Centre for
Data Analytics
and Cognition

Online version

Hands-on Deep Learning for Industrial Informatics Applications

Presented by:

- Rashmika Nawaratne
- Achini Adikari
- Daswin De Silva

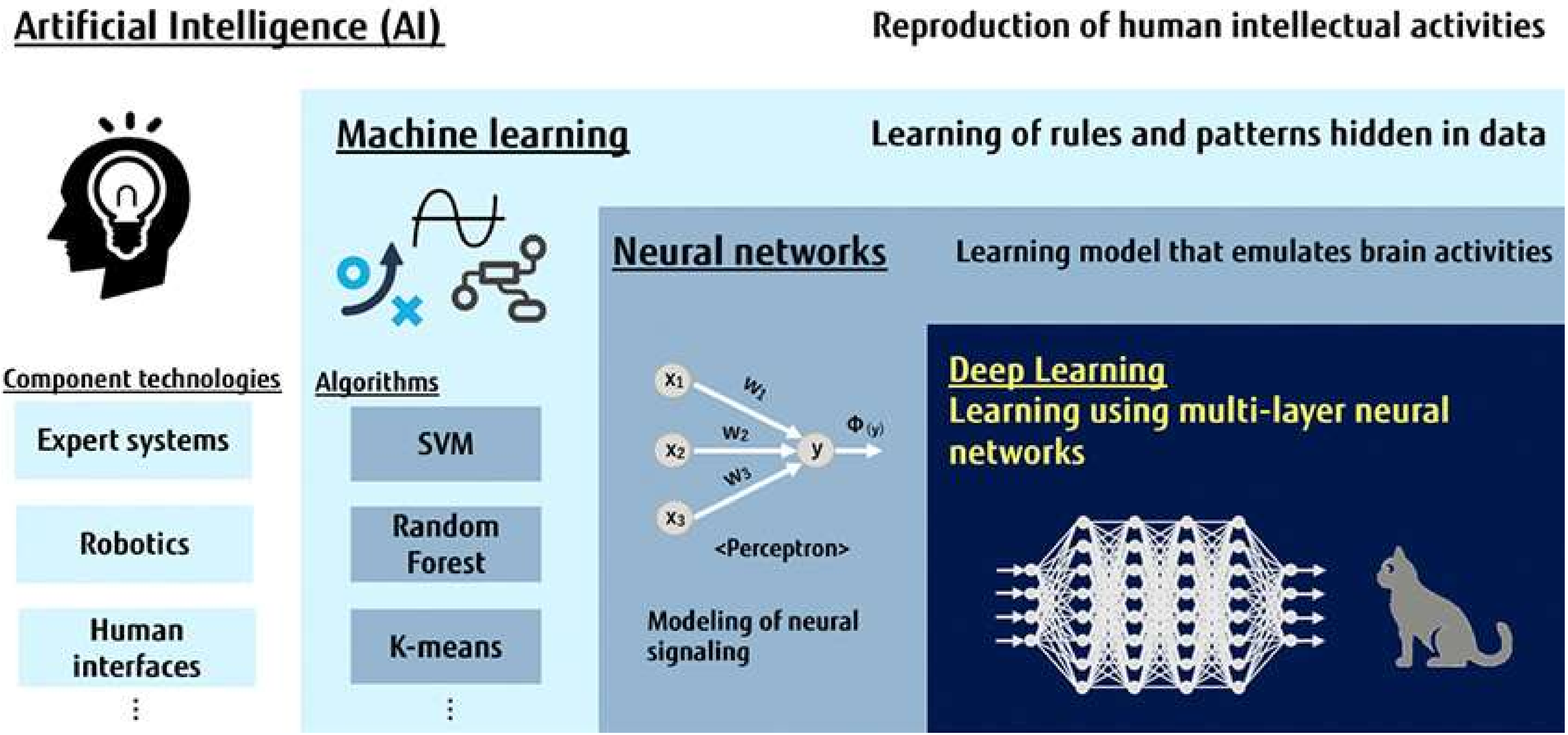
Centre for Data Analytics and Cognition
La Trobe University, Melbourne, Australia.

RESEARCH CENTRE FOR DATA, ANALYTICS AND COGNITION (CDAC)

LA TROBE UNIVERSITY | Centre for Data Analytics and Cognition

- **Vision: to revolutionise human engagement and interaction in data intensive environments by advancing a new paradigm of AI, self-structuring artificial intelligence, inspired by human cognition.**

- Research – world class research in AI, Machine Learning and Data Analytics.
  - Novel ideas, high impact publications, research collaborations and research funding
- Technology - transformation and integration of research into tools and platforms.
  - Software (API) libraries and platforms developed for industry and commercial applications
- Collaboration - with academia, industry and government.
  - More than $3 million of government and industry funding (2014-2019)
- Training - postgraduate teaching and executive education in data analytics and AI.
  - More than 100 full-time equivalent coursework students with increasing rates of employment.



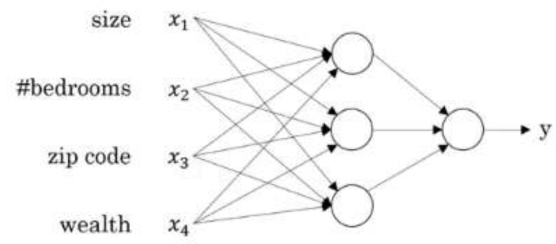
Sato et al., “Overview of Deep Learning”, 2018

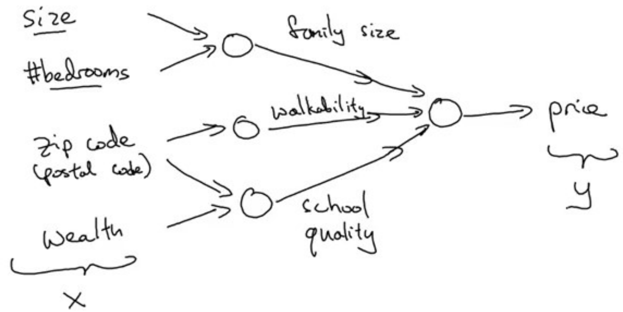
A Neural Network for multiple attributes

Centre for Data Analytics and Cognition

- A single neuron is sufficient when just the single attributes (size of the house) is available.
- But when there are multiple features such as number of bedrooms, bathrooms, parking, postcode, etc., more than a single neuron is needed to model the non-linear relationship across all attributes.
- Solution is a network of neurons (neural network).
- Think about what new features can be derived based on these features.
- Family size, Walkability, School quality – sounds logical?
- A neural network is capable of automatically identifying hidden features from the input features, and finally a prediction.
- Let us further explore these “hidden features”.

price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition
2.161300e+04	21613.000000	21613.000000	21613.000000	2.161300e+04	21613.000000	21613.000000	21613.000000	21613.000000
5.400881e+05	3.370842	2.114757	2079.899736	1.510697e+04	1.494309	0.007542	0.234303	3.409430
3.671272e+05	0.930062	0.770163	918.440897	4.142051e+04	0.539989	0.086517	0.766318	0.650743
7.500000e+04	0.000000	0.000000	290.000000	5.200000e+02	1.000000	0.000000	0.000000	1.000000
3.219500e+05	3.000000	1.750000	1427.000000	5.040000e+03	1.000000	0.000000	0.000000	3.000000
4.500000e+05	3.000000	2.250000	1910.000000	7.618000e+03	1.500000	0.000000	0.000000	3.000000
6.450000e+05	4.000000	2.500000	2550.000000	1.068800e+04	2.000000	0.000000	0.000000	4.000000
7.700000e+06	33.000000	8.000000	13540.000000	1.651359e+06	3.500000	1.000000	4.000000	5.000000





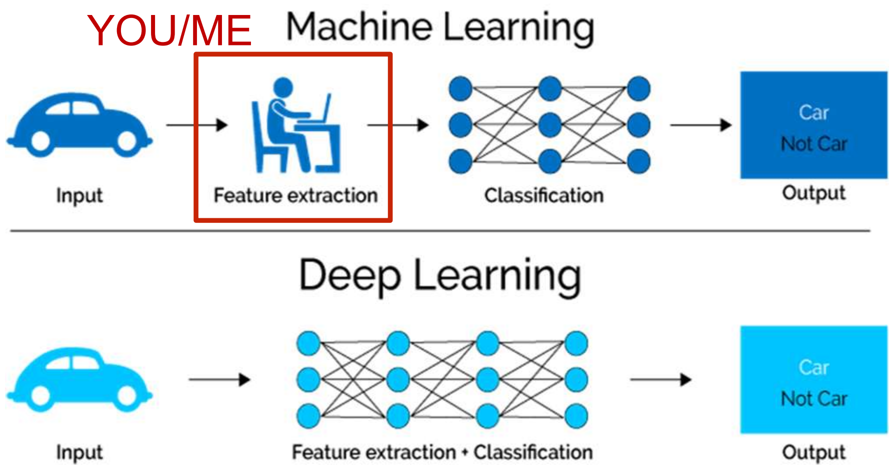
DeepLearning.ai, “Neural Networks”, 2018.

Extracting hidden features

LA TROBE UNIVERSITY
Centre for Data Analytics and Cognition

- Feature engineering - in traditional machine learning the extraction of hidden features is usually conducted by humans as a pre-processing step.
- Even though traditional neural networks have the capability to automatically extract features, it requires multiple layers of intermediary neurons.
- As the number of multiple layers of intermediary neurons increases, so does the training time and computational complexities. This is why **traditional neural networks are limited to 2 or 3 hidden layers**.
- Deep neural networks consist of more than 3 layers with a large number of hidden units, which facilitates automatic feature extraction.

A deep neural network is much more effective because it can automatically extract features and learn/train based on these features.



Satyaseel et al., “Introduction to Neural Network”, 2018.

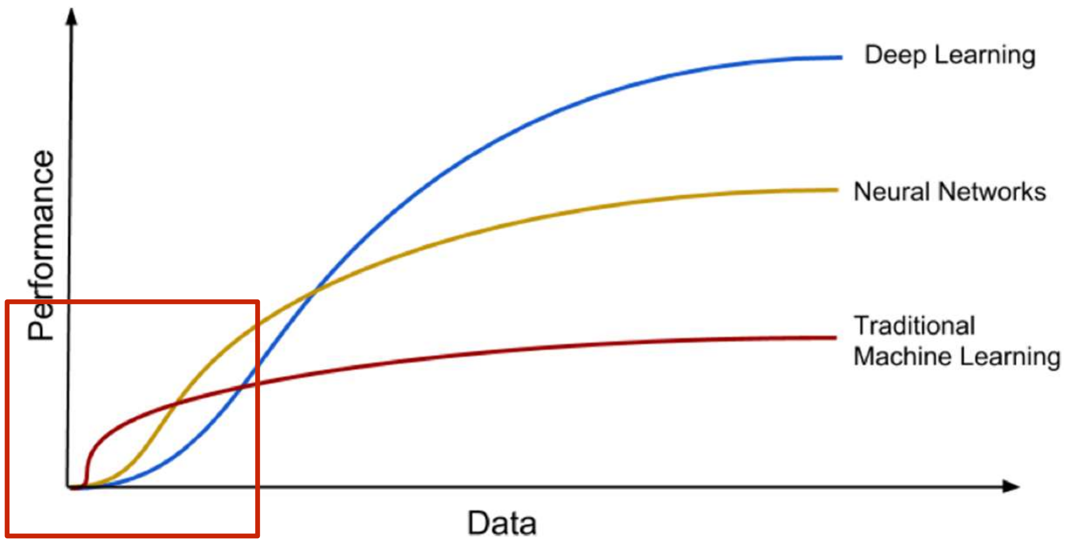
Is there a hype in Deep Learning?

LA TROBE UNIVERSITY
Centre for Data Analytics and Cognition

- The performance of traditional machine learning algorithms has not kept up with exponential increases in the volume, velocity and variety of data.
- The learning performance saturates very early into the training phase.
- In the case of neural networks, the performance of the model increases as the availability of the data increases.

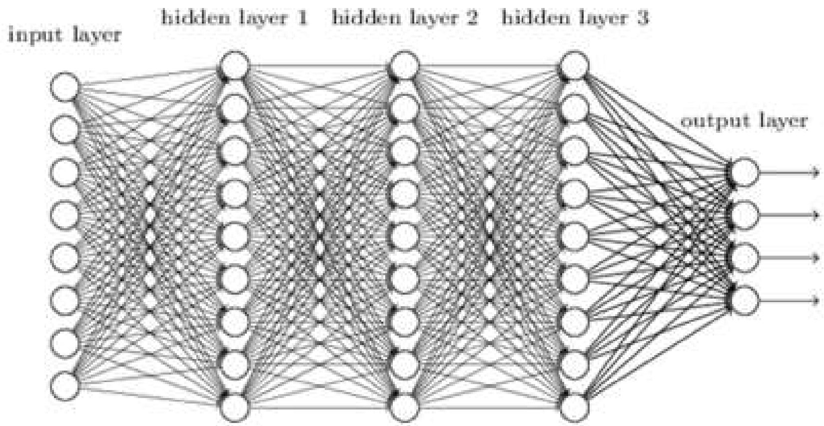
Deep learning vs conventional machine learning:

1. Automated feature extraction
2. Utilization of Big Data for training
3. High performance computing infrastructure that accommodates the computational complexity of advanced activations
4. New brain-inspired learning algorithms for optimization and regularization

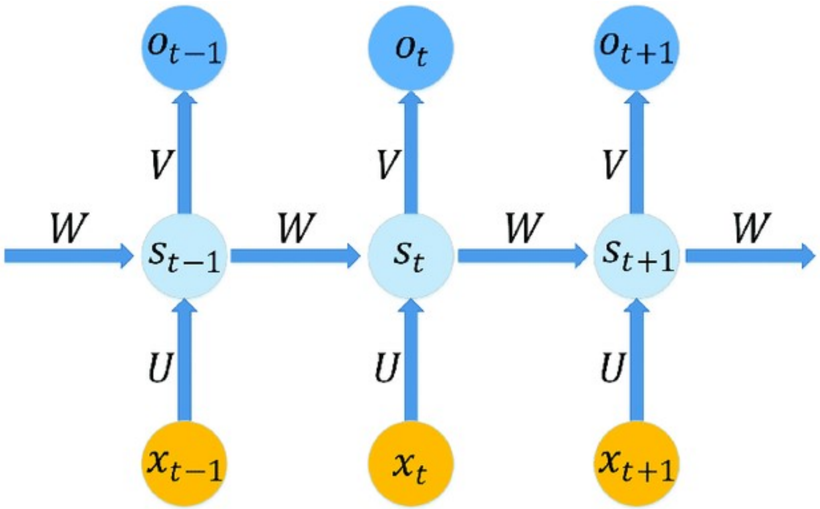


Source: Online

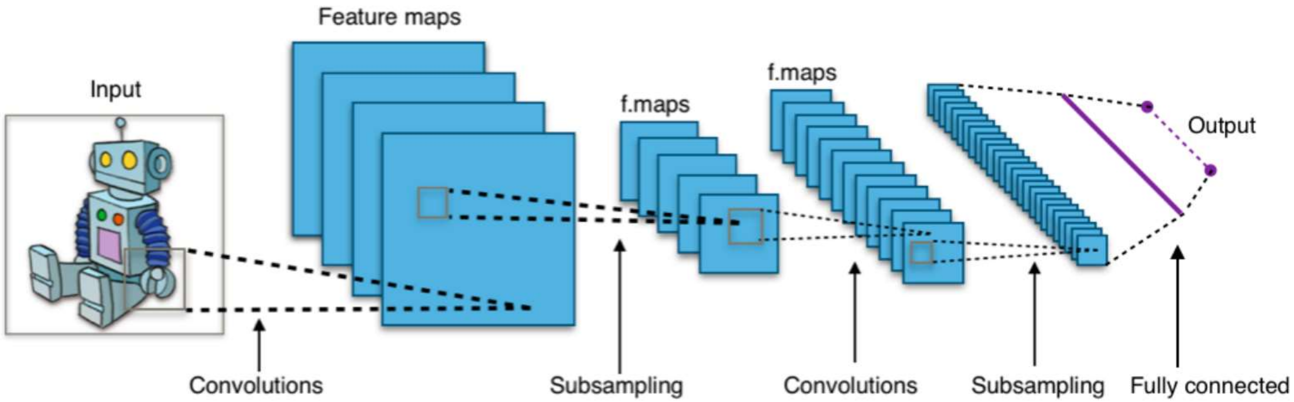
Widely used deep neural network types



Deep Neural Network



Recurrent Neural Network



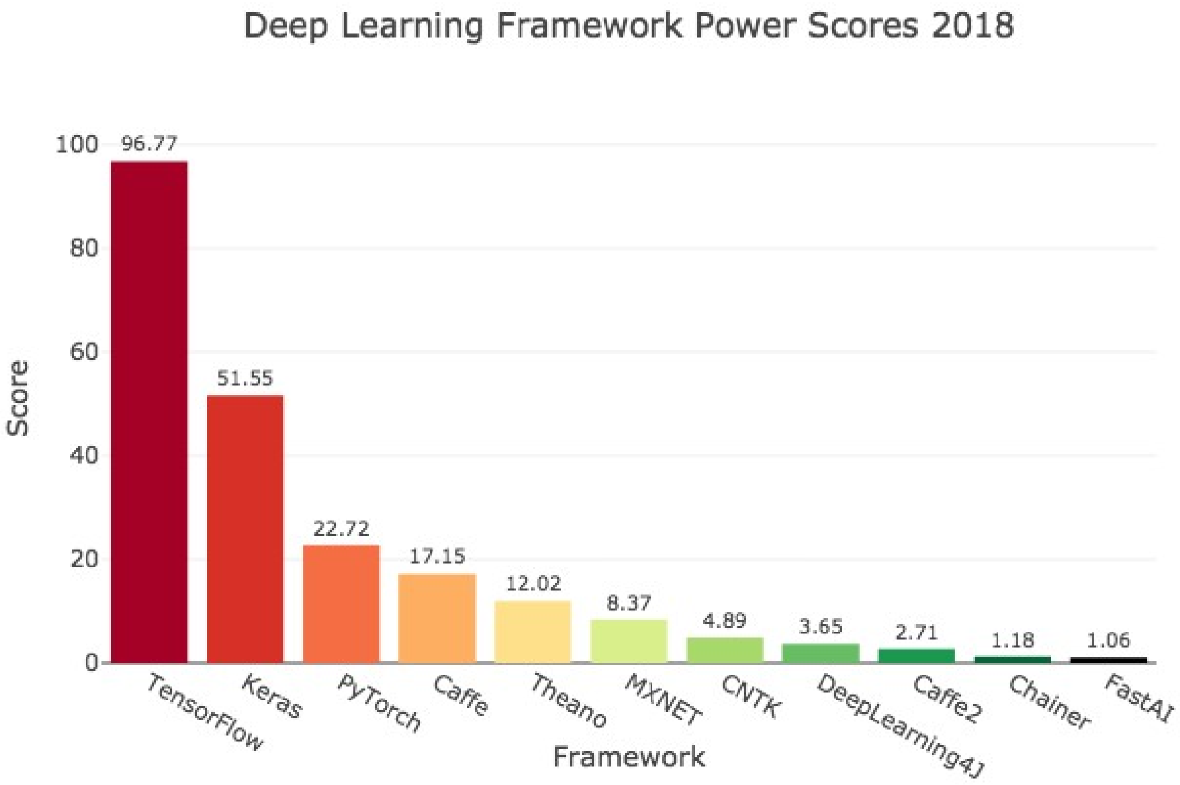
Convolutional Neural Network

Source: Online

Many options for developing a deep learning algorithm

Centre for Data Analytics and Cognition

Deep Learning Framework Power Scores by KDnuggets, Sept, 2018

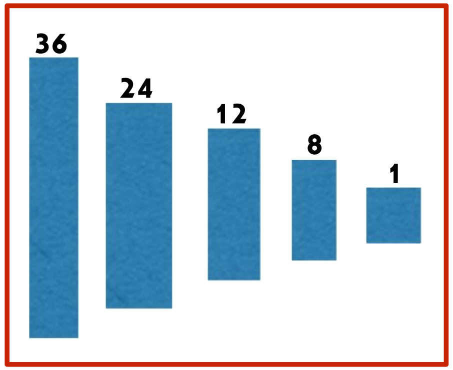


DNN – "Hello World"

LA TROBE UNIVERSITY
Centre for Data Analytics and Cognition

```
X_train, X_test, y_train, y_test = train_test_split(df_numerical, Y_scaled,
                                        test_size=0.3, random_state=2)
NN_model = Sequential()

NN_model.add(Dense(36, input_dim = X_train.shape[1], activation='relu'))

NN_model.add(Dense(24, activation='relu'))
NN_model.add(Dense(12, activation='relu'))
NN_model.add(Dense(8, activation='relu'))

NN_model.add(Dense(1, activation='linear'))

# Compile the DNN
NN_model.compile(loss='mean_absolute_error', optimizer='sgd', metrics=['mse'])
# Train the model
NN_model.fit(X_train, y_train, epochs=epochs, batch_size=batch_size,
                    validation_split = validation_split)
```

Model Architecture

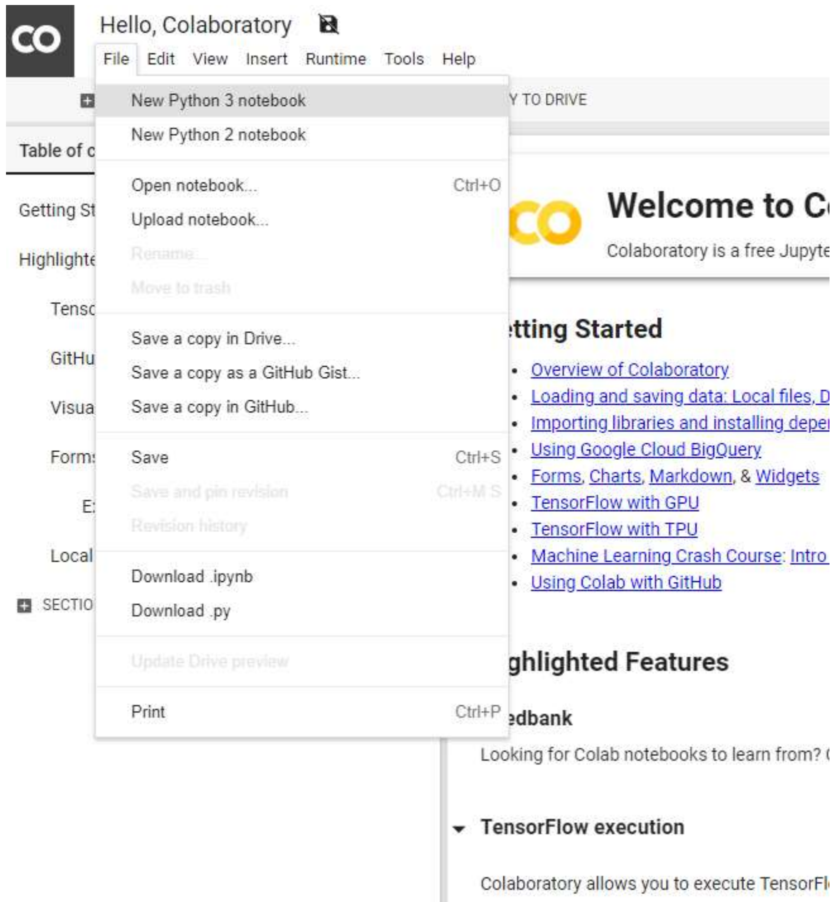


Hyper-params

Python on Google Colab

Centre for
Data Analytics
and Cognition

- Google Colaboratory (Colab) is a free Jupyter notebook environment that requires no setup.
- This runs on Google Cloud Servers.
- Go to https://colab.research.google.com/
  (Tested on Chrome, Firefox and Safari)
- Sign in to your Gmail account
- Create a new Python 3 Notebook
- Hands on …



Exercise 1: Deep Neural Networks for House Price Prediction

Centre for Data Analytics and Cognition

- Standard application of DNN to predict house price based on multiple attributes.
- Using the Ames Housing dataset[1]
- The dataset contains 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa.
- In this workbook we will,
  1. Pre-process the data
  2. Develop a deep neural network model for price prediction
  3. Train the model
  4. Evaluate the model
- [https://github.com/CDAC-lab/deep-learning-tutorial]

[1] D. De Cock, "Ames, Iowa: Alternative to the Boston Housing Data as an End of Semester Regression Project," Journal of Statistics Education, vol. 19, no. 3, p. null-null, Nov. 2011.

Sequence in datasets

Centre for Data Analytics and Cognition

- Sequences are data structures where each sample could be seen as a series of data points.
- E.g., “I am excited to present the DL workshop at ETFA 2020.”
  This sentence is an example that consists of multiple words and each word depends on the previous (one or many) terms.
- Same sequencing applies to instances such as IoT readings, audio, video and time-series data.
- Some application areas:
  - Motor traffic prediction
  - Energy consumption prediction
  - Human activity monitoring
  - Image recognition and annotation
  - Asset monitoring and lifecycle management
  - Speech recognition (Siri, Alexa, Google)
  - Language translation
  - Stock market prediction

Structured data

Year	City	Country	Nations
1896	Athens	Greece	14
1900	Paris	France	24
1904	St. Louis	USA	12
...	...	...	...
2004	Athens	Greece	201
2008	Beijing	China	204
2012	London	UK	204

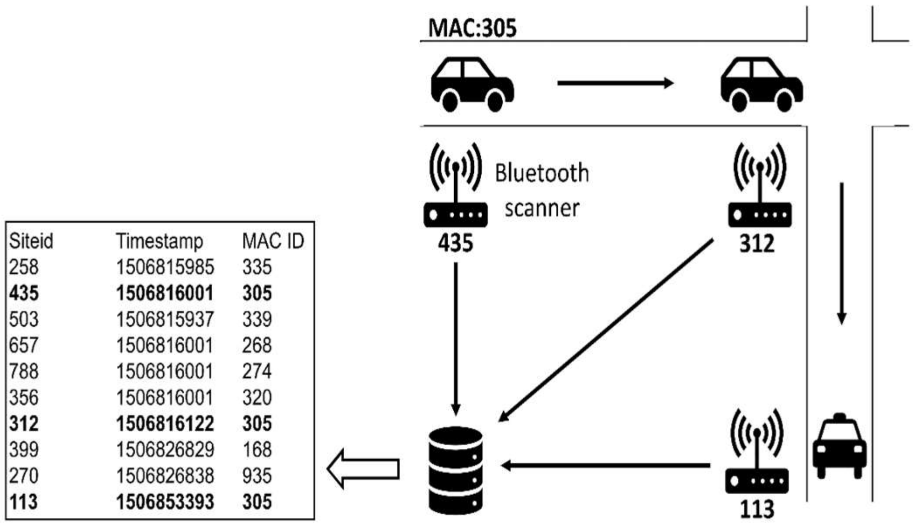
Sequence data
e.g., IoT, Text, Time Series

Date	Open	Sales
2013-09-13	1	8610
2013-07-31	1	10895
2014-10-26	0	0
2015-05-12	1	6043
2013-10-19	1	3568
2014-01-15	1	5163
2014-09-09	1	5978
2013-12-20	1	12352

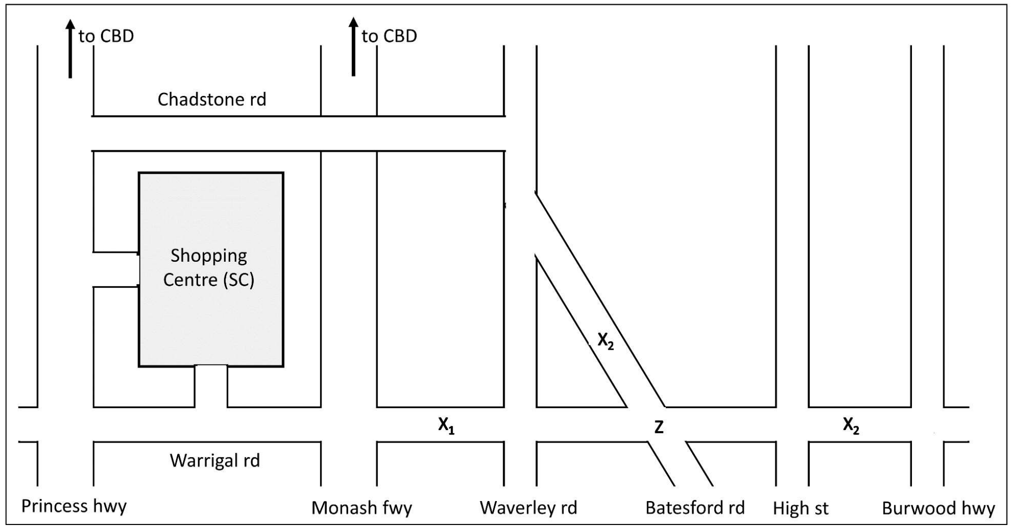
Exercise 2: Traffic Prediction using Recurrent Neural Networks

LA TROBE UNIVERSITY
Centre for Data Analytics and Cognition

- Hands on Recurrent Neural Networks (LSTM).
- Experiment with Bluetooth traffic monitoring data from Victoria roads authority, Australia.[1]
- Uses RNN to predict the traffic flow (vehicle count) at intersection Z, based on its neighboring road segments.
- [https://github.com/CDAC-lab/deep-learning-tutorial]



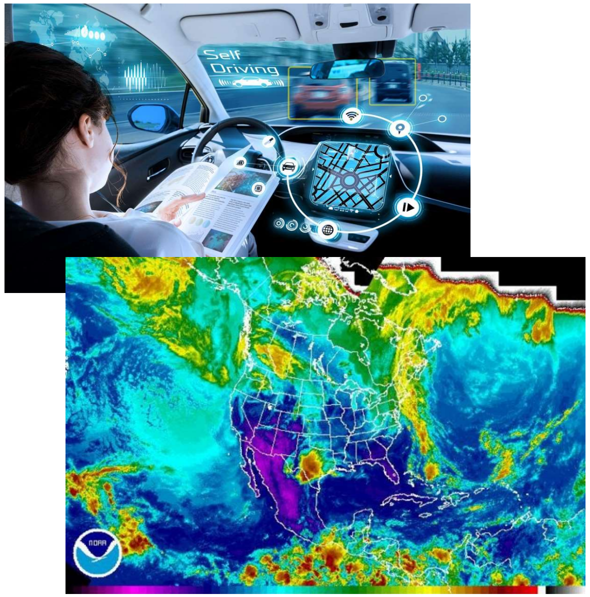
Siteid	Timestamp	MAC ID
258	1506815985	335
435	**1506816001**	**305**
503	1506815937	339
657	1506816001	268
788	1506816001	274
356	1506816001	320
312	**1506816122**	**305**
399	1506826829	168
270	1506826838	935
113	**1506853393**	**305**

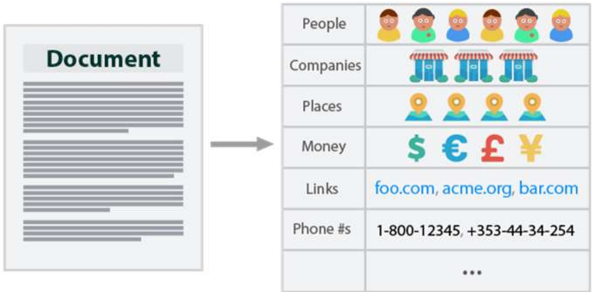


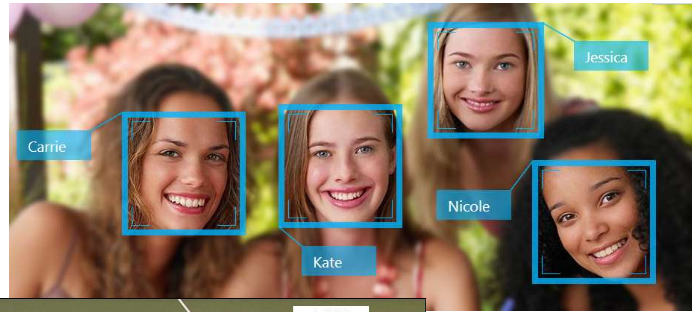
[1] D. Nallaperuma, R. Nawaratne, T. Bandaragoda, A. Adikari, S. Nguyen, T. Kempitiya, D. De Silva, D. Alahakoon, D. Pothuhera. "Online Incremental Machine Learning Platform for Big Data-Driven Smart Traffic Management." IEEE Transactions on Intelligent Transportation Systems, pp. 1–12, 2019.

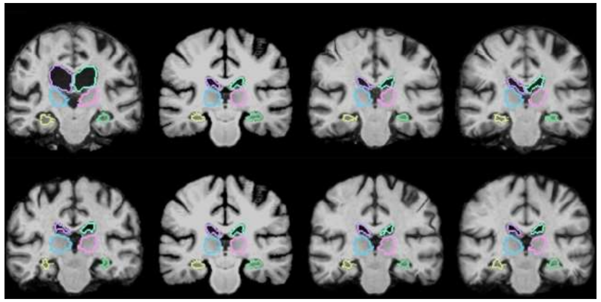
Unstructured data - images and video

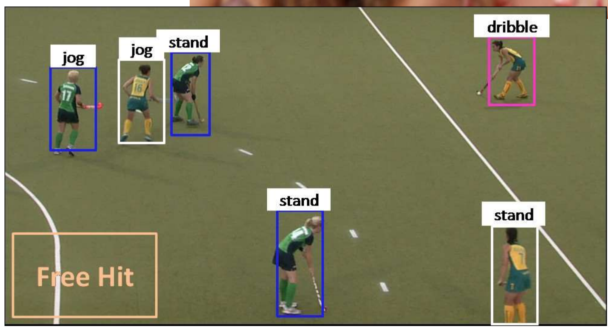
- DNN can be effectively used to analyse (classify/predict) numerical attributes and sequential datasets (using RNN).
- However, there are a number of application areas such defense and security, weather analysis, telecommunication, transportation, entertainment, etc., that generate **high-dimensional** data such as images and videos.







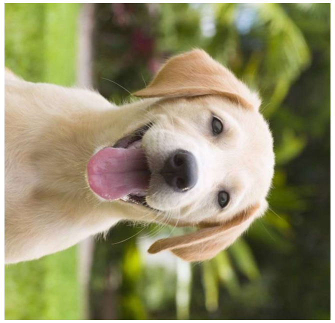
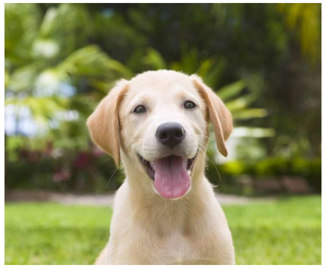
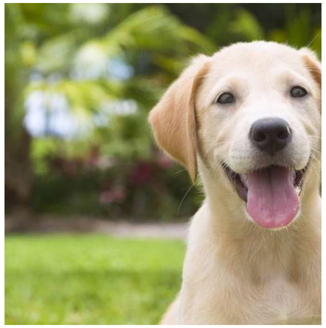




Source: Online

Oversimplification of treating images as digits

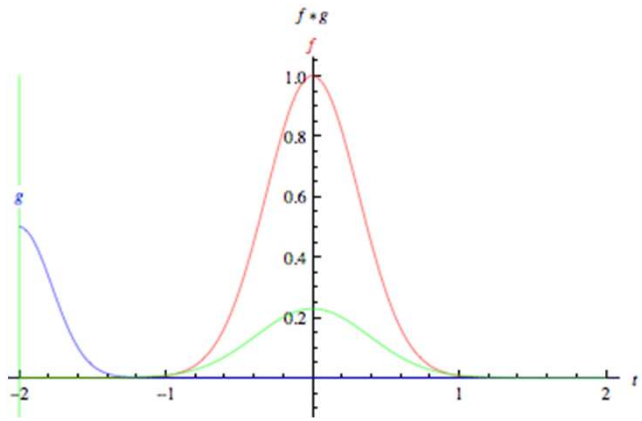
- High dimensionality
  - Size: 200 x 200 pixels (40,000 attributes)
  - Colour, 3 channels for red, greed and blue (40,000 x 3 = 120,000 attributes)
- Rotation
- Scale
- Location

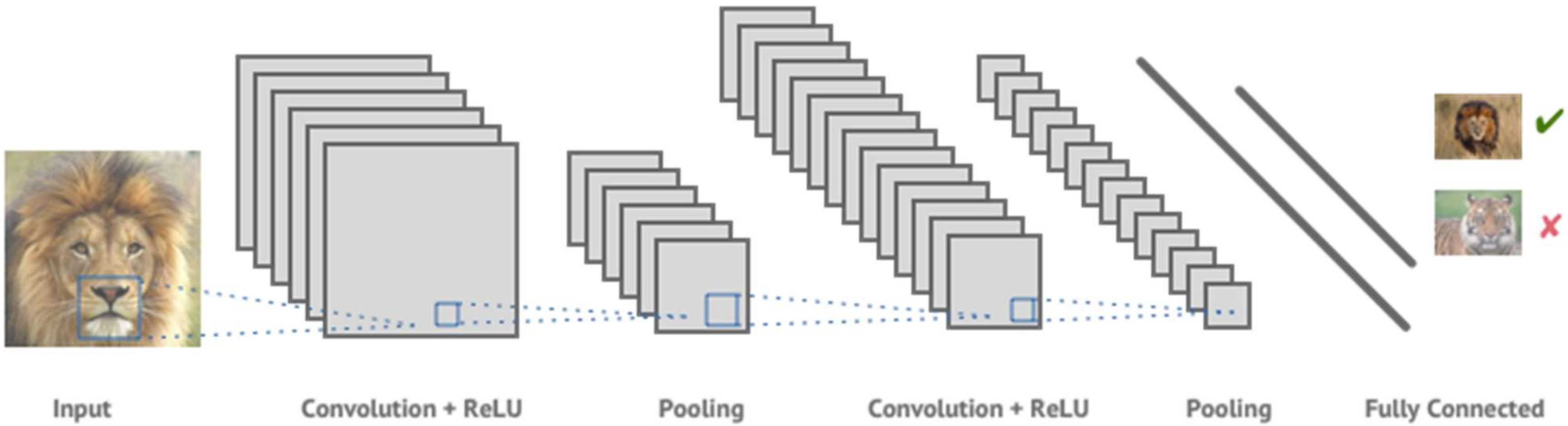





Source: Online

Convolutional Neural Networks (CNN)

- Convolution: Latin for “to convolve” means to roll together.
- Convolution can be thought of as the mix of two functions by multiplication.
- The static red curve is the input image, and the mobile function is a filter which picks up a signal or feature in the image. The two functions relate through multiplication.

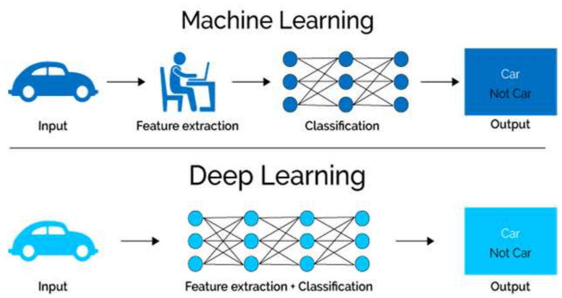


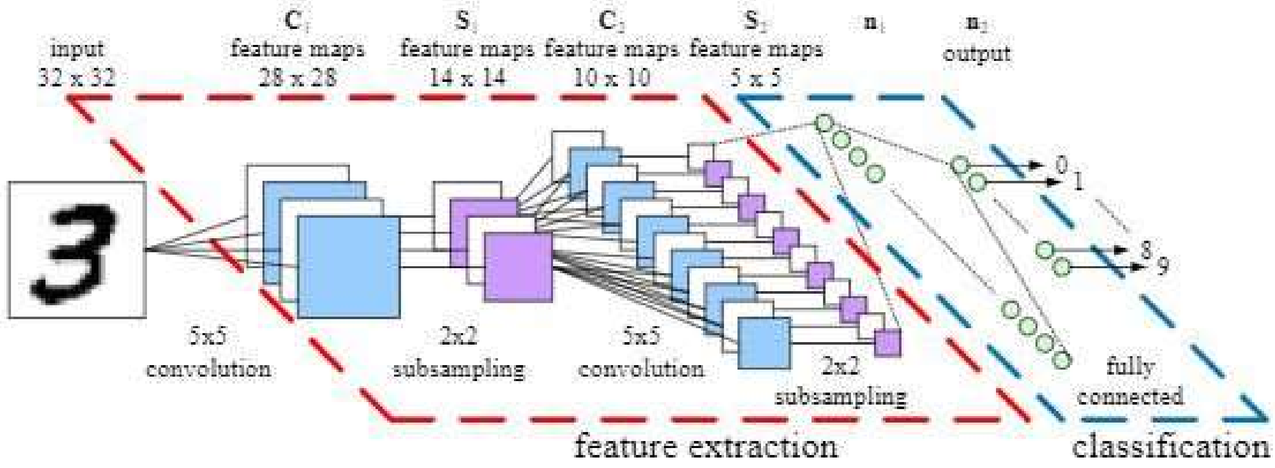


Source: Online

Workings of CNN

- A CNN is composed of two major parts; **feature extraction** and **classification**.
- *Feature extraction* is the process of automatically extracting features from the images.
  - Features can be understood as attributes of the image, e.g., an image of a cat might have features like whiskers, two ears, four legs etc. A handwritten digit image might have features as horizontal and vertical lines or loops and curves.
- *Classification* is the aspect related to prediction of an outcome.



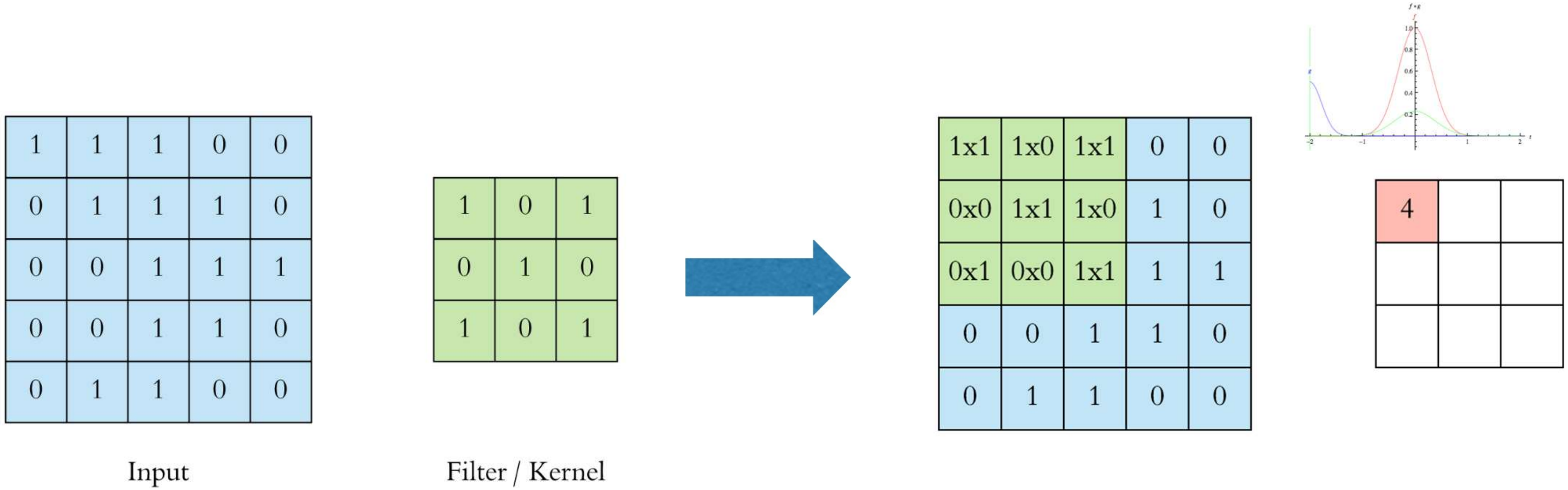


Source: Mesman et al. 2011

CNN: Feature Extraction

LA TROBE UNIVERSITY
Centre for Data Analytics and Cognition

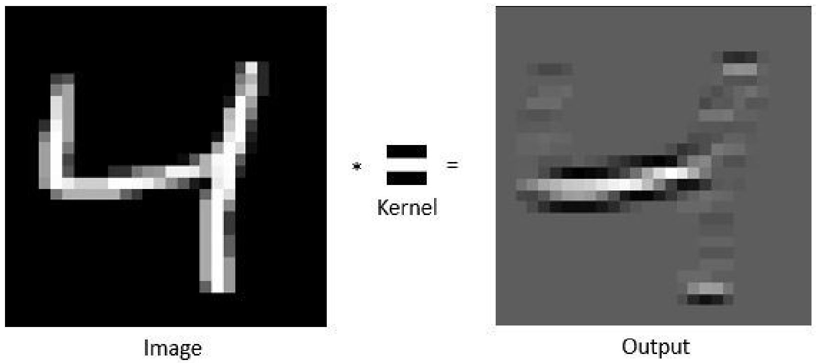
- Feature extraction of a CNN is conducted using three types of operations. 1) Convolution, 2) Non-Linearity and 3) Pooling.
- Convolution is the mathematical operation which is central to the efficacy of this algorithm.
- Convolution operation is performed on an input image using a filter or a kernel.



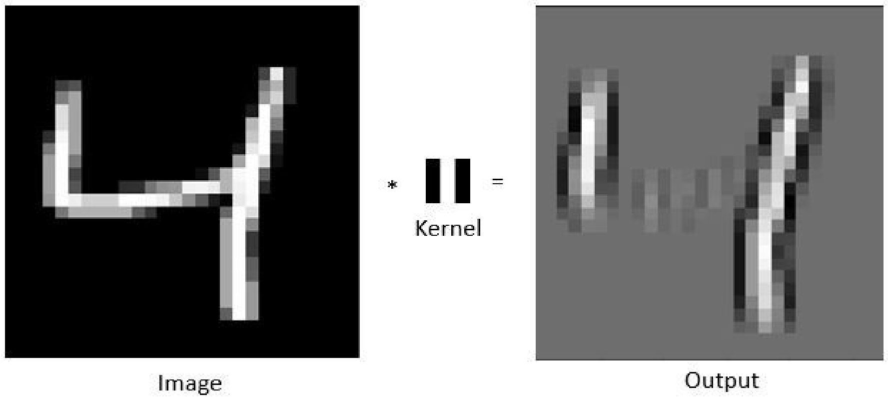
Source: Karkare 2019

CNN: Feature Extraction – Convolution Operation

LA TROBE UNIVERSITY
Centre for Data Analytics and Cognition



Horizontal Edge Filter



Vertical Edge Filter

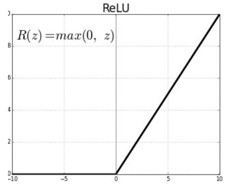
Source: Karkare 2019

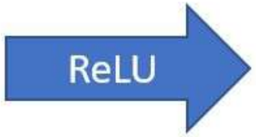
CNN: Feature Extraction – Non-Linearity

LA TROBE UNIVERSITY
Centre for Data Analytics and Cognition

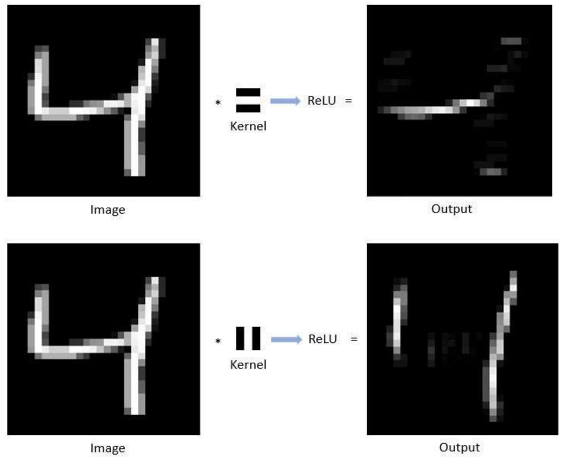
- After sliding our filter over the original image, the output which we get is passed through another mathematical function which is called an activation function.
- This is the same activation function ReLU (Rectified Linear Unit) we used for the Deep Neural Network.
- Makes this computationally inexpensive than the standard activation functions, such as sinh or tanh.

1	14	-9	4
-2	-20	10	6
-3	3	11	1
2	54	-2	80





1	14	0	4
0	0	10	6
0	3	11	1
2	54	0	80

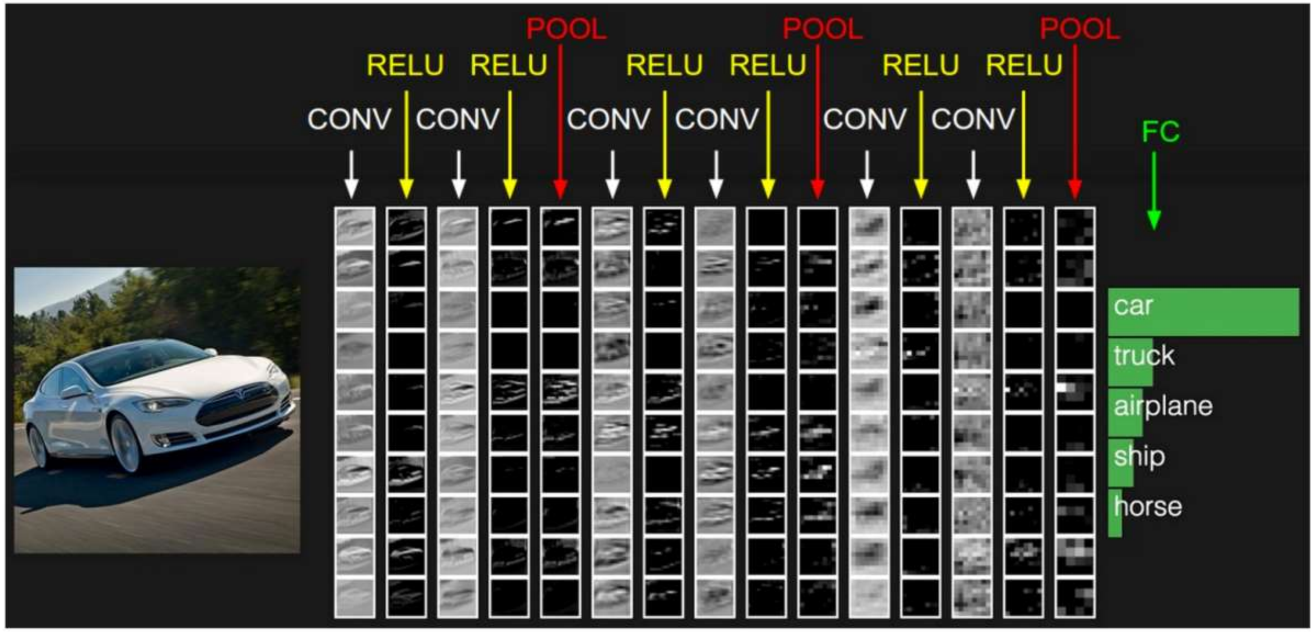


Source: Karkare 2019

CNN: Classification

Centre for Data Analytics and Cognition

- In the CNN, feature extraction is conducted using convolution, activation (non-linearity) and pooling layers.
- Now the extracted features are fed into a 2 or 3 hidden layers (similar to layers in a Deep Neural Network). This is called **Fully Connected** layers in the CNN.



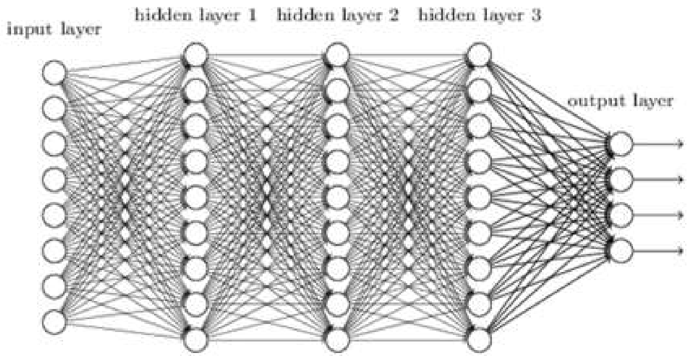
Source: Kdnuggets 2015

Deep Learning so far...

Centre for Data Analytics and Cognition

Structured data

Year	City	Country	Nations
1896	Athens	Greece	14
1900	Paris	France	24
1904	St. Louis	USA	12
...	...	...	...
2004	Athens	Greece	201
2008	Beijing	China	204
2012	London	UK	204

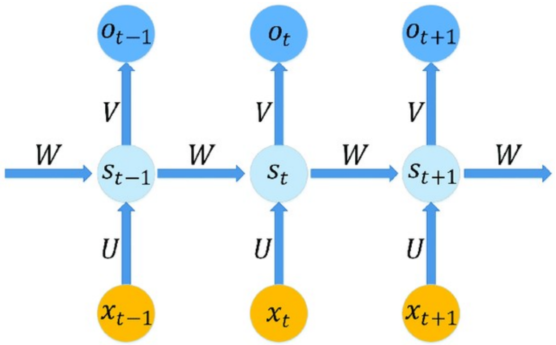


Deep Neural Network

Sequence data e.g., IoT, Text

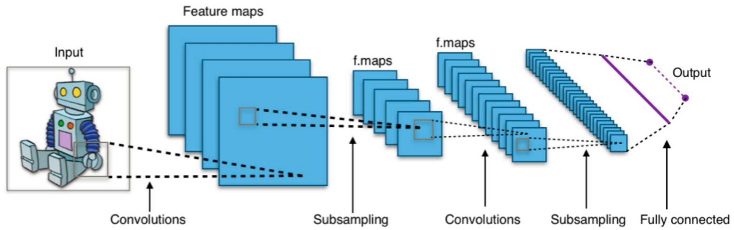
	tweet	wc
0	@user when a father is dysfunctional and is s...	
1	@user @user thanks for #lyft credit i can't us...	
2	bihday your majesty	
3	#model i love u take with u all the time in ...	
4	factsguide: society now #motivation	

Date	Open	Sales
2013-09-13	1	8610
2013-07-31	1	10895
2014-10-26	0	0
2015-05-12	1	6043
2013-10-19	1	3568
2014-01-15	1	5163
2014-09-09	1	5978
2013-12-20	1	12352



Recurrent Neural Network

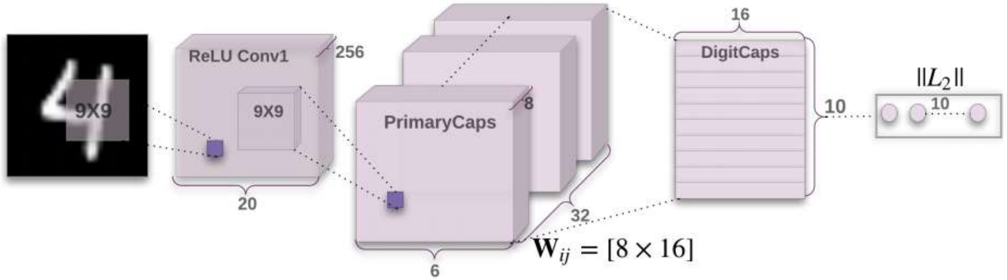
Image and Video data

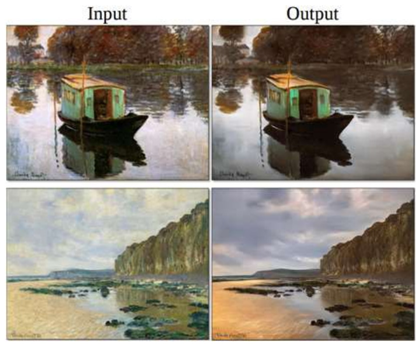


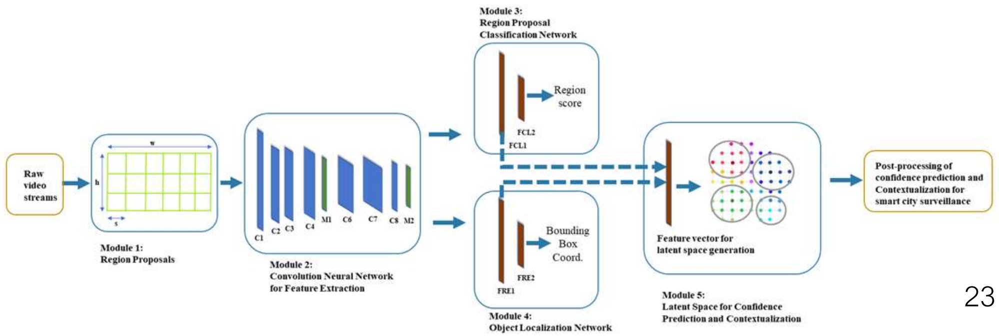
Convolutional Neural Network

Other DNN techniques and architectures

- CapsNet
  - add structures called “capsules” to a CNN to reuse output from several of those capsules to form more stable representations for higher order capsules.
- Generative Adversarial Networks
  - a deep neural net architectures comprised of two nets, pitting one against the other
  - Image generation (e.g., https://thispersondoesnotexist.com/)
- Latent spaces
  - Utilize/train intermediate layers of deep network as feature layer for complementary (secondary) machine learning task
  - E.g., Autoencoder, GSOM







Happy (deep) coding ☺

Keep in touch with our learning materials:
https://www.linkedin.com/company/centre-for-data-analytics-and-cognition-cdac

Contact us:
Rashmika Nawaratne
cdac@latrobe.edu.au
http://latrobe.edu.au/cdac