Introduction to Deep Learning

Inna Skarga-Bandurova
Computer Science and Engineering Depatrment
V. Dahl East Ukrainian National University

Funded by the
Erasmus+ Programme
of the European Union

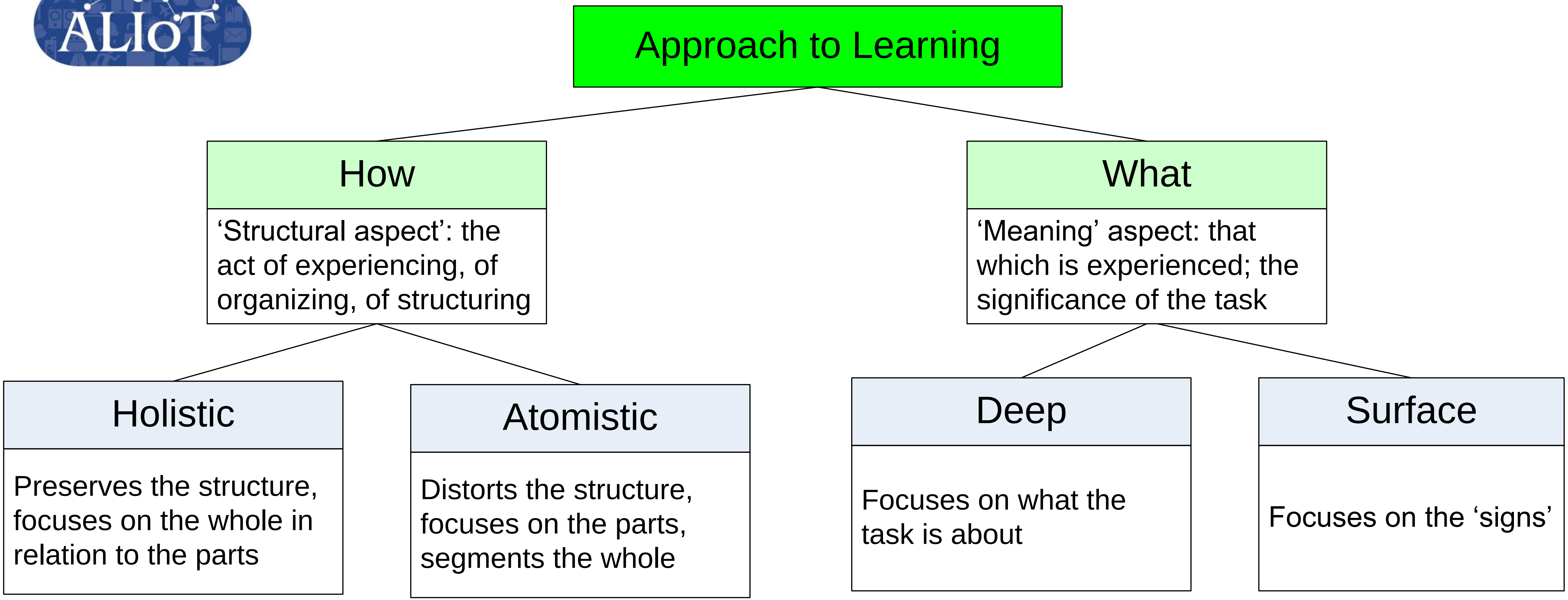
Kharkiv, 17-19 April 2018

Structure of Lectures

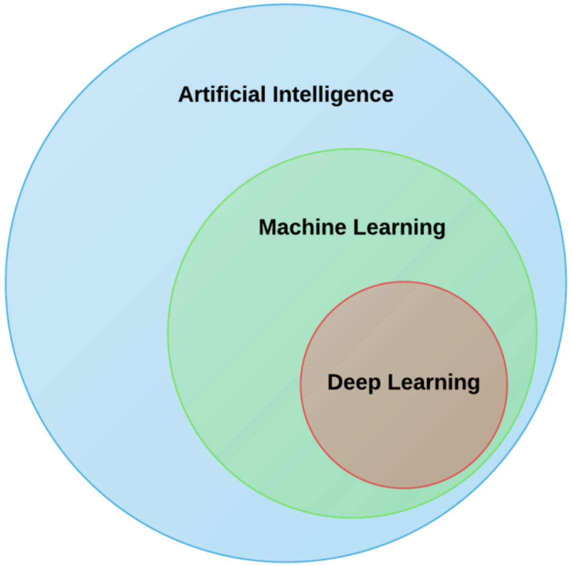
- **Today:** Introduction to Deep Learning
  - Approaches to learning
  - DL taxonomy and key principles of DL
  - IoT applications and foundation services
  - A general stack of DL models as a service in the cloud platforms
  - Common Data Sets for Deep Learning in IoT
- **Tomorrow:** Deep Learning and Recommendation Systems
- **Overmorrow:** Deep Learning for Human-Computer Interaction

This is a lecture series about the challenges (and new opportunities) for ML/DL

Funded by the Erasmus+ Programme of the European Union


Approach to Learning
How
'Structural aspect': the act of experiencing, of organizing, of structuring
What
'Meaning' aspect: that which is experienced; the significance of the task
Holistic
Preserves the structure, focuses on the whole in relation to the parts
Atomistic
Distorts the structure, focuses on the parts, segments the whole
Deep
Focuses on what the task is about
Surface
Focuses on the 'signs'

Deep Learning Taxonomy



Artificial Intelligence

Machine Learning

Other methods

Supervised Learning (labeled data: classification)

Unsupervised Learning (unlabeled data: pattern recognition)

Reinforcement Learning

Deep Learning (5-20 layers)

Neural Nets (NN)

Other methods

Convolutional Nets (images)

Recurrent Nets (text, speech)

Shallow Learning (1-2 layers)

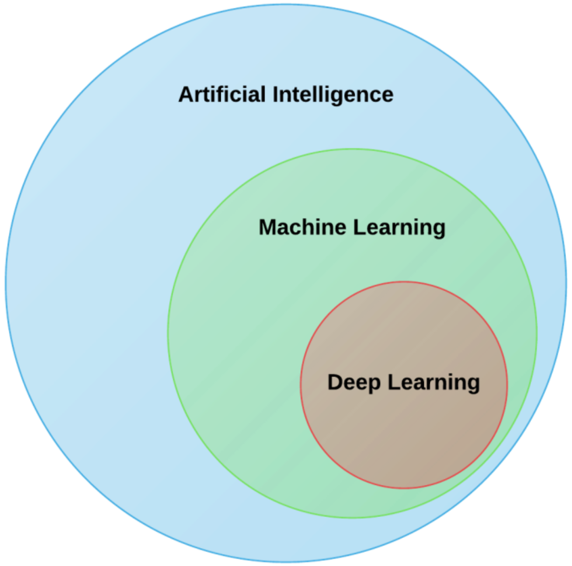
Bayesian inference

Support Vector Machines

Decision trees

K-means, K-nearest neighbor

Deep Learning Taxonomy



Artificial Intelligence

Machine Learning

Other methods

Supervised Learning (labeled data: classification)

Unsupervised Learning (unlabeled data: pattern recognition)

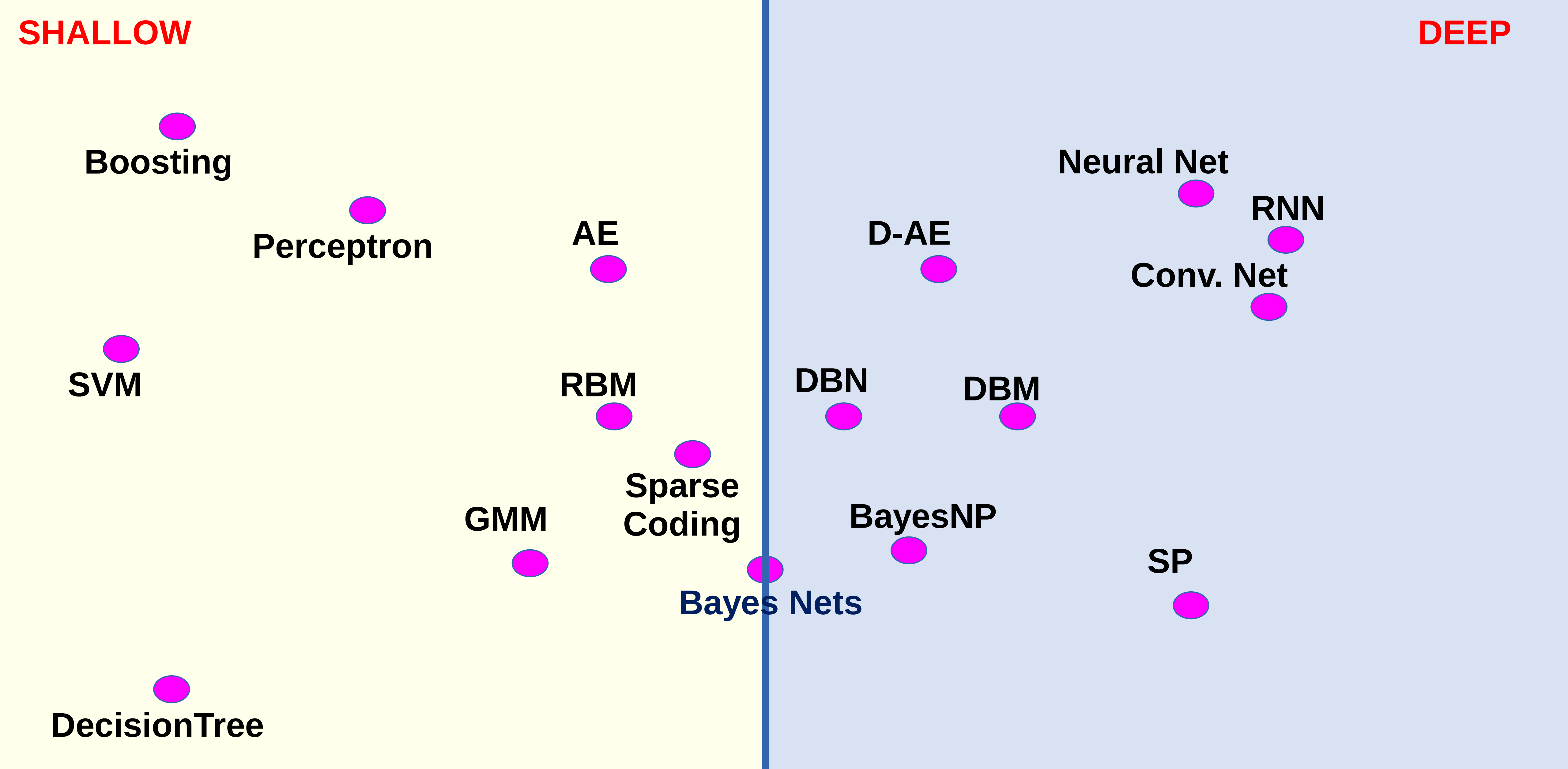
Reinforcement Learning

Deep Learning (5-20 layers)

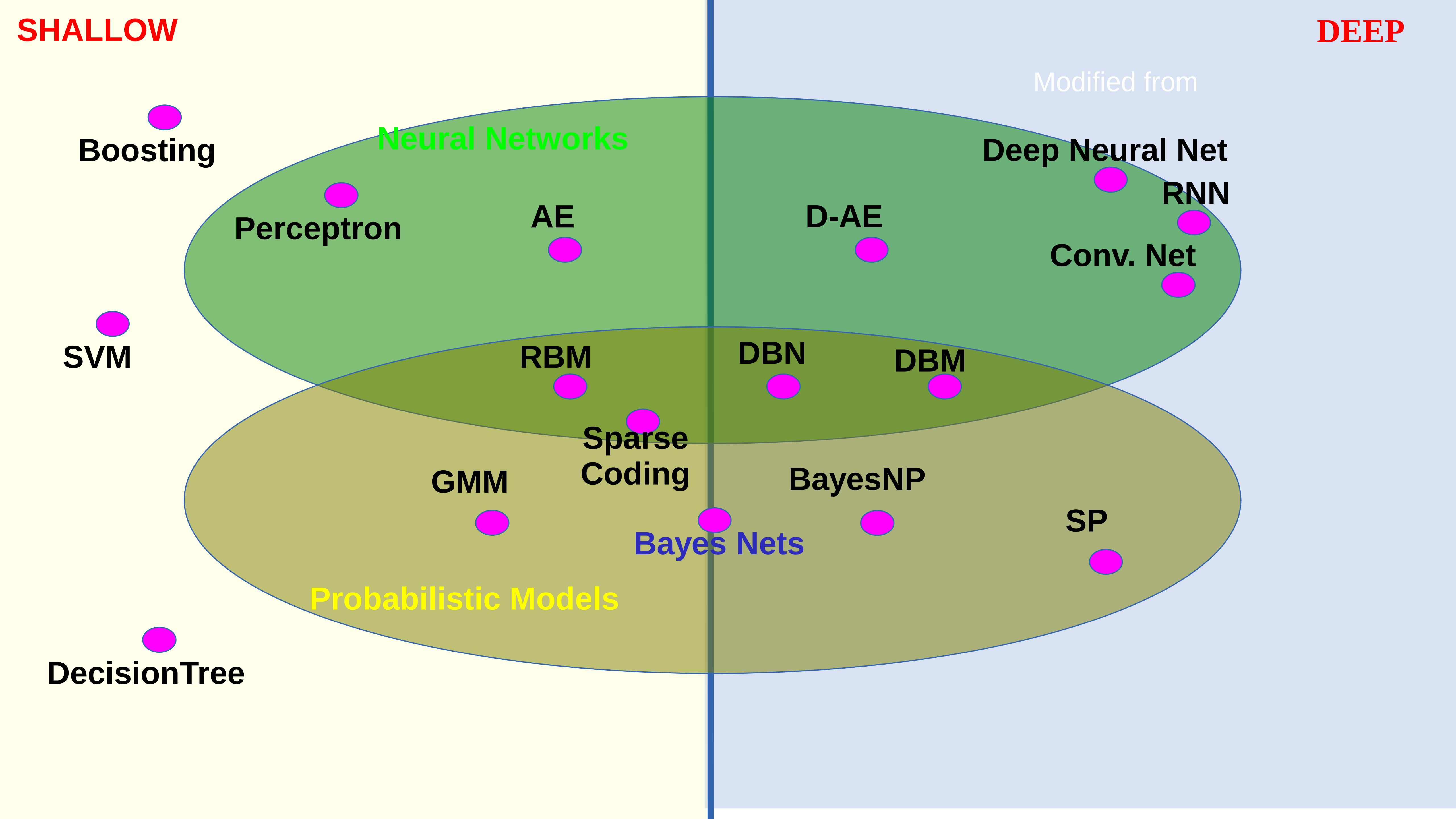
- Neural Nets (NN)
- Other methods
- Convolutional Nets (images)
- Recurrent Nets (text, speech)

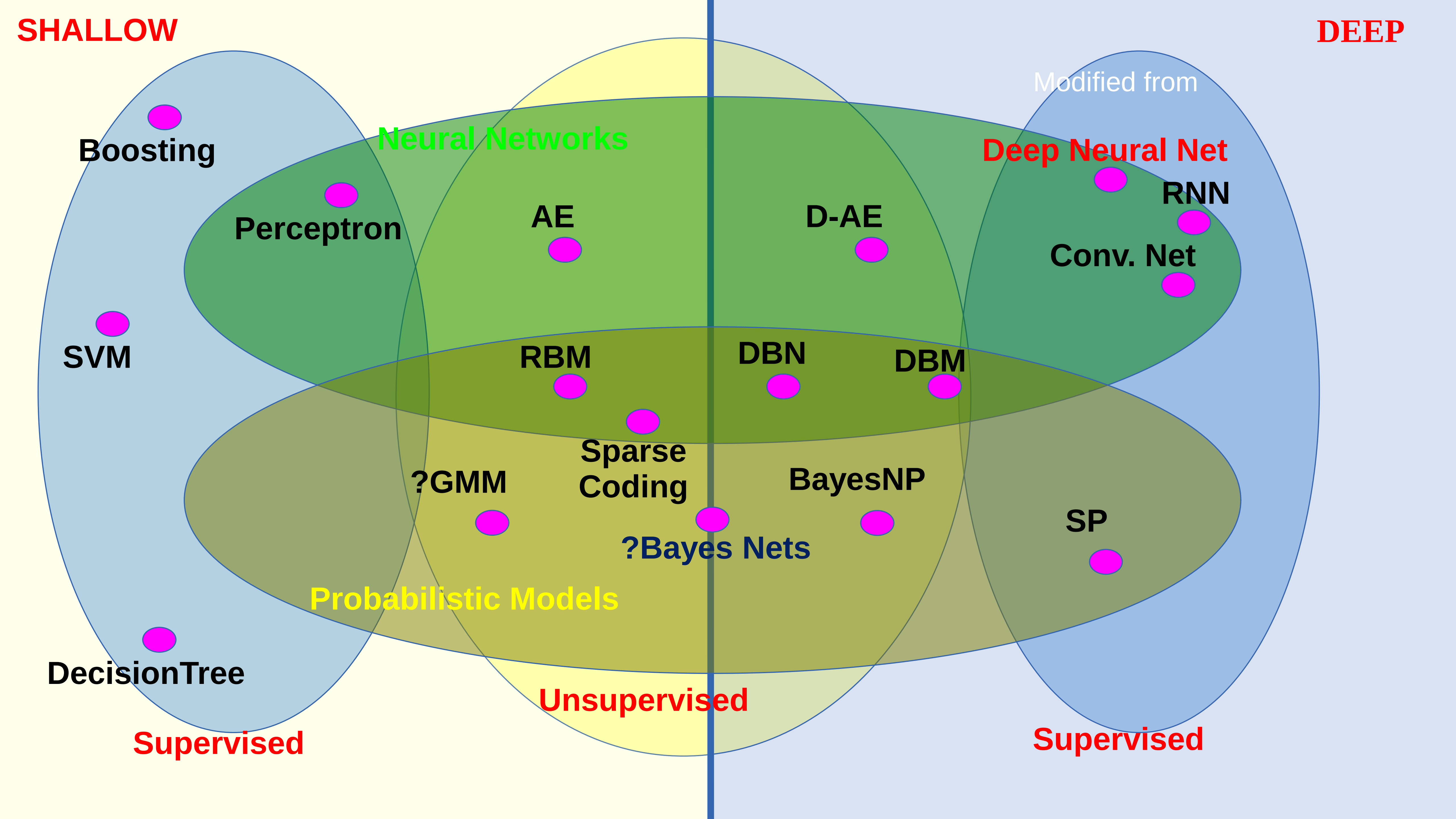
Shallow Learning (1-2 layers)

- Bayesian inference
- Support Vector Machines
- Decision trees
- K-means, K-nearest neighbor



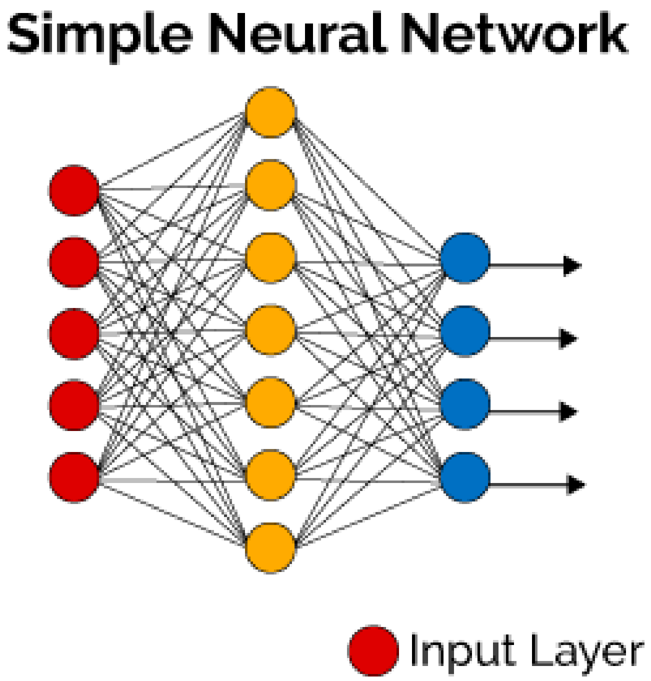
Modified from MA Ranzato

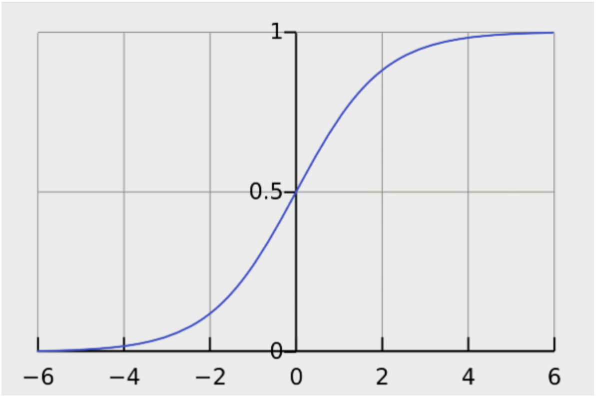

SHALLOW
DEEP
Modified from
Boosting
Neural Networks
Deep Neural Net
RNN
Perceptron
AE
D-AE
Conv. Net
SVM
RBM
DBN
DBM
Sparse
Coding
GMM
BayesNP
SP
Bayes Nets
Probabilistic Models
DecisionTree



SHALLOW
DEEP
Modified from
Neural Networks
Deep Neural Net
Boosting
Perceptron
AE
D-AE
RNN
Conv. Net
SVM
RBM
DBN
DBM
Sparse Coding
?GMM
BayesNP
SP
?Bayes Nets
Probabilistic Models
DecisionTree
Unsupervised
Supervised
Supervised

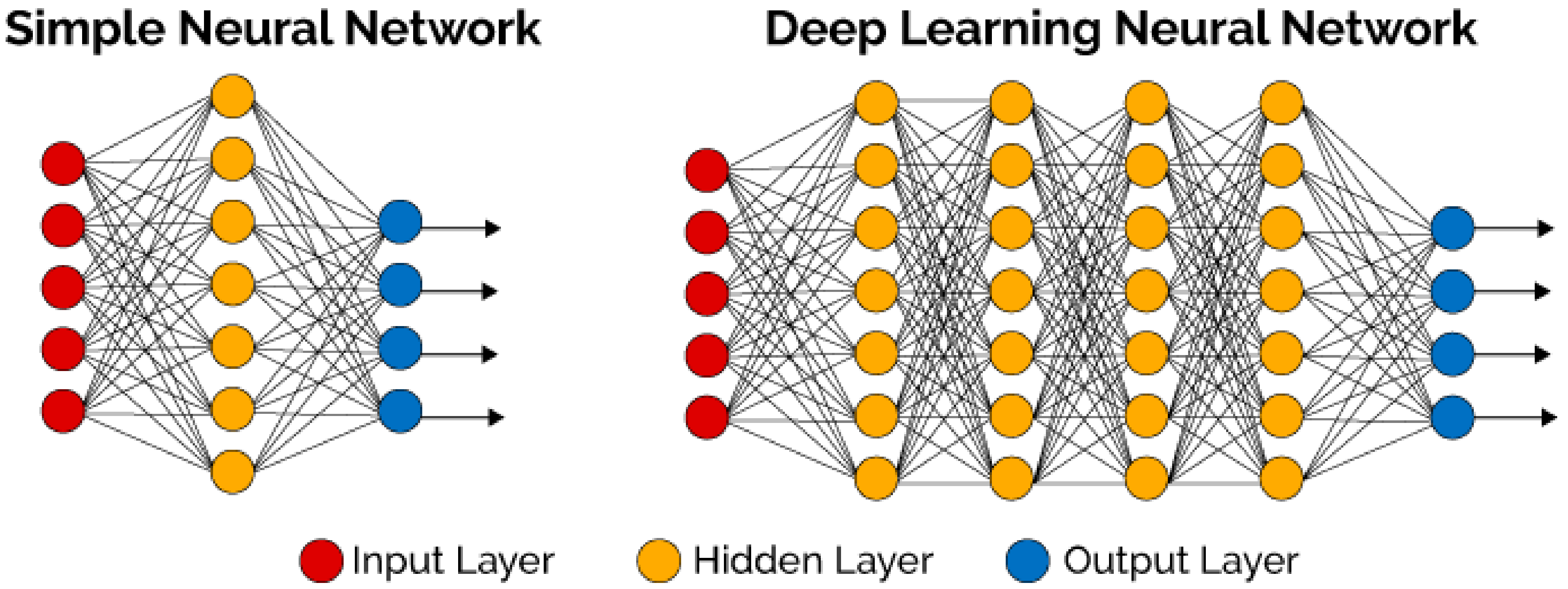
DL : 1970......





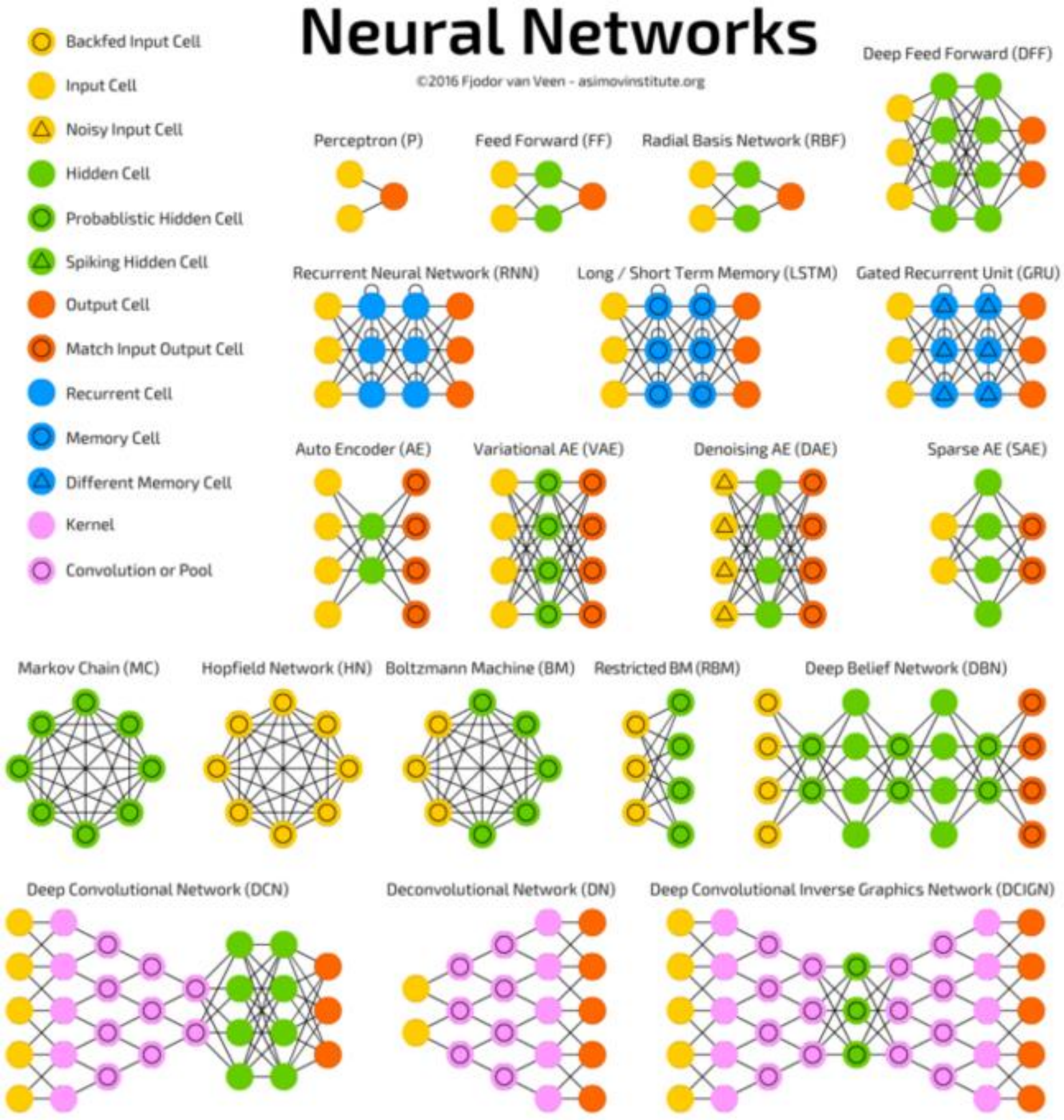
- 1970 “NN with hidden liyers" - DL
- 1990 - DL - math puzzle without practical purpose
- D. Hinton, R Salahutdinov
- 2010 - DL and big data

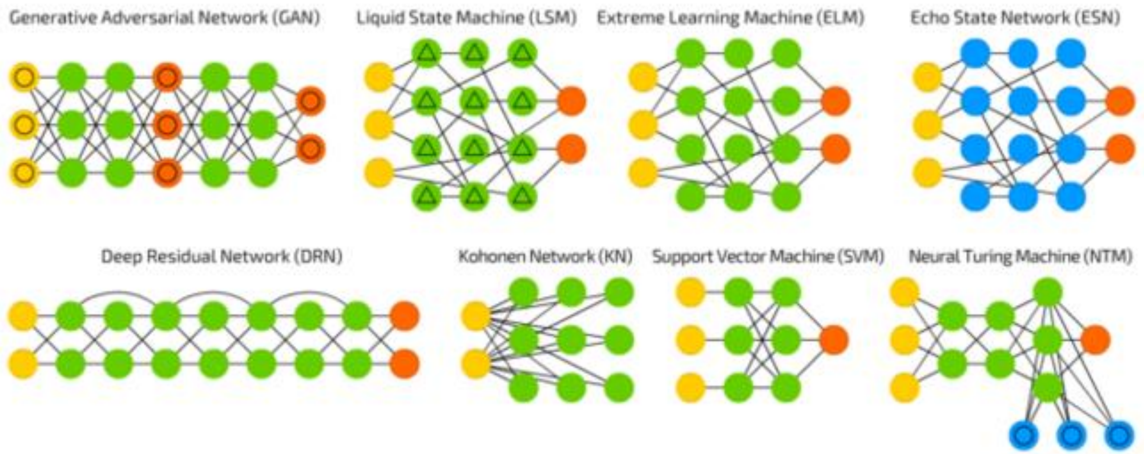
DL : 1970……



- 1970 “NN with hidden liyers" - DL
- **Deep Learning** - is a class of machine learning algorithms in the form of a neural network that uses a cascade of layers (tiers) of processing units to extract features from data and make predictive guesses about new data.
- 1990 - DL - math puzzle without practical purpose
- D. Hinton, R Salahutdinov 2010 - DL and big data

A mostly complete chart of NN

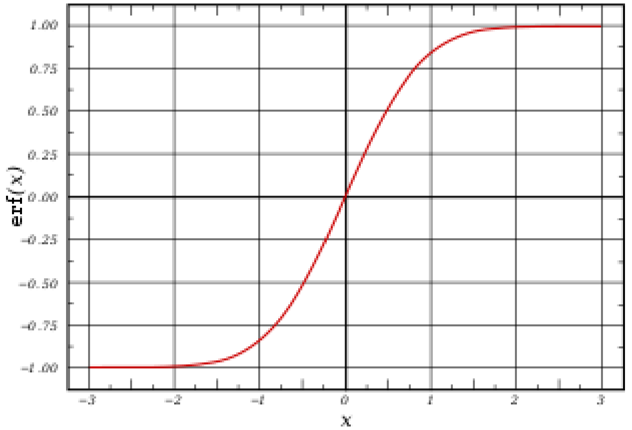




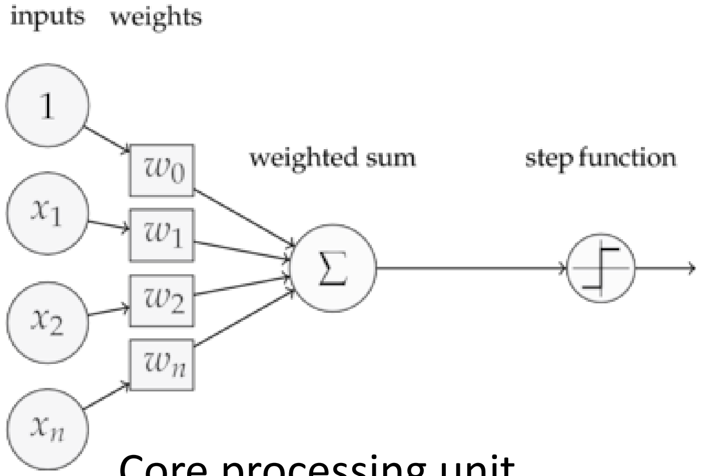
https://tproger.ru/translations/neural-network-zoo-1/

Three Key Technical Principles of DL

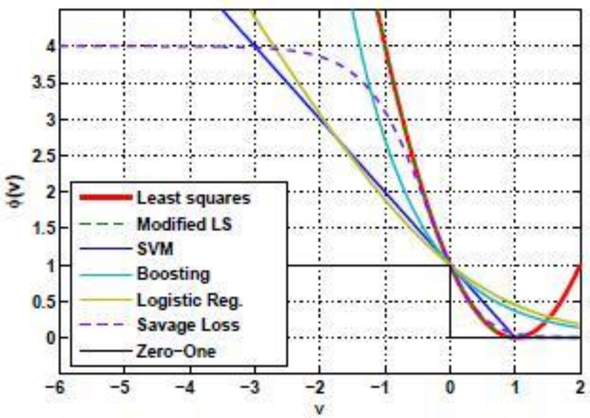
Sigmoid Function



Perceptron Structure



Loss Function



What

Squash values into probability function (Sigmoid (0-1); Tanh ((-1)-1))

Core processing unit input-processing-output) Levers: weights and bias

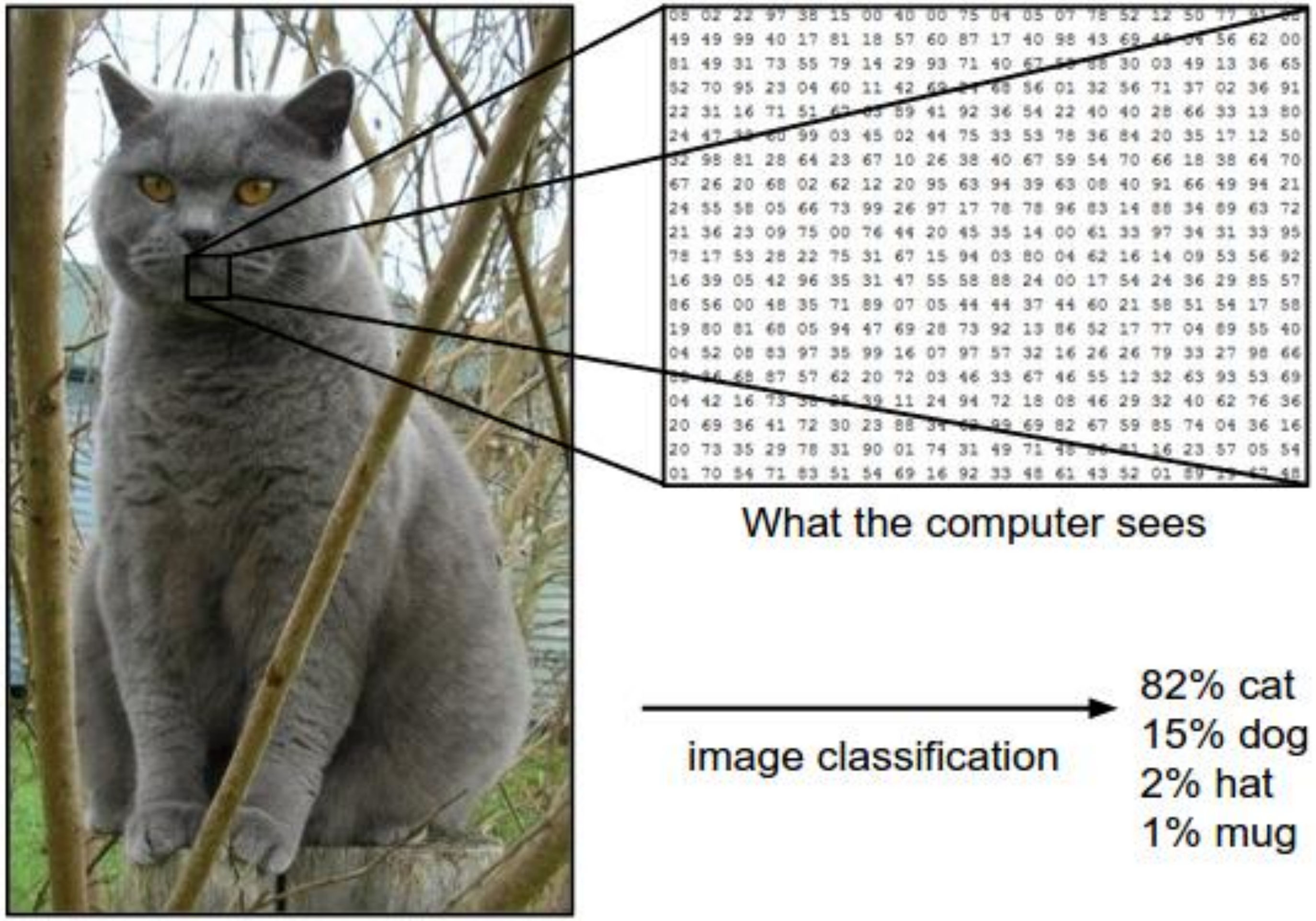
Reduce combinatoric dimensionality

Why

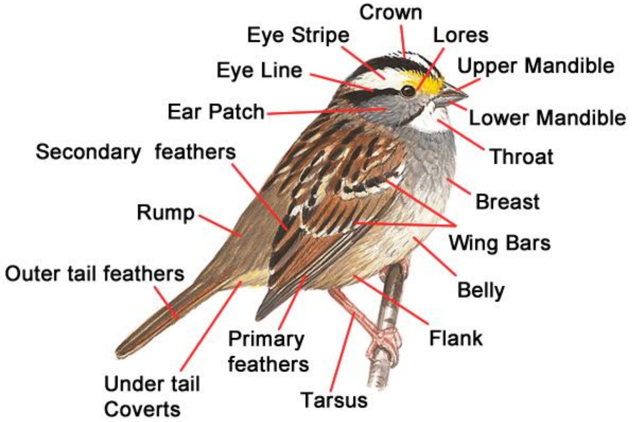
Formulate as a logistic regression problem for greater mathematical manipulation

"Dumb" system learns by adjusting parameters and checking against outcome

Loss function optimizes efficiency of solution



- keep in mind that to a computer an image is represented as one large 3-dimensional array of numbers.
- 248 pixels wide, 400 pixels tall, and has three color channels Red, Green, Blue (RGB).
- Therefore, the image consists of 248 x 400 x 3 numbers, or a total of 297,600 numbers.
- Each number is an integer that ranges from 0 (black) to 255 (white).
- Our task is to turn this quarter of a million numbers into a single label, such as *"cat"*.



Rule based systems: Input → Manual Model Deveopment → Вихід

Classical Machine Learning: Input → Manual Feature Extraction → Feature → Output

Representation Learning: Input → Features (Fe) → Feature Representation → Output

Deep Learning: Input → Simple Feature → Complex Features → Feature Representation → Output

Supervised Learning

Given a set of labeled training examples: $\{\mathbf{x}^{(t)}, y^{(t)}\}$, we perform Empirical Risk Minimization:

$$\arg\min_{\boldsymbol{\theta}} \frac{1}{T} \sum_t \underbrace{l(f(\mathbf{x}^{(t)};\boldsymbol{\theta}), y^{(t)})}_{\text{Loss function}} + \lambda\Omega(\boldsymbol{\theta})$$

where

$f(\mathbf{x}^{(t)};\theta)$ (non-linear) function mapping inputs to outputs, parameterized by $\theta \rightarrow$ Non-convex optimization

$l(\mathbf{f}(\mathbf{x}^{(t)};\boldsymbol{\theta}), y^{(t)})$ is the loss function.

Supervised Learning

Given a set of labeled training examples: $\{\mathbf{x}^{(t)}, y^{(t)}\}$, we perform Empirical Risk Minimization:

$$\arg\min_{\boldsymbol{\theta}} \frac{1}{T} \sum_t \underbrace{l(f(\mathbf{x}^{(t)}; \boldsymbol{\theta}), y^{(t)})}_{\text{Loss function}} + \underbrace{\lambda\Omega(\boldsymbol{\theta})}_{\text{Regularizer}}$$

where

$\arg\min_{\boldsymbol{\theta}} \frac{1}{T} \sum_t l(f(\mathbf{x}^{(t)}; \boldsymbol{\theta}), y^{(t)}) + \lambda\Omega(\boldsymbol{\theta})$ (non-linear) function mapping inputs to outputs, parameterized by θ → Non-convex optimization

$l(\mathbf{f}(\mathbf{x}^{(t)}; \boldsymbol{\theta}), y^{(t)})$ is the loss function.

$\Omega(\boldsymbol{\theta})$ is a regularization term.

Supervised Learning

Given a set of labeled training examples: $\{\mathbf{x}^{(t)}, y^{(t)}\}$, we perform Empirical Risk Minimization:

$$\arg\min_{\boldsymbol{\theta}} \frac{1}{T} \sum_t \underbrace{l(f(\mathbf{x}^{(t)}; \boldsymbol{\theta}), y^{(t)})}_{\text{Loss function}} + \underbrace{\lambda\Omega(\boldsymbol{\theta})}_{\text{Regularizer}}$$

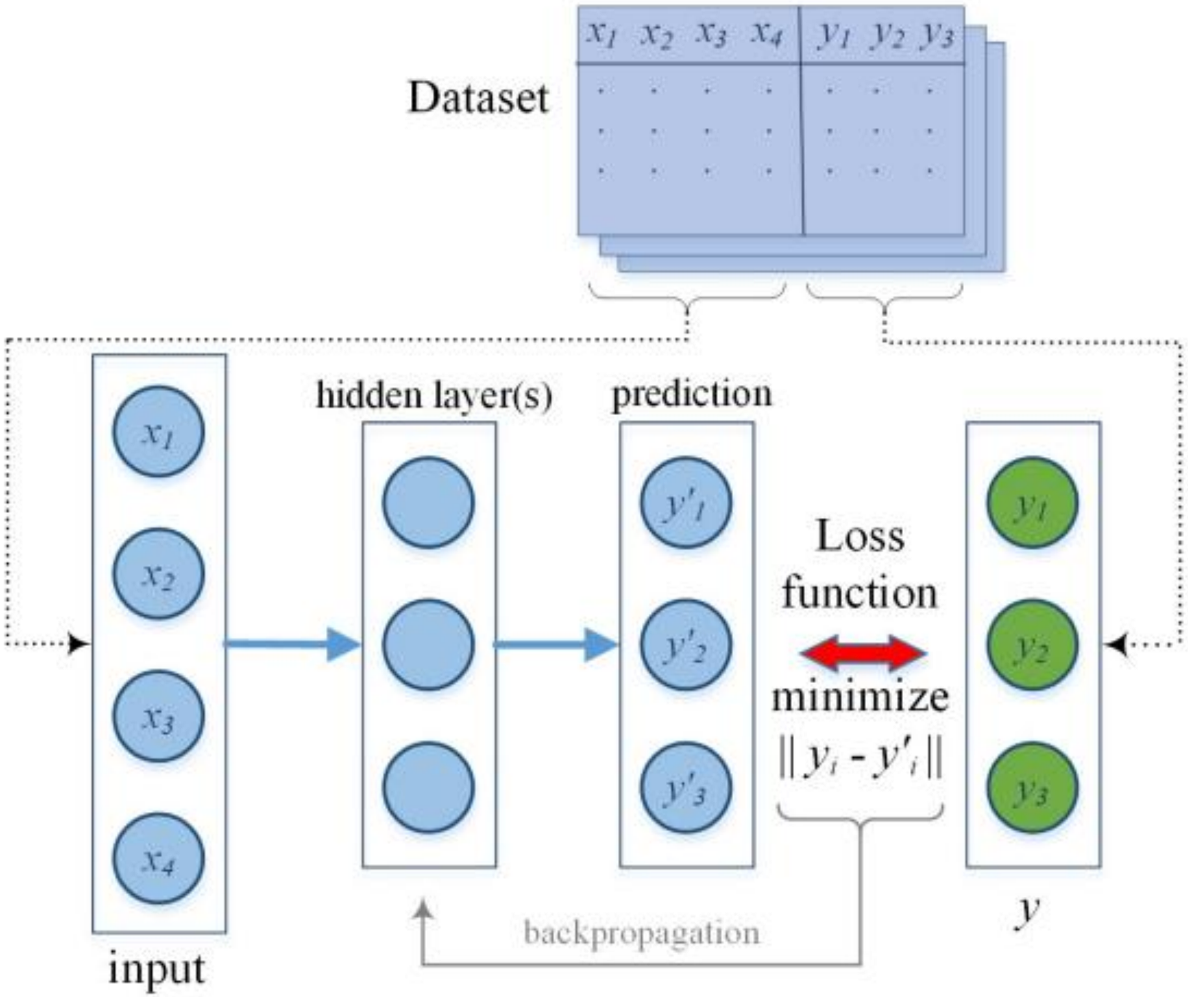
Learning is cast as optimization.

For classification problems, we would like to minimize classification error.

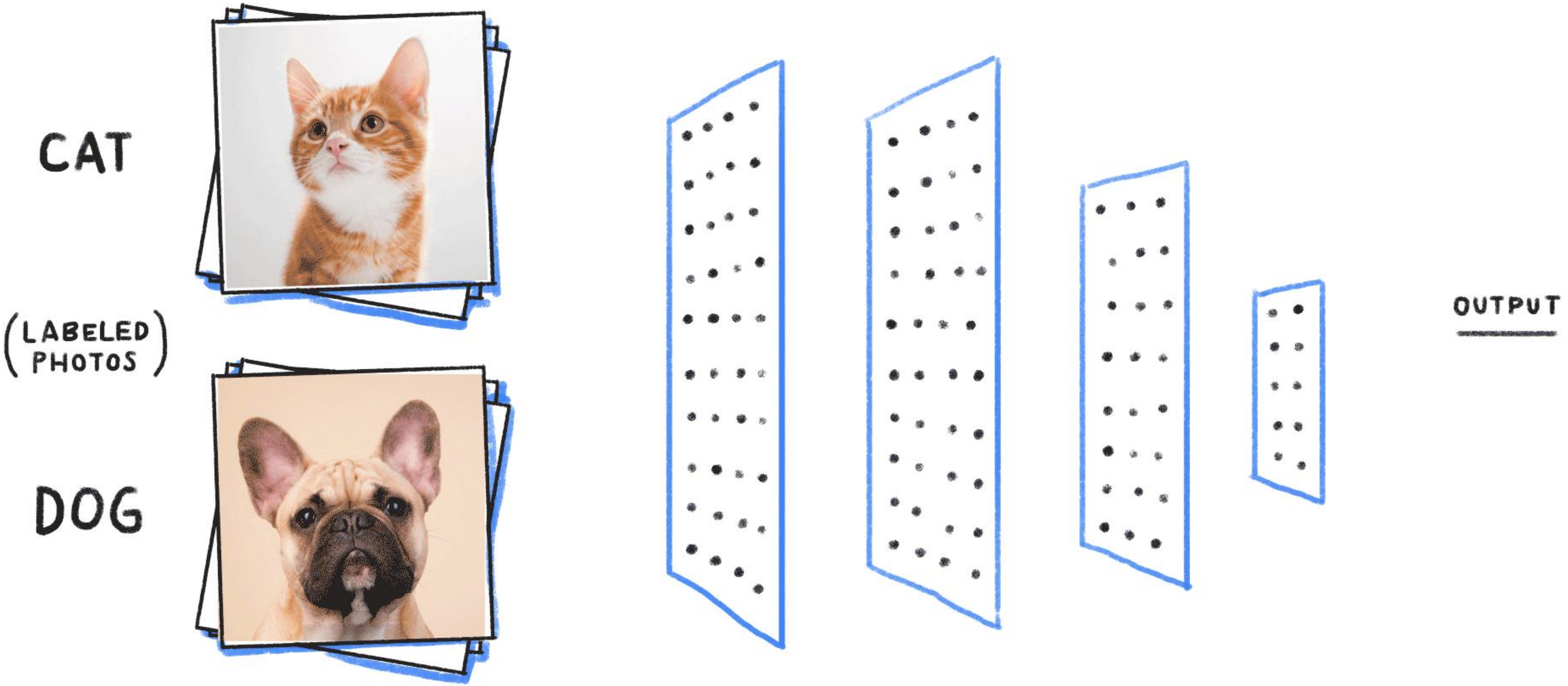
Loss function can sometimes be viewed as a surrogate for what we want to optimize (e.g. upper bound)

The overall mechanism of training of a DL model

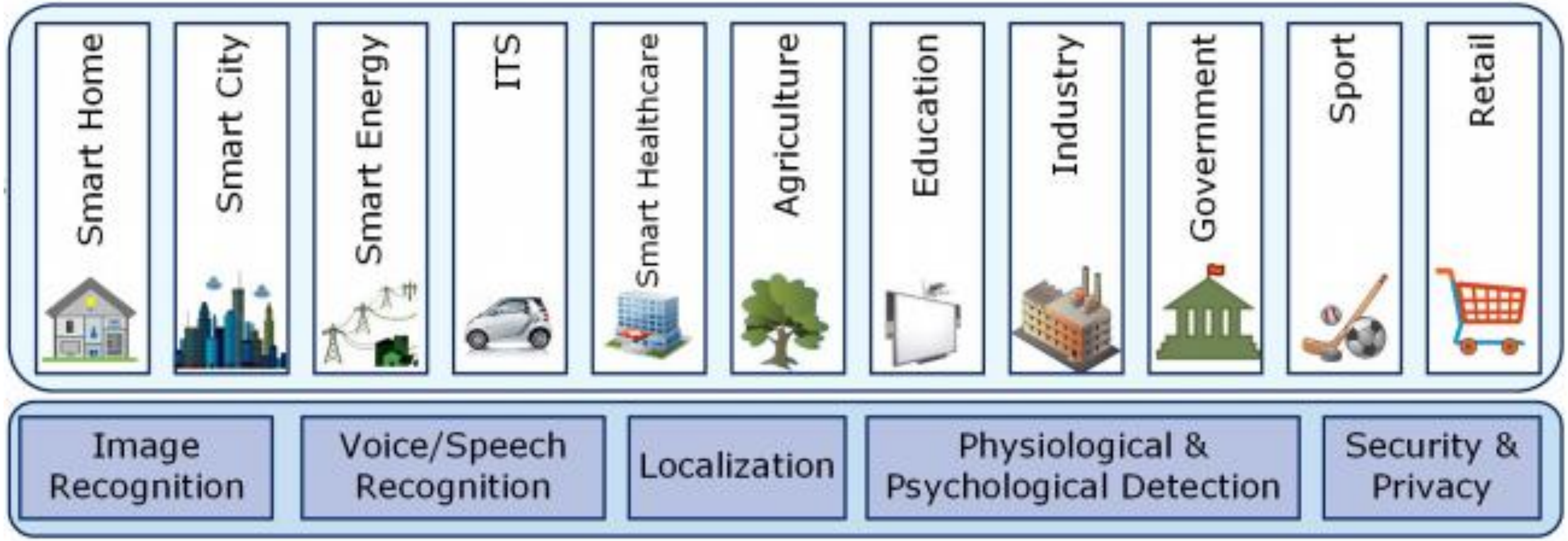
- The input layer assigns (usually randomly) weights to the input training data and passes it to the next layer.
- Each subsequent layer also assigns weights to their input and produces their output, which serves as the input for the following layer.
- At the last layer, the final output representing the model prediction will be produced.
- A loss function determines how right or wrong is this prediction by computing the error rate between the prediction and true value.
- The error rate is propagated back across the network to the input layer.
- The network then repeats this training cycle, after balancing the weights on each neuron in each cycle, until the error rate falls below a desired threshold.



They are very expensive to train, but once the training is finished it is very cheap to classify a new test example.


CAT
(LABELED PHOTOS)
DOG
OUTPUT

IoT Applications and Foundation Services

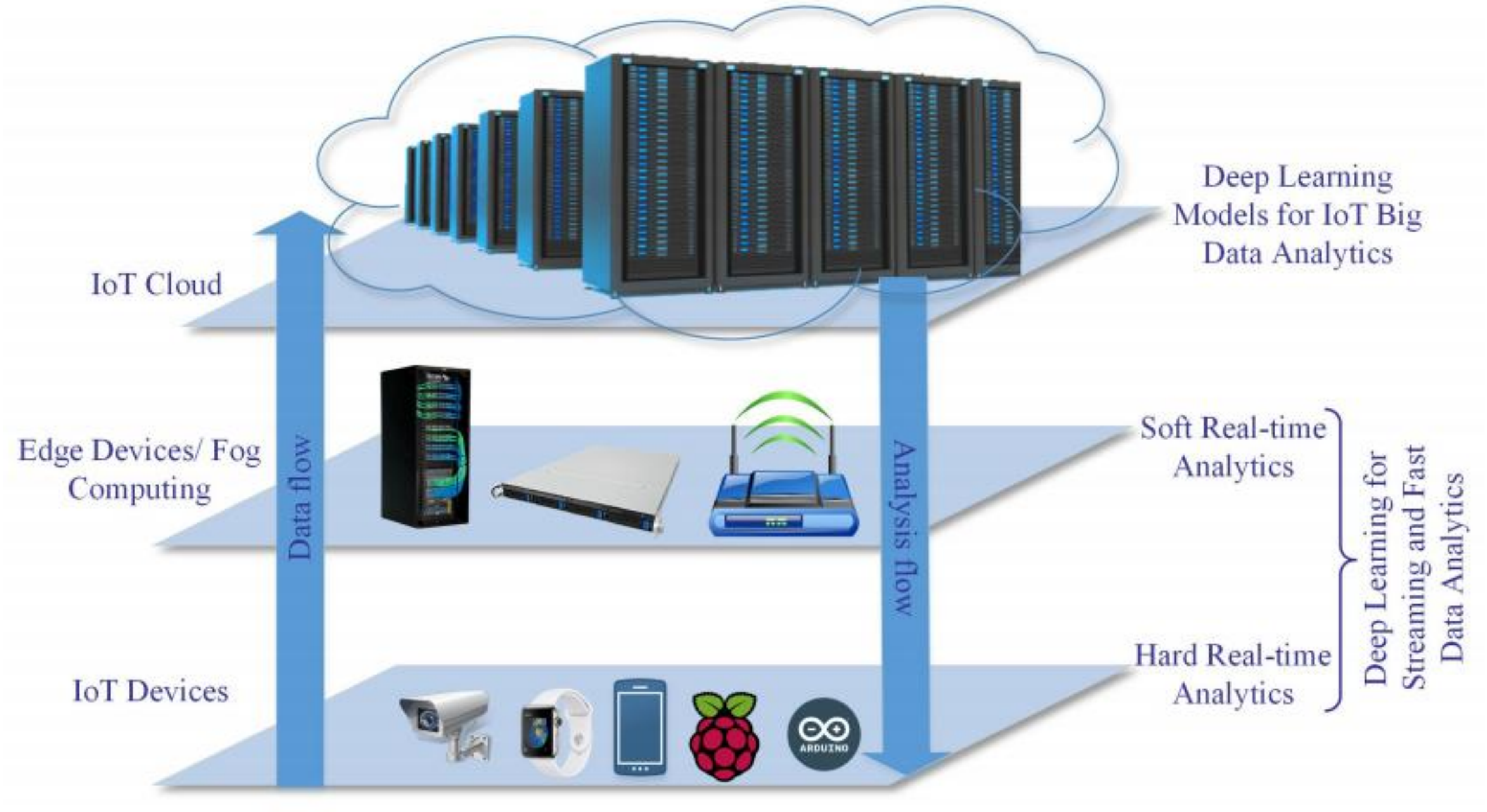


[Mohammadi M, 2017]

Typical IoT-Based Services in Smart City

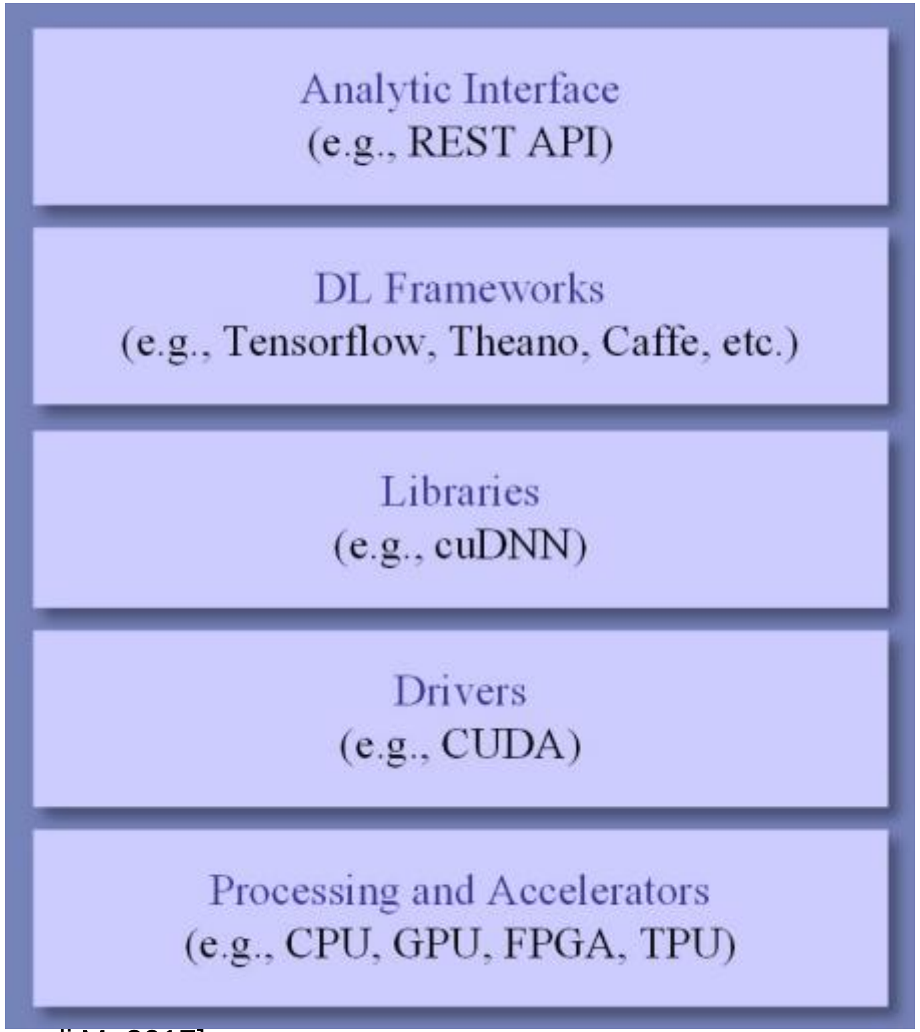
Service	Reference	Input data	DL architecture
Crowd density / transportation prediction	X. Song, H. Kanasugi, and R. Shibasaki, "Deeptransport: Prediction and simulation of human mobility and transportation mode at a citywide level." IJCAI, 2016.	GPS / transition mode	LSTM
	V. C. Liang, R. T. Ma, W. S. Ng, L. Wang, M. Winslett, H. Wu, S. Ying, and Z. Zhang, "Mercury: Metro density prediction with recurrent neural network on streaming cdr data," in Data Engineering (ICDE), 2016 IEEE 32nd International Conference on. IEEE, 2016, pp. 1374–1377.	Telecommunication data / CDR	RNN
Parking IoT	G. Amato, F. Carrara, F. Falchi, C. Gennaro, C. Meghini, and C. Vairo, "Deep learning for decentralized parking lot occupancy detection," Expert Systems with Applications, 2017. S. Valipour, M. Siam, E. Stroulia, and M. Jagersand, "Parking-stall vacancy indicator system, based on deep convolutional neural networks," in 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT), 2016, pp. 655–660	Images of parking spaces	CNN
Waste Management	G. Mittal, K. B. Yagnik, M. Garg, and N. C. Krishnan, "Spotgarbage: smartphone app to detect garbage using deep learning," in Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2016, pp. 940–945.	Garbage images	CNN

IoT Data Generation at Different Levels and DL Models to address their knowledge abstraction



[Mohammadi M et al., 2017]

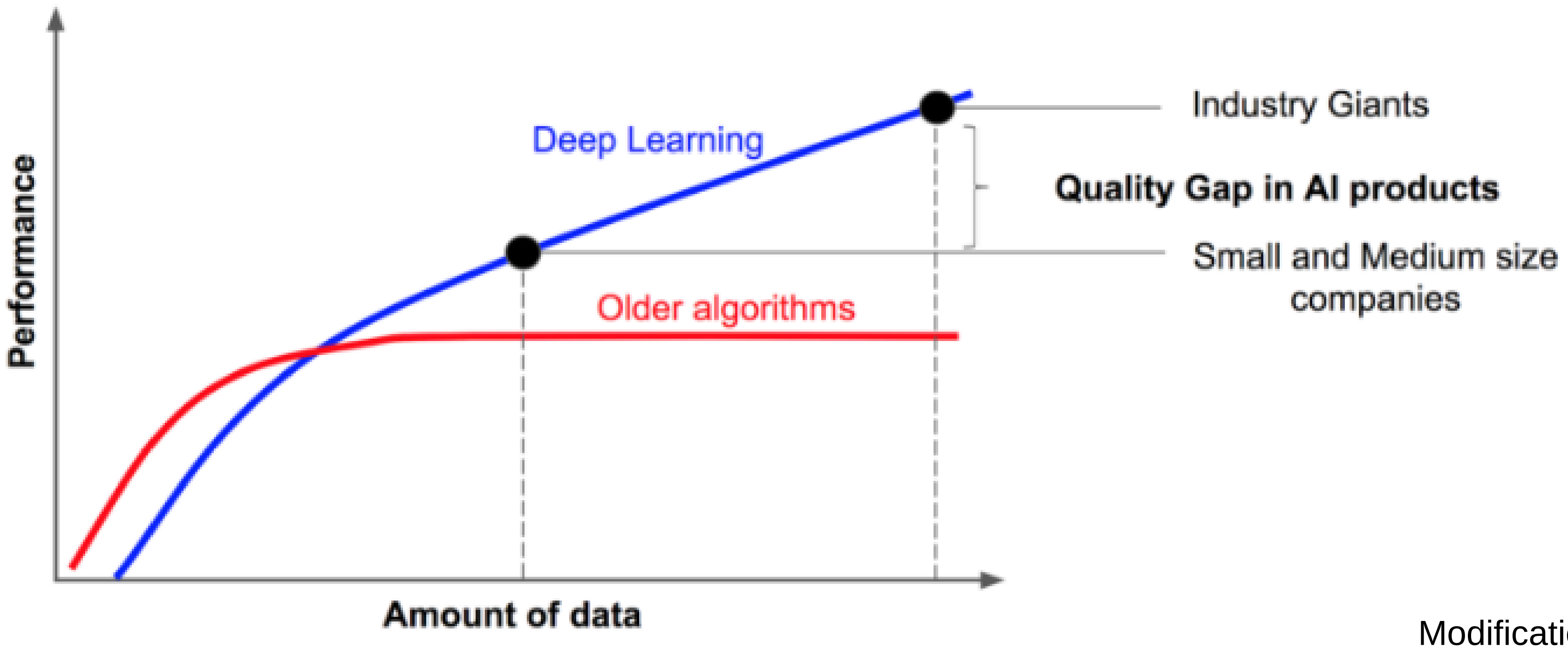
A general stack of DL models as a service in the cloud platforms



[Mohammadi M, 2017]

- Currently, the integration of DL analytics into IoT applications is limited to RESTful APIs, which are based on the HTTP protocol. While there exist several other application protocols that are extensively used in IoT applications, such as Message Queue Telemetry Transport (MQTT), CoAP, XMPP, AMQP, the integration of these protocols with the DL analytic interfaces calls for enhancing their compatibility with the aforementioned protocols to eliminate the need for message conversion proxies, which imposes extra overhead on the analytics response time
- There is a need for mechanisms and approaches to make DL models accessible through APIs, in order to be easily integrated into IoT applications.
- This aspect has not been investigated much, and only a few products are available, such as Amazon's AWS DL AMIs4 , Google cloud ML5 , and IBM Watson6 . This creates opportunities for cloud providers to offer "DL models as a service" as a new sub-category of Software as a Service (SaaS). However, this imposes several challenges for cloud providers, since DL tasks are computationally intensive and may starve other cloud services

Big Picture: More Data – Smarter AI



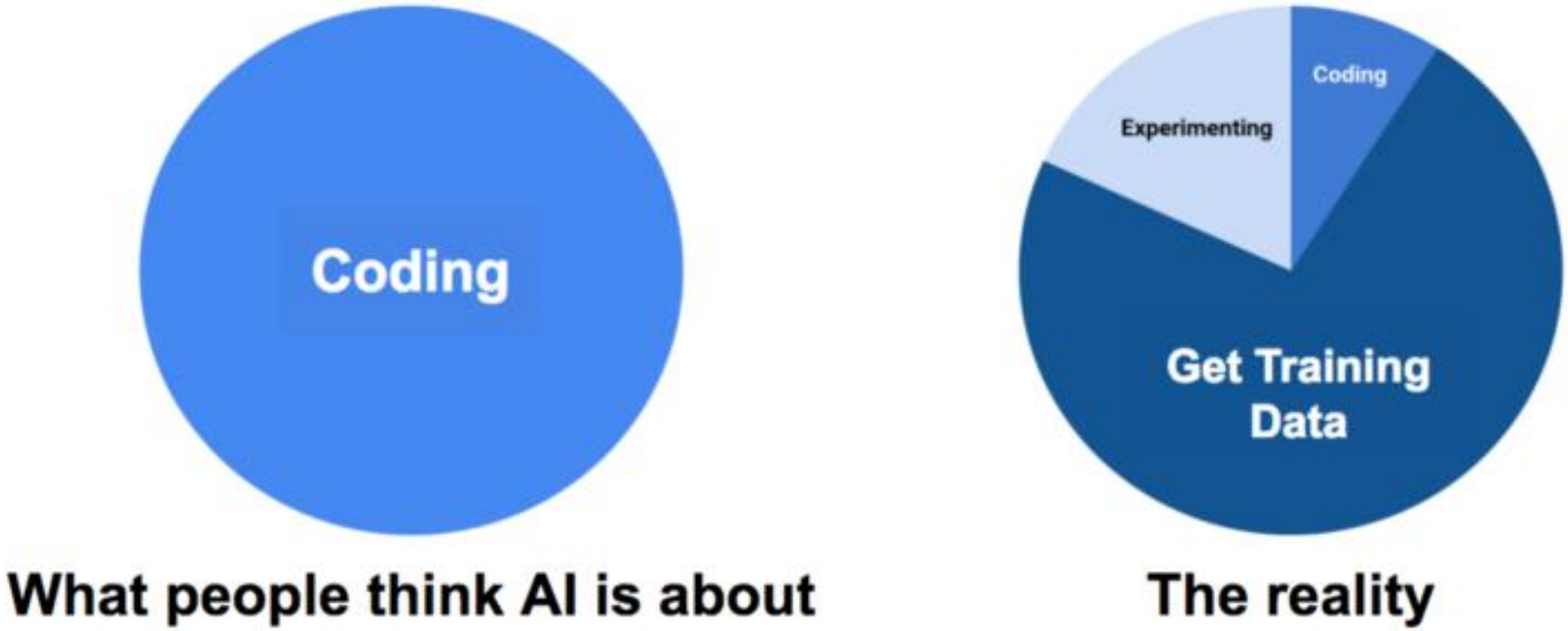
Modification from Andrew Ng.

Conclusion 0: AI products need data.
Conclusion 1: the more data we have—the smarter AI will be.
Conclusion 2: industry giants have much more data than others.
Conclusion 3: quality gap in AI products is defined by amount of data.

Funded by the Erasmus+ Programme of the European Union

https://hackernoon.com/%EF%B8%8F-big-challenge-in-deep-learning-training-data-31a88b97b282

Common mistake: DL is all about building neural nets



https://hackernoon.com/%EF%B8%8F-big-challenge-in-deep-learning-training-data-31a88b97b282

Common Data Sets for DL in IoT

Dataset Name	Domain	Provider	Notes	Address/Link
Taxi Service Trajectory	Transportation	Prediction Challenge, ECML PKDD 2015	Trajectories performed by all the 442 taxis running in the city of Porto, in Portugal.	http://www.geolink.pt/ecmlpkdd2015-challenge/dataset.html
GeoLife GPS Trajectories	Transportation	Microsoft	A GPS trajectory by a sequence of time-stamped points	https://www.microsoft.com/en-us/download/details.aspx?id=52367
T-Drive trajectory data	Transportation	Microsoft	Contains a one-week trajectories of 10,357 taxis	https://www.microsoft.com/en-us/research/publication/t-drive-trajectory-data-sample/
Chicago Bus Traces data	Transportation	M. Doering	Bus traces from the Chicago Transport Authority for 18 days with a rate between 20 and 40 seconds.	http://www.ibr.cs.tu-bs.de/users/mdoering/bustraces/
Uber trip data	Transportation	FiveThirty-Eight	About 20 million Uber pickups in New York City during 12 months.	https://github.com/fivethirtyeight/uber-tlc-foil-response
Traffic Sign Recognition	Transportation	K. Lim	Three datasets: Korean daytime, Korean nighttime, and German daytime traffic signs based on Vienna traffic rules.	https://figshare.com/articles/Traffic_Sign_Recognition_Testsets/4597795
DDD17	Transportation	J. Binas	End-To-End DAVIS Driving Dataset.	http://sensors.ini.uzh.ch/databases.html

[Mohammadi M et al., 2017]

Common Data Sets for Deep Learning in IoT

CityPulse Dataset Collection	Smart City	CityPulse EU FP7 project	Road Traffic Data, Pollution Data, Weather, Parking	http://iot.ee.surrey.ac.uk:8080/datasets.html
Open Data Institute - node Trento	Smart City	Telecom Italia	Weather, Air quality, Electricity, Telecommunication	http://theodi.fbk.eu/openbigdata/
Málaga datasets	Smart City	City of Malaga	A broad range of categories such as energy, ITS, weather, Industry, Sport, etc.	http://datosabiertos.malaga.eu/dataset
Gas sensors for home activity monitoring	Smart home	Univ. of California San Diego	Recordings of 8 gas sensors under three conditions including background, wine and banana presentations.	http://archive.ics.uci.edu/ml/datasets/Gas+sensors+for+home+activity+monitoring
CASAS datasets for activities of daily living	Smart home	Washington State University	Several public datasets related to Activities of Daily Living (ADL) performance in a two-story home, an apartment, and an office settings.	http://ailab.wsu.edu/casas/datasets.html
ARAS Human Activity Dataset	Smart home	Bogazici University	Human activity recognition datasets collected from two real houses with multiple residents during two months.	https://www.cmpe.boun.edu.tr/aras/
MERLSense Data	Smart home, building	Mitsubishi Electric Research Labs	Motion sensor data of residual traces from a network of over 200 sensors for two years, containing over 50 million records.	http://www.merl.com/wmd

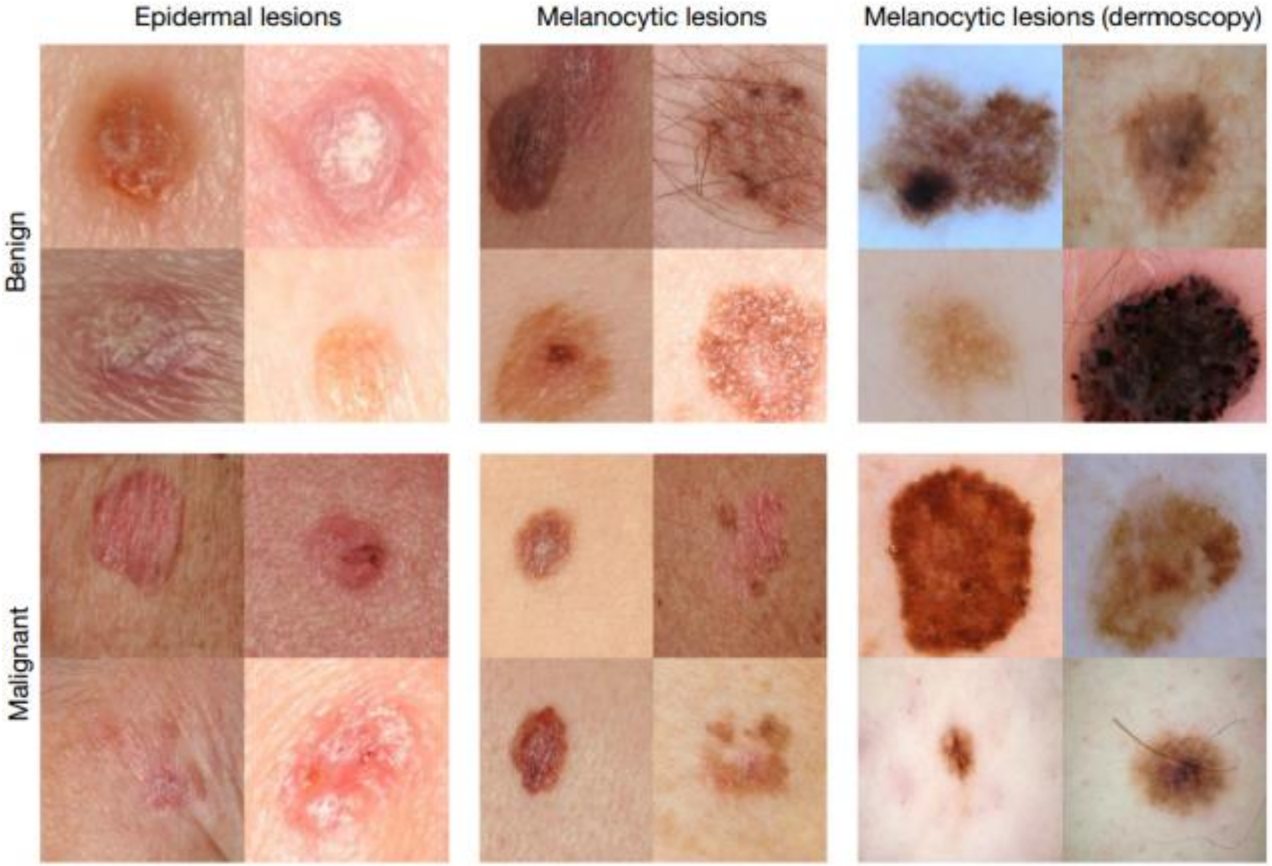
Application

Question: What kind of machine/deep learning problems are you facing?

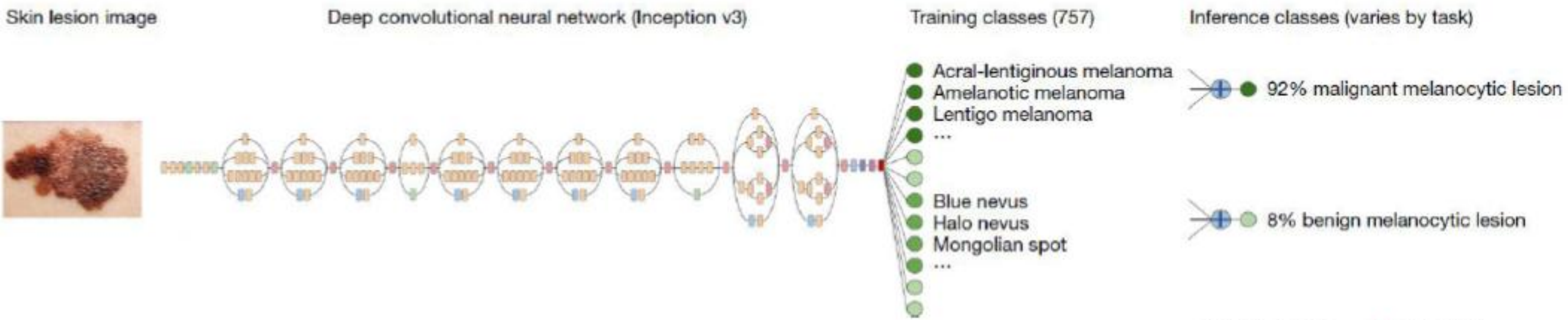
- Prediction?
- Classification?
- Anomaly detection?
- Mimicry/simulation?

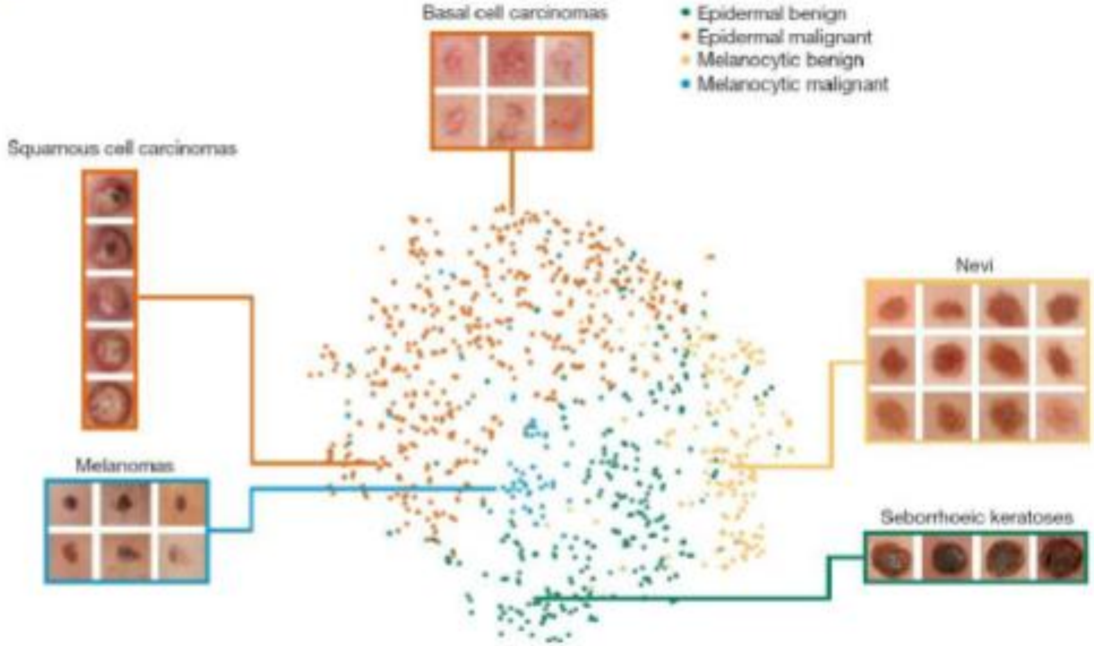
DL algorithm for diagnosis skin cancer

- There's no huge dataset of skin cancer that we can just train our algorithms on, so we had to make our own.
- We gathered images from the internet and worked with the medical school to create a nice taxonomy out of data that was very messy.
- But in the end, the researchers amassed about 130,000 images of skin lesions representing over 2,000 different diseases.



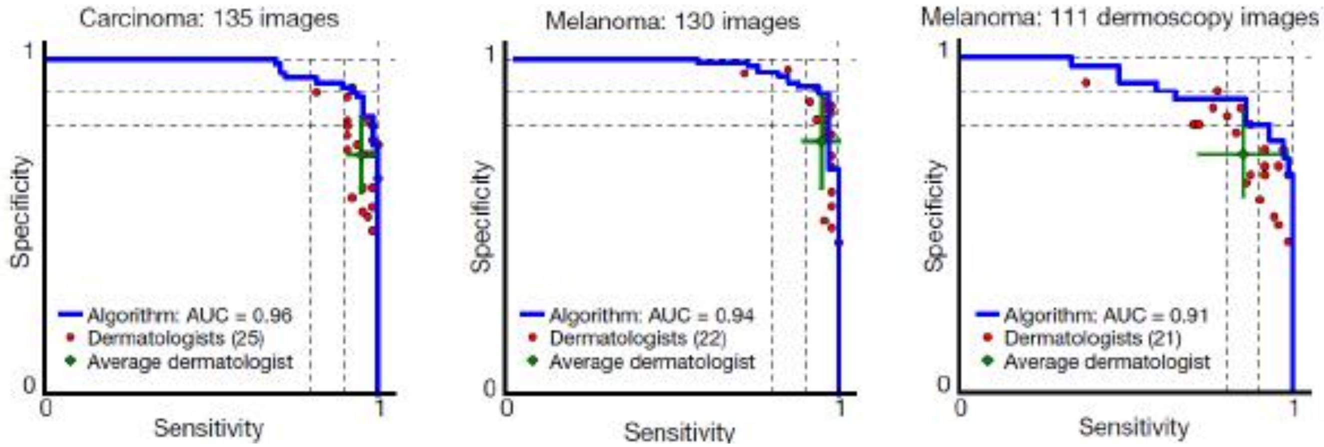
https://www.extremetech.com/extreme/243352-deep-learning-algorithm-diagnoses-skin-cancer-seasoned-dermatologists





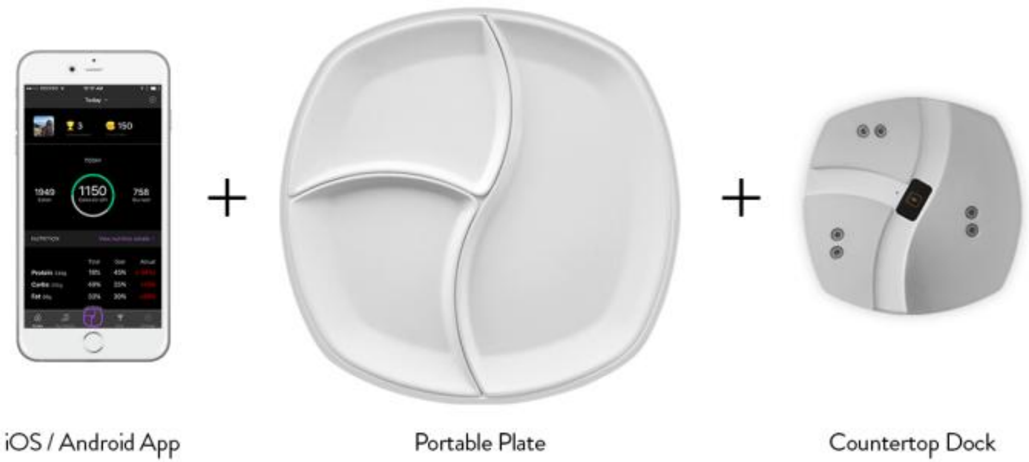
- Learned features

- Comparision to dermatilogists



[Esteva et al. Nature 2017]

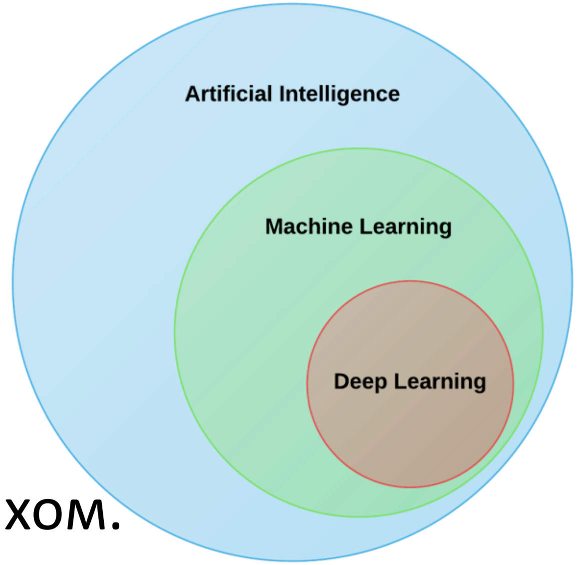
Intelligent Nutrition Platform that instantly analyzes everything you eat.



https://www.indiegogo.com/projects/smart-plate-topview-your-personal-nutritionist-fitness#/

Food Recognition Technology

Кордони



- DL - єдине поле штучного інтелекту (AI), яке пішло вірусним шляхом.
- Платні та безкоштовні курси DL складають 100 000 студентів усіх вікових груп.
- Занадто багато стартапів та продуктів називаються "глибоким чимось", так само як і словосполученням: дуже мало хто з них використовує DL.
- Більшість ігнорують той факт, що DL становить 1% поля "Машинного навчання" (ML), а ML - це 1% поля AI.
- Залишок 99% - те, що використовується практично для більшості завдань. "Експерт, що спеціалізується лише на DL" не є "цілим експертом з експертів"

Image recognition

Machine-generated (but turker prefered)	a man holding a tennis racquet on a tennis court
Human-annotated (but turker not prefered)	the man is on the tennis court playing a game

Machine-generated (but turker prefered)	a group of motorcycles parked next to a motorcycle
Human-annotated (but turker not prefered)	two girls wearing are wearing short skirts and one of them sits on a motorcycle while the other stands nearby

Кордони

- Ліміт DL - це те, що істиною вважається те, що відображається у даних частіше.
- Справедливість DL відбувається не від самої DL, а від людей, які вибирають та готують дані DL.
- DL може читати тексти та перекладати між текстами, але не "людським шляхом".
- Якщо в моделі DL навчено понад 100 книг: 40, які говорять про те, що війни, смерть та руйнування є поганими, і 60 книжок, які вказують на те, що ідеї гітлерівських нацистів були правильними, DL стане 100% нацистом

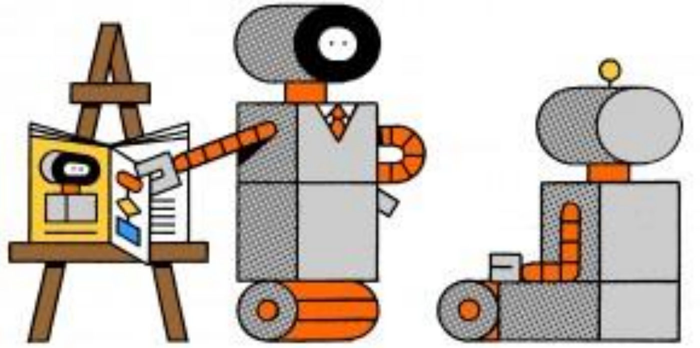
"Я найчастіше читав, що нацизм правий, тому він має бути правильним".

Кордони

- DL навчиться і наслідує найбільш недосконалу логіку без з'ясування недоліків, включаючи тероризм.
- Навіть маленькі діти самостійно усвідомлюють, хто є поганими хлопцями у фільмі, але не DL, якщо люди не навчають це в першу чергу.
- Збір даних про громадян 28 європейських країн підпадає під вимоги General Data Protection Regulation (GDPR)
- 25 травня 2018 року. Це дата, коли DL відмовиться від кількох додатків в ЄС, завдяки чому початківці AI швидко замінять DL на що завгодно, або ризикують отримати штраф.
- Спеціальні засоби DL, такі як градієнтний спуск із зворотним поширенням, а також власні апаратні засоби DL - переважно статистика та геометрія, ймовірно зникнуть в AI 2037 року.

DL вражає, але DL не є синонімом AI

- Найбільш рекламовані AI інструменти від Google, Facebook переважно або тільки DL, тому широка громадськість вважає, що всі нові записи AI є і будуть виконуватися тільки з DL.
- Decision Trees, такі як XGBoost, не роблять заголовки, але тихо побили DL на багатьох змаганнях з табличних даних Kaggle.
- ЗМІ мали на увазі, що AlphaGo - це лише DL, але це дерево пошуку Монте-Карло + DL, що свідчить про те, що чистого DL було недостатньо, щоб виграти.



When bread + shoes = pampers?

+

+

=

TARGET

Funded by the
Erasmus+ Programme
of the European Union

Thank you.

Inna Skarga-Bandurova
Computer Science and Engineering Depatrment
V. Dahl East Ukrainian National University
skarga_bandurova@snu.edu.ua