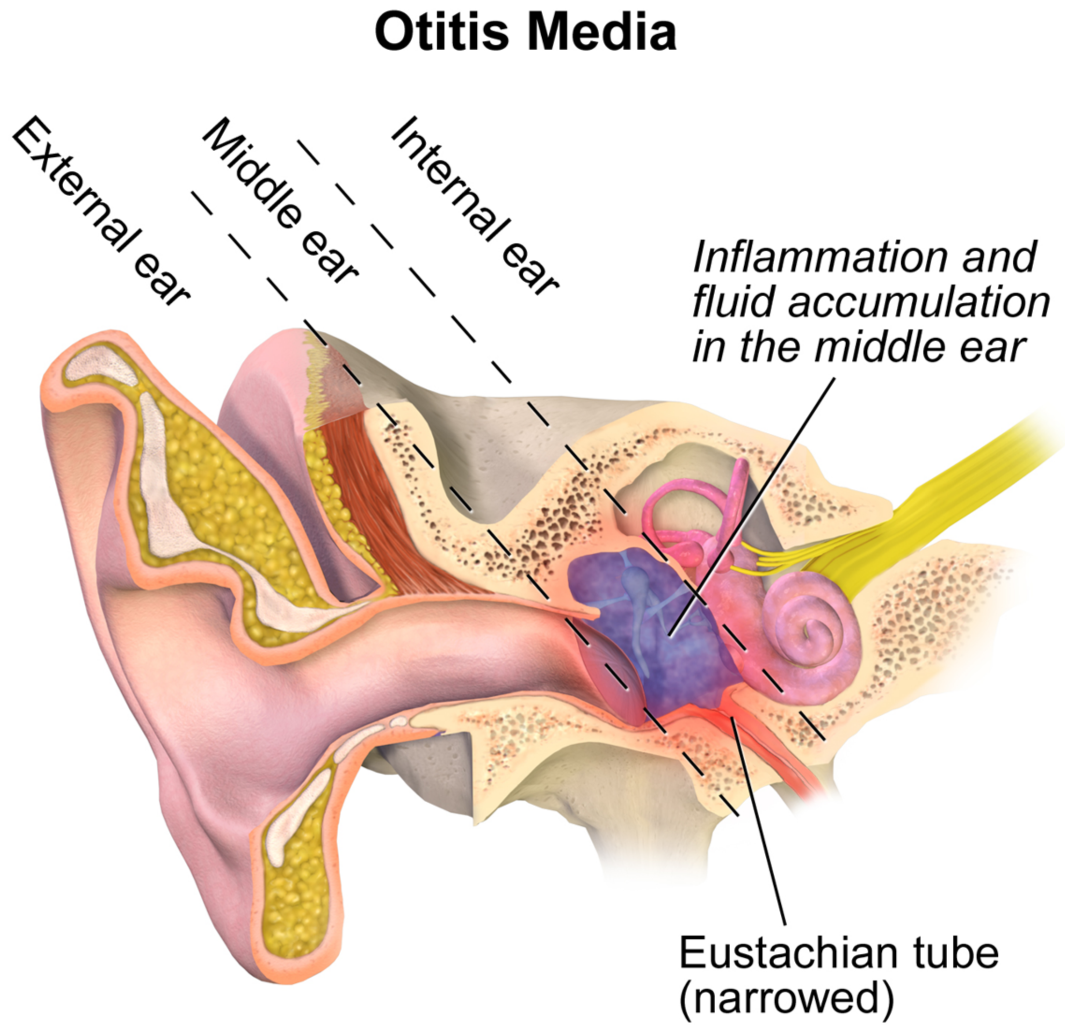
[4] Società Servizi Telematici – Pedianet

Outline

1. **Overview**: otitis
2. **Database and Task**: PEDIANET
3. **Programming paradigms**: classical and machine learning
4. **Classical text classification** search string and regular expression
5. **Shallow machine learning**: document-text matrix
6. **Neural networks**: basics
7. **Embeddings**: dense representation of text
8. **Deep networks**: convolutions, sequential/recurrent
9. **Task & metrics**: multi-class, scores and human-performances
10. **Analyses plan**: flowchart
11. **Our approach**: architectures explored
12. **Results, final remarks and conclusions**

Overview
classification of otitis

- one of the most common infections in pediatrics
- the main cause of antibiotic prescriptions
- challenging diagnosis
- frequently, little attention is paid to the guidelines
- continuing interest in defining the incidence and burden of AOM



https://commons.wikimedia.org/wiki/File:Otitis_Media.png

*Marchisio P. et al. "Burden of acute otitis media in primary care pediatrics in Italy: a secondary data analysis from the Pedianet database", **BMC Pediatrics** 2012.

*Spiro DM, Arnold, DH. "The concept and practice of a wait-and-see approach to acute otitis media.", **Current Opinion in Pediatrics** 2008

Database and Task

PEDIANET

A. Kao, S. Poteet (2007) **Text Mining: the discovery and the extraction of interesting, non-trivial knowledge from free or unstructured text.**

snapshot considerd: from 1st January 2004 to 23rd August 2017

records: $6,903,035$ ($297,373$ filtered by a search string)

pediatricians: 144 (troughout Italy)

children: $216,976$

fields (all free-text, Italian-language):

- diagnosis
- signs-and-symptoms
- diary
- prescription
- visit description
- visit result

2019 PEDIANET investigation

- data from 2010 to 2015
- **on primary diagnosis only**

Adding even the diaries in a traditional manual human-driven analysis proved to be **too costly** in terms both of person-time and economic resources

Antibiotic prescriptions in acute otitis media and pharyngitis in Italian pediatric outpatients

E. Barbieri, D. Donà, A. Cantarutti, R. Lundin, A. Scamarcia, G. Corrao, L. Cantarutti & C. Giaquinto

Italian Journal of Pediatrics 45, Article number: 103 (2019) | Download Citation

Abstract

Background

Acute otitis media (AOM) and pharyngitis are very common infections in children and adolescents. Italy is one of the European countries with the highest rate of antibiotic prescriptions. The aim of this study is to describe first-line treatment approaches for AOM and pharyngitis in primary care settings in Italy over six years, including the prevalence of 'wait and see' for AOM, where prescription of antibiotics is delayed 48 h from presentation, and differences in prescribing for pharyngitis when diagnostic tests are used.

Methods

The study is a secondary data analysis using Pedianet, a database including data at outpatient level from children aged 0–14 in Italy. Prescriptions per antibiotic group, per age group and per calendar year were described as percentages. "Wait and see" approach rate was described for AOM and pharyngitis prescriptions were further grouped according to the diagnostic test performed and test results.

Results

We identified 120,338 children followed by 125 family pediatricians between January 2010 and December 2015 for a total of 923,780 person-years of follow-up. Among them 30,394 (mean age 44 months) had at least one AOM diagnosis (n = 54,943) and 52,341 (mean age 5 years) had at least one pharyngitis diagnosis (n = 126,098). 82.5% of AOM diagnoses were treated with an antibiotic within 48 h (mainly amoxicillin and amoxicillin/clavulanate) and the "wait and see" approach was adopted only in 17.5% of cases. The trend over time shows an increase in broad spectrum antibiotic prescriptions in the last year (2015). 79,620 (63%) cases of pharyngitis were treated and among GABHS pharyngitis confirmed by rapid test 56% were treated with amoxicillin. The ones not test confirmed were treated mainly with broad spectrum antibiotics.

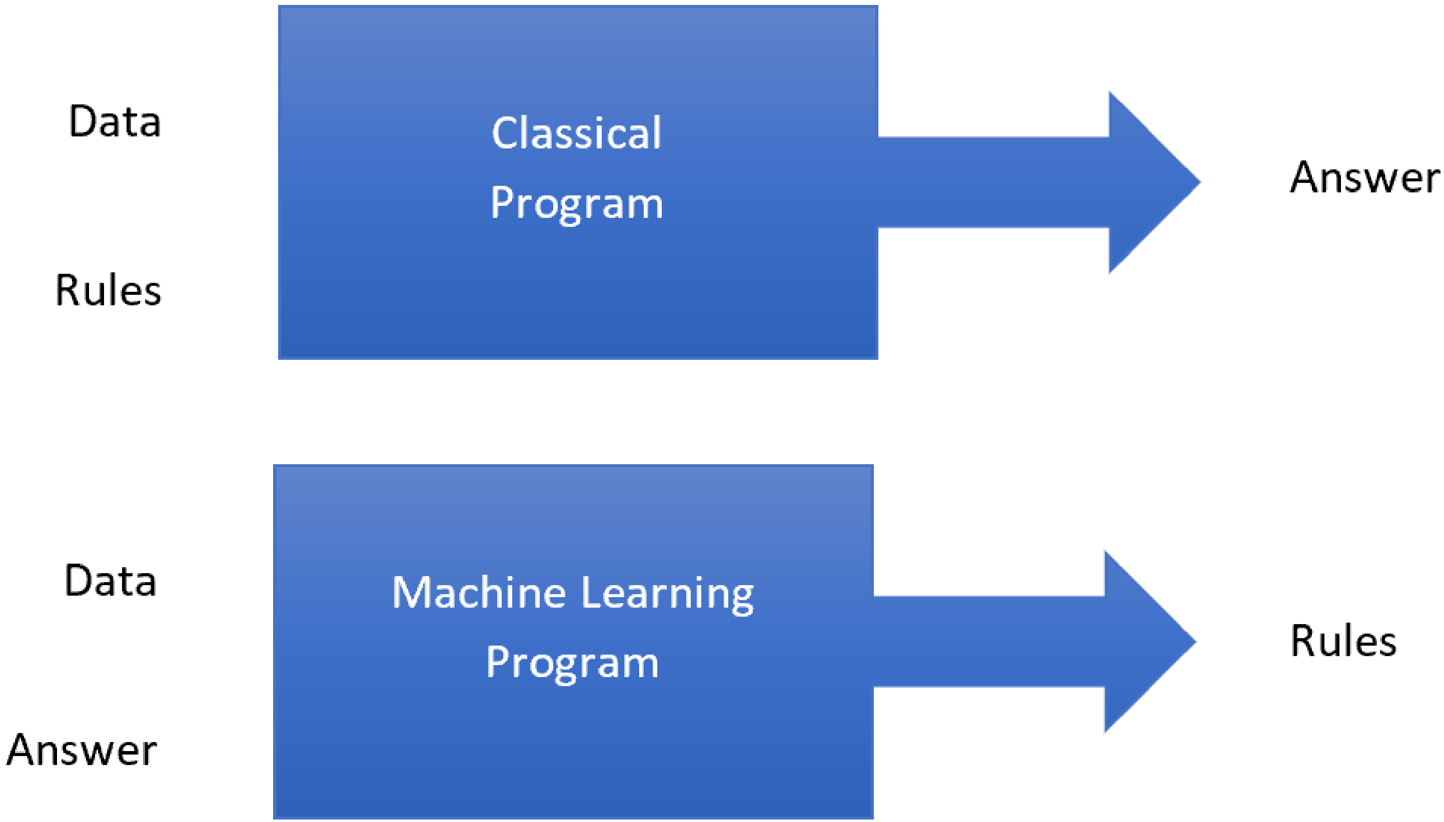
Conclusions

Despite guidance to use the 'wait and see' approach in the age group analyzed, this strategy is not often used for AOM, as previously noted in other studies in hospital settings. Broad-spectrum antibiotic prescription was more frequent when pharyngitis was not confirmed by rapid test, in keeping with evidence from other studies that diagnostic uncertainty leads to overuse of antibiotics.

It is necessary to develop an **accurate** system able to classify all the PEDIANET records **automatically** investigating **all the textual fields** in the database.

*Barbieri et.al "Antibiotic prescriptions in acute otitis media and pharyngitis in Italian pediatric outpatients", **Italian J. of Pediatrics** 2019

Programming paradigms
classical and machine learning

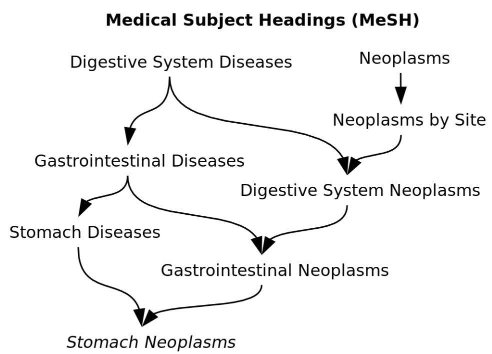


Arthur Samuel (1959). **Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.**

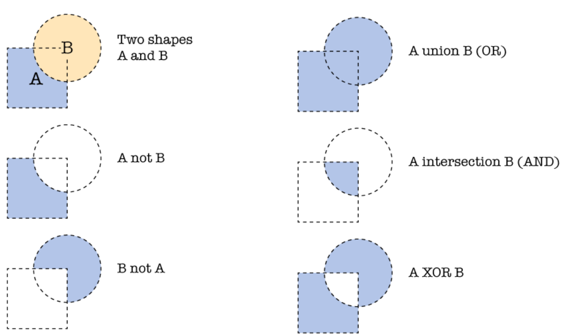
Classical text classification
Search string and regular expression

Search string

One or several strings (also called patterns) founded within a larger string or text.



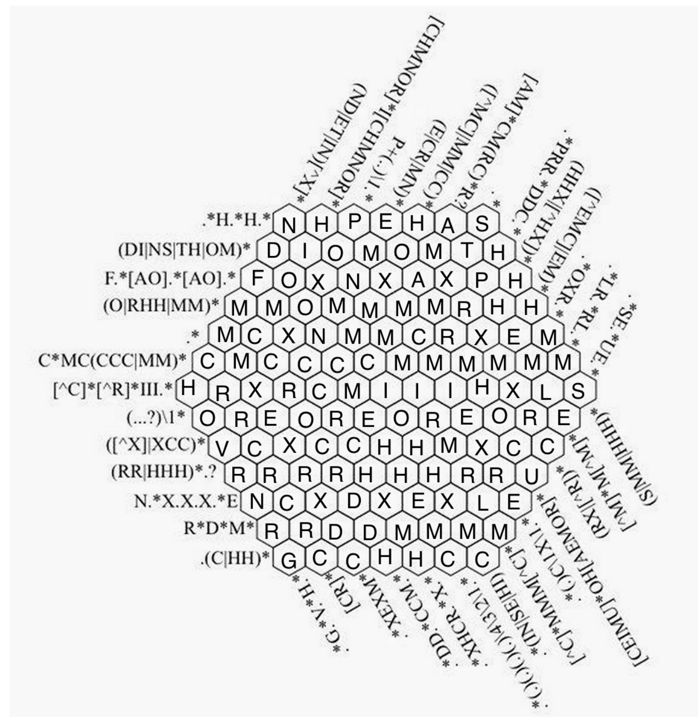
https://upload.wikimedia.org/wikipedia/commons/thumb/8/85/MeSH-example.svg/1024px-MeSH-example.svg.png



https://it.wikipedia.org/wiki/File:Boolean_operations_on_shapes.png

Regular expressions

A sequence of characters that define a search pattern



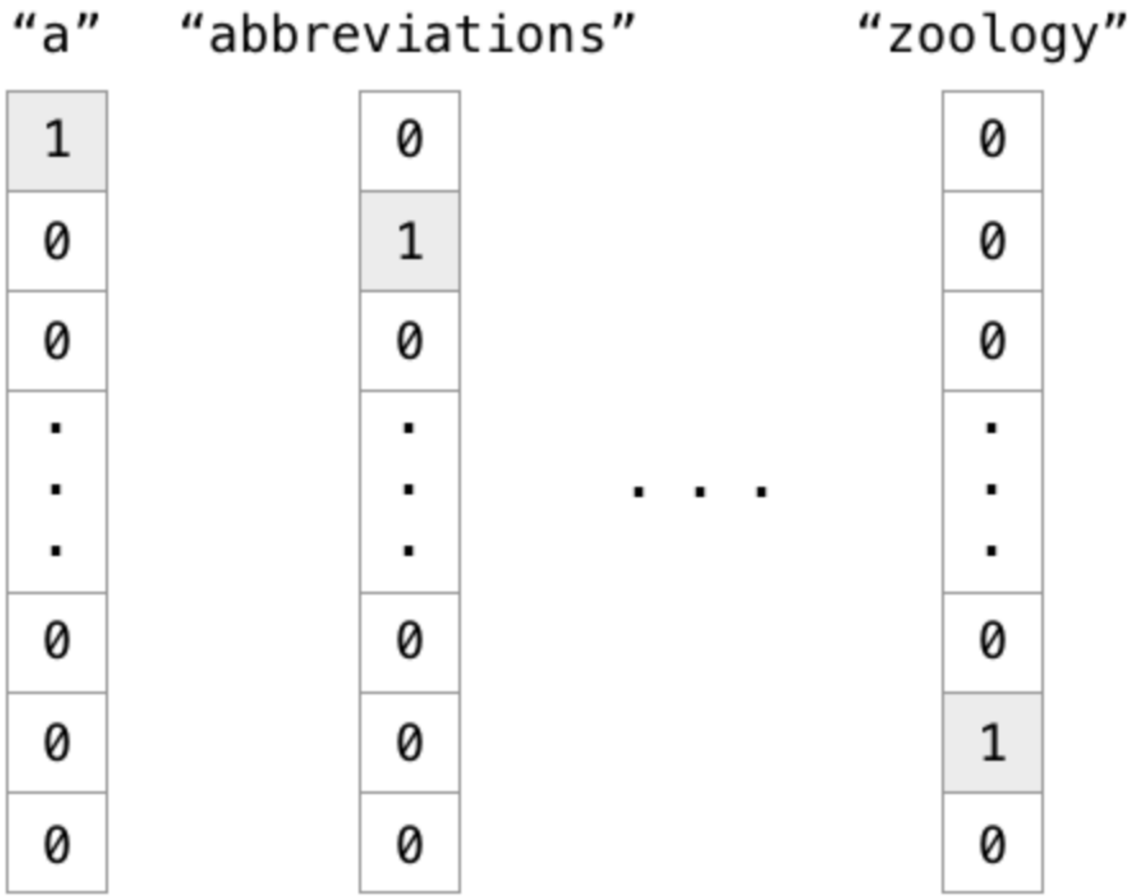
https://www.flickr.com/photos/bluesmoon/8458343489/in/photostream/

Shallow machine learning
document-text matrix

Preprocessing

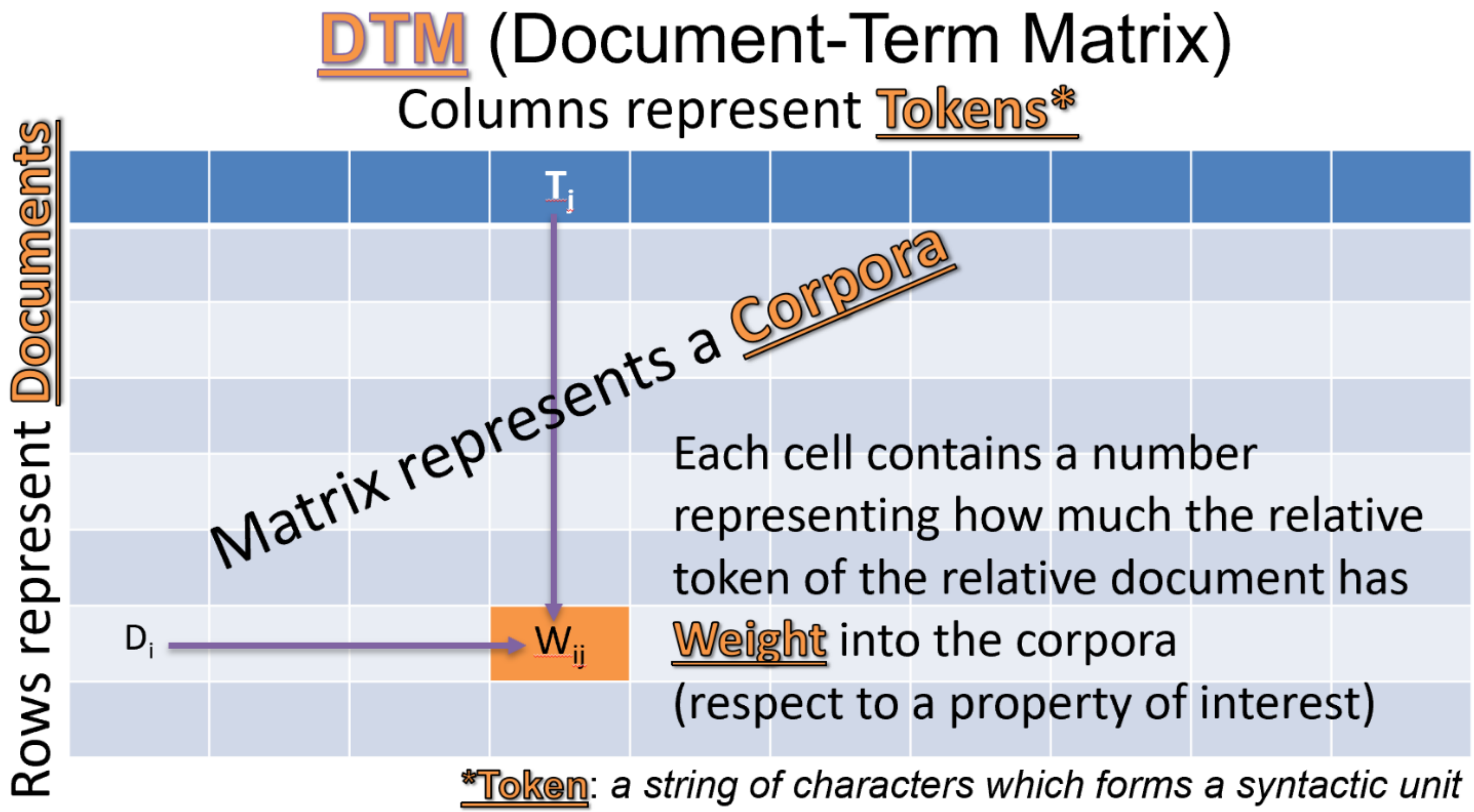
- **removing** (removes noise):
  - remove non-word text
  - remove stopwords
  - strip whitespace
- **merging** (reduce the risk to allow important information to become noise because they are dispersed):
  - lowering
  - stemming
  - lemmatization
- **producing** (from words to “tokens”, i.e. single indivisible piece of information to bring information which could be lost otherwise, e.g. negations):
  - n-Grams (consecutive sequences of n words)

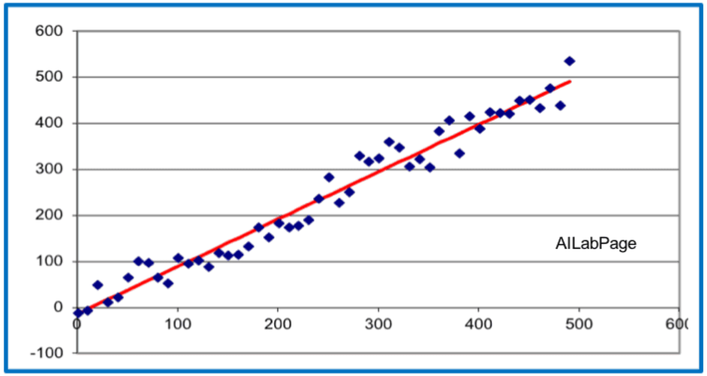
Sparse representation

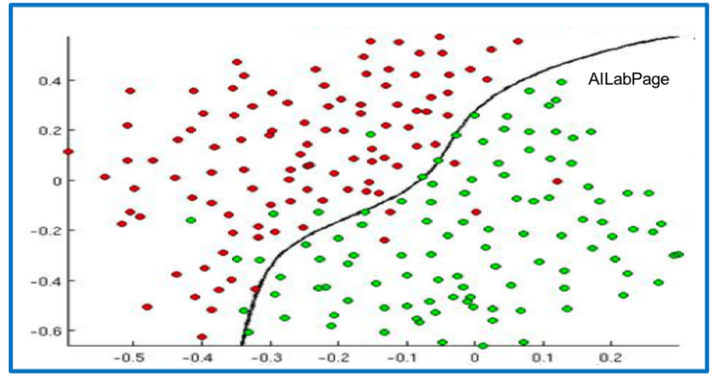


https://towardsdatascience.com/word-embedding-with-word2vec-and-fasttext-a209c1d3e12c

Document-term matrix







Regression

1. The system attempts to predict a value for an input based on past data.
2. Real number / Continuous numbers – Regression problem
3. Example – 1. Temperature for tomorrow

Classification

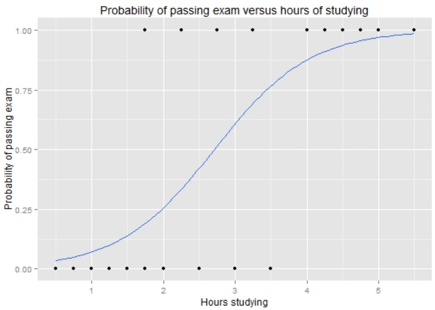
1. In classification, predictions are made by classifying them into different categories.
2. Discreate / categorical variable – Classification problem
3. Example – 1. Type of cancer 2. Cancer Y/N

AILabPage

https://vinodsblog.com/2018/11/08/classification-and-regression-demystified-in-machine-learning/

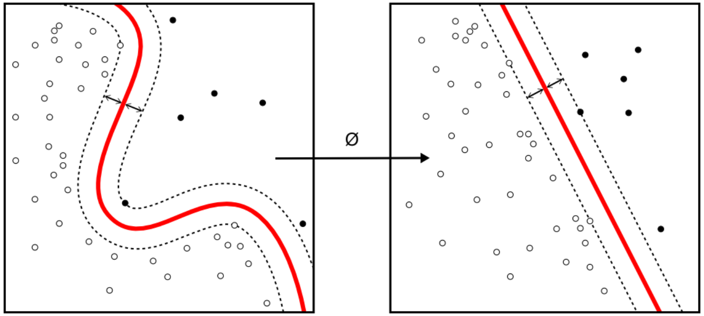
(Shallow) Model definition

Logistic regression



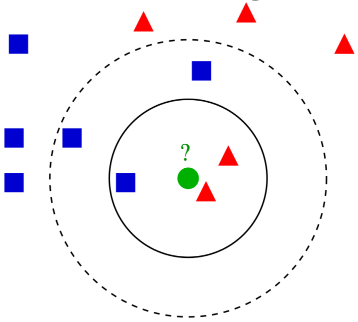
https://commons.wikimedia.org/wiki/File:Exam_pass_logistic_curve.jpeg

support-vector machine



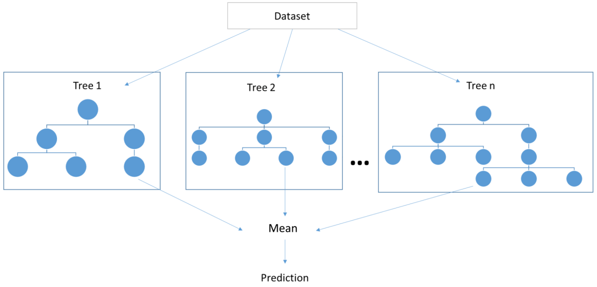
https://commons.wikimedia.org/wiki/File:Kernel_Machine.png

k-nearest neighbor



https://it.m.wikipedia.org/wiki/File:KnnClassification.svg

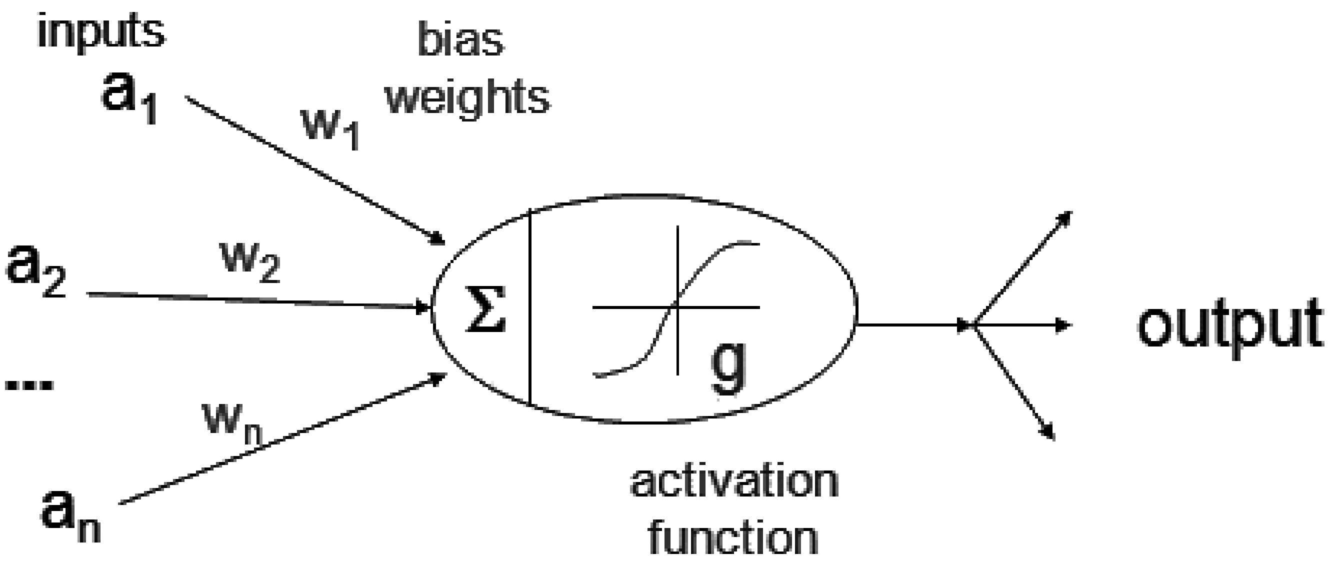
random forest (decision tree)



https://towardsdatascience.com/random-forests-and-decision-trees-from-scratch-in-python-3e4fa5ae4249

Neural networks
basic structure

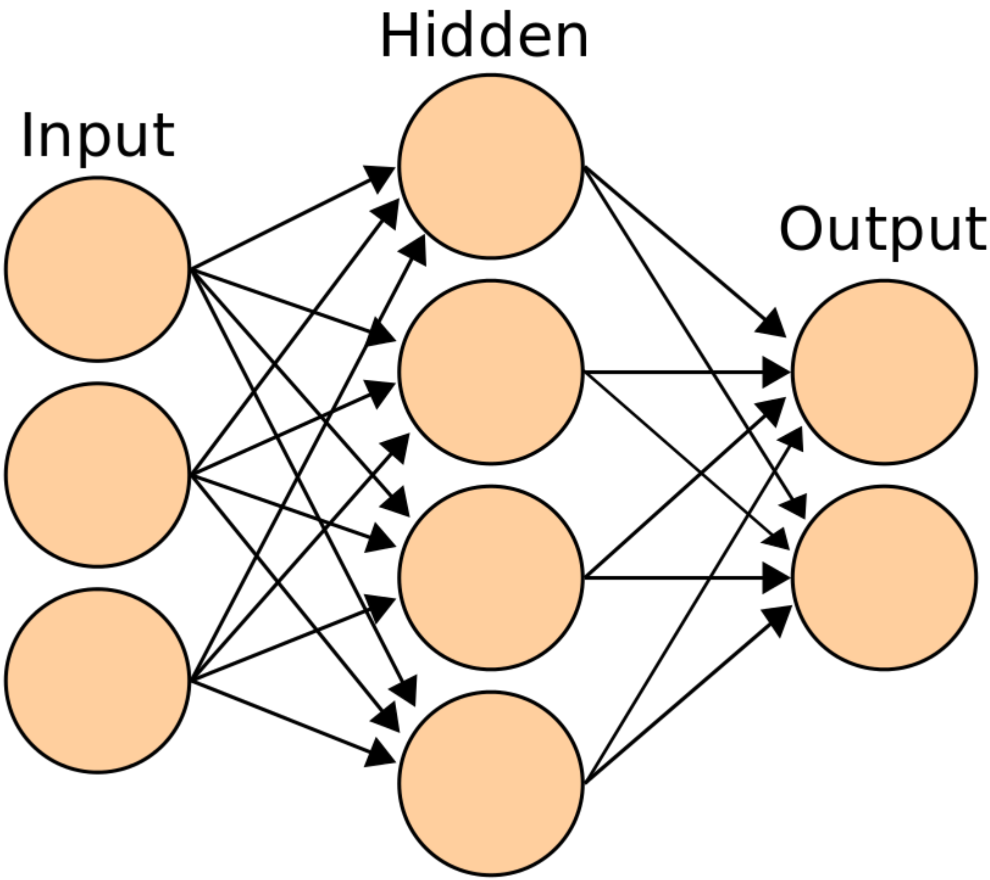
(Artificial) Neuron



https://www.cs.iusb.edu/~danav/teach/c463/12_nn.html

$$= g(\sum(w_i a_i))$$

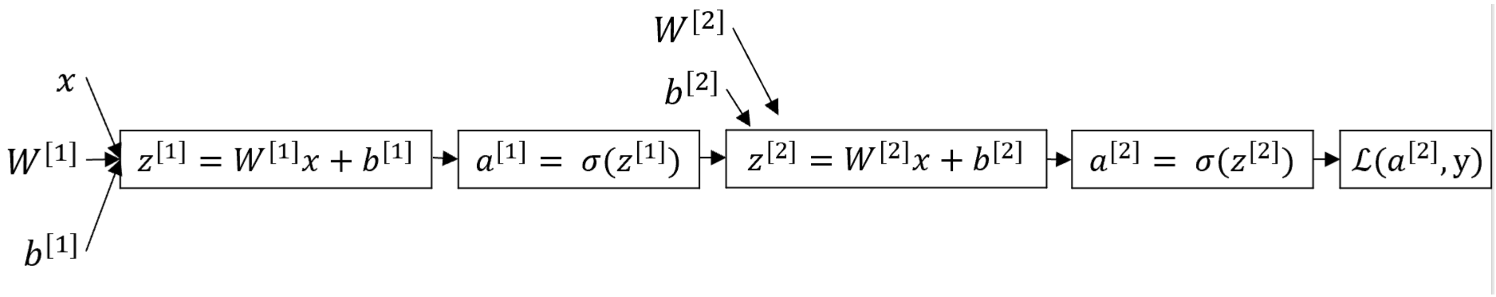
Fully connect



https://upload.wikimedia.org/wikipedia/commons/thumb/e/e4/Artificial_neural_network.svg/860px-Artificial_neural_network.svg.png

Learning

Tom Mitchell (1998). **Well-posed Learning Problem: A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.**

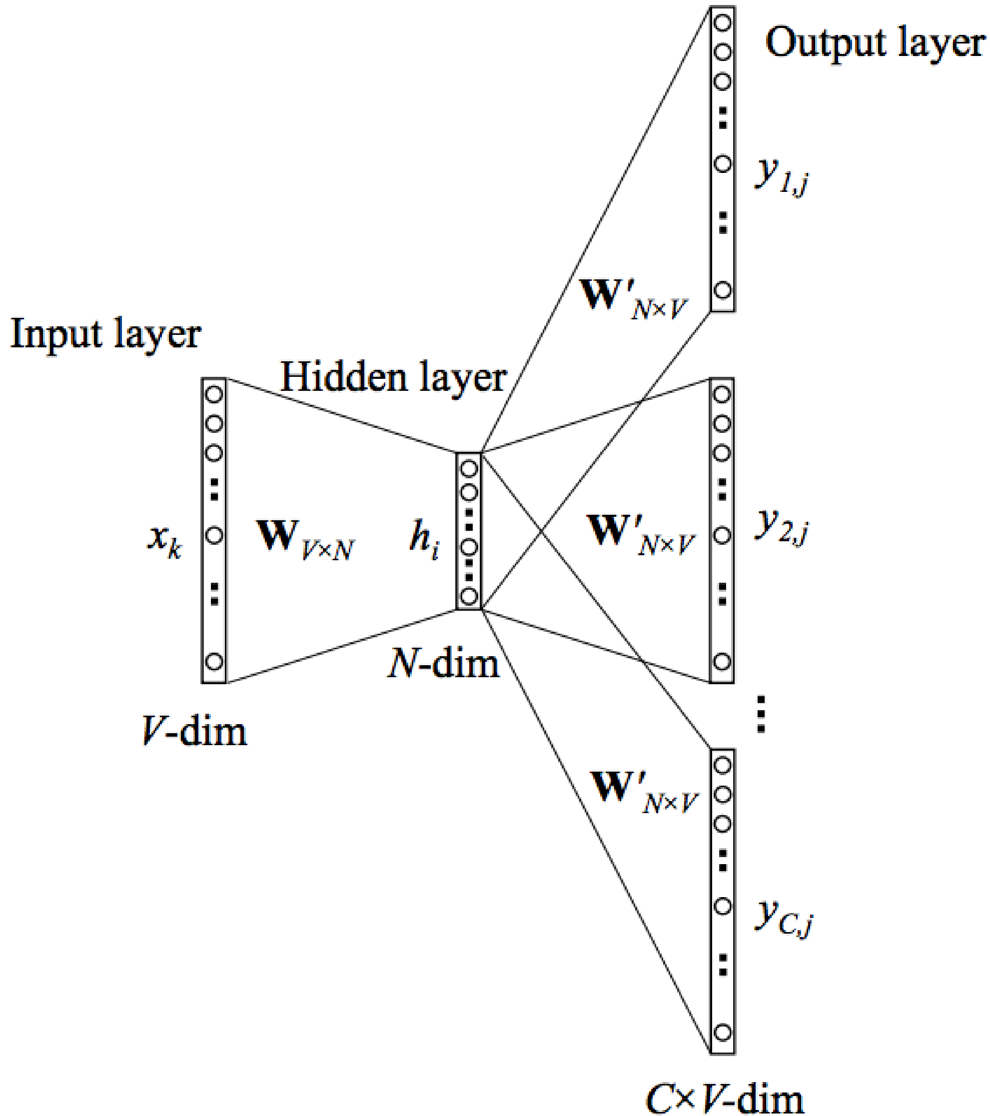


1. Random inizialization
2. Forward propagation
3. Loss evaluation
4. Back propagation
5. Optimization and update

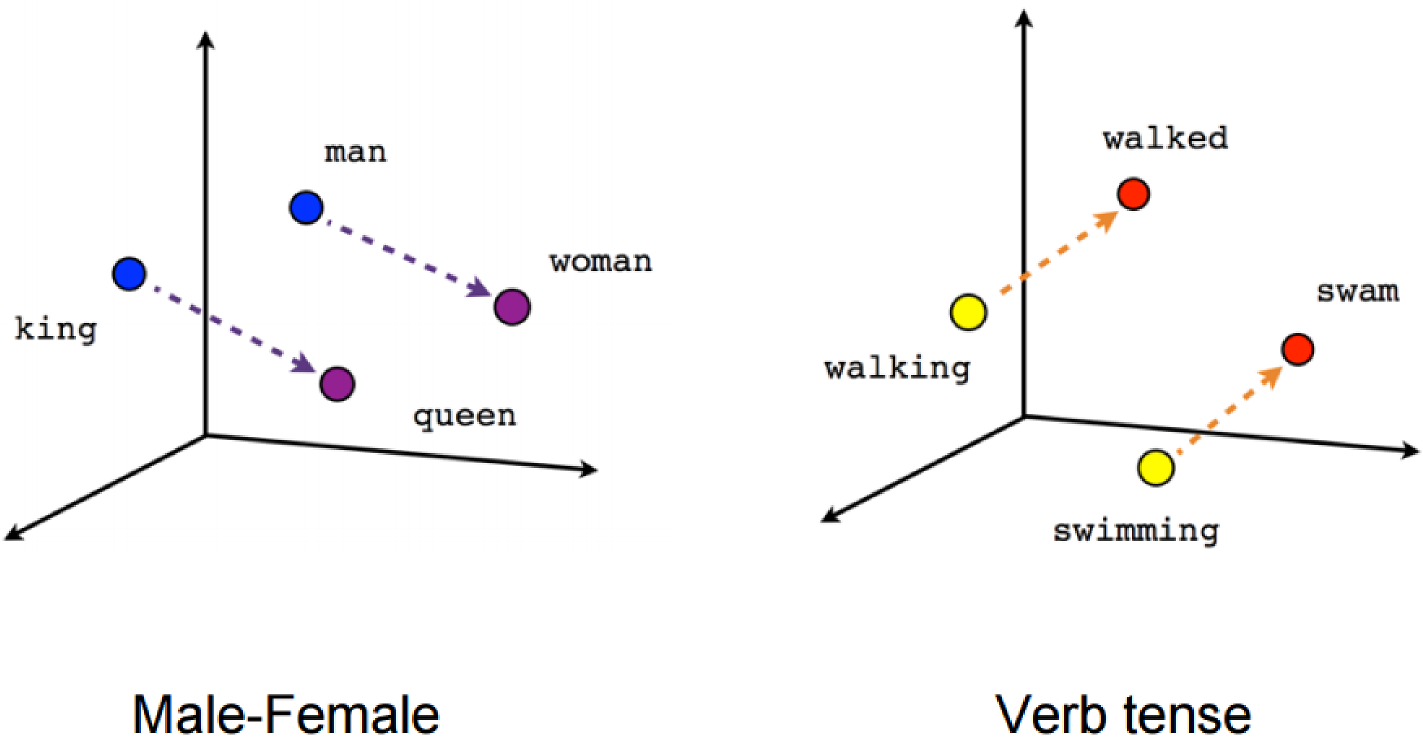
https://upload.wikimedia.org/wikipedia/commons/a/a3/Gradient_descent.gif

Embeddings

dense representation of text



https://www.analyticsvidhya.com/blog/2017/06/word-embeddings-count-word2veec/



https://www.tensorflow.org/images/linear-relationships.png

Embeddings
sumulation

https://ronxin.github.io/wevi/

Deep networks
convolutions, sequential/recurrent

Convolution

2 ^{3}	3 ^{4}	7 ^{4}	4	6	2	9
6 ^{1}	6 ^{0}	9 ^{2}	8	7	4	3
3 $^{-1}$	4 ^{0}	8 ^{3}	3	8	9	7
7	8	3	6	6	3	4
4	2	1	8	3	4	6
3	2	4	1	9	8	3
0	1	3	9	2	1	4

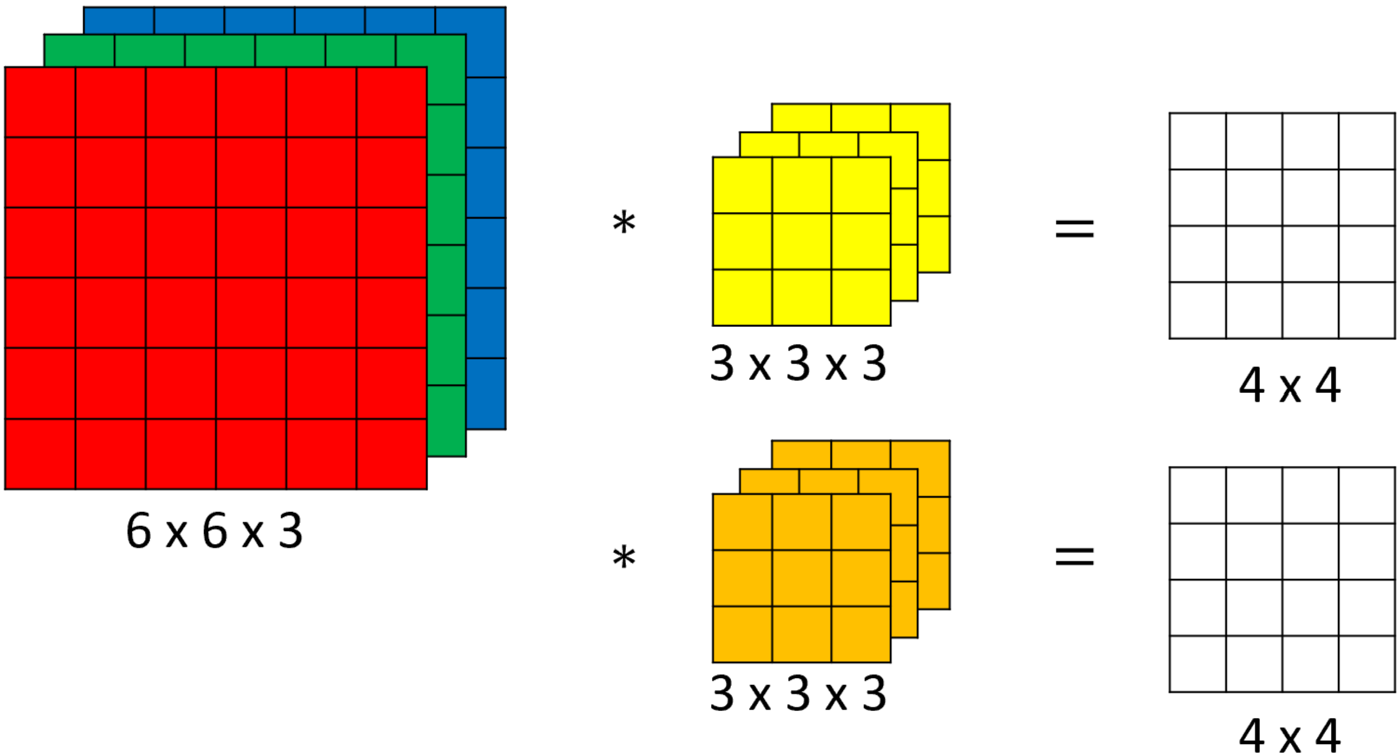
$*$

3	4	4
1	0	2
-1	0	3

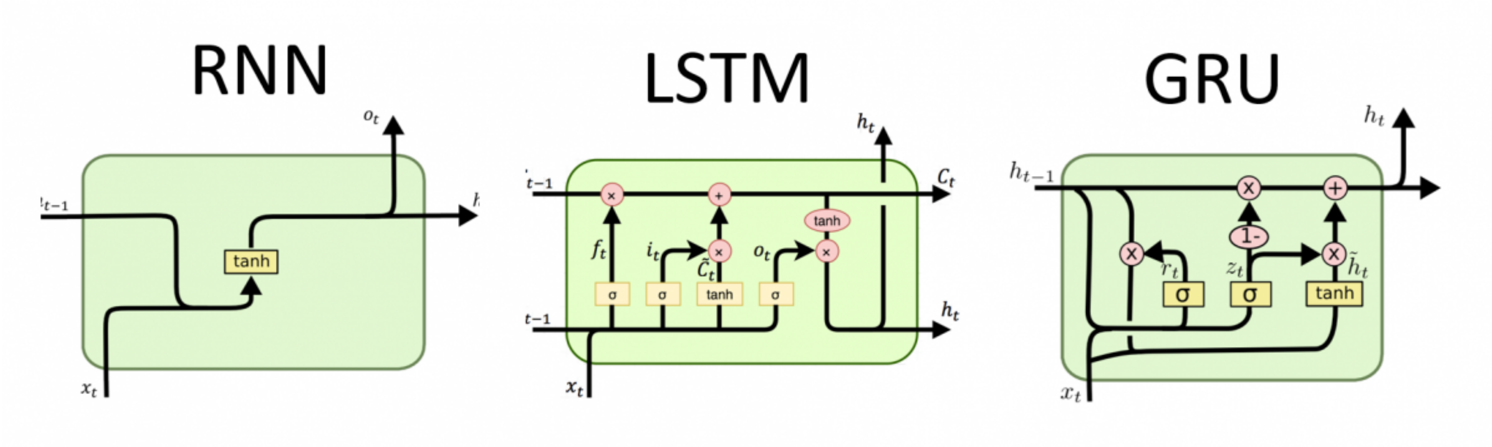
$=$

91	**100**	**83**
69	91	127
44	72	74

Convolution

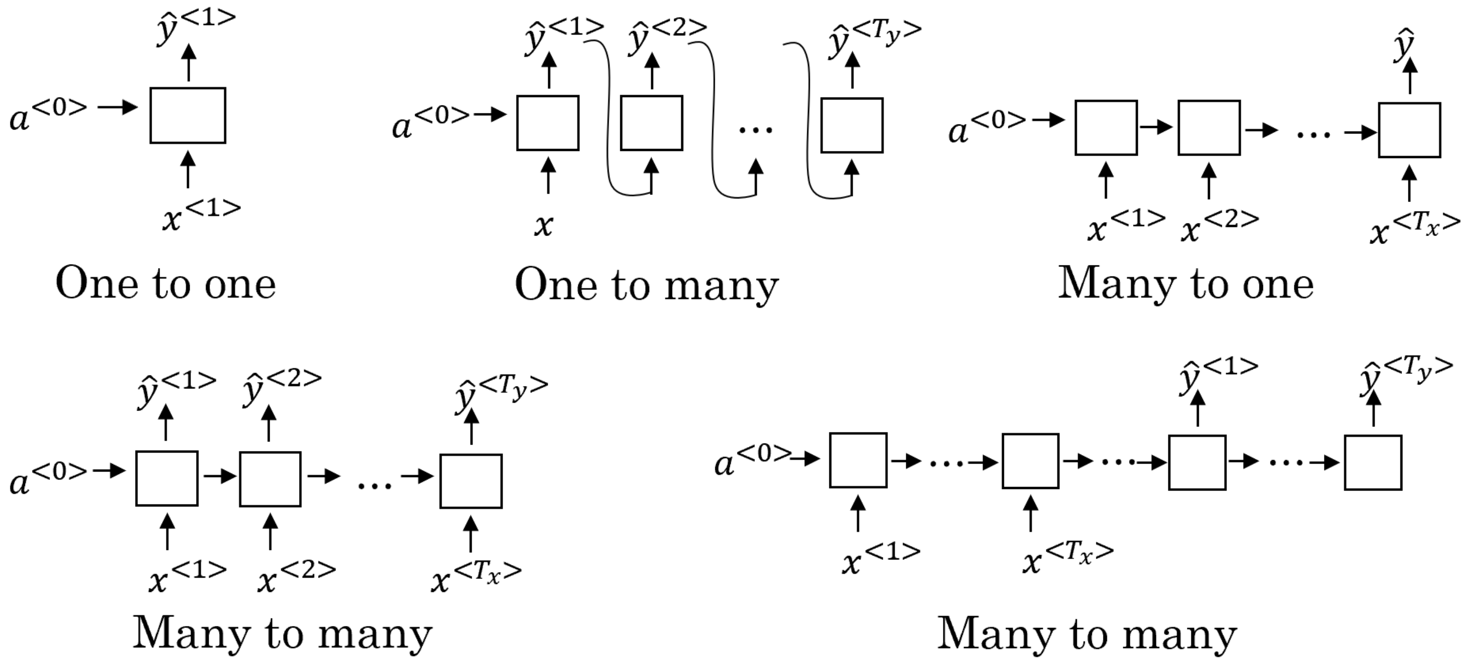


Sequential/Recurrent



http://dprogrammer.org/rnn-lstm-gru

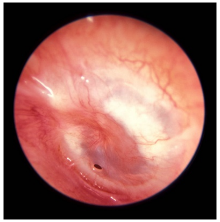
Sequential/Recurrent



Task

multi-class classification

Based on data provided in the EHRs at the visit level, decide if it reports:

0. whatever other than an otitis case
1. an otitis case which is not media (whatever severity)
2. a media otitis which is not acute
3.  an AOM (w/o tympanic membrane perforation, nor recurrent)
4.  an AOM with tympanic membrane perforation
5. a recurrent AOM

http://otitismedia.hawkelibrary.com

Metrics

scores and human-performances

Gold-standard definition:

- two independent expert annotators (weighted Cohen's Kappa = 0.89)
- one pediatrician specialized in infectious diseases decided where the experts shown disagreement:*

Expert annotators	Accuracy [%]	Balanced F1 [%]
A	95.91	93.47
B	95.80	90.12

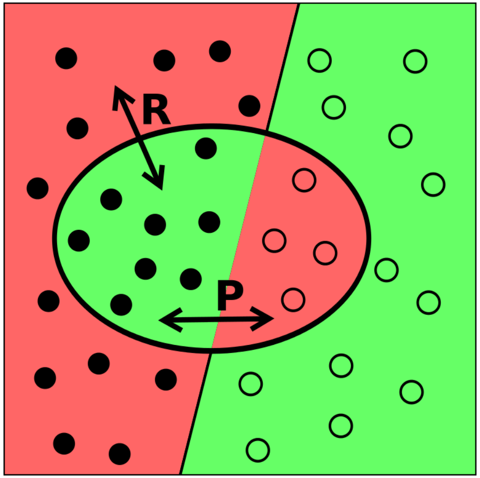
$$\text{Accuracy} = \frac{|\text{true classified}|}{|\text{records}|}$$

*Performances reported for the test set.

$$\text{Balanced F1} = \frac{\text{balanced precision} \cdot \text{balanced recall}}{\text{balanced precision} + \text{balanced recall}}$$

$$\text{Balanced precision} = \frac{\sum_{i \in \text{classes}} \frac{|\text{true classified as } i|}{|\text{labelled as } i|}}{|\text{classes}|}$$

$$\text{Balanced recall} = \frac{\sum_{i \in \text{classes}} \frac{|\text{true classified as } i|}{|\text{class } i|}}{|\text{classes}|}$$



Analysis plan

flowchart

Gold Standard

Train

Years: $2004 - 2007$

Records: $4,926$

Validation

Years: $2008 - 2017$

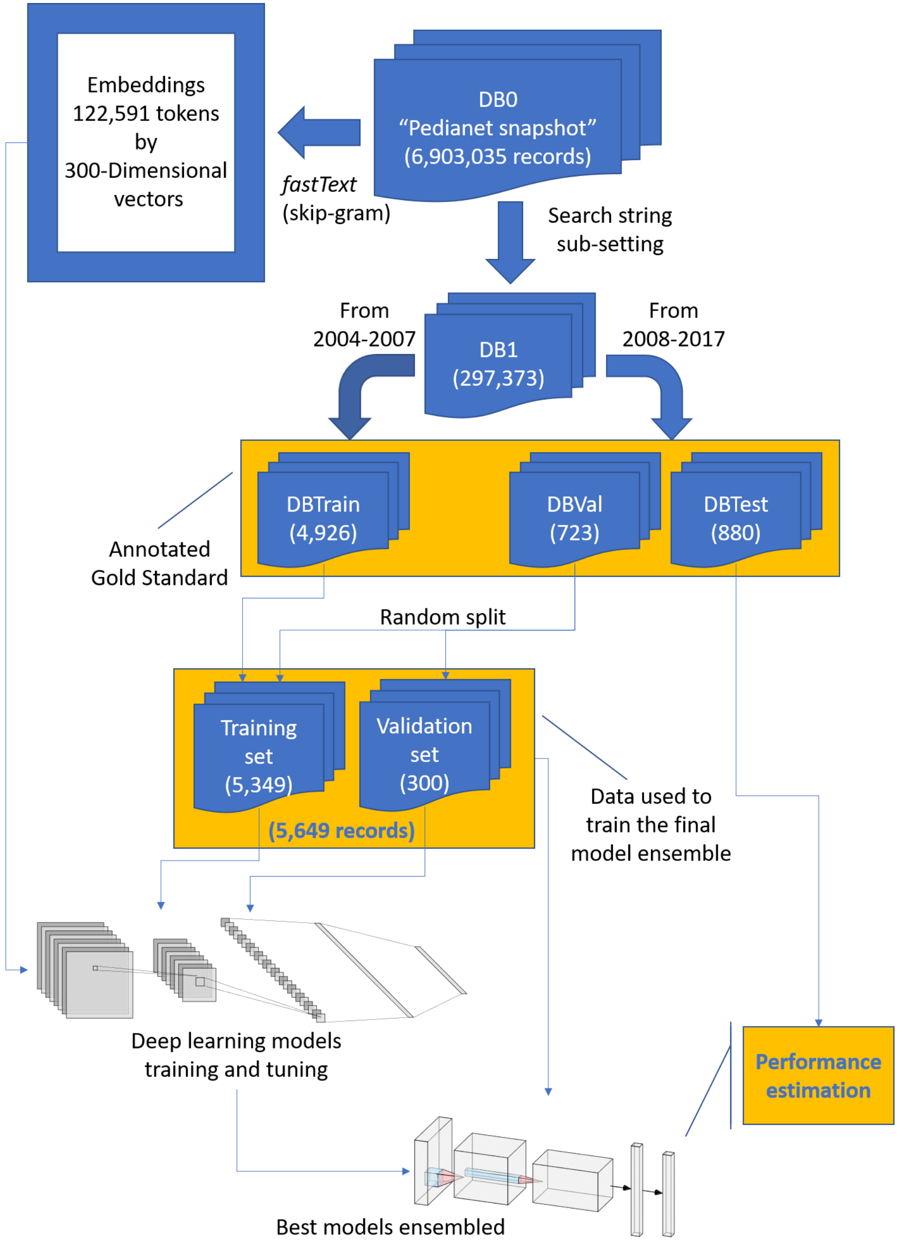
Records: 723

Test

Years: $2008 - 2017$

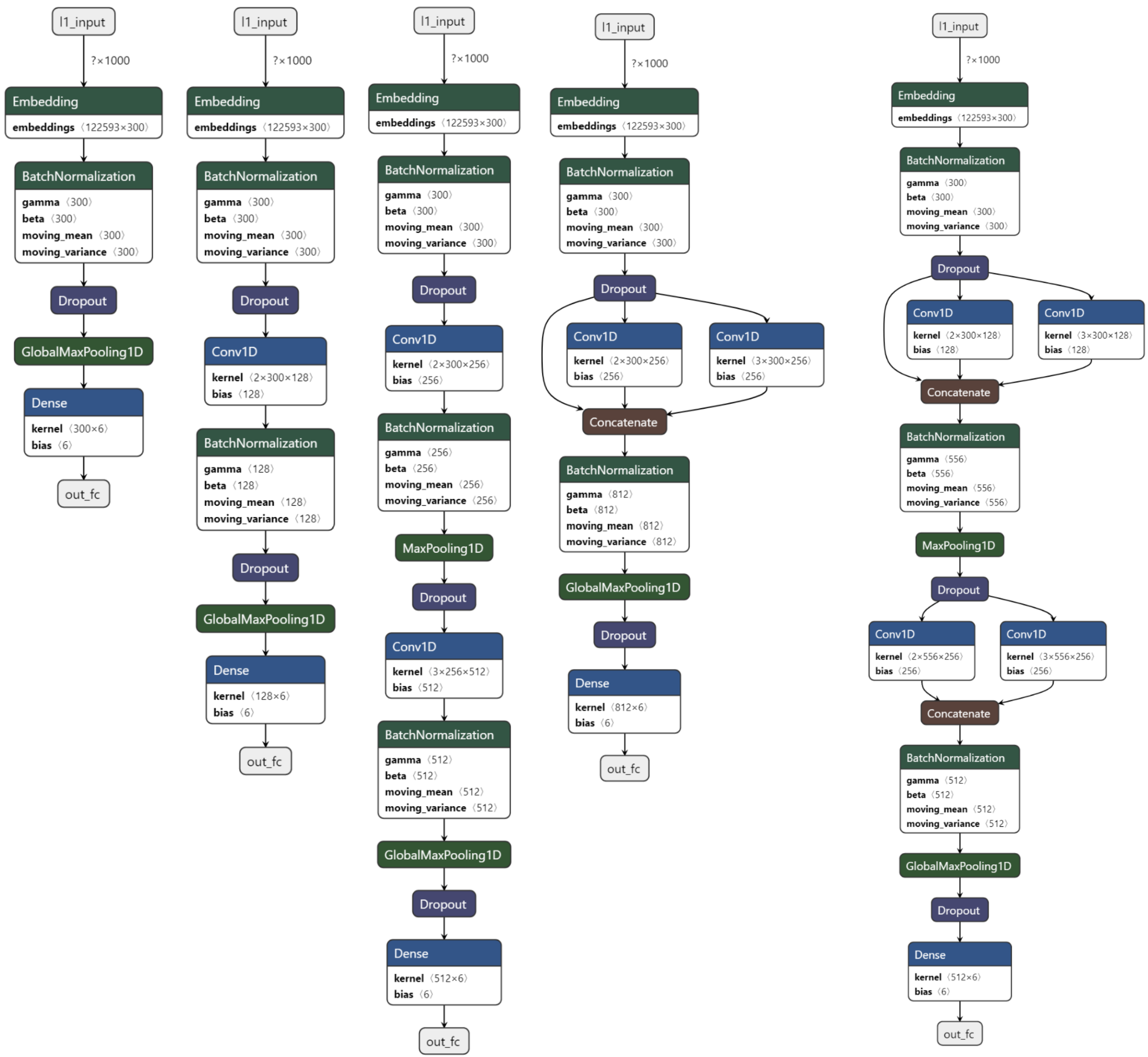
Records: 880

(NOTE: 4 months of annotatotion, by 2 independent experts)


Embeddings
122,591 tokens
by
300-Dimensional
vectors
DB0
“Pedianet snapshot”
(6,903,035 records)
fastText
(skip-gram)
Search string
sub-setting
From
2004-2007
DB1
(297,373)
From
2008-2017
DBTrain
(4,926)
DBVal
(723)
DBTest
(880)
Annotated
Gold Standard
Random split
Training
set
(5,349)
Validation
set
(300)
(5,649 records)
Data used to
train the final
model ensemble
Deep learning models
training and tuning
Performance
estimation
Best models ensembled

Deep-learning approach

architecture explored


l1_input
?×1000
Embedding
embeddings (122593×300)
BatchNormalization
gamma (300)
beta (300)
moving_mean (300)
moving_variance (300)
Dropout
GlobalMaxPooling1D
Dense
kernel (300×6)
bias (6)
out_fc
l1_input
?×1000
Embedding
embeddings (122593×300)
BatchNormalization
gamma (300)
beta (300)
moving_mean (300)
moving_variance (300)
Dropout
Conv1D
kernel (2×300×128)
bias (128)
BatchNormalization
gamma (128)
beta (128)
moving_mean (128)
moving_variance (128)
Dropout
GlobalMaxPooling1D
Dense
kernel (128×6)
bias (6)
out_fc
l1_input
?×1000
Embedding
embeddings (122593×300)
BatchNormalization
gamma (300)
beta (300)
moving_mean (300)
moving_variance (300)
Dropout
Conv1D
kernel (2×300×256)
bias (256)
BatchNormalization
gamma (256)
beta (256)
moving_mean (256)
moving_variance (256)
MaxPooling1D
Dropout
Conv1D
kernel (3×256×512)
bias (512)
BatchNormalization
gamma (512)
beta (512)
moving_mean (512)
moving_variance (512)
GlobalMaxPooling1D
Dropout
Dense
kernel (512×6)
bias (6)
out_fc
l1_input
?×1000
Embedding
embeddings (122593×300)
BatchNormalization
gamma (300)
beta (300)
moving_mean (300)
moving_variance (300)
Dropout
Conv1D
kernel (2×300×256)
bias (256)
Conv1D
kernel (3×300×256)
bias (256)
Concatenate
BatchNormalization
gamma (812)
beta (812)
moving_mean (812)
moving_variance (812)
GlobalMaxPooling1D
Dropout
Dense
kernel (812×6)
bias (6)
out_fc
l1_input
?×1000
Embedding
embeddings (122593×300)
BatchNormalization
gamma (300)
beta (300)
moving_mean (300)
moving_variance (300)
Dropout
Conv1D
kernel (2×300×128)
bias (128)
Conv1D
kernel (3×300×128)
bias (128)
Concatenate
BatchNormalization
gamma (556)
beta (556)
moving_mean (556)
moving_variance (556)
MaxPooling1D
Dropout
Conv1D
kernel (2×556×256)
bias (256)
Conv1D
kernel (3×556×256)
bias (256)
Concatenate
BatchNormalization
gamma (512)
beta (512)
moving_mean (512)
moving_variance (512)
GlobalMaxPooling1D
Dropout
Dense
kernel (512×6)
bias (6)
out_fc

Results

Network	Balanced precision	Balanced recall	Accuracy	Balanced F1
Simple Embedding	84.51	68.63	81.70	75.75
Single Kernel	92.60	*91.87*	94.66	*92.23*
Sequential CNN	*95.94*	81.26	93.64	87.99
Parallel CNN	**96.95**	*94.78*	**96.59**	**95.86**
Deep CNN	**96.38**	*93.36*	**96.25**	**94.85**
Ensemble (w/o Simple Embeddings)	**97.03**	*93.97*	**96.59**	**95.47**

* **Bold face** = over the maximum
* *italic* = over the mean

Annotators	Balanced precision	Balanced recall	Accuracy	Balanced F1
A	91.70	95.30	95.91	93.47
B	96.33	84.66	95.80	90.12
(mean)	(94.02)	(89.98)	(95.86)	(91.80)

Final remarks

Strenghts

- embedding: no more needs to hand-craft features
- deep learning:
  - automatical detection and modeling of non linearities and interactions
  - update models w/ new data
  - use pre-trained or merge multiple models
  - can take advantage of more data than shallow models

Weakness

- our human-performance estimation is not based at the professional levels
- missing of computational power (i.e, GPUs) for deeper networks, e.g., recurrent

Possible improvement

- improve the gold standard:
  - quality (human-performance level)
  - accuracy (error analyses)
  - quantity (more training records / active learning)
- deeper networks / advanced architectures
- different weighting schemes for the ensamble

Conclusions

Deep Learning Approach to Text-Mining EHR

- Can be used to identifying and classifying diagnosis from (huge ammount of) free text
- Quality comparable with human-performances
- Trained models can be adopted on other health care databases, different from the original one
- It can improve healthcare research limiting human errors and time, speeding up databases interrogations

Thank you
for the attention

Questions?

Slides created via the R package **xaringan** powered by remark.js, **knitr**, and R Markdown.