

Machine Learning - Big Picture



“Umelá Inteligencia je budúcnosť nielen pre Rusko, ale pre celé ľudstvo.

Je to obrovská príležitosť, ale aj hrozba ktorú je ťažko predpovedať.

Ten kto sa stane lídrom v tejto oblasti sa stane vládcom sveta.”

--Vladimir Putin, 2017

ML Applications

Multimedia
Security
Image processing
Natural language processing
Information retrieval
Bioinformatics
Gene expression
Intrusion detection system

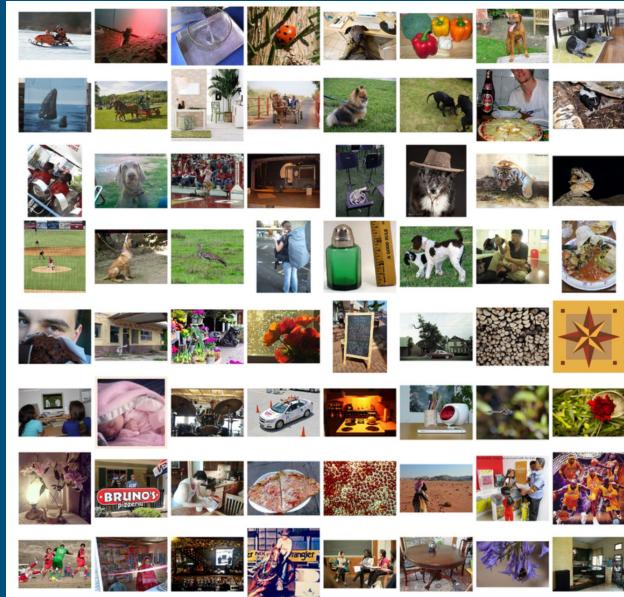
Handwriting recognition
Recommender systems
CRM (Customer relationship management)
Text summarization
Computer Security
Sentiment analysis
Search engine
Face recognition
Game
Collaborative filtering
Fraud detection

Face detection
Personalization
Marketing
Manufacturing
Market basket analysis
Speech recognition
Anomaly detection
Human interaction
Object recognition
Spam

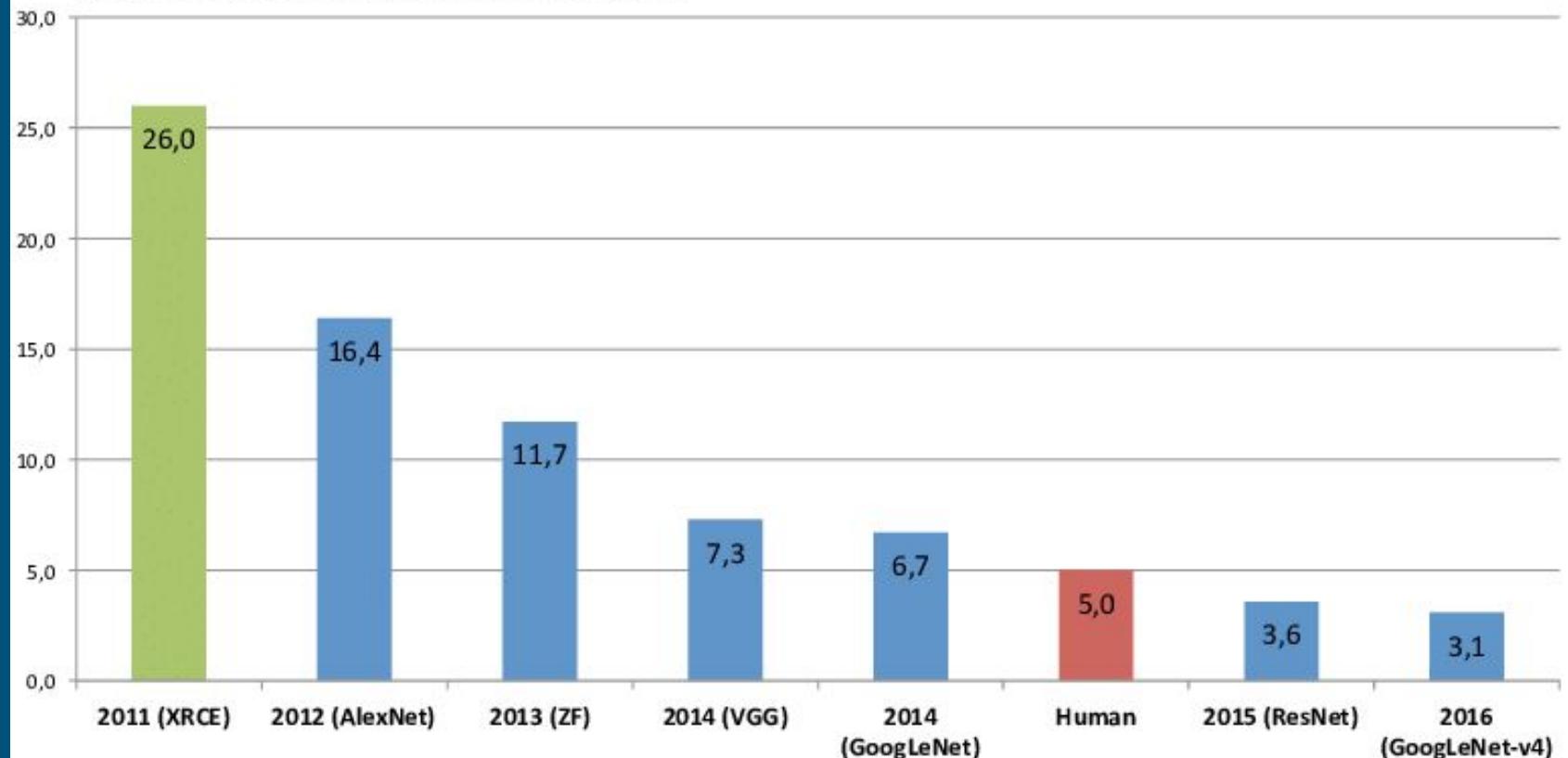
Diagnosis
Medicine

ImageNet Large Scale Visual Recognition Challenge

- Zaradiť obrázok do jednej z 1000 daných kategórií
(90 plemien psov)
- Chyba pri 5 pokusoch.



ImageNet Classification Error (Top 5)

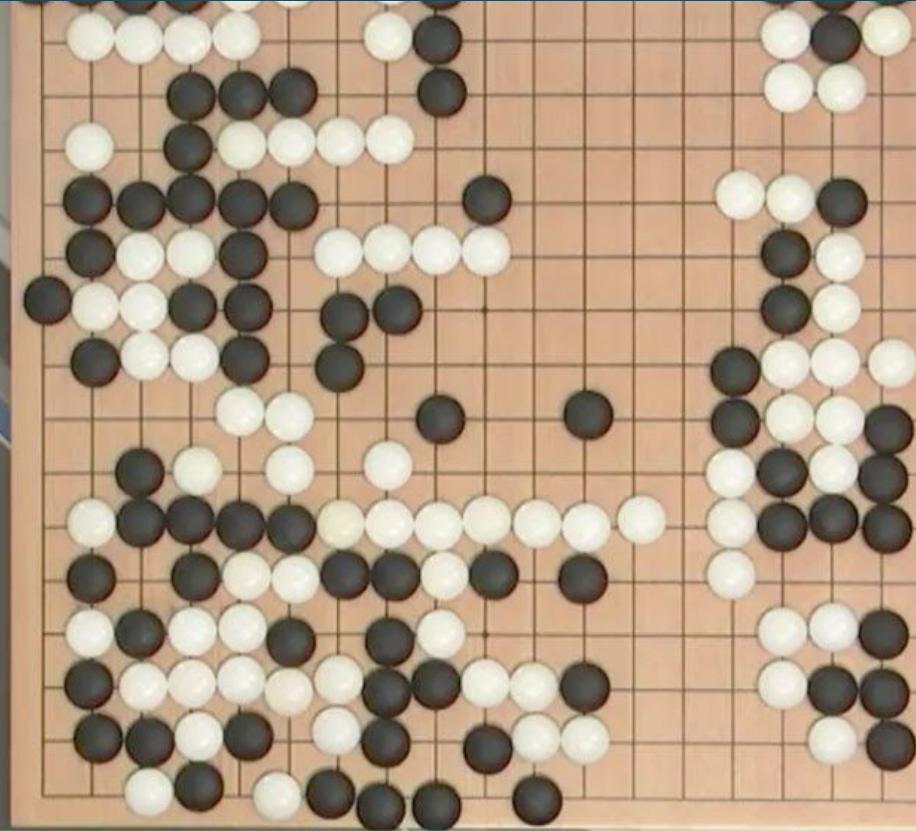


1997: Deep Blue vs. Garry Kasparov, 3½–2½

Garry Kasparov vs



2016: AlphaGo vs. Lee Sedol, 4–1



● LEE SEDOL

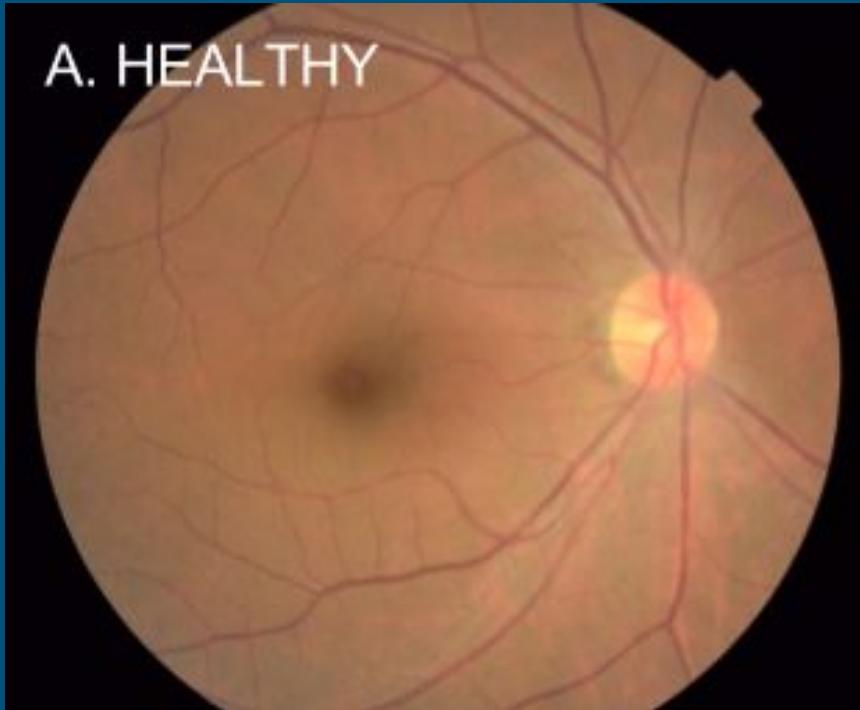
2017: AlphaZero vs. AlphaGo 100–0

- Všeobecná technika pre hranie hier:
 - Go, Šach, Shogi, ...
- Žiadna preexistujúca vedomosť o hre (okrem pravidiel).
- AlphaZero vs. Stockfish: 64–36 (28 výhier, 0 prehier, 72 remíz).

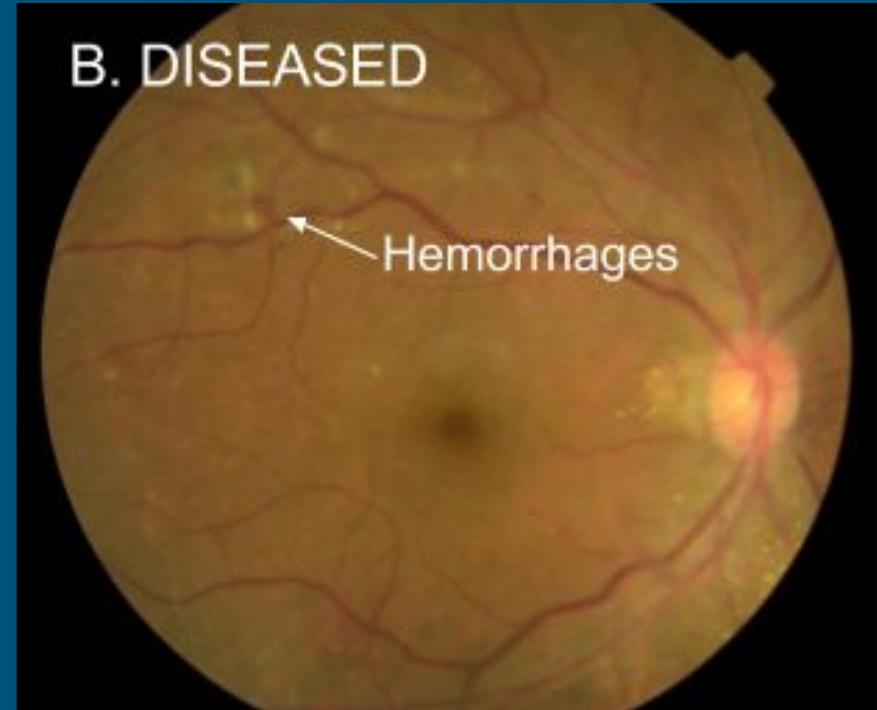
“Vždy som sa zamýšľal aké by to bolo, keby nejaký nadradený druh pristál na zemi a ukázal nám ako hrajú šach. Teraz už viem.” --Peter Heine Nielsen, GM

Diagnostika z lekárskych snímkov

A. HEALTHY



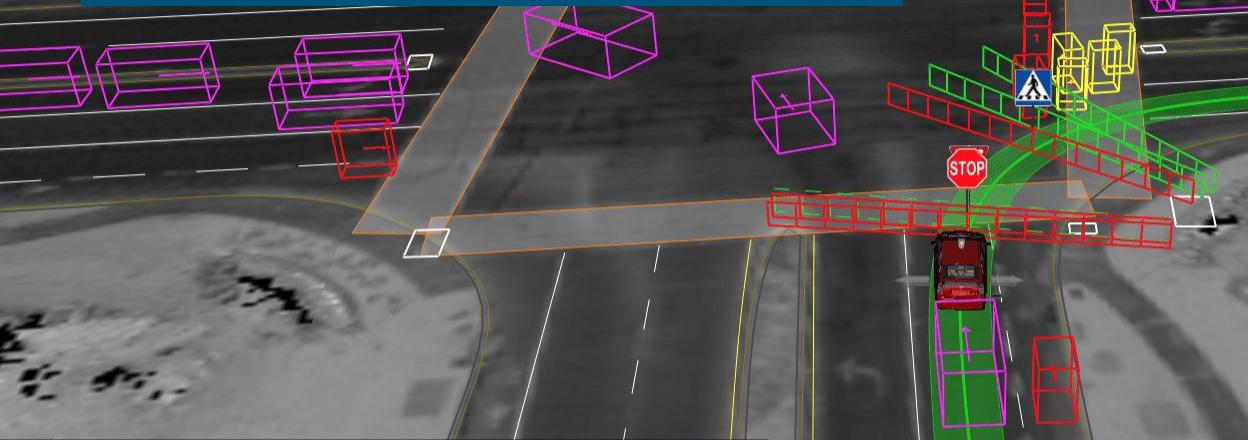
B. DISEASED

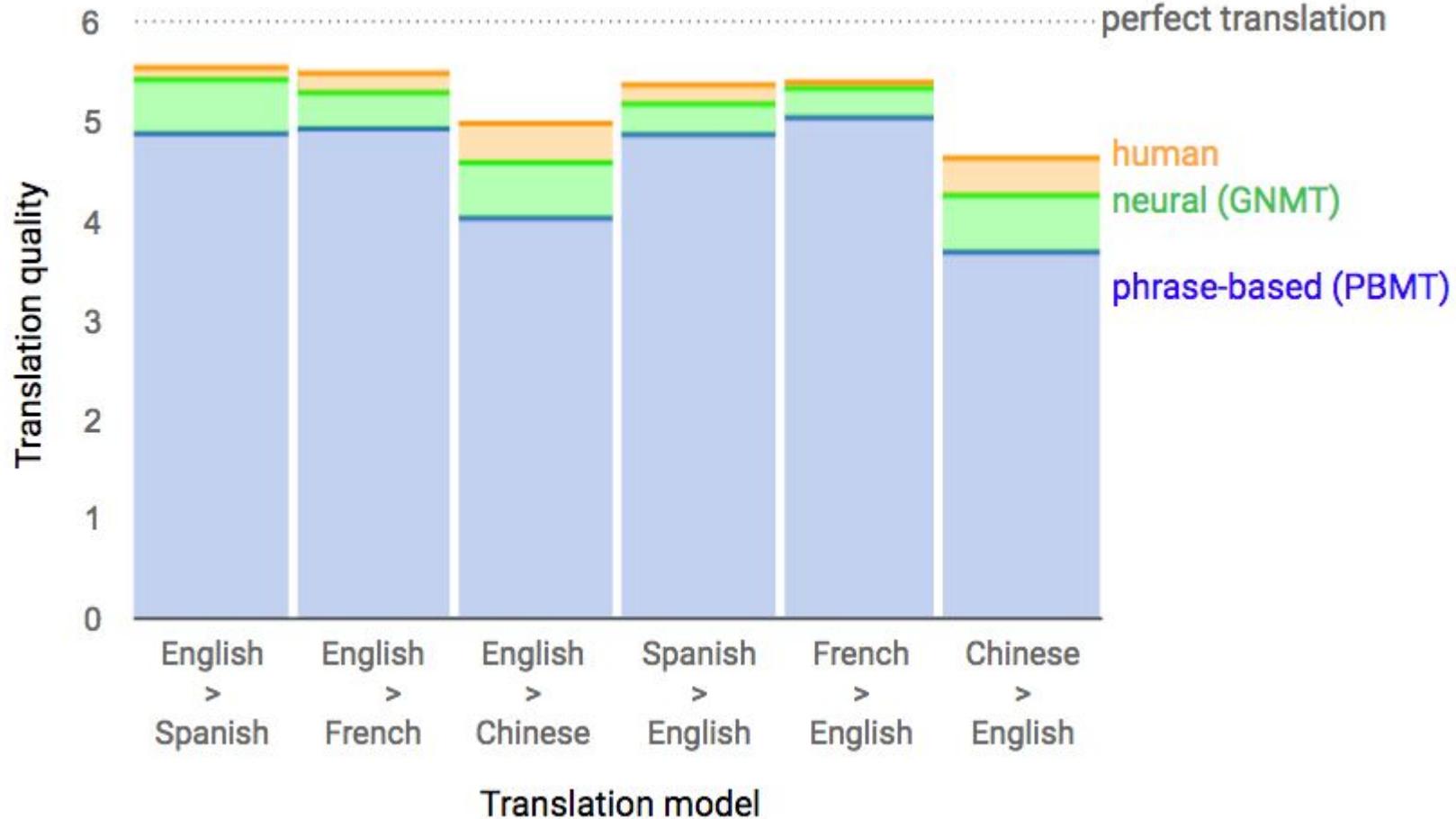


Cruise

0 MPH

Samojazdiace autá





Umenie (imitácia)



SOFTWARE

Algorithms in Decision Making

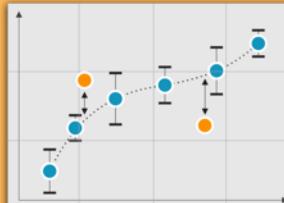
RULE-BASED DECISION MAKING

...
IF condition fulfilled **THEN**
 activity 1
ELSE
 activity 2
...

Boolean data
(yes or no)

- Examples:**
- phone notification
 - time- or threshold-based alarms
 - simple pattern matching

STATISTICAL REASONING



Numerical data
allowing for
curve fitting

- Examples:**
- extra- and interpolation
 - outlier detection
 - predictive maintenance

FUZZY BOUNDARIES

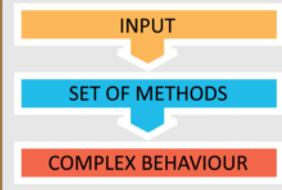
MACHINE LEARNING



Arbitrary data
that needs to be
abstracted into
numbers

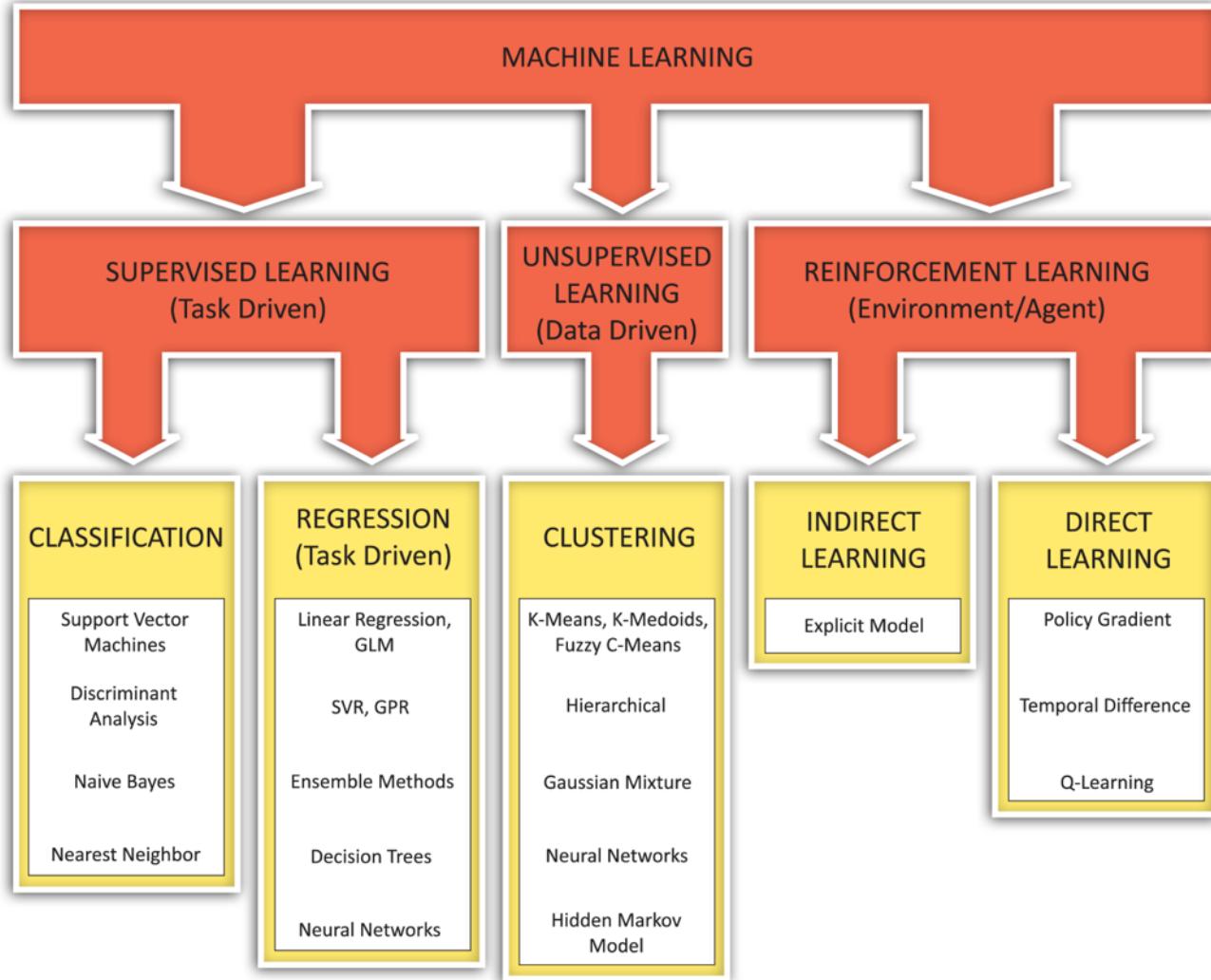
- Examples:**
- identification of relevant features from large input datasets
 - quality control using various metrics

ARTIFICIAL INTELLIGENCE

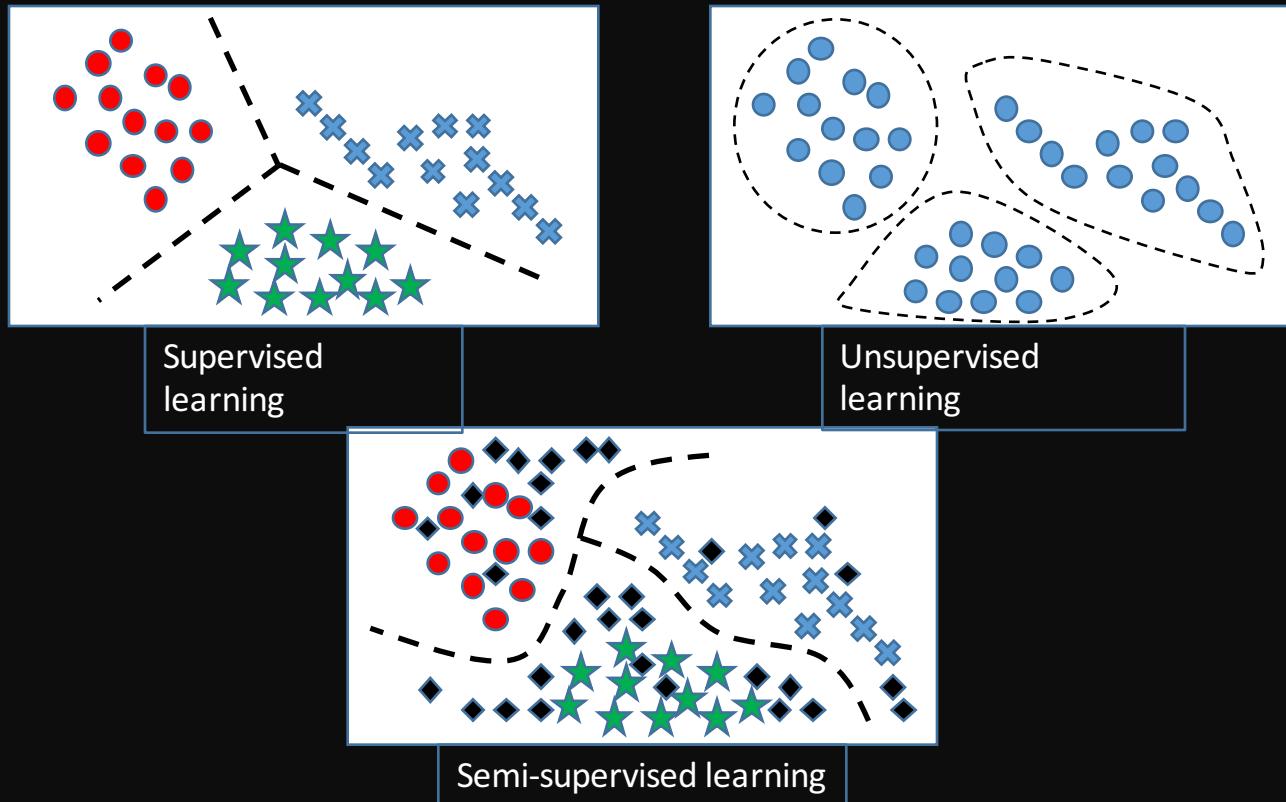


Autonomous selection
of best methodology
when presented with
arbitrary data

- Examples:**
- autonomous vehicles
 - human-like conversational skills
 - intelligent digital assistant



Algorithms

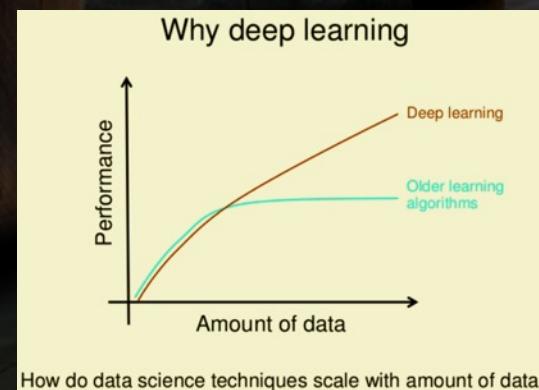


Machine Learning

- „Ručný“ feature engineering – potrebný data preprocessing
- Hardwarovo menej náročné (CPU,RAM)
- Na tréning stačí aj menej dát
- Menej komplexné modely – limitovanejšia performance
- Zvyčajne dobre interpretovateľné

Deep Learning

- Menší feature engineering – možnosť „surovejších“ dát
- Hardwarovo viac náročné (CPU/GPU,RAM)
- Na tréning potrebuje veľa dát
- Komplexnejšie modely - v mnohých úlohách lepší performance
- Zvyčajne ľahké na interpretáciu



Example 1: Spam classification

Input Attributes Target Attribute

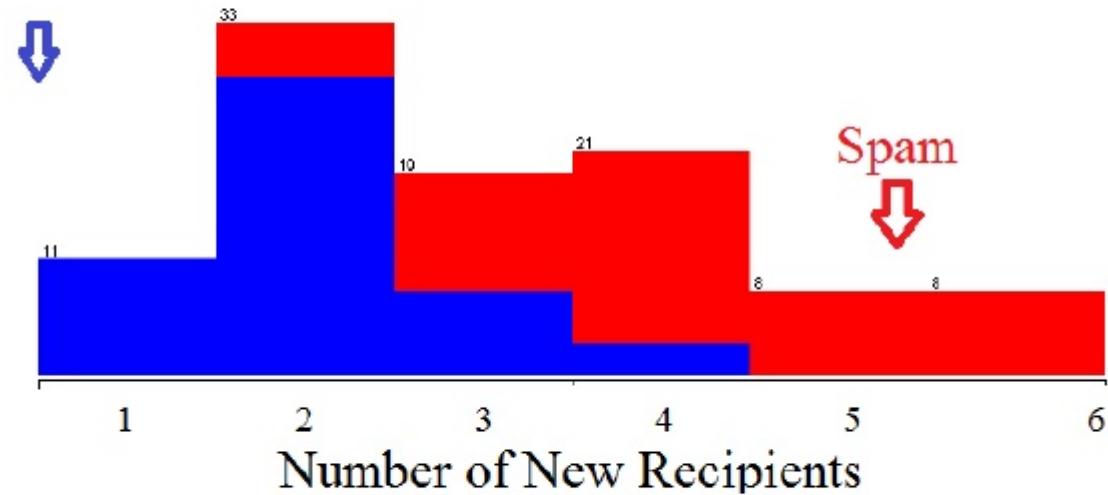
Number of new Recipients	Email Length (K)	Country (IP)	Customer Type	Email Type
0	2	Germany	Gold	Ham
1	4	Germany	Silver	Ham
5	2	Nigeria	Bronze	Spam
2	4	Russia	Bronze	Spam
3	4	Germany	Bronze	Ham
0	1	USA	Silver	Ham
4	2	USA	Silver	Spam

Instances

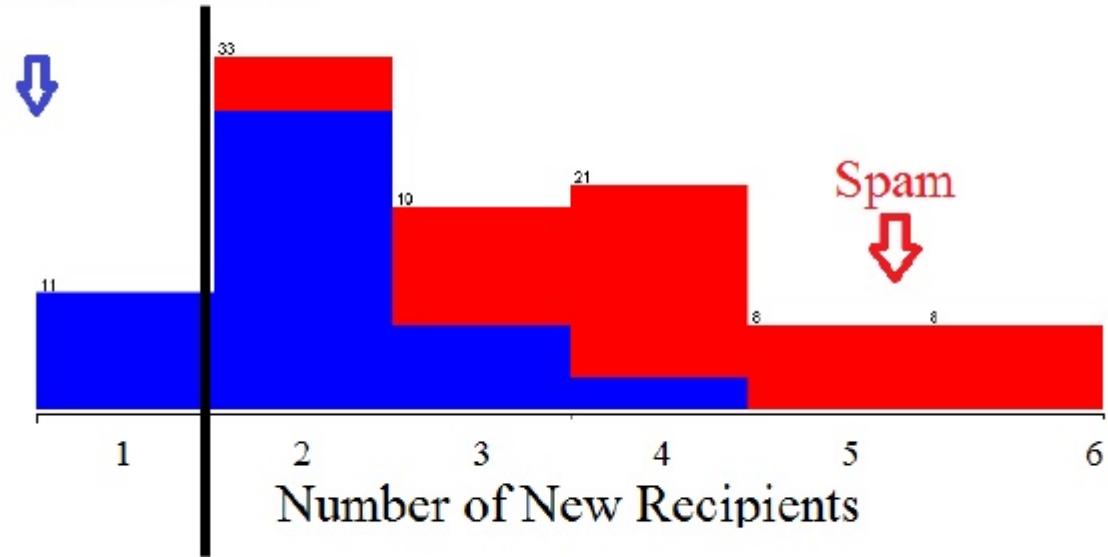
Numeric Nominal Ordinal

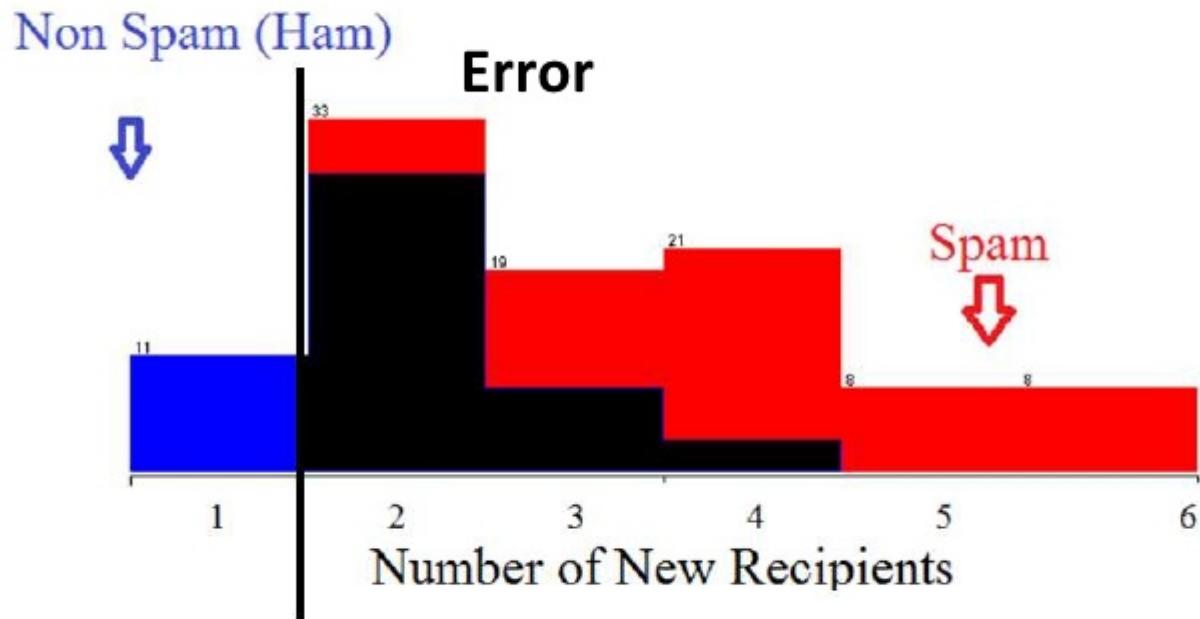
Diagram illustrating the classification of email instances based on input attributes and target attribute. Instances are grouped by a bracket labeled "Instances". Input attributes include Number of new Recipients (Numeric), Email Length (K) (Nominal), Country (IP) (Nominal), Customer Type (Ordinal), and Email Type (Nominal). A bracket under the first four columns is labeled "Input Attributes" and a bracket under the last column is labeled "Target Attribute". Below the table, arrows point from the attribute names to their respective data types: Number of new Recipients to "Numeric", Email Length (K) to "Nominal", and Customer Type to "Ordinal".

Non Spam (Ham)

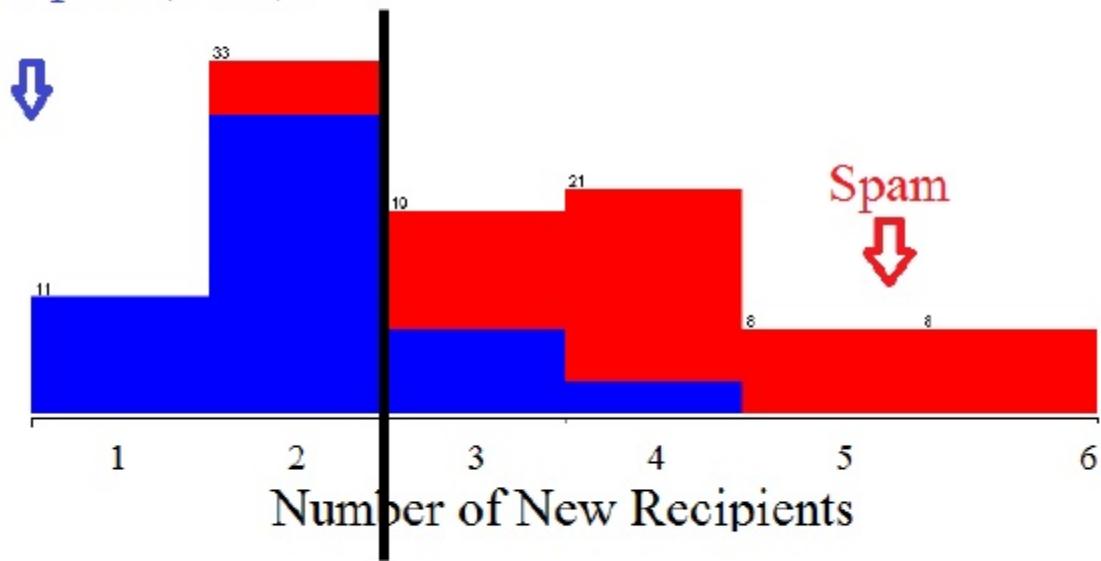


Non Spam (Ham)

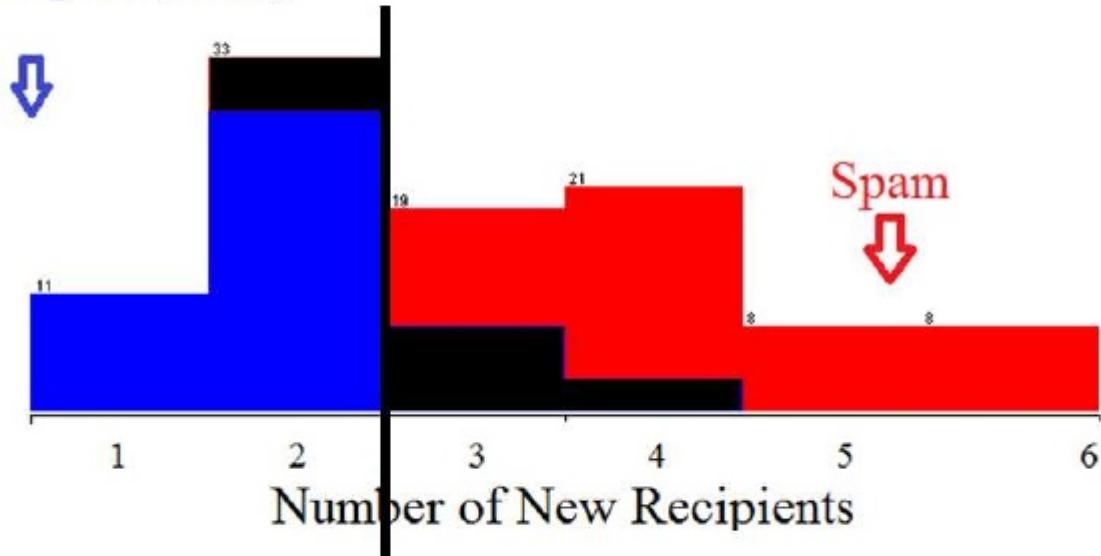


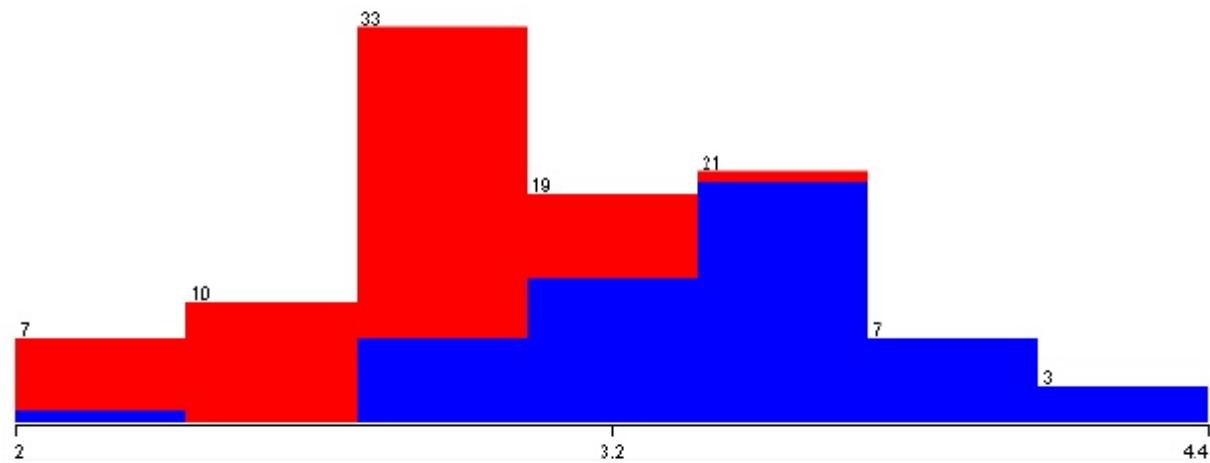


Non Spam (Ham)



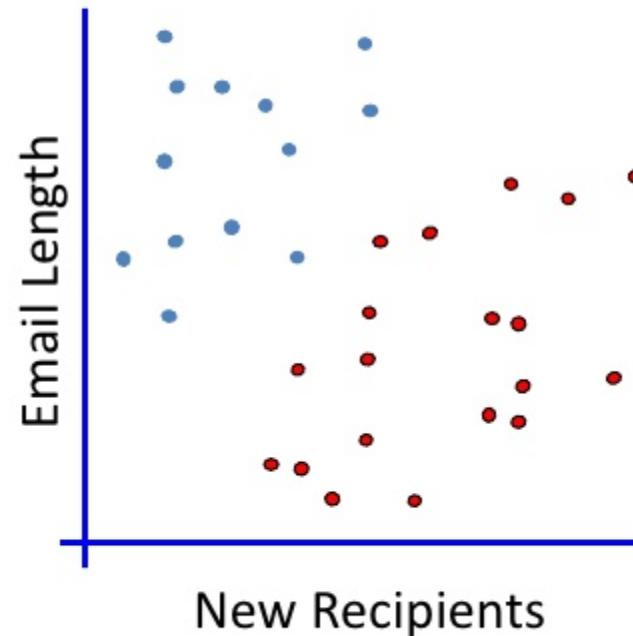
Non Spam (Ham)





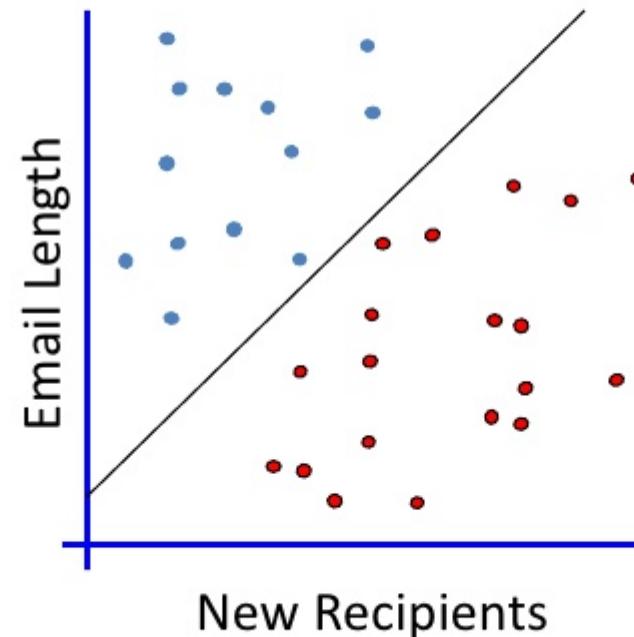
Email Length

Linear Classifiers



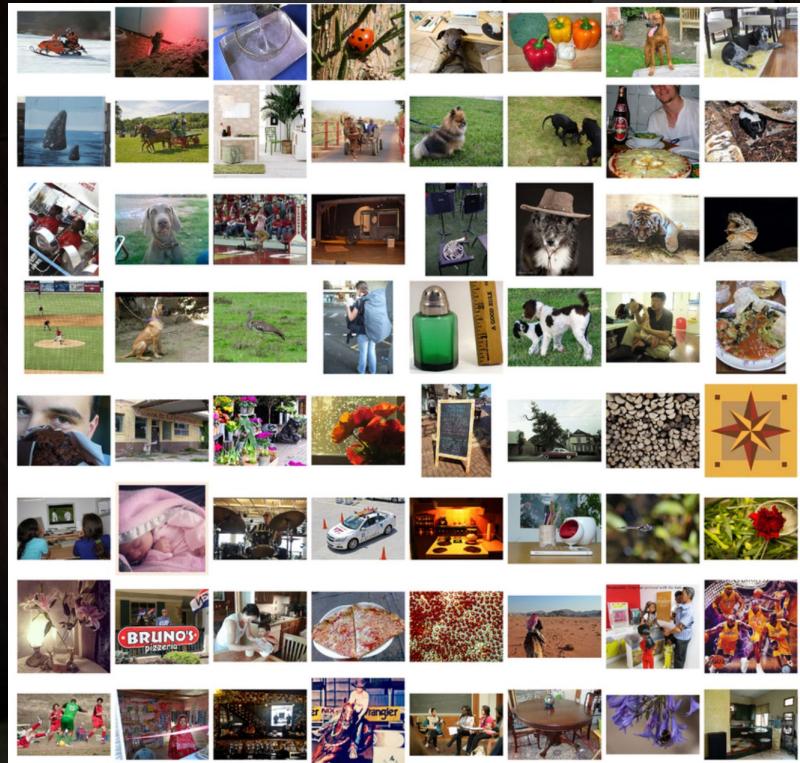
How would you
classify this data?

Linear Classifiers



How would you
classify this data?

Example 2: Image classification



Traditional approach

-1	0	+1
-2	0	+2
-1	0	+1

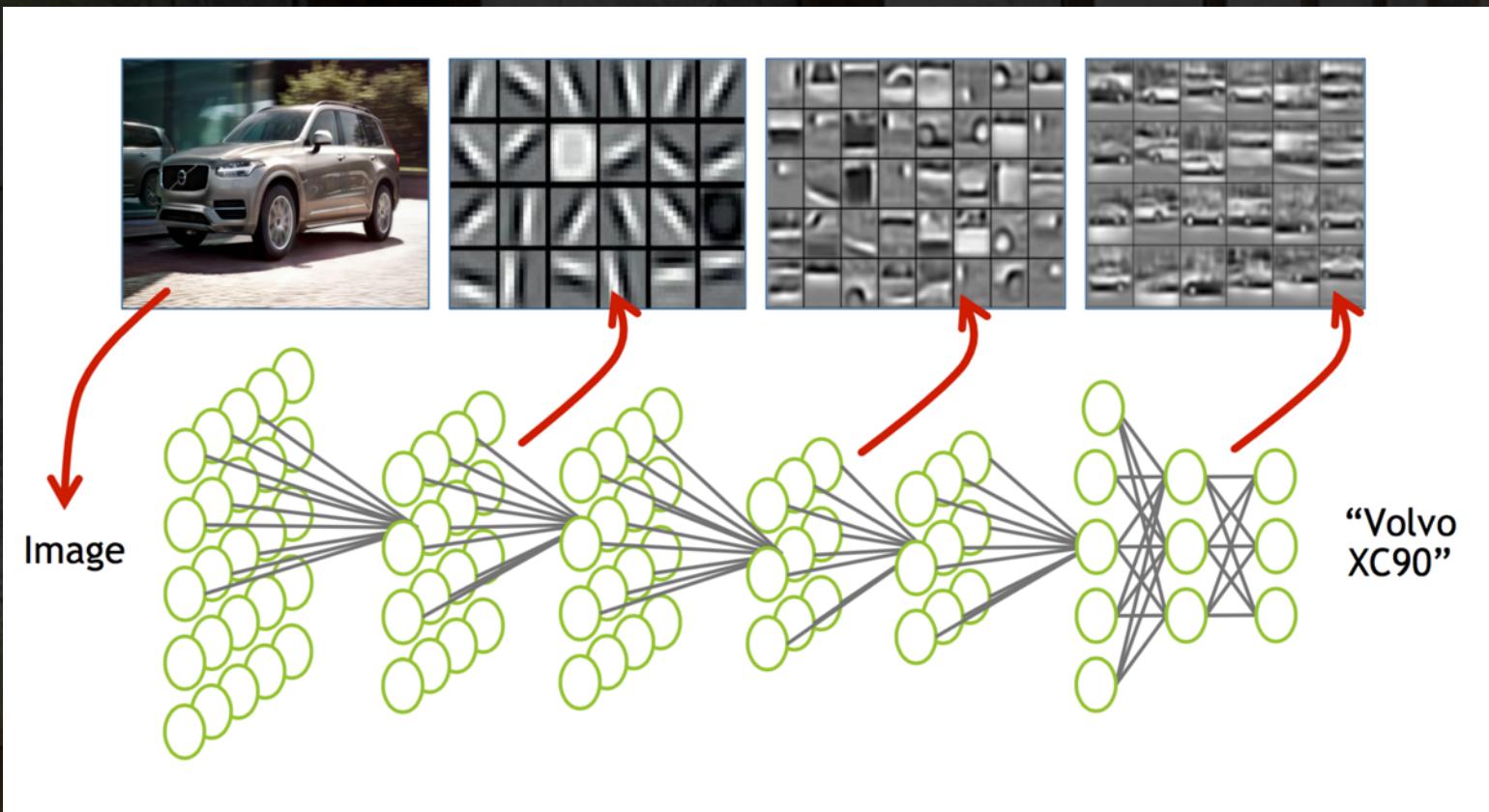
x filter

+1	+2	+1
0	0	0
-1	-2	-1

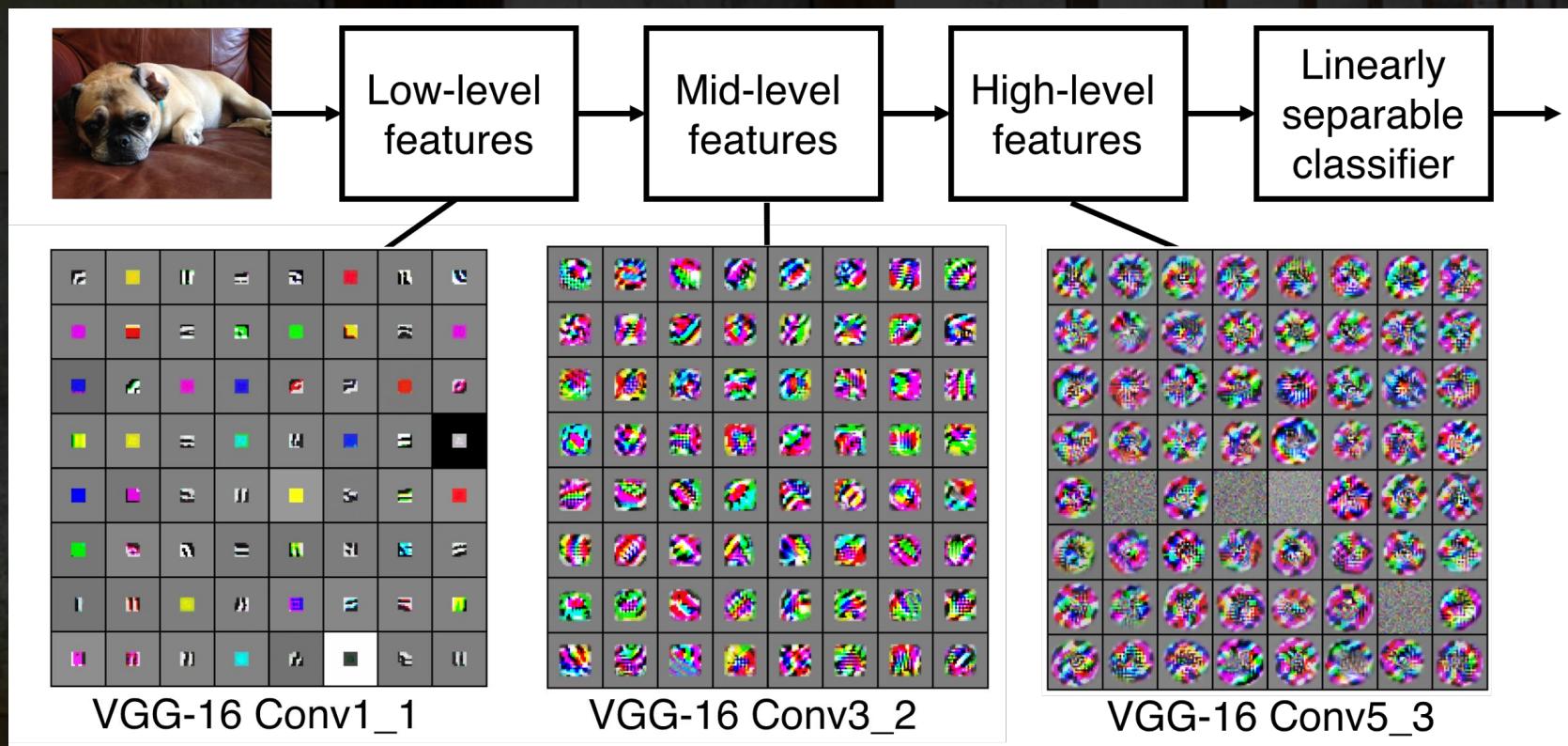
y filter



Neural networks



Convolutional neural networks



Čo stroje (zatiaľ) nevedia?

- Umelé neurónové siete stále omnoho menšie ako ľudský mozog:
 - potrebujeme ešte viac výpočtovej sily.
- Začínajú trochu rozumieť svetu (napr. že Rím je pre Taliansko rovnaké ako Paríž pre Francúzsko) ale porozumenie je stále plytké a krehké.
- Vstup: Zadanie úlohy z programátorskej súťaže.
Výstup: Efektívny program riešiaci danú úlohu.