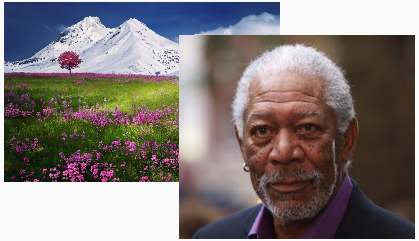
Machine Learning

A Cybersecurity Perspective

Jakub Tomczak
AMLAB, Universiteit van Amsterdam

Era of Big Data

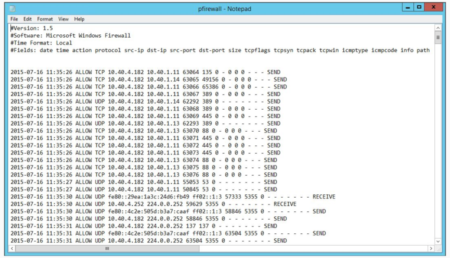
We live in (Big) Data Era



Images

Sound

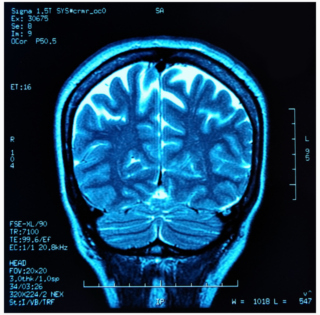
Transactions



```
#Version: 1.5
#Software: Microsoft Windows Firewall
#Time Format: Local
#Fields: date time action protocol src-ip dst-ip src-port dst-port size tcpflags tcpsyn tcpack tcpwin icmptype icmpcode info path


2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63064 135 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.14 63065 49156 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63066 65386 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63067 389 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW UDP 10.40.4.182 10.40.1.14 62292 389 0 - - - - - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63068 389 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63069 445 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW UDP 10.40.4.182 10.40.1.13 62293 389 0 - - - - - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63070 88 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63071 445 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63072 445 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63073 445 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63074 88 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63075 88 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63076 88 0 - 0 0 0 - - - SEND
2015-07-16 11:35:27 ALLOW UDP 10.40.4.182 10.40.1.11 55053 53 0 - - - - - - - SEND
2015-07-16 11:35:27 ALLOW UDP 10.40.4.182 10.40.1.11 50845 53 0 - - - - - - - SEND
2015-07-16 11:35:30 ALLOW UDP fe80::29ea:1a3c:24d6:fb49 ff02::1:3 57333 5355 0 - - - - - - - RECEIVE
2015-07-16 11:35:30 ALLOW UDP 10.40.4.252 224.0.0.252 59629 5355 0 - - - - - - - RECEIVE
2015-07-16 11:35:30 ALLOW UDP fe80::4c2e:505d:b3a7:caaf ff02::1:3 58846 5355 0 - - - - - - - SEND
2015-07-16 11:35:30 ALLOW UDP 10.40.4.182 224.0.0.252 58846 5355 0 - - - - - - - SEND
2015-07-16 11:35:31 ALLOW UDP 10.40.4.182 224.0.0.252 137 137 0 - - - - - - - SEND
2015-07-16 11:35:31 ALLOW UDP fe80::4c2e:505d:b3a7:caaf ff02::1:3 63504 5355 0 - - - - - - - SEND
2015-07-16 11:35:31 ALLOW UDP 10.40.4.182 224.0.0.252 63504 5355 0 - - - - - - - SEND
```

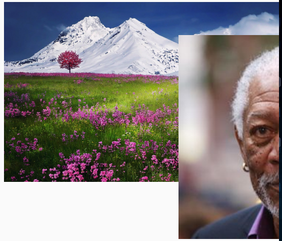
Logs



Medical images

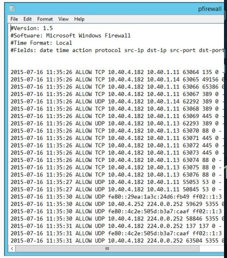
Social media

We live in (Big) Threat Era



Images

Transactions



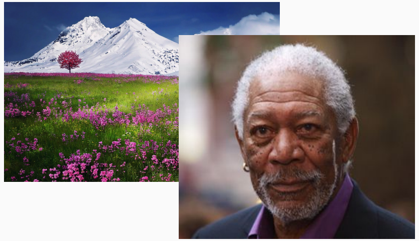
```
#Version: 1.5
#Software: Microsoft Windows Firewall
#Time Format: Local
#Fields: date time action protocol src-ip dst-ip src-port dst-port


2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63064 135 0 -
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.14 63065 49156 0
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63066 65386 0
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63067 389 0 -
2015-07-16 11:35:26 ALLOW UDP 10.40.4.182 10.40.1.14 62292 389 0 -
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63068 389 0 -
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63069 445 0 -
2015-07-16 11:35:26 ALLOW UDP 10.40.4.182 10.40.1.13 62293 389 0 -
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63070 88 0 -
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63071 445 0 -
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63072 445 0 -
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63073 445 0 -
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63074 88 0 -
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63075 88 0 -
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63076 88 0 -
2015-07-16 11:35:27 ALLOW UDP 10.40.4.182 10.40.1.11 55053 53 0 -
2015-07-16 11:35:27 ALLOW UDP 10.40.4.182 10.40.1.11 50845 53 0 -
2015-07-16 11:35:30 ALLOW UDP fe80::29ea:1a3c:24d6:fb49 ff02::1:3
2015-07-16 11:35:30 ALLOW UDP 10.40.4.252 224.0.0.252 59629 5355 0
2015-07-16 11:35:30 ALLOW UDP fe80::4c2e:505d:b3a7:caaf ff02::1:3
2015-07-16 11:35:30 ALLOW UDP 10.40.4.182 224.0.0.252 58846 5355 0
2015-07-16 11:35:31 ALLOW UDP 10.40.4.182 224.0.0.252 137 137 0 -
2015-07-16 11:35:31 ALLOW UDP fe80::4c2e:505d:b3a7:caaf ff02::1:3
2015-07-16 11:35:31 ALLOW UDP 10.40.4.182 224.0.0.252 63504 5355 0
```

Lo

Social media

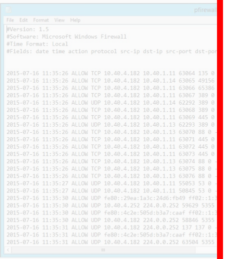
We live in (Big) Threat Era



Images

Sound

Transactions



```
#Version: 1.5
#Software: Microsoft Windows Firewall
#Time Format: Local
#Fields: date time action protocol src-ip dst-ip src-port dst-por


2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63064 135 0
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.14 63065 49156
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63066 65386
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63067 389 0
2015-07-16 11:35:26 ALLOW UDP 10.40.4.182 10.40.1.14 62292 389 0
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63068 389 0
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63069 445 0
2015-07-16 11:35:26 ALLOW UDP 10.40.4.182 10.40.1.13 62293 389 0
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63070 88 0 -
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63071 445 0
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63072 445 0
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63073 445 0
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63074 88 0 -
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63075 88 0 -
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63076 88 0 -
2015-07-16 11:35:27 ALLOW UDP 10.40.4.182 10.40.1.11 55053 53 0 -
2015-07-16 11:35:27 ALLOW UDP 10.40.4.182 10.40.1.11 50845 53 0 -
2015-07-16 11:35:30 ALLOW UDP fe80::29ea:1a3c:24d6:fb49 ff02::1:3
2015-07-16 11:35:30 ALLOW UDP 10.40.4.252 224.0.0.252 59629 5355
2015-07-16 11:35:30 ALLOW UDP fe80::4c2e:505d:b3a7:caaf ff02::1:3
2015-07-16 11:35:30 ALLOW UDP 10.40.4.182 224.0.0.252 58846 5355
2015-07-16 11:35:31 ALLOW UDP 10.40.4.182 224.0.0.252 137 137 0 -
2015-07-16 11:35:31 ALLOW UDP fe80::4c2e:505d:b3a7:caaf ff02::1:3
2015-07-16 11:35:31 ALLOW UDP 10.40.4.182 224.0.0.252 63504 5355
```

Lo

Stealing authorship

Stealing identity

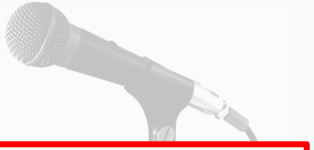
Manipulating with facts (fake news)

...

Social media

We live in (Big) Threat Era

Images

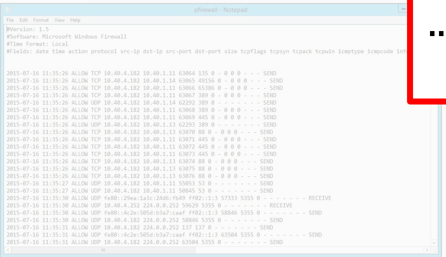


Stealing fragile information

Phishing

…

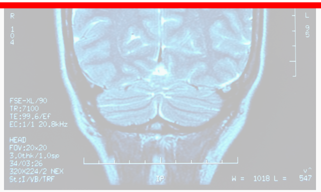
Transactions



```
#Version: 1.5
#Software: Microsoft Windows Firewall
#Time Format: Local
#Fields: date time action protocol src-ip dst-ip src-port dst-port size tcpflags tcpsyn tcpack tcpwin icmptype icmpcode inf


2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63064 135 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.14 63065 49156 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63066 65386 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63067 389 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW UDP 10.40.4.182 10.40.1.14 62292 389 0 - - - - - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63068 389 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63069 445 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW UDP 10.40.4.182 10.40.1.13 62293 389 0 - - - - - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63070 88 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63071 445 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63072 445 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63073 445 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63074 88 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63075 88 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63076 88 0 - 0 0 0 - - - SEND
2015-07-16 11:35:27 ALLOW UDP 10.40.4.182 10.40.1.11 55053 53 0 - - - - - - - SEND
2015-07-16 11:35:27 ALLOW UDP 10.40.4.182 10.40.1.11 50845 53 0 - - - - - - - SEND
2015-07-16 11:35:30 ALLOW UDP fe80::29ea:1a3c:24d6:fb49 ff02::1:3 57333 5355 0 - - - - - - - RECEIVE
2015-07-16 11:35:30 ALLOW UDP 10.40.4.252 224.0.0.252 59629 5355 0 - - - - - - - RECEIVE
2015-07-16 11:35:30 ALLOW UDP fe80::4c2e:505d:b3a7:caaf ff02::1:3 58846 5355 0 - - - - - - - SEND
2015-07-16 11:35:30 ALLOW UDP 10.40.4.182 224.0.0.252 58846 5355 0 - - - - - - - SEND
2015-07-16 11:35:31 ALLOW UDP 10.40.4.182 224.0.0.252 137 137 0 - - - - - - - SEND
2015-07-16 11:35:31 ALLOW UDP fe80::4c2e:505d:b3a7:caaf ff02::1:3 63504 5355 0 - - - - - - - SEND
2015-07-16 11:35:31 ALLOW UDP 10.40.4.182 224.0.0.252 63504 5355 0 - - - - - - - SEND
```

Logs



Medical images

Social media

We live in (Big) Threat Era

Images

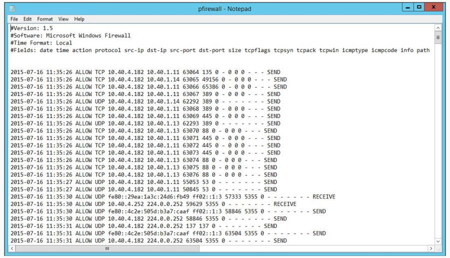
Transactions

Viruses, worms, trojan horses, bots

Spams, packet sniffing

Stealing passwords, zombie computers

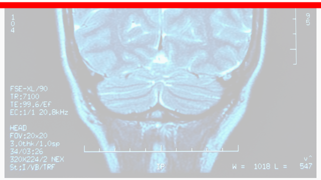
…



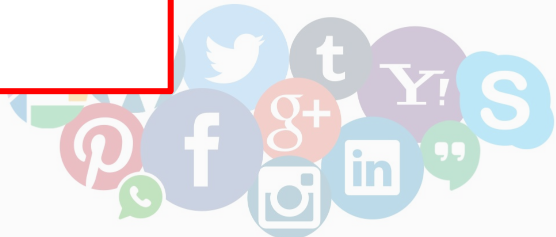
```
#Version: 1.5
#Software: Microsoft Windows Firewall
#Time Format: Local
#Fields: date time action protocol src-ip dst-ip src-port dst-port size tcpflags tcpsyn tcpack tcpwin icmptype icmpcode info path


2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63064 135 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.14 63065 49156 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63066 65386 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63067 389 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW UDP 10.40.4.182 10.40.1.14 62292 389 0 - - - - - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63068 389 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63069 445 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW UDP 10.40.4.182 10.40.1.13 62293 389 0 - - - - - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63070 88 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63071 445 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63072 445 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63073 445 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63074 88 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63075 88 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63076 88 0 - 0 0 0 - - - SEND
2015-07-16 11:35:27 ALLOW UDP 10.40.4.182 10.40.1.11 55053 53 0 - - - - - - - SEND
2015-07-16 11:35:27 ALLOW UDP 10.40.4.182 10.40.1.11 50845 53 0 - - - - - - - SEND
2015-07-16 11:35:30 ALLOW UDP fe80::29ea:1a3c:24d6:fb49 ff02::1:3 57333 5355 0 - - - - - - - RECEIVE
2015-07-16 11:35:30 ALLOW UDP 10.40.4.252 224.0.0.252 59629 5355 0 - - - - - - - RECEIVE
2015-07-16 11:35:30 ALLOW UDP fe80::4c2e:505d:b3a7:caaf ff02::1:3 58846 5355 0 - - - - - - - SEND
2015-07-16 11:35:30 ALLOW UDP 10.40.4.182 224.0.0.252 58846 5355 0 - - - - - - - SEND
2015-07-16 11:35:31 ALLOW UDP 10.40.4.182 224.0.0.252 137 137 0 - - - - - - - SEND
2015-07-16 11:35:31 ALLOW UDP fe80::4c2e:505d:b3a7:caaf ff02::1:3 63504 5355 0 - - - - - - - SEND
2015-07-16 11:35:31 ALLOW UDP 10.40.4.182 224.0.0.252 63504 5355 0 - - - - - - - SEND
```

Logs



Medical images



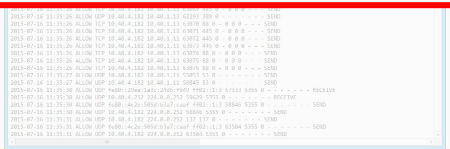
Social media

We live in (Big) Threat Era

Stealing patient information

Misuse of personal information

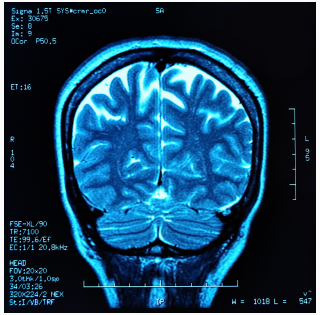
...



```
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63069 445 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW UDP 10.40.4.182 10.40.1.13 62293 389 0 - - - - - - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63070 88 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63071 445 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63072 445 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.11 63073 445 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63074 88 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63075 88 0 - 0 0 0 - - - SEND
2015-07-16 11:35:26 ALLOW TCP 10.40.4.182 10.40.1.13 63076 88 0 - 0 0 0 - - - SEND
2015-07-16 11:35:27 ALLOW UDP 10.40.4.182 10.40.1.11 55053 53 0 - - - - - - - - SEND
2015-07-16 11:35:27 ALLOW UDP 10.40.4.182 10.40.1.11 50845 53 0 - - - - - - - - SEND
2015-07-16 11:35:30 ALLOW UDP fe80::29ea:1a3c:24d6:fb49 ff02::1:3 57333 5355 0 - - - - - - - - RECEIVE
2015-07-16 11:35:30 ALLOW UDP 10.40.4.252 224.0.0.252 59629 5355 0 - - - - - - - - RECEIVE
2015-07-16 11:35:30 ALLOW UDP fe80::4c2e:505d:b3a7:caaf ff02::1:3 58846 5355 0 - - - - - - - - SEND
2015-07-16 11:35:30 ALLOW UDP 10.40.4.182 224.0.0.252 58846 5355 0 - - - - - - - - SEND
2015-07-16 11:35:31 ALLOW UDP 10.40.4.182 224.0.0.252 137 137 0 - - - - - - - - SEND
2015-07-16 11:35:31 ALLOW UDP fe80::4c2e:505d:b3a7:caaf ff02::1:3 63504 5355 0 - - - - - - - - SEND
2015-07-16 11:35:31 ALLOW UDP 10.40.4.182 224.0.0.252 63504 5355 0 - - - - - - - - SEND
```

Logs

Sound



Medical images

Transactions

Social media

We live in (Big) Threat Era

Stealing identity and private information

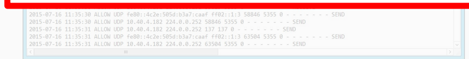
Stealing fragile information

Taking control over a person or an organization

...

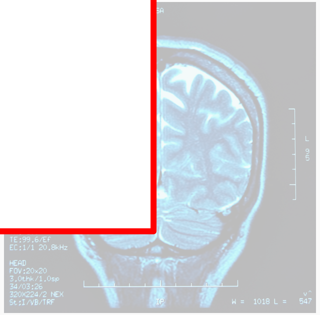
Sound

Transactions



```
2015-07-16 11:35:30 ALLOW UDP fe80::4c2e:505d:b3a7:caaf ff02::1:3 58846 5355 0 - - - - - - - - SEND
2015-07-16 11:35:30 ALLOW UDP 10.40.4.182 224.0.0.252 58846 5355 0 - - - - - - - - SEND
2015-07-16 11:35:31 ALLOW UDP 10.40.4.182 224.0.0.252 137 137 0 - - - - - - - - SEND
2015-07-16 11:35:31 ALLOW UDP fe80::4c2e:505d:b3a7:caaf ff02::1:3 63504 5355 0 - - - - - - - - SEND
2015-07-16 11:35:31 ALLOW UDP 10.40.4.182 224.0.0.252 63504 5355 0 - - - - - - - - SEND
```

Logs



Medical images

Social media

Machine Learning for the rescue!

What is Machine Learning?

Machine Learning

Statistics

- Probabilistic modelling
- Estimators

Optimization

- Optimization methods
- Convex programming

(Big) Data

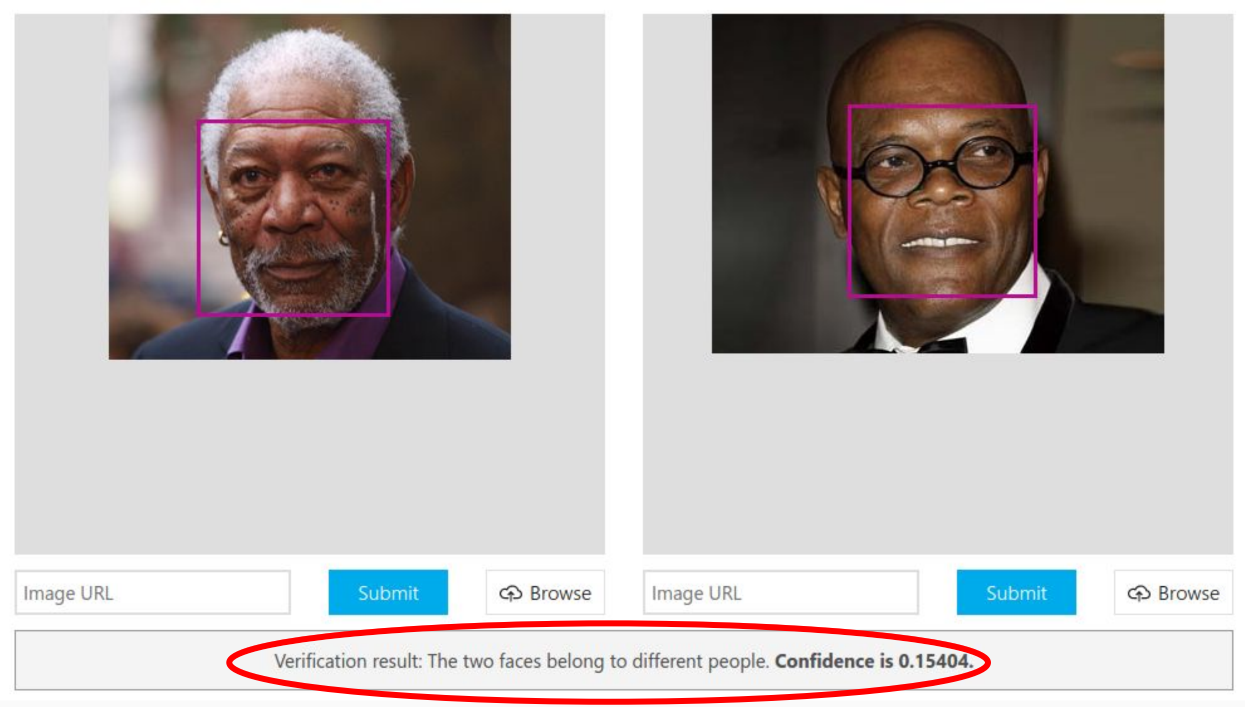
- Image, Sound, Text ...
- Countless data sources

Machine learning: A remedy for cyberattacks

Identity identification (static data)

- Face recognition
- Face comparison

https://azure.microsoft.com/en-us/services/cognitive-services/face/



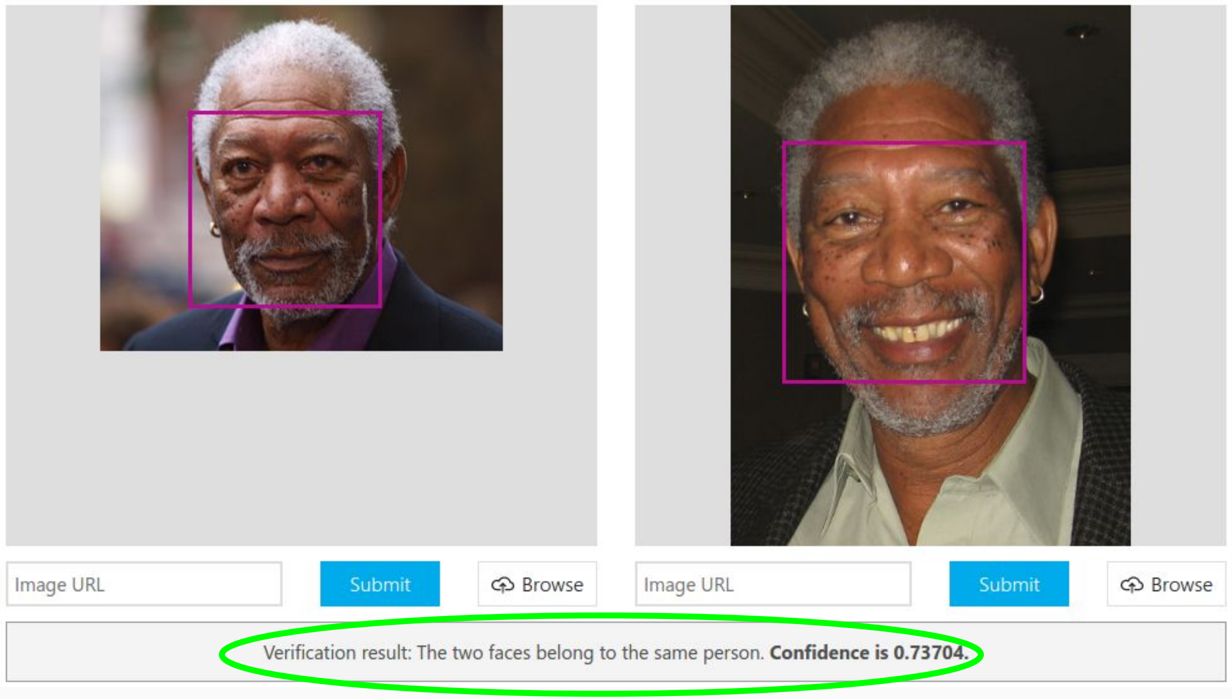
Verification result: The two faces belong to different people. **Confidence is 0.15404.**

Machine learning: A remedy for cyberattacks

Identity identification (static data)

- Face recognition
- Face comparison

https://azure.microsoft.com/en-us/services/cognitive-services/face/



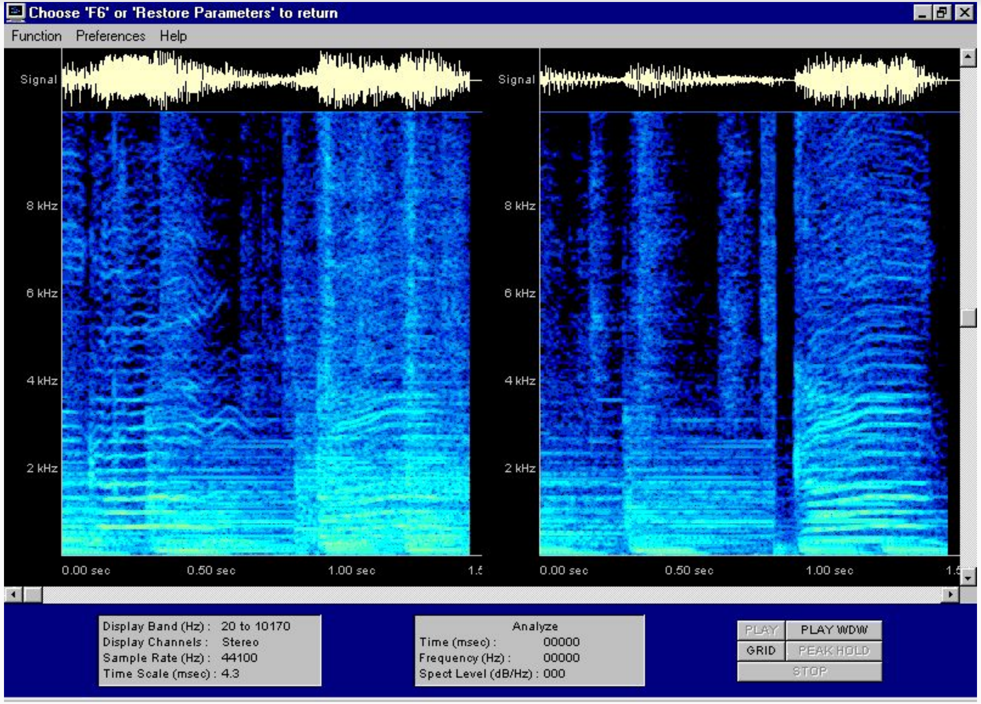
Verification result: The two faces belong to the same person. **Confidence is 0.73704.**

Machine learning: A remedy for cyberattacks

Identity identification

(sequential data)

- Voice recognition
- Voice comparison



```
Display Band (Hz) :   20 to 10170
Display Channels :    Stereo
Sample Rate (Hz) :    44100
Time Scale (msec) : 4.3
```

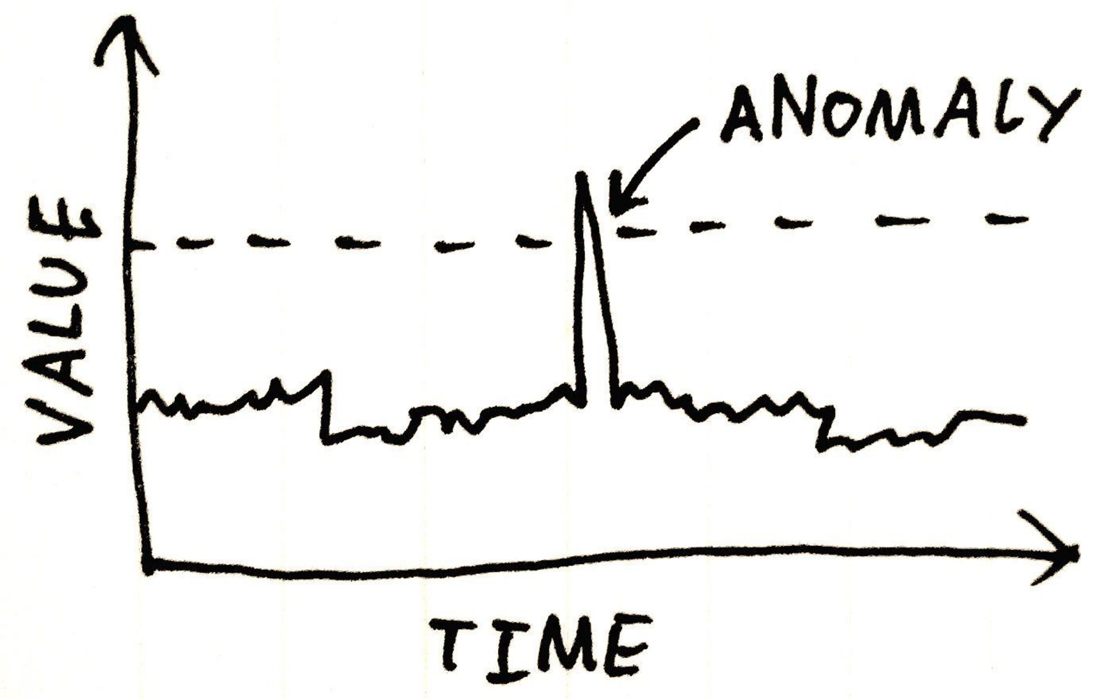
```
                Analyze
Time (msec) :           00000
Frequency (Hz) :        00000
Spect Level (dB/Hz) : 000
```

Machine learning: A remedy for cyberattacks

Behavior analysis

(temporal data)

- *E.g.*: expenditures
- Anomaly detection

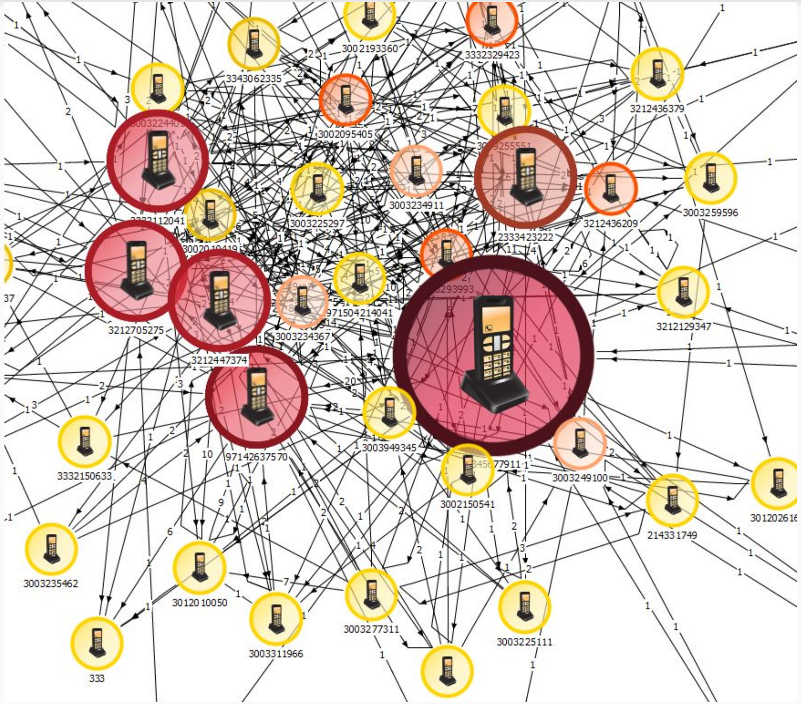


Machine learning: A remedy for cyberattacks

Network analysis

(network data)

- *E.g.*: mobile network
- Hubs identification



Delving into machine learning: Typical tasks

Typical tasks of machine learning

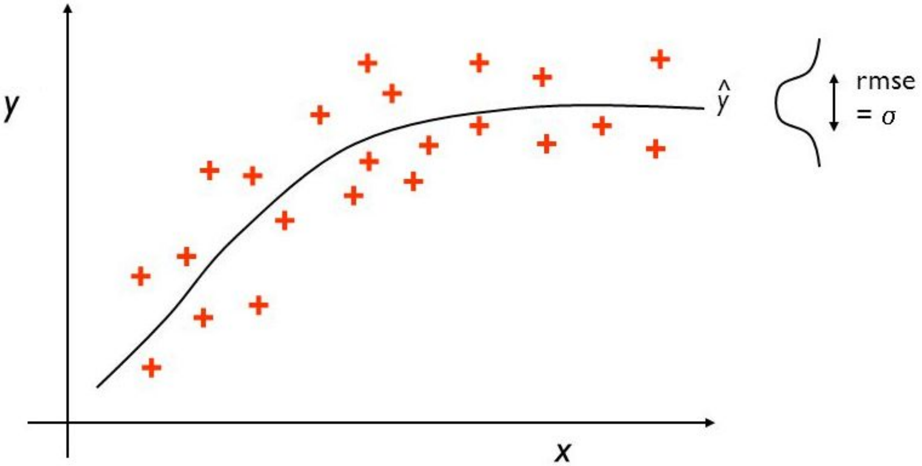
Machine Learning

- **Supervised learning**
- **Unsupervised learning**
- **Semi-supervised learning**
- **Reinforcement learning**

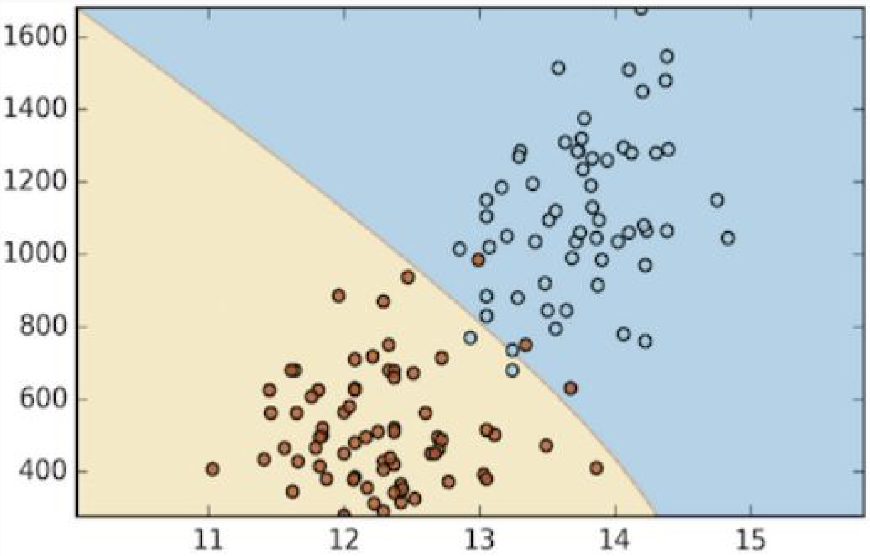
Supervised learning

- **Input** (object) and **target** are **known**.
- **Aim**: train a model to **predict** the target for a new input.
- Two cases:
  - regression
  - classification

Regression



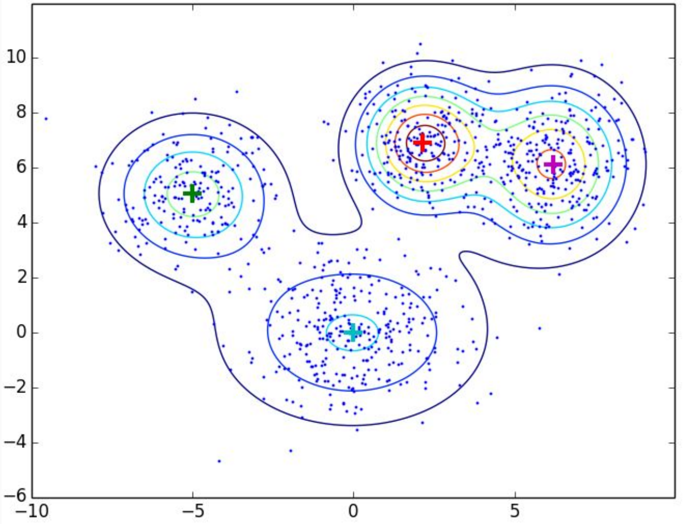
Classification



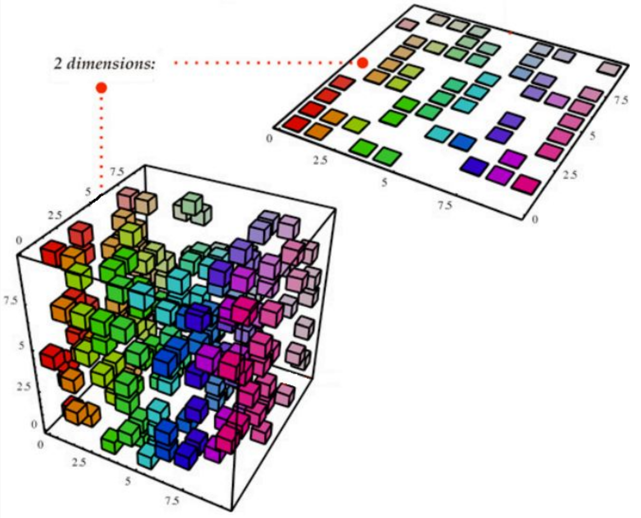
Unsupervised learning

- **Input** (object) is **known**. **Target** is **unknown**.
- **Aim**: **density estimation**.
- Typical tasks:
  - clustering
  - dimensionality reduction

Clustering



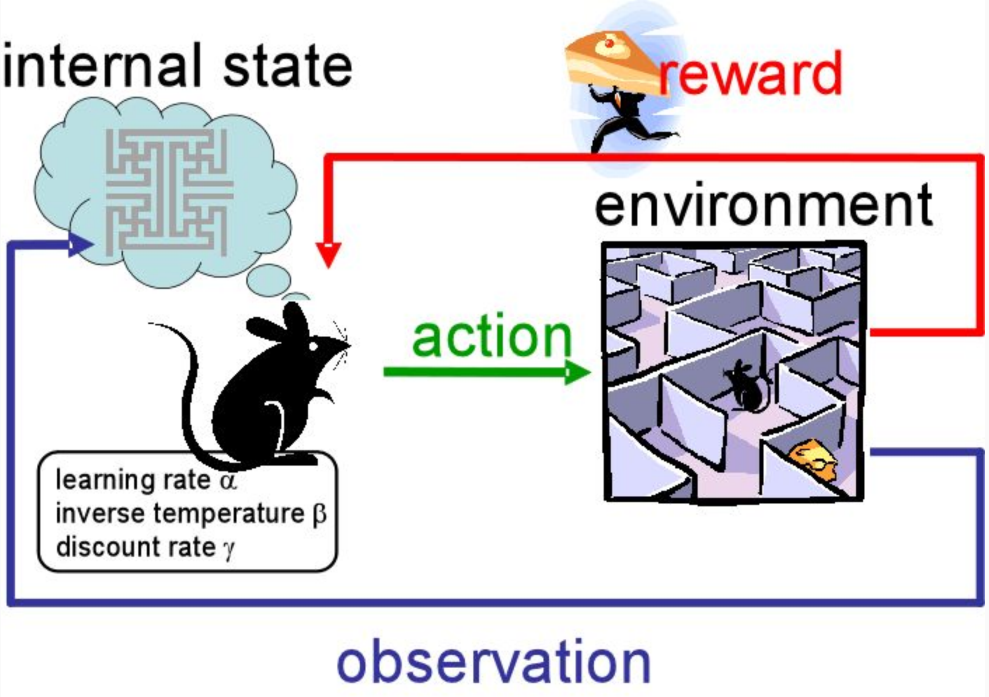
Dimensionality reduction



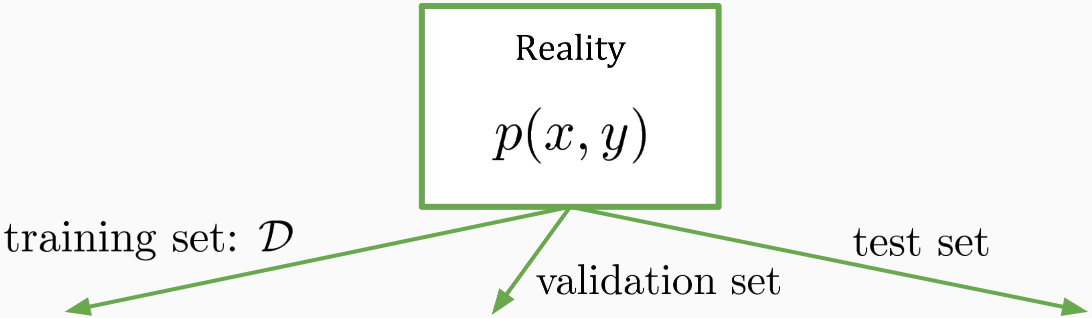
Reinforcement learning

- **Agent** interacts with **environment** to achieve a **goal.**

- **Aim**: training a **policy** (a series of actions to achieve the goal).



Delving into machine learning: Main components



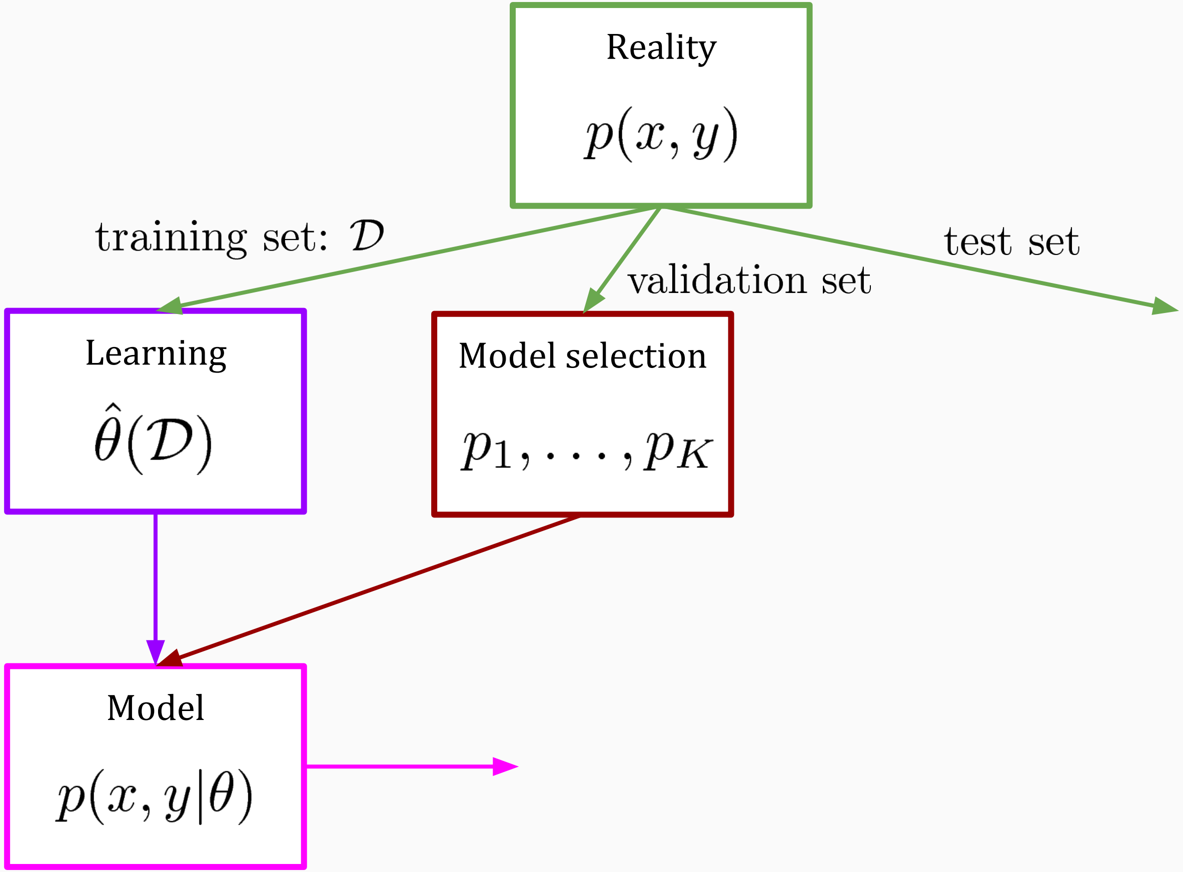
Reality

$p(x, y)$

training set: $\mathcal{D}$

validation set

test set



Reality

$p(x, y)$

training set: $\mathcal{D}$

validation set

test set

Learning

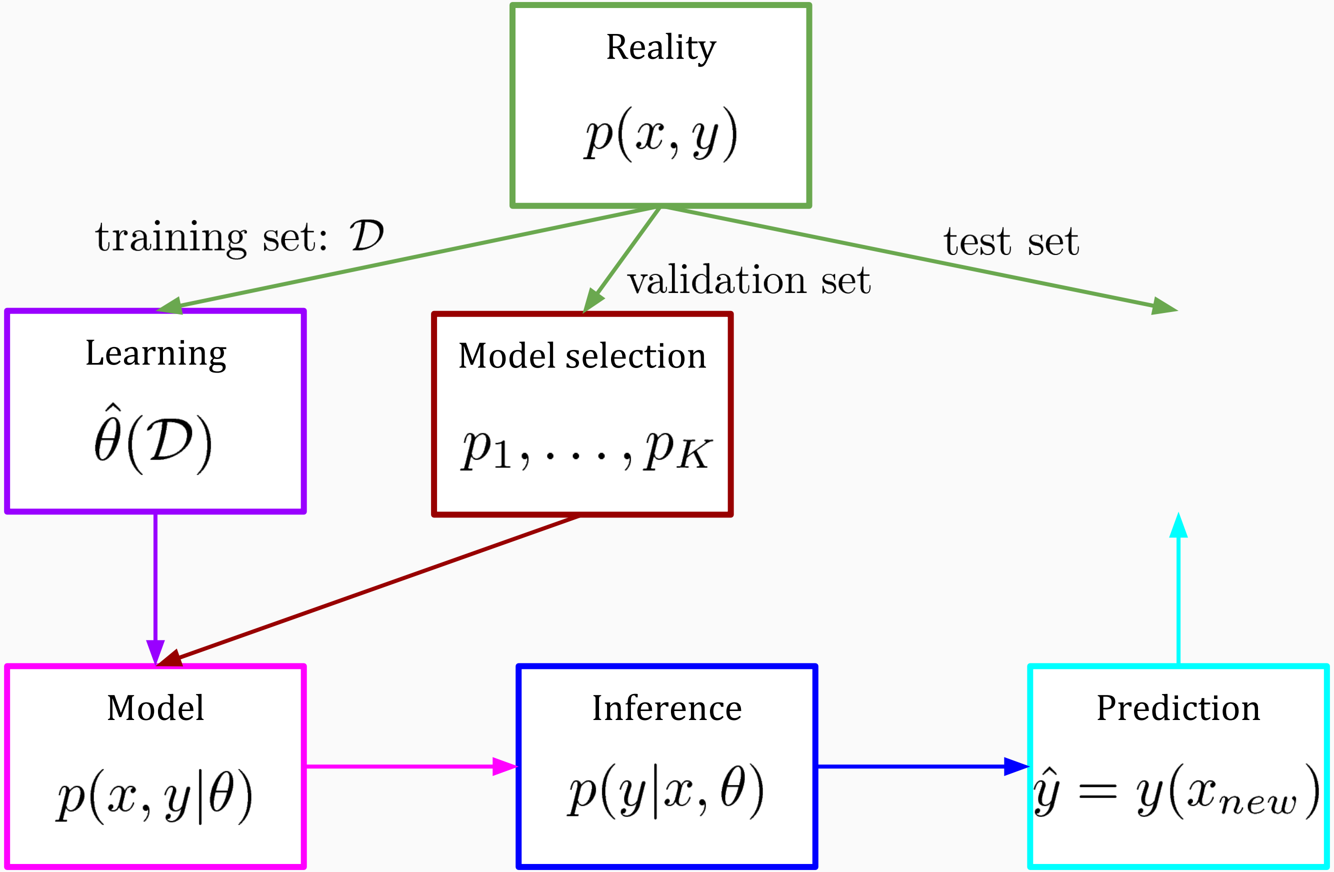
$\hat{\theta}(\mathcal{D})$

Model selection

$p_1, \ldots, p_K$

Model

$p(x, y|\theta)$



Reality

$p(x, y)$

training set: $\mathcal{D}$

validation set

test set

Learning

$\hat{\theta}(\mathcal{D})$

Model selection

$p_1, \ldots, p_K$

Model

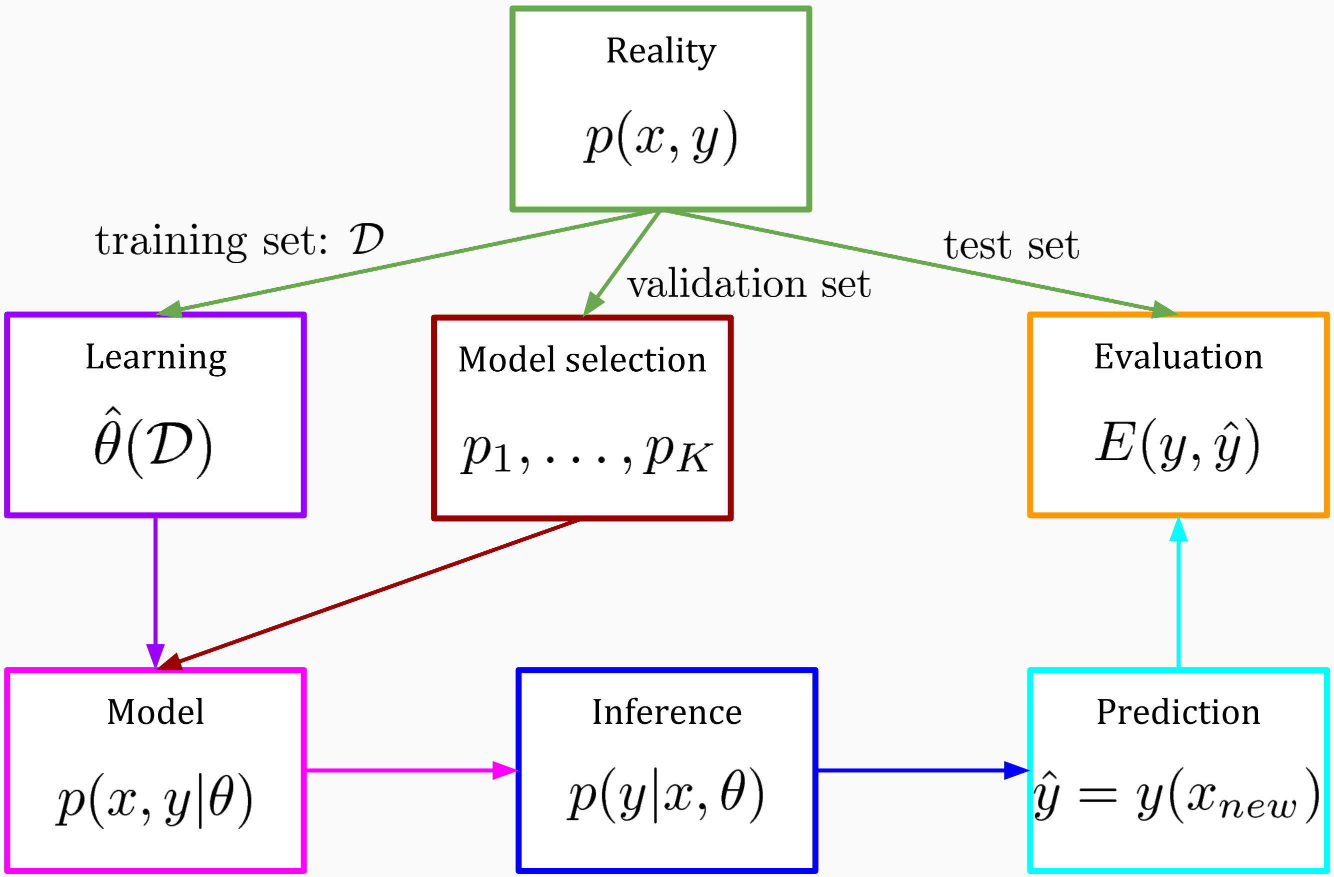
$p(x, y|\theta)$

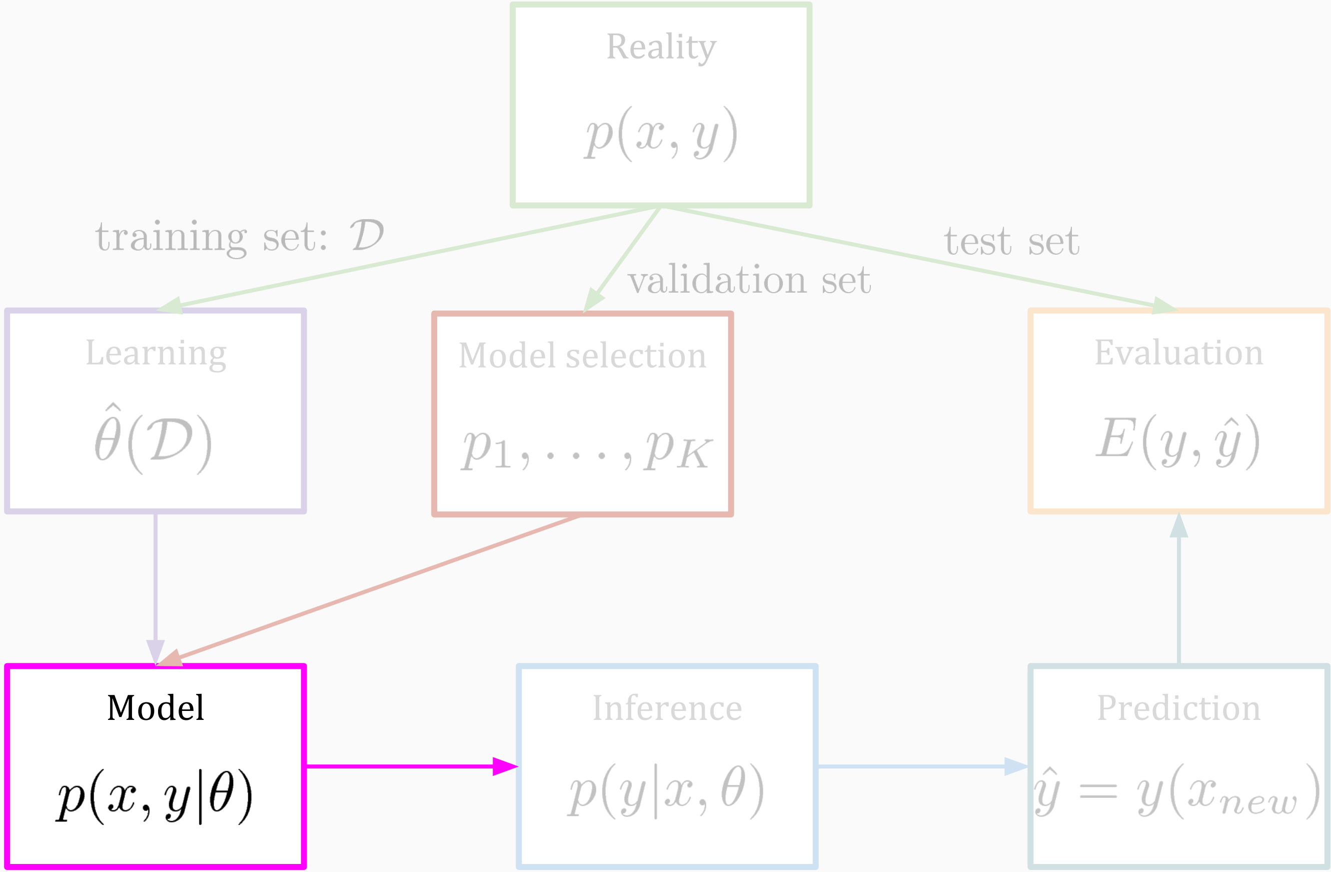
Inference

$p(y|x, \theta)$

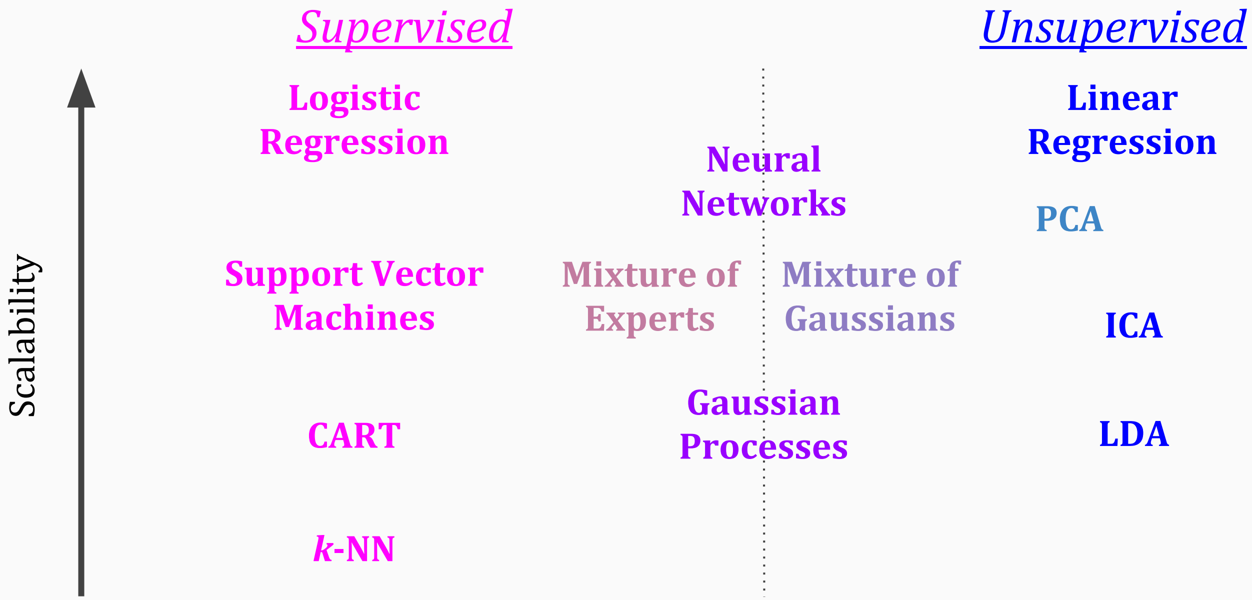
Prediction

$\hat{y} = y(x_{new})$


Reality
$p(x, y)$
training set: $\mathcal{D}$
validation set
test set
Learning
$\hat{\theta}(\mathcal{D})$
Model selection
$p_1, \dots, p_K$
Evaluation
$E(y, \hat{y})$
Model
$p(x, y|\theta)$
Inference
$p(y|x, \theta)$
Prediction
$\hat{y} = y(x_{new})$


Reality
$p(x,y)$
training set: $\mathcal{D}$
validation set
test set
Learning
$\hat{\theta}(\mathcal{D})$
Model selection
$p_1,\ldots,p_K$
Evaluation
$E(y,\hat{y})$
Model
$p(x,y|\theta)$
Inference
$p(y|x,\theta)$
Prediction
$\hat{y}=y(x_{new})$

Machine learning: Models



Supervised

Unsupervised

Scalability

Logistic Regression

Support Vector Machines

CART

***k*-NN**

Neural Networks

Mixture of Experts

Mixture of Gaussians

Gaussian Processes

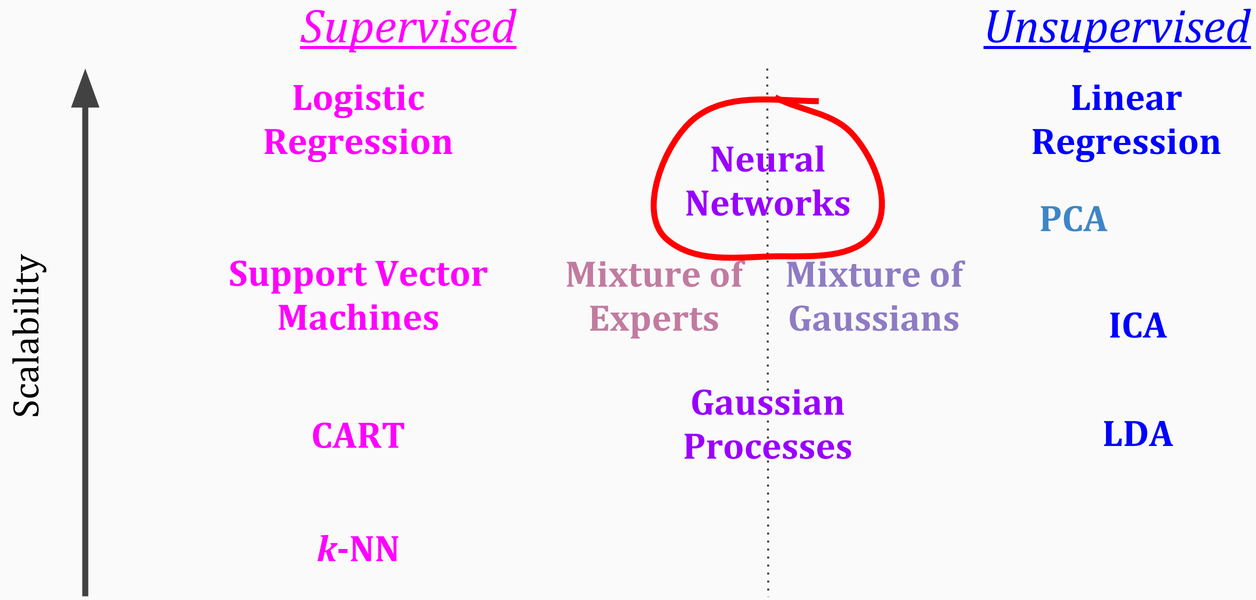
Linear Regression

PCA

ICA

LDA

Machine learning: Models



Supervised

Unsupervised

Scalability

Logistic Regression

Support Vector Machines

CART

***k*-NN**

Neural Networks

Mixture of Experts

Mixture of Gaussians

Gaussian Processes

Linear Regression

PCA

ICA

LDA

Delving into machine learning: Deep Learning (neural networks)

- The INTERNET
- Unlabeled data
- ImageNet database

14,197,122 images

IMAGENET

- Parallel computing
- GPU, FPGA
- Deep learning frameworks

theano

Caffe

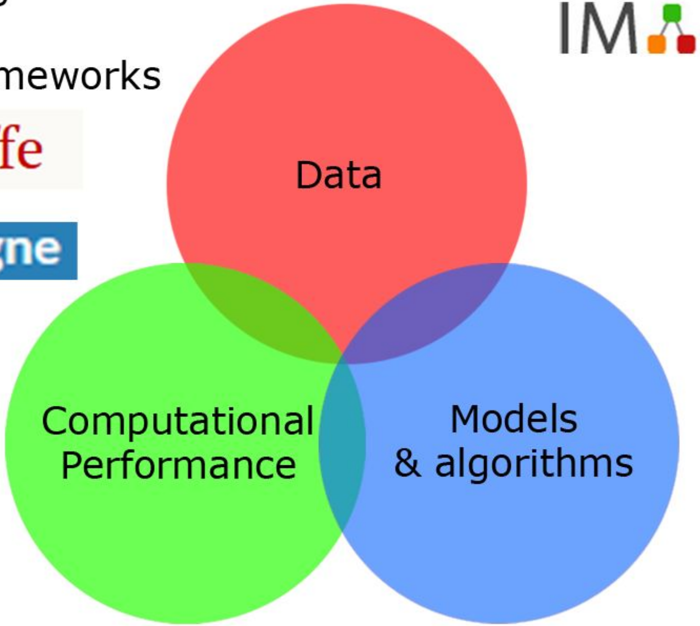
TensorFlow

Lasagne

torch

PYTORCH

cuDNN



NVIDIA

Google

Microsoft Research

facebook

Baidu 百度

Alibaba Group 阿里巴巴集团

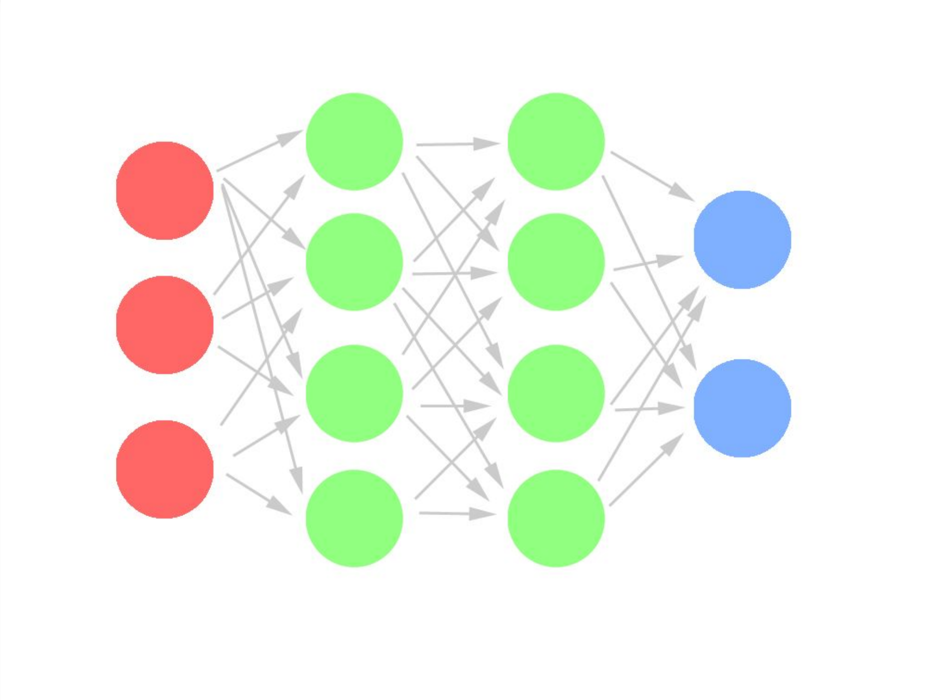
NETFLIX

amazon

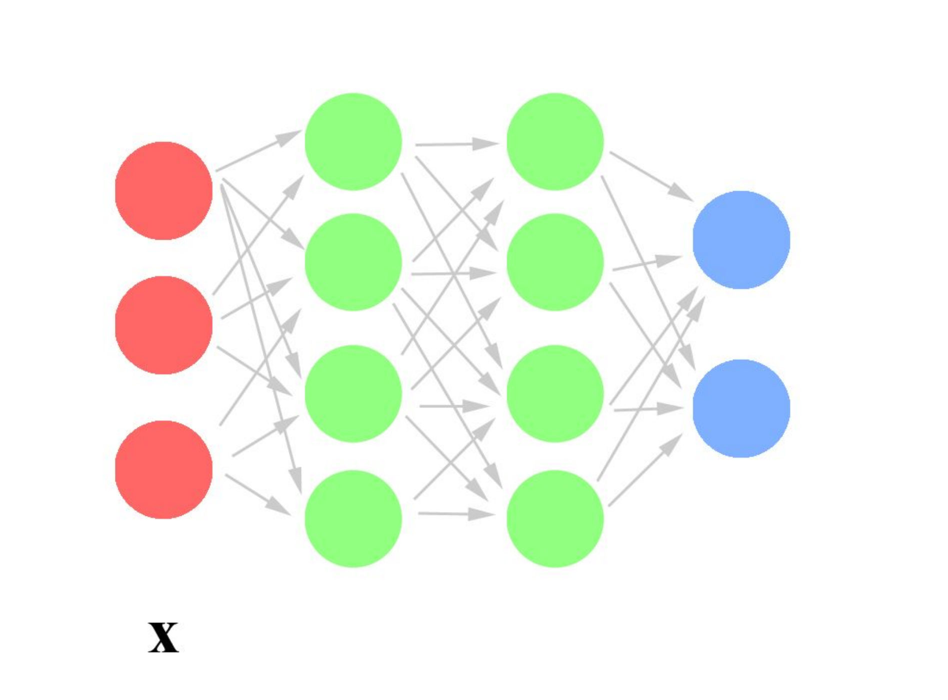
MERCK

NOVARTIS

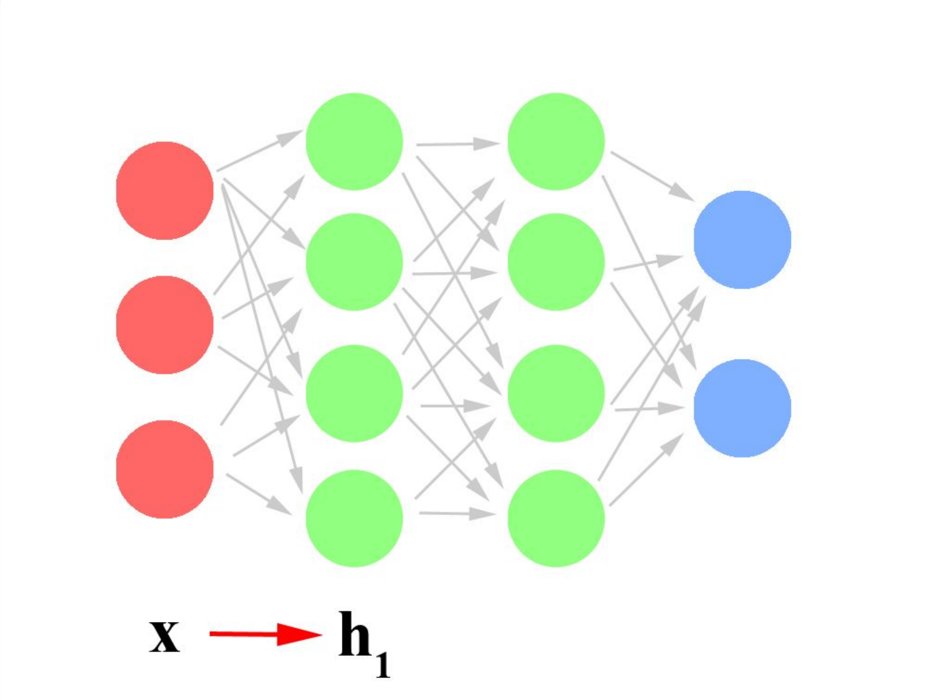
Deep learning: Multilayer Perceptron (MLP)



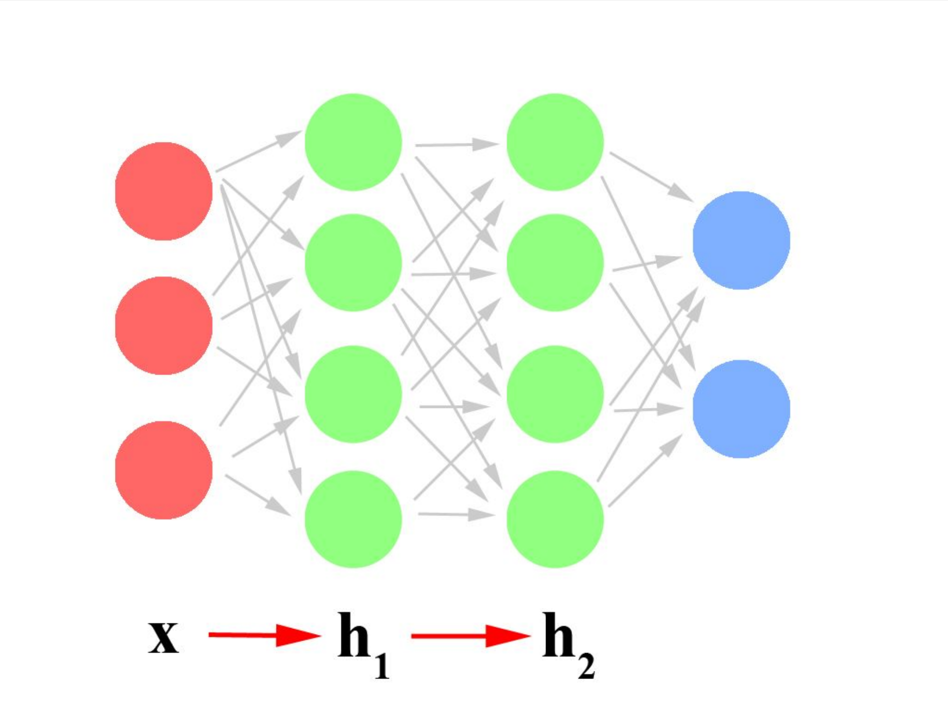
Deep learning: Multilayer Perceptron (MLP)



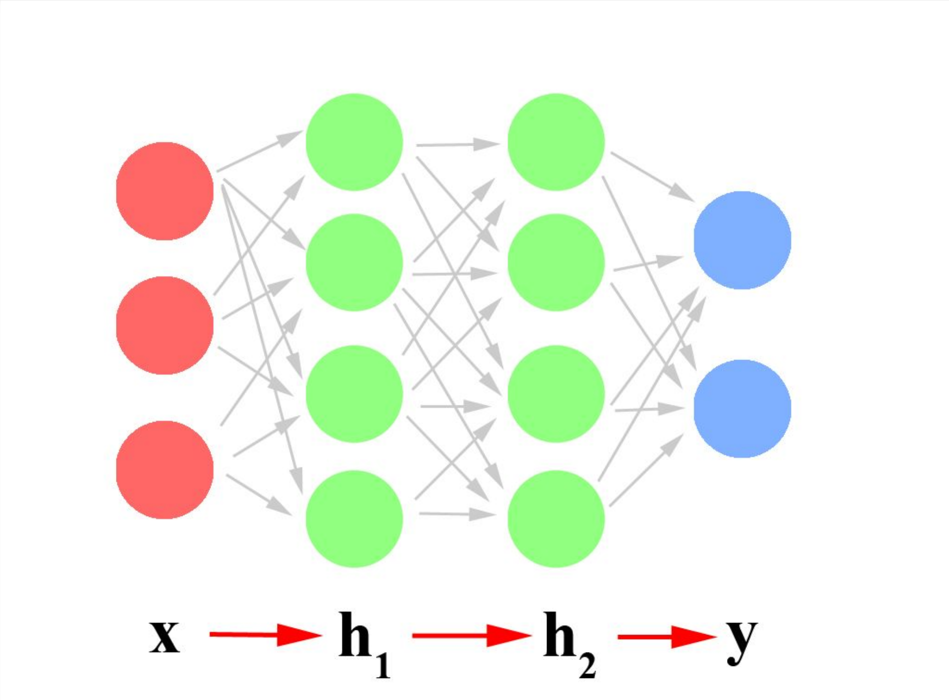
Deep learning: Multilayer Perceptron (MLP)



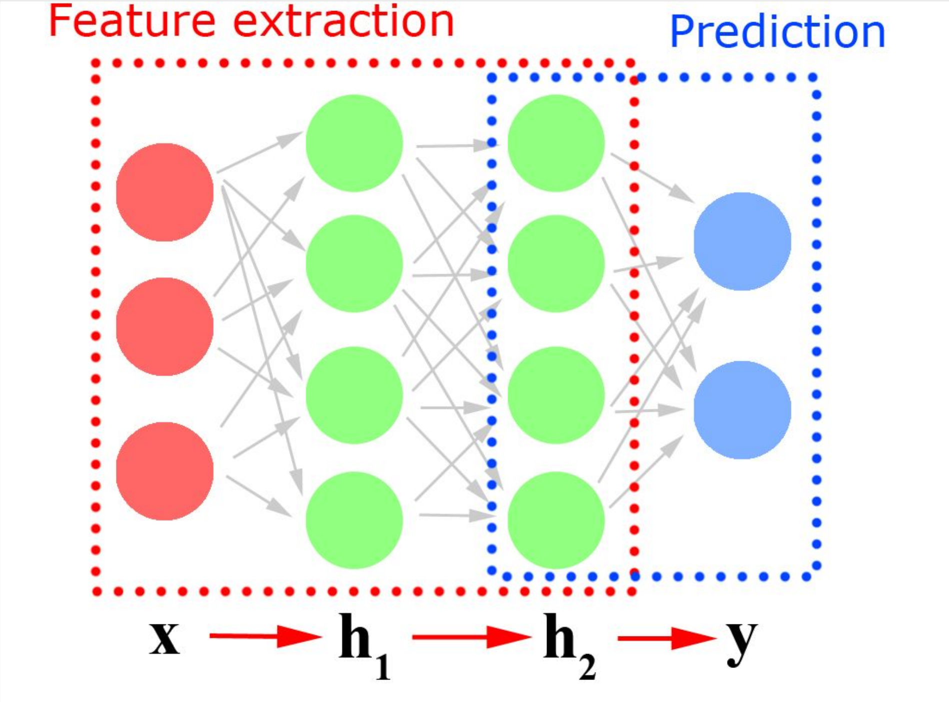
Deep learning: Multilayer Perceptron (MLP)


$\mathbf{x} \rightarrow \mathbf{h}_1 \rightarrow \mathbf{h}_2$

Deep learning: Multilayer Perceptron (MLP)


$\mathbf{x} \rightarrow \mathbf{h}_1 \rightarrow \mathbf{h}_2 \rightarrow \mathbf{y}$

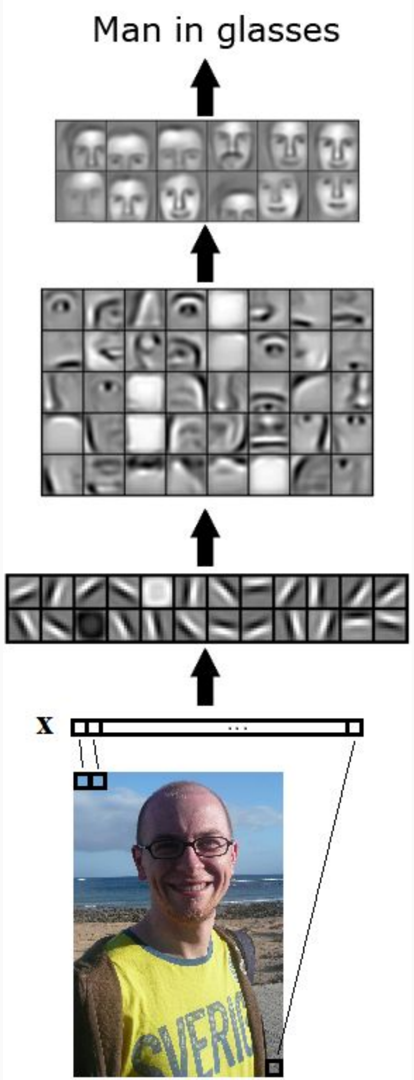
Deep learning: Multilayer Perceptron (MLP)



Automatic feature extraction

- Feature in successive layers represent **higher level of abstraction.**

- Good features should be:
  - **informative**
  - **robust**
  - **invariant**



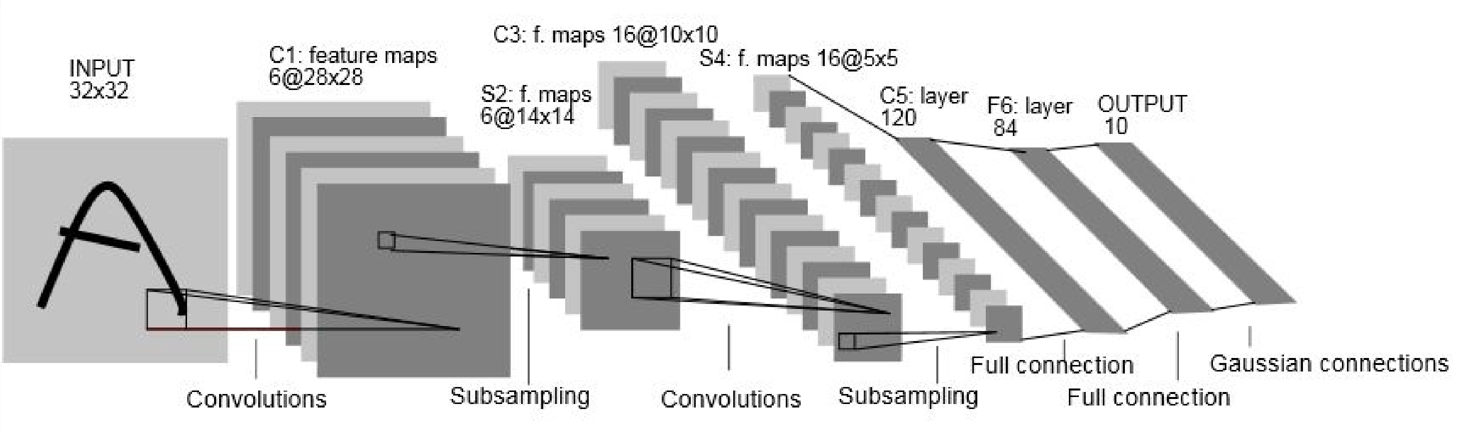
Highest-level features

Higher-level features

Low-level features

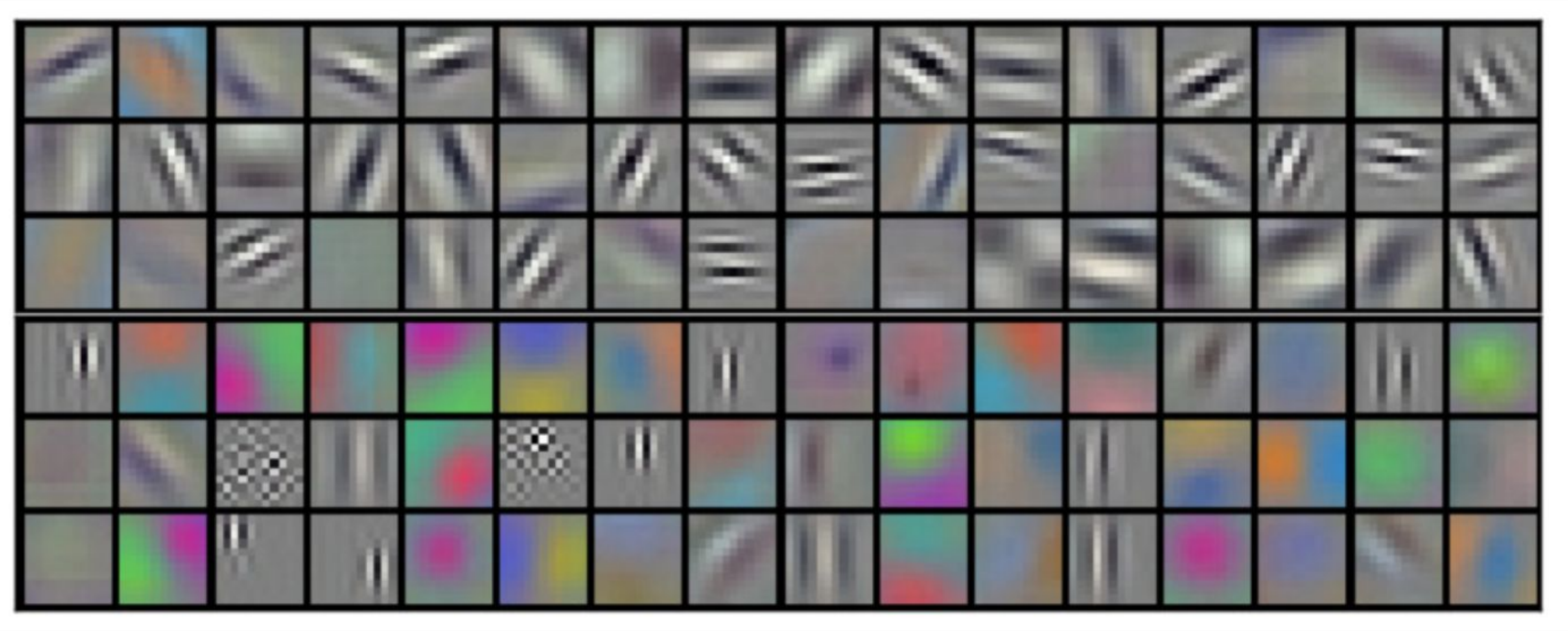
Raw data

Deep learning: Convolutional Networks

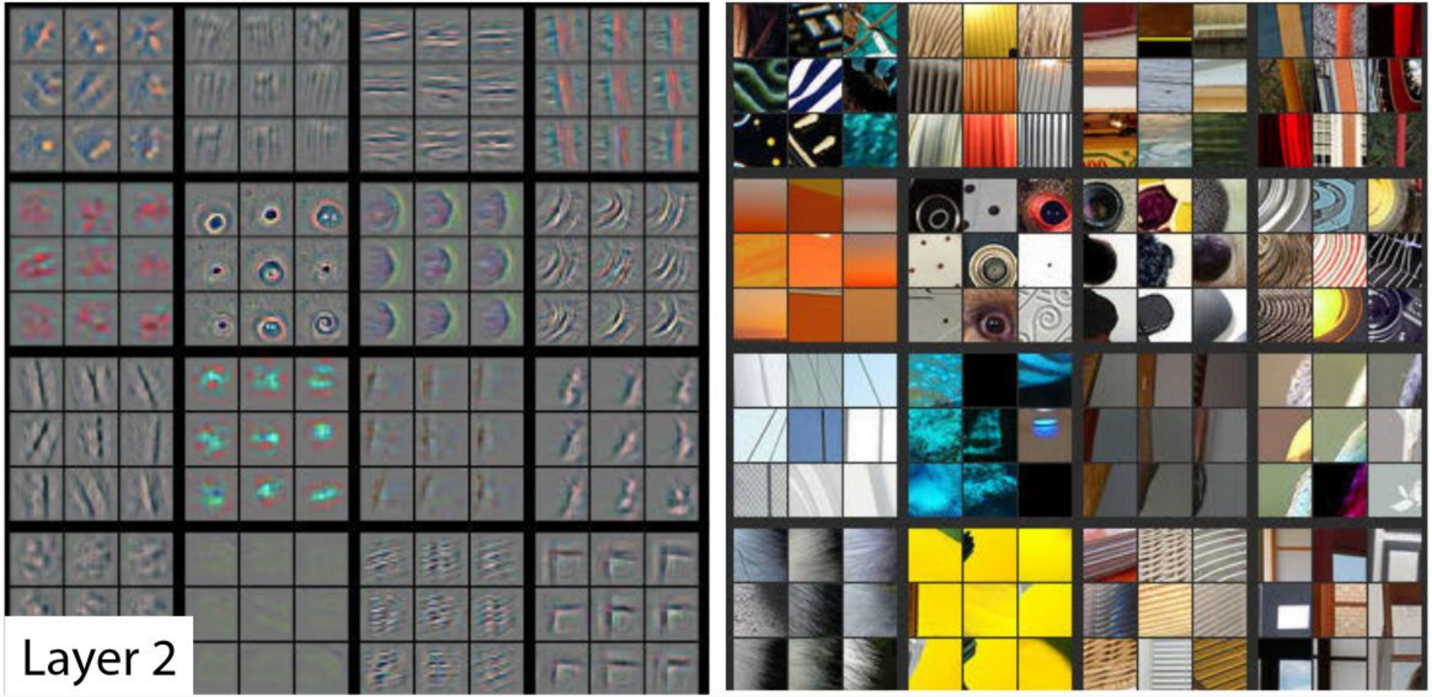


- **Local** connectivity.
- **Invariance to translations**.
- Current state-of-the-art architectures for image analysis and text processing.

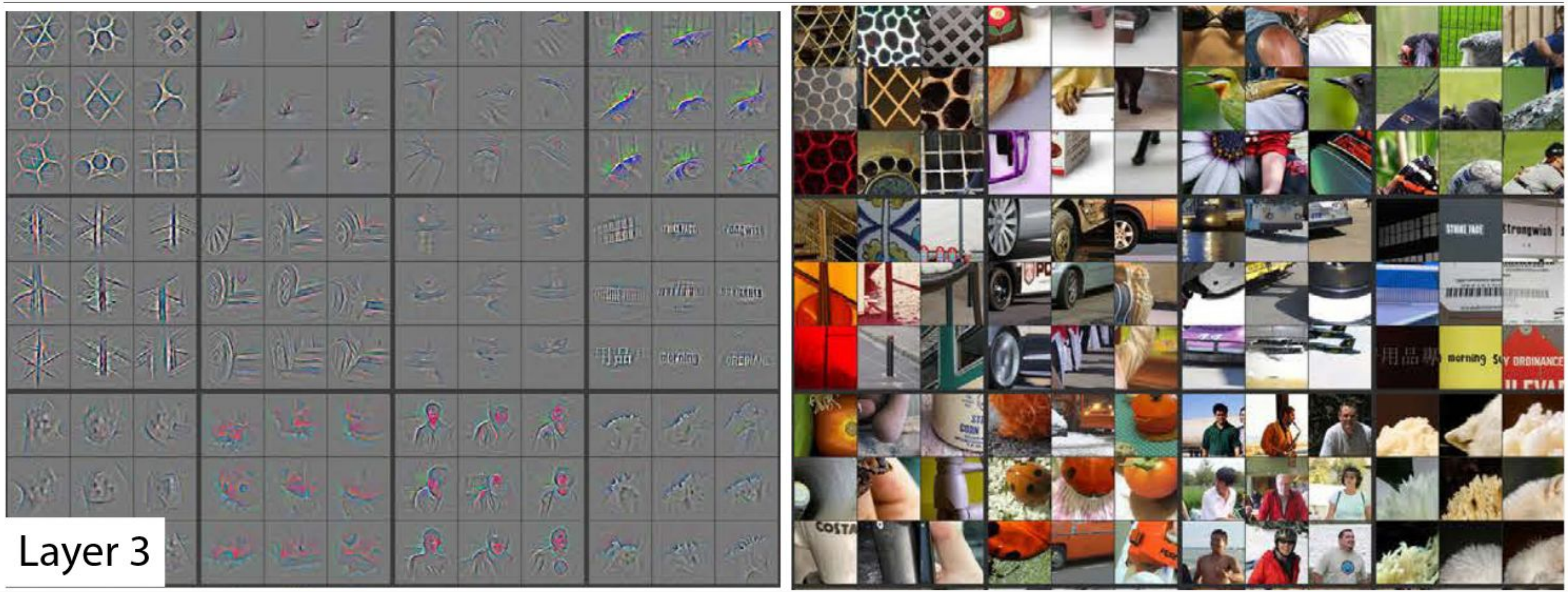
Deep learning: Convolutional Networks



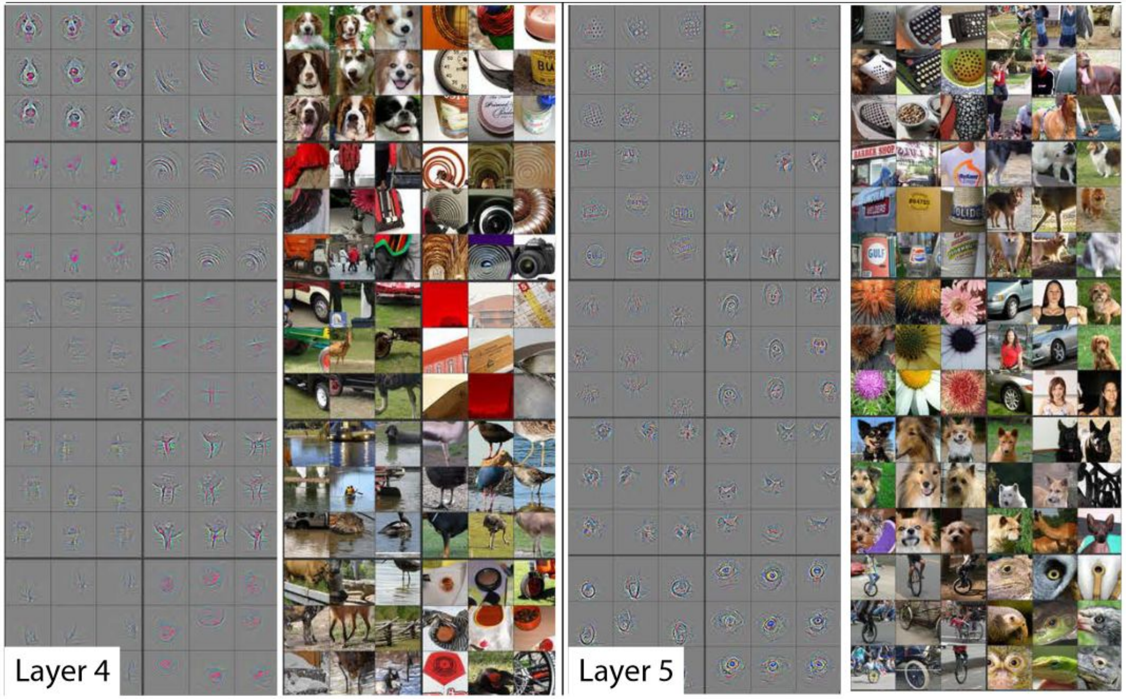
Deep learning: Convolutional Networks



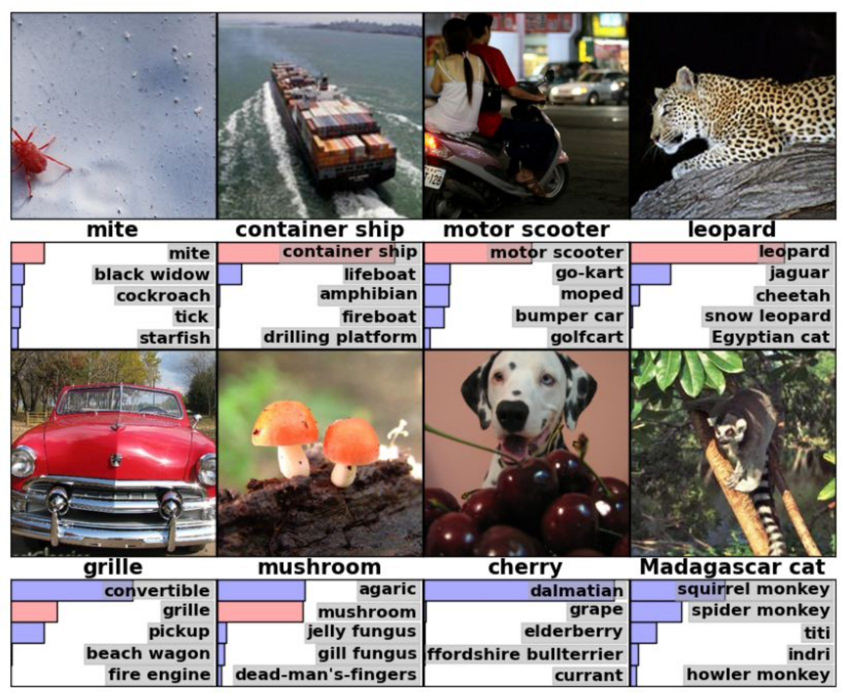
Deep learning: Convolutional Networks



Deep learning: Convolutional Networks



Deep learning: Convolutional Networks



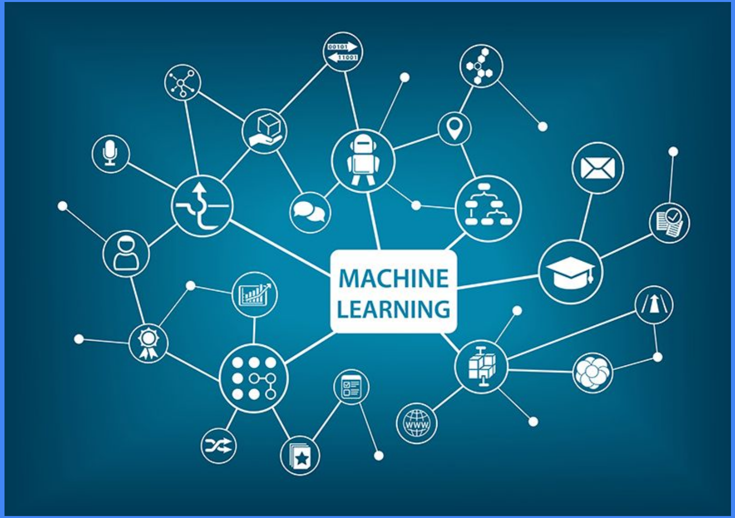
Conclusion

601,705

This is a number of downloads from anaconda.org of the following machine learning packages: scikit-learn, PyTorch, Tensorflow, Theano.

Machine Learning is a breakthrough

In order to handle Big Data, we need scalable and efficient tools.



“All models are wrong but some are useful”

- **Box**, G. E. P. (1979), "Robustness in the strategy of scientific model building"

Machine learning is a
remedy
for cyberattacks.

Thanks!

Contact information:

J.M.Tomczak@uva.nl
jakubmkt@gmail.com
https://jmtomczak.github.io

Code on github:
https://github.com/jmtomczak

UNIVERSITY OF AMSTERDAM

RESEARCH & INNOVATION
Marie Skłodowska-Curie actions

The research carried out by Jakub Tomczak if funded by the European Commission within the Marie Skłodowska-Curie Individual Fellowship (Grant No. 702666, "Deep learning and Bayesian inference for medical imaging").