DIVERGENCE ACADEMY

An Introduction to Deep Learning

Drew Minkin
Lead Data Science Instructor
Jan 5 2023

Pythoneers

Topics Overview

1. The Basics of Machine Learning
2. Machine Learning , Deep Learning & Artificial Intelligence
3. Starting with Deep Learning: Keras
4. TensorFlow vs PyTorch and Model Production
5. ONNX and Model Consumption

Who’s Drew?

Previous Incarnations

SQL Support/Consulting 2000-06

Data Scientist 2010 – Now

5 Startups 2007-19

Analytics Architect/Dev

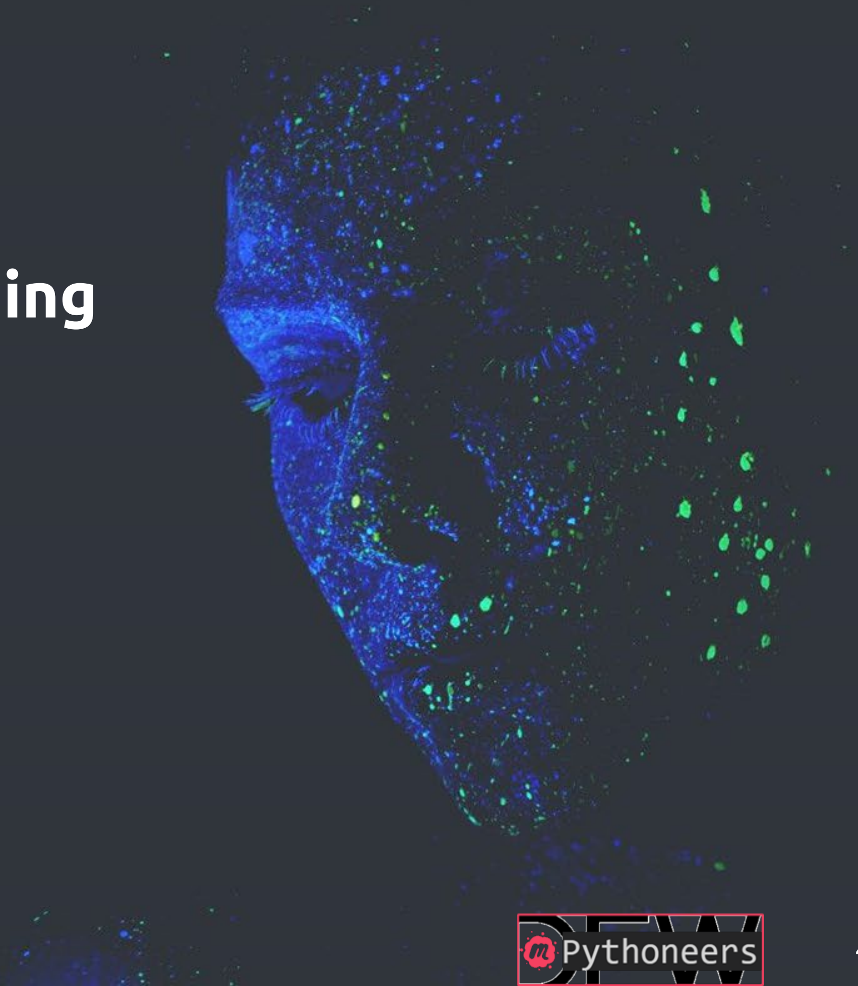
Entrepreneur

Current Incarnation

MCT, Data Science Instructor DIVERGENCE ACADEMY

DIVERGENCE ACADEMY

The Basics of Machine Learning

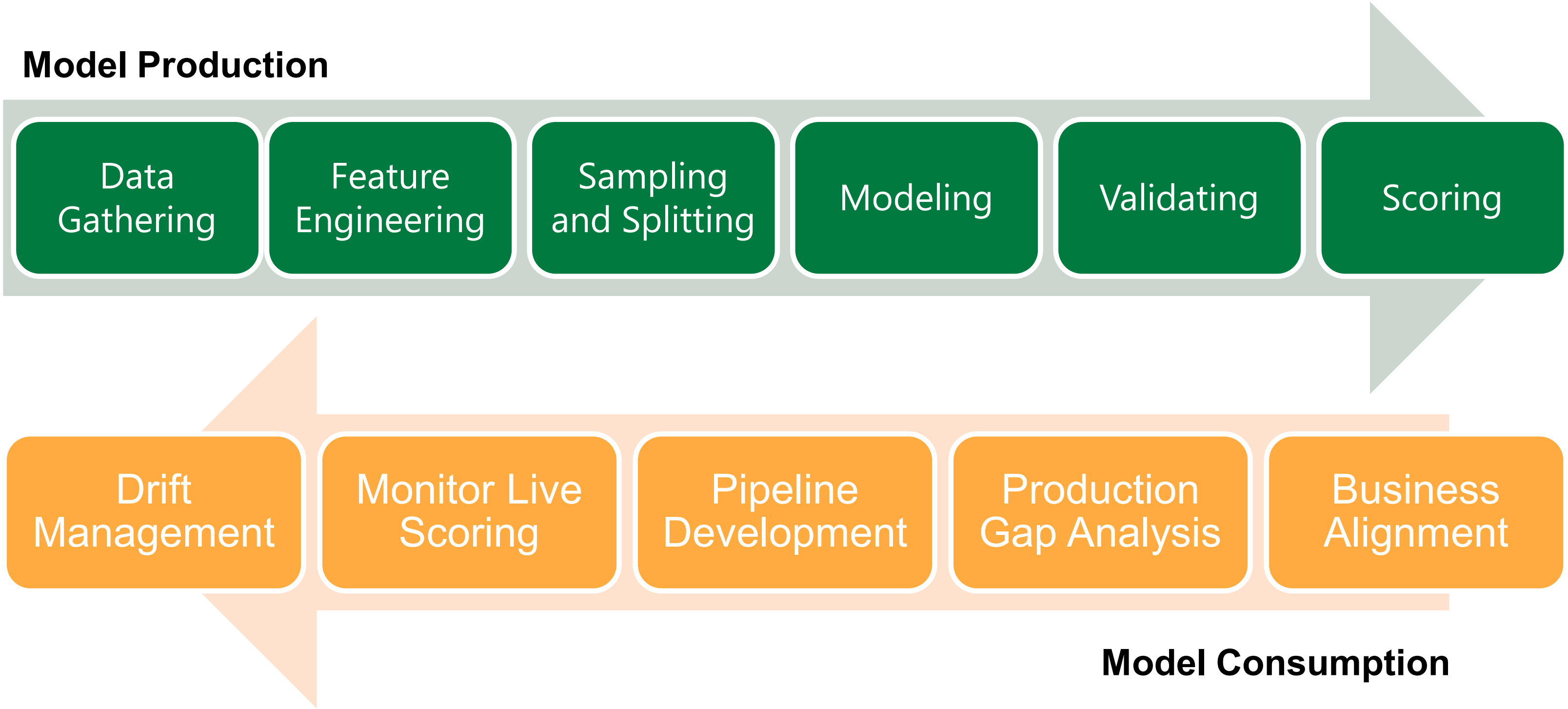


Pythoneers

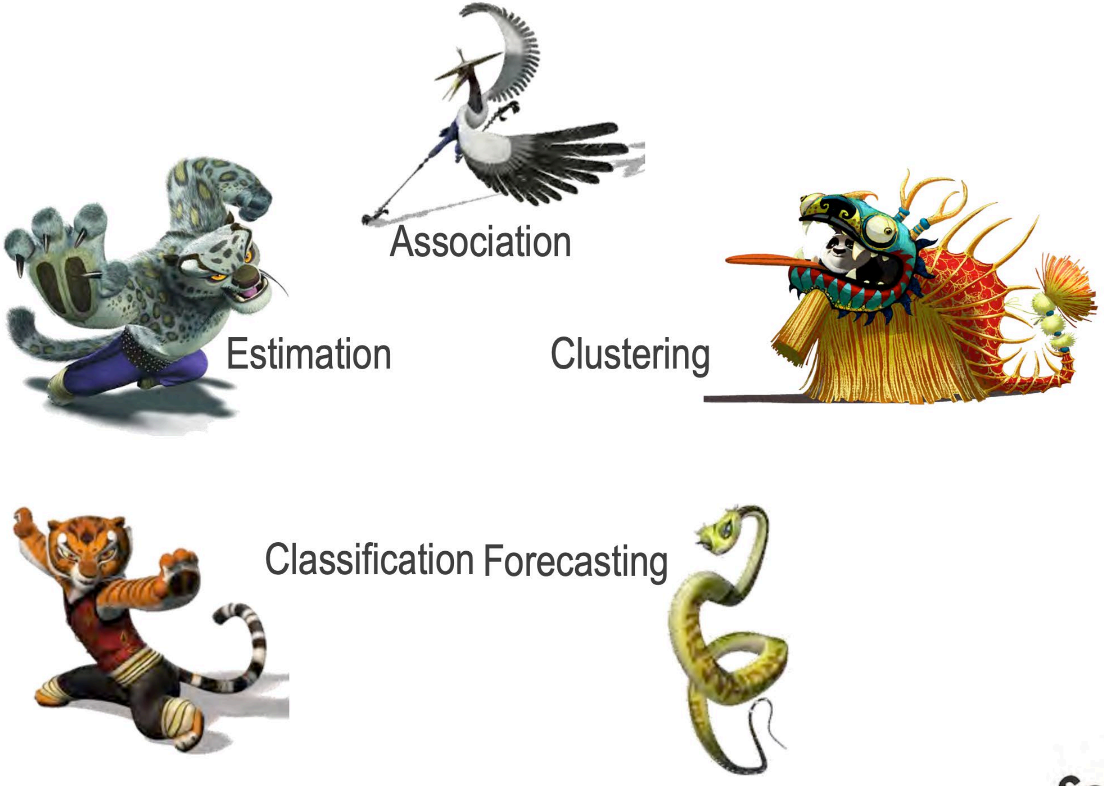
DIVERGENCE ACADEMY

Machine Learning , Deep Learning & Artificial Intelligence

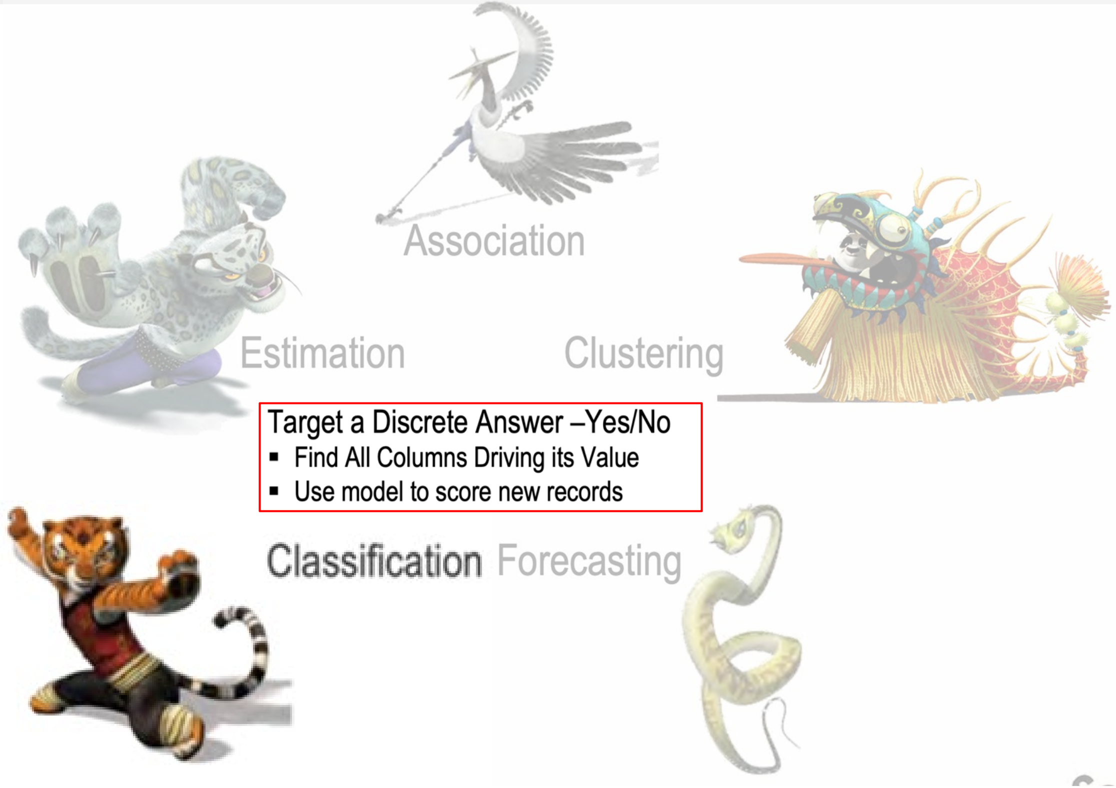
Machine Learning Process



Machine Learning Modeling Families

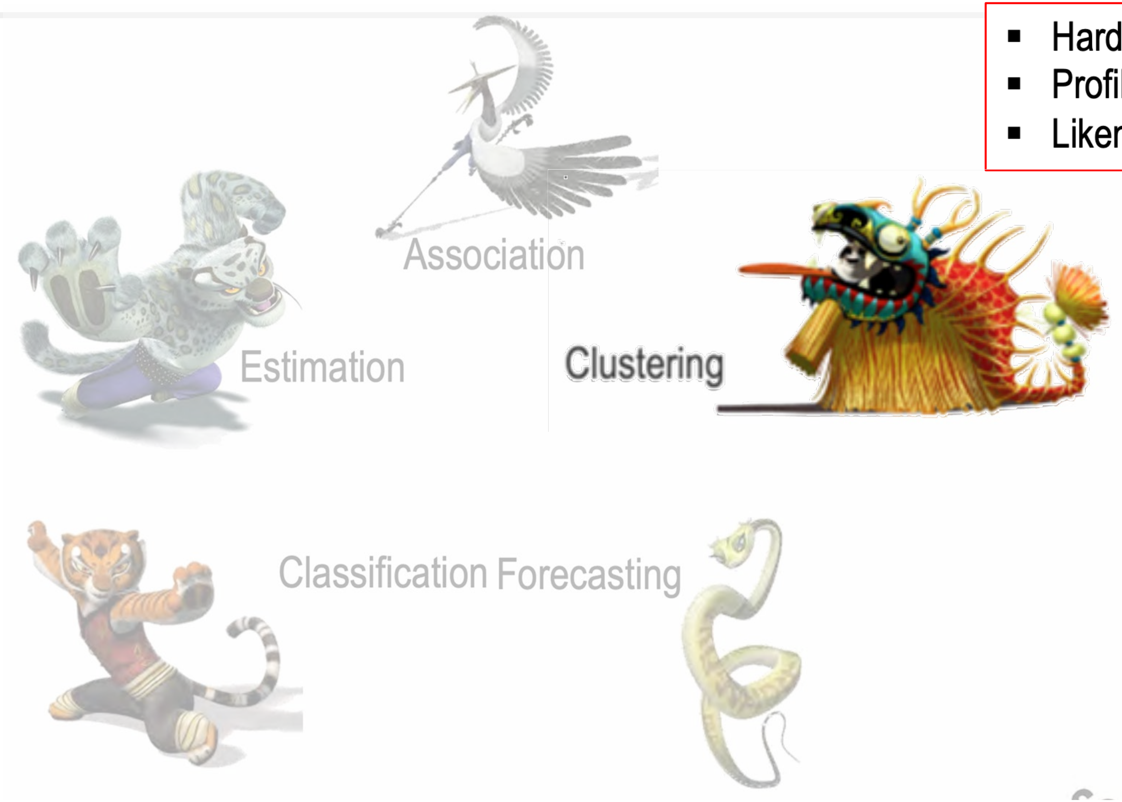


Machine Learning Modeling Families



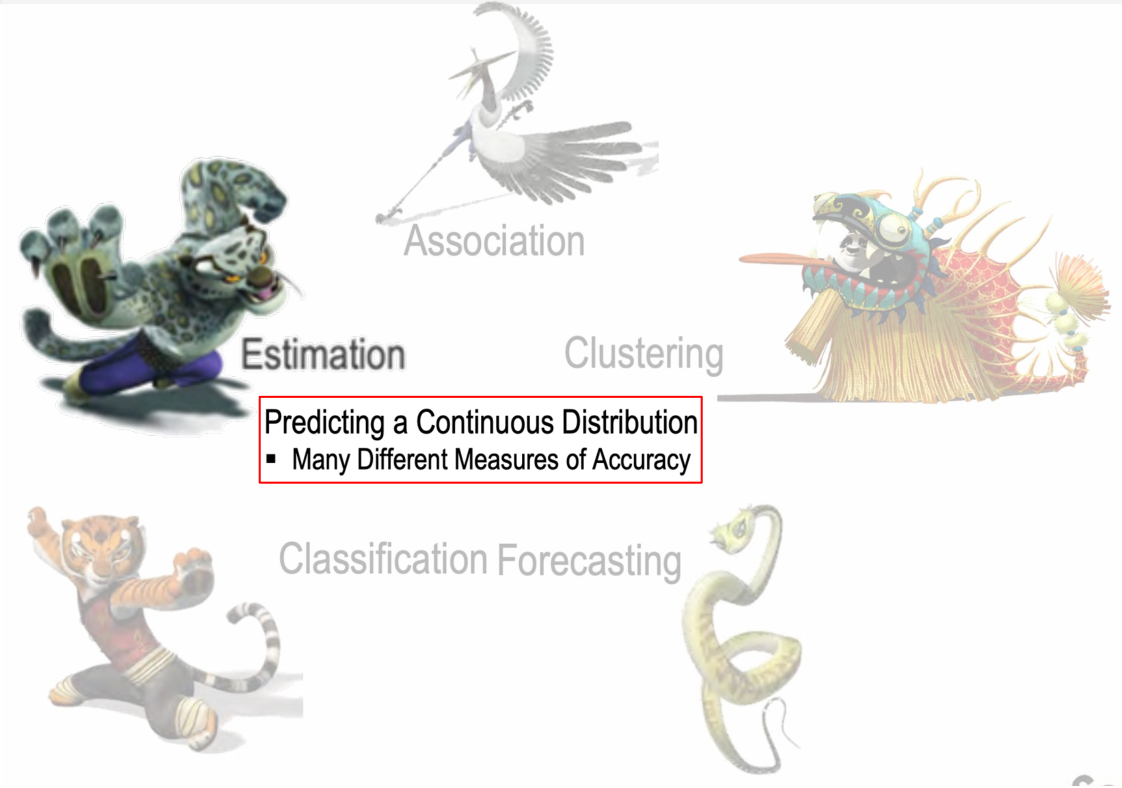
Machine Learning Modeling Families

- Hard and Soft Groupings
- Profiles of Subgroups
- Likenesses and Differences

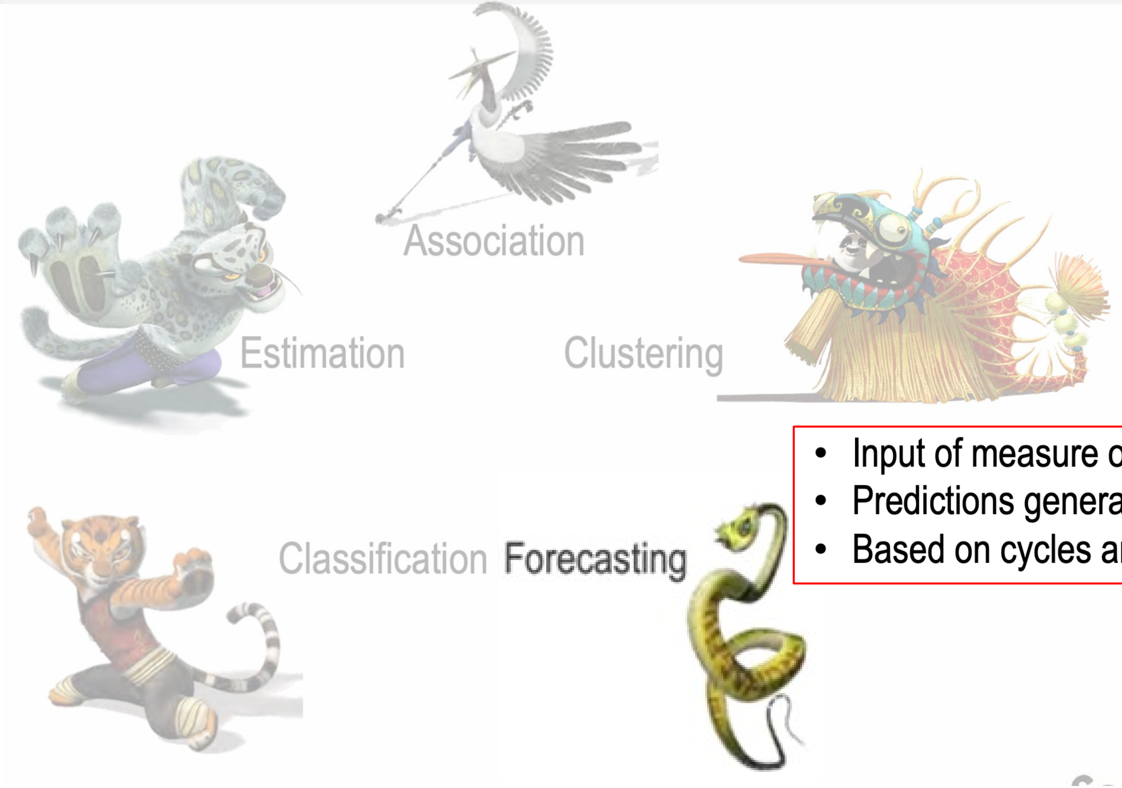


DIVERGENCE ACADEMY
Pythoneers

Machine Learning Modeling Families

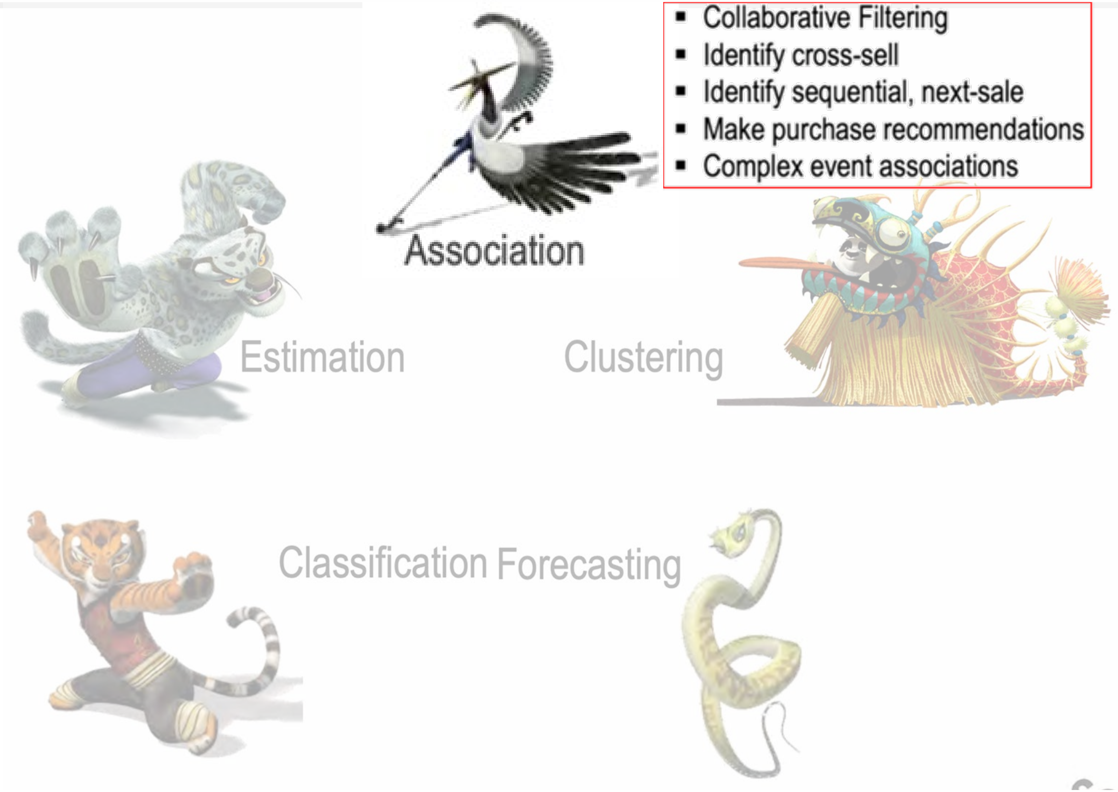


Machine Learning Modeling Families



- Input of measure over time and related series
- Predictions generated for short term trends
- Based on cycles and events

Machine Learning Modeling Families



What is Artificial Intelligence?

- Software that exhibits human-like capabilities, such as:

Visual Perception

Text Analysis

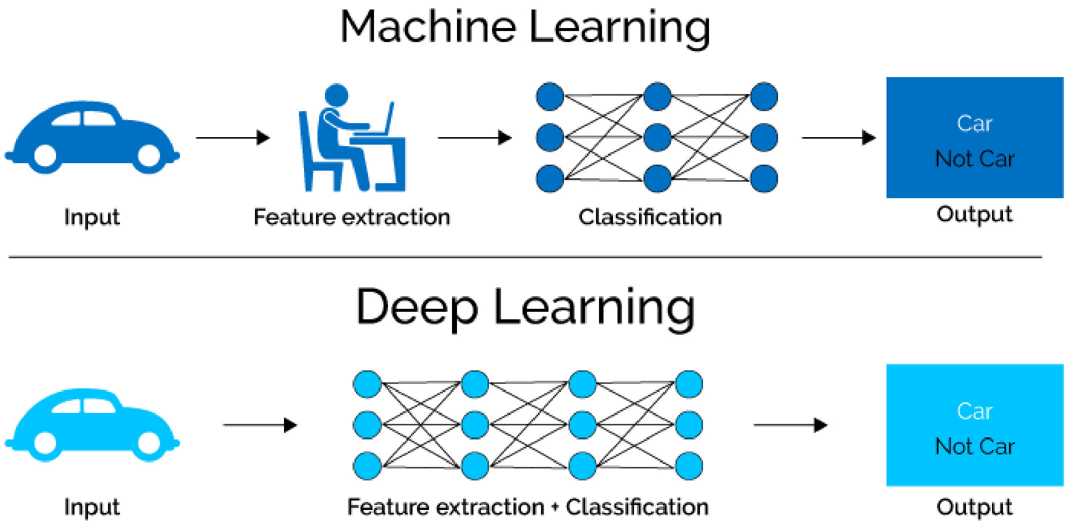
Speech

Decision Making

© Copyright Microsoft Corporation. All rights reserved.

Pythoneers

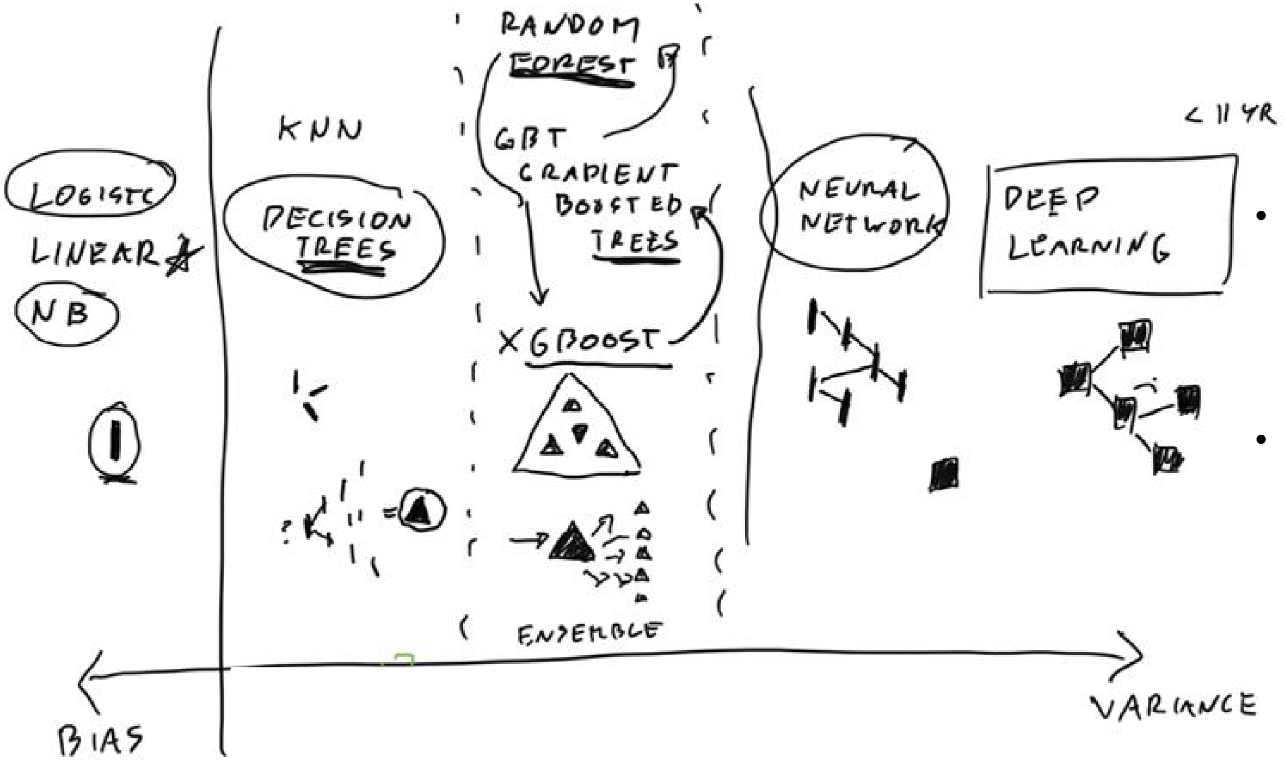
Machine Learning vs Deep Learning



- A more complex representation of data.
- Processing of complex feature spaces
- Feature Engineering baked into Models
- Deep layering of algorithm, not wide.

https://www.xenonstack.com/blog/static/public/uploads/media/machine-learning-vs-deep-learning.png

Deep Learning in Machine Learning Algorithm Context



- LR, NN and other algorithms part of **hierarchical layers** of DL model
- **Transfer Learning** Learning models on same type of data

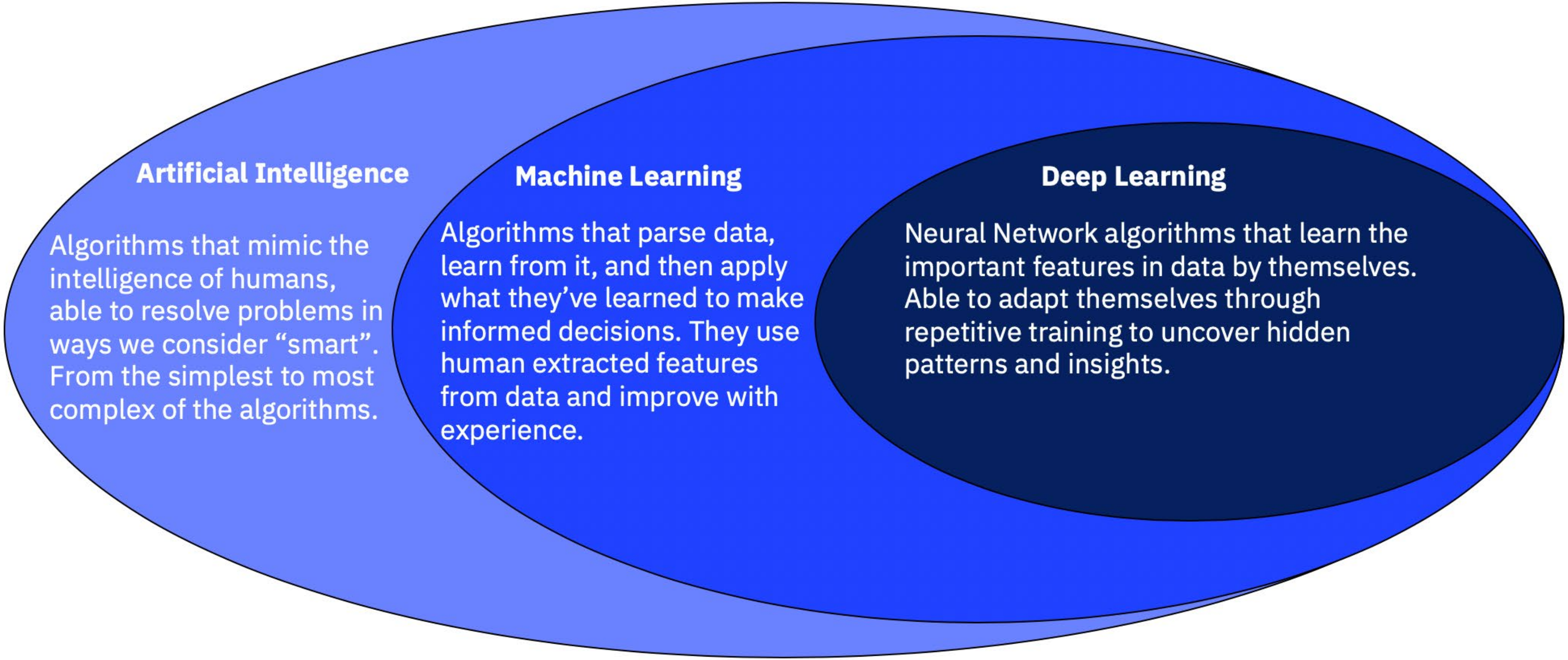
Data Science, Machine Learning, and AI

Artificial Intelligence
Intelligent software apps and agents

Machine Learning
Use of data and algorithms to train predictive models

Data Science
Application of mathematical and statistical techniques to analyze data

Deep Learning, Machine Learning, and AI



Azure Deep Learning Platforms – Azure Only

AZURE
MACHINE LEARNING STUDIO

AZURE
SYNAPSE ANALYTICS

AZURE
COGNITIVE SERVICES

Azure Deep Learning Platforms - Databricks

AZURE DATABRICKS

- Git Support
- Deeper Spark IP
- Cross-cloud support
- Single node support
- Maturer autoscale
- Tight MLFlow integration
- Tighter Delta Lake in Spark SQL
- Spark SQL Live Tables
- Component level Azure Data Factory integration
- Notebook as Dashboard

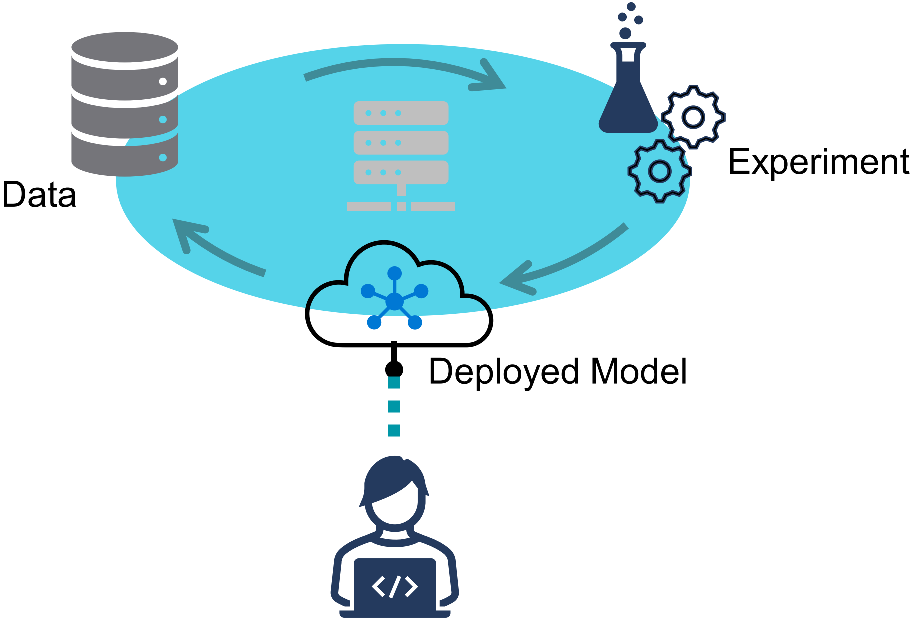
Hive Support
Java/Scala/Python
Spark Scaleout
GPU Support
SparkML Support

AZURE SYNAPSE

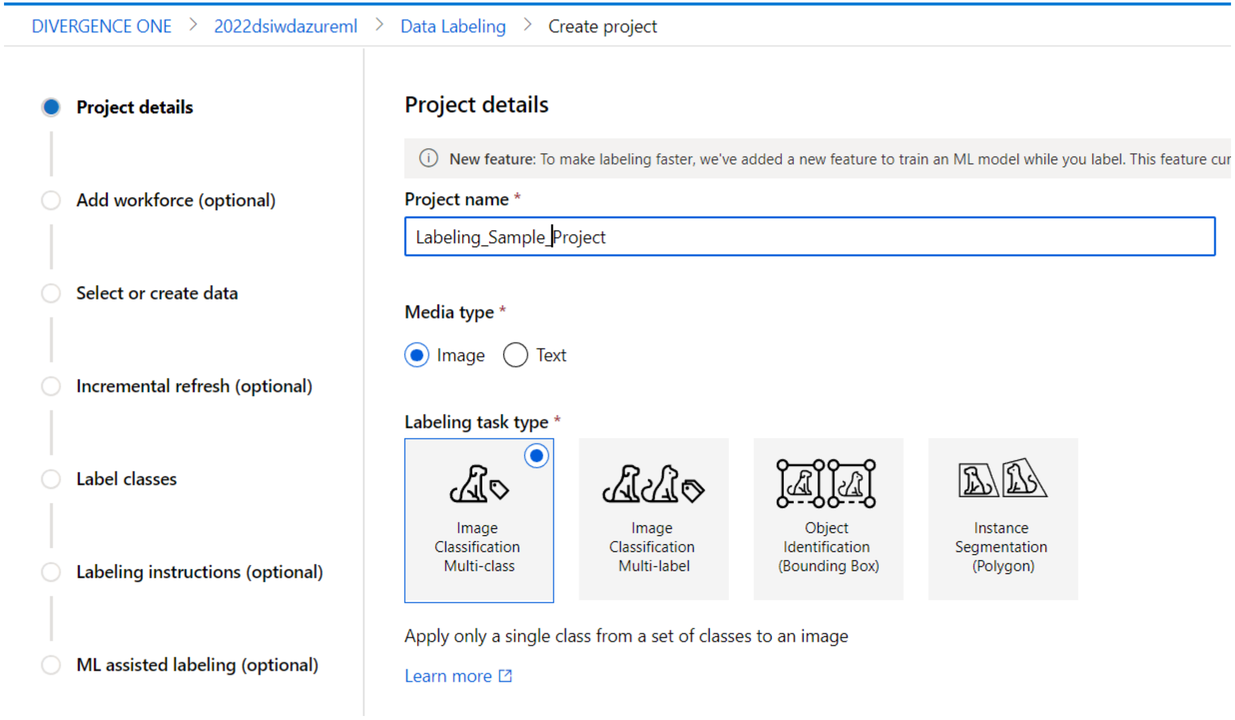
- SynapseML support
- Serverless or Dedicated SQL Pool
- Data warehousing
- Data Lake Tables
- Tighter Power BI integration
- .NET language support
- Big Data Scale Model Scoring with Zero Data Movement
- Transact-SQL compatibility
- Azure Purview Integration
- Azure Data Factory workspace integration

Azure Machine Learning

- Cloud platform for creating and operating machine learning solutions

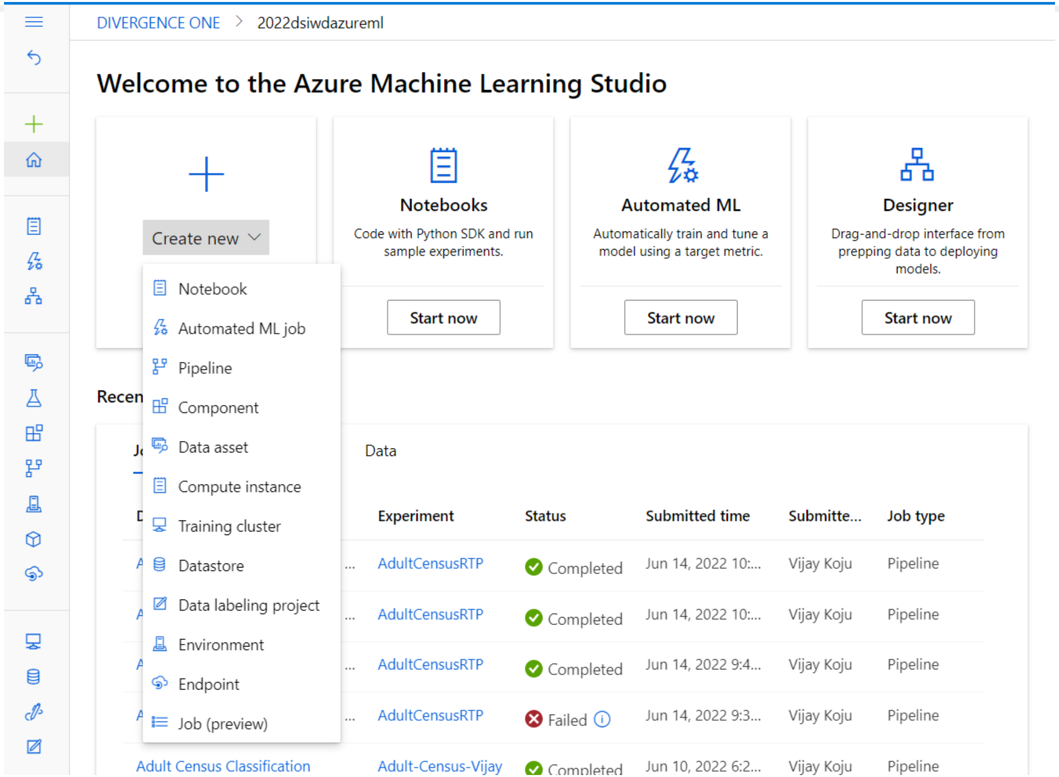


Azure Machine Learning – Deep Learning Benefits



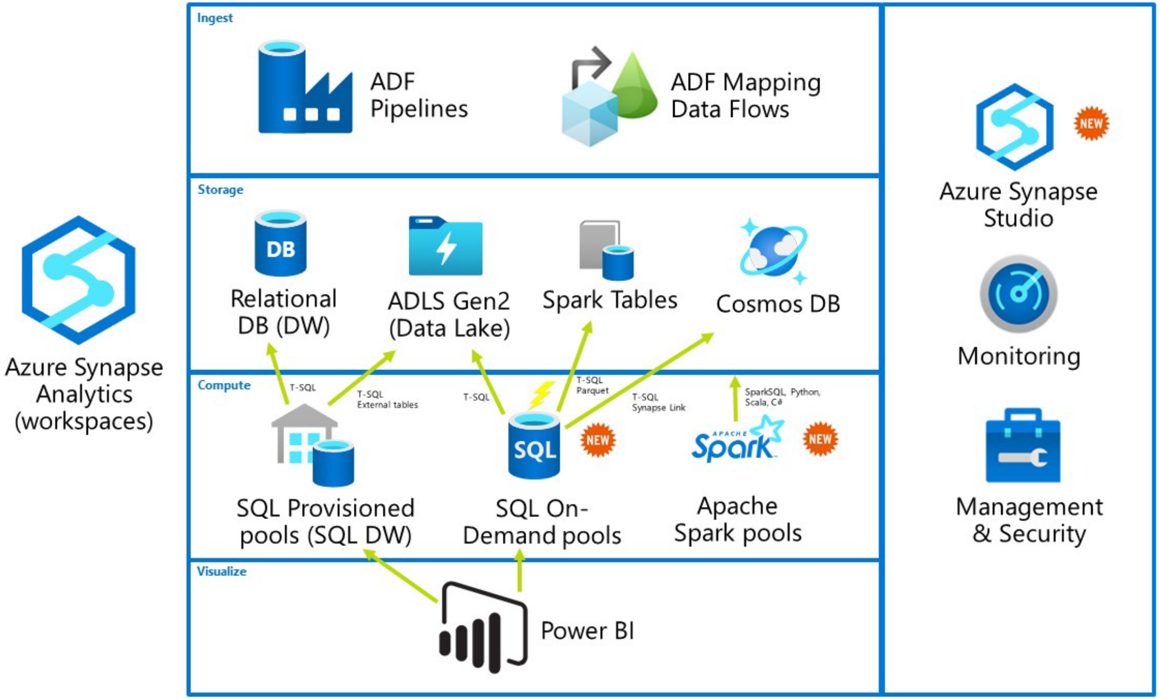
- Labeling is a common technique to add target columns to raw data acquired prior to formal deep learning
- ML assisted = Semi-supervised learning

Azure Machine Learning – Deep Learning Benefits

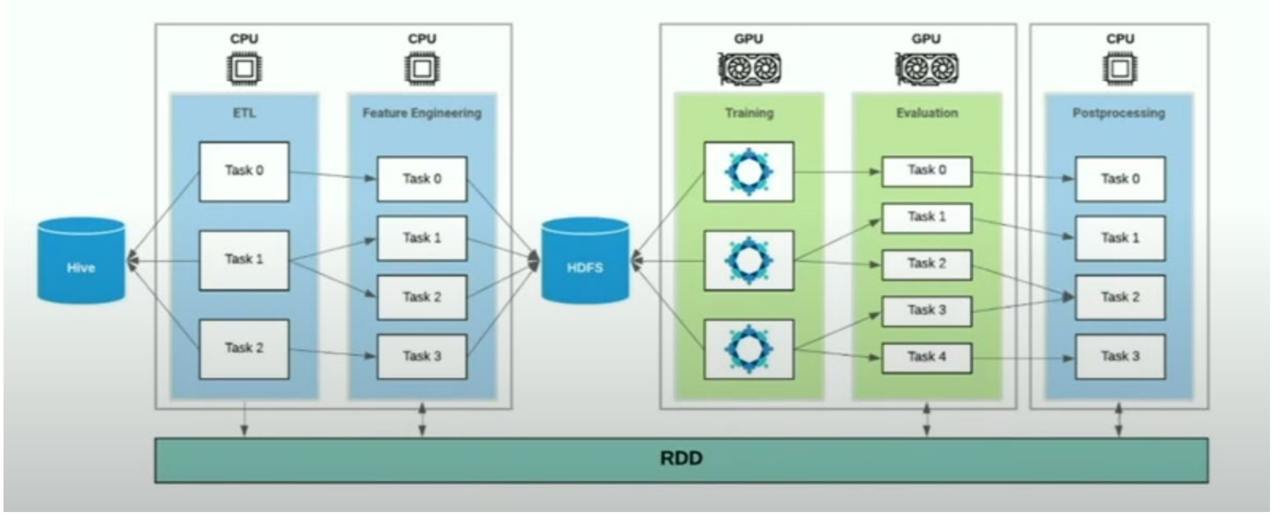


- Scale-up
- Azure ML cluster of attached compute
- JupyterHub integration
- R Studio VM integration
- Easy integration with full ML lifecycle

Azure Synapse Analytics - Architecture



Azure Synapse Analytics – Deep Learning Benefits



- Scale-out
- Spark integration built-in
- Easier Azure DevOps & Azure Data Factory integration
- .NET language, Transact-SQL integration

https://www.youtube.com/watch?v=jbbnZIpCu-U

Azure Cognitive Services

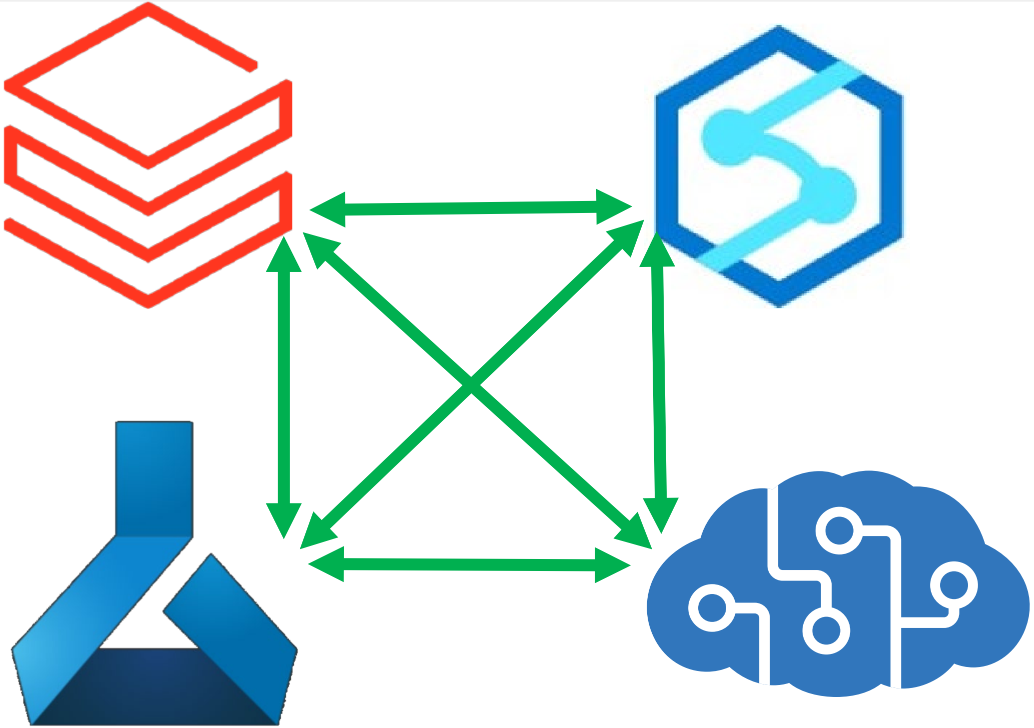
Prepackaged AI services you can integrate into solutions

Language	Speech	Vision	Decision
• Text analysis • Question answering • Language understanding • Translation	• Speech recognition • Speech synthesis • Speech Translation • Speaker Recognition	• Image analysis • Video analysis • Image classification • Object detection • Facial analysis • Optical character recognition	• Anomaly detection • Content moderation • Content personalization

Azure Applied AI Services

- Form Recognizer
- Metrics Advisor
- Video Analyzer for Media
- Immersive Reader
- Bot Service
- Cognitive Search

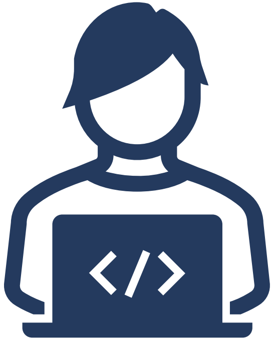
Azure Deep Learning Platforms – Hybrid Answers



AI for Software Engineers

Software Development Skills

- Coding (C#, Python, Node.js, …)
- Consuming APIs (REST or SDKs)
- DevOps (source control, CI/CD)



Conceptual AI Understanding

- Model training and inferencing
- Probability and confidence scores
- Responsible AI and ethics

Considerations for Responsible AI

Fairness	Reliability & Safety	Privacy & Security
Inclusiveness	Transparency	Accountability

https://www.microsoft.com/ai/responsible-ai

Big Questions on Cognitive Services vs BYOM

DO YOU HAVE
THE RIGHT SPONSORS
AND TALENT FOR THE TASK?

WHAT LEVEL OF RISK & REWARD
ARE YOU ASSUMING BUILDING
EVERYTHING ON YOUR OWN?

DO YOU HAVE A PESSIMISTIC
ENOUGH BUDGET
EITHER WAY?

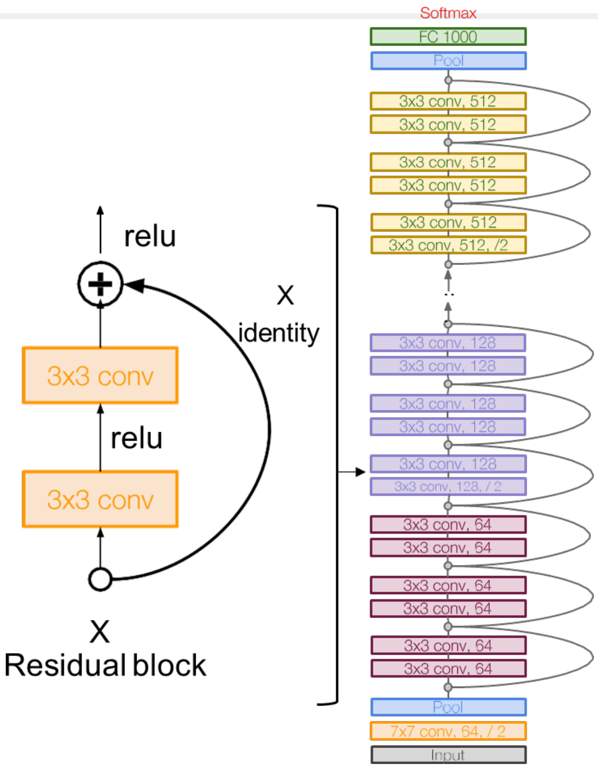
DIVERGENCE ACADEMY

Deep Learning
Deeper Dive

Deep Learning and Activation Functions

Activation function	Equation	Example	1D Graph
Unit step (Heaviside)	$\phi(z) = \begin{cases} 0, & z < 0, \\ 0.5, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Sign (Signum)	$\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Linear	$\phi(z) = z$	Adaline, linear regression	
Piece-wise linear	$\phi(z) = \begin{cases} 1, & z \geq \frac{1}{2}, \\ z + \frac{1}{2}, & -\frac{1}{2} < z < \frac{1}{2}, \\ 0, & z \leq -\frac{1}{2}, \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, Multi-layer NN	
Hyperbolic tangent	$\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	Multi-layer Neural Networks	
Rectifier, ReLU (Rectified Linear Unit)	$\phi(z) = max(0, z)$	Multi-layer Neural Networks	
Rectifier, softplus	$\phi(z) = \ln(1 + e^z)$	Multi-layer Neural Networks	

Copyright © Sebastian Raschka 2016
(http://sebastianraschka.com)



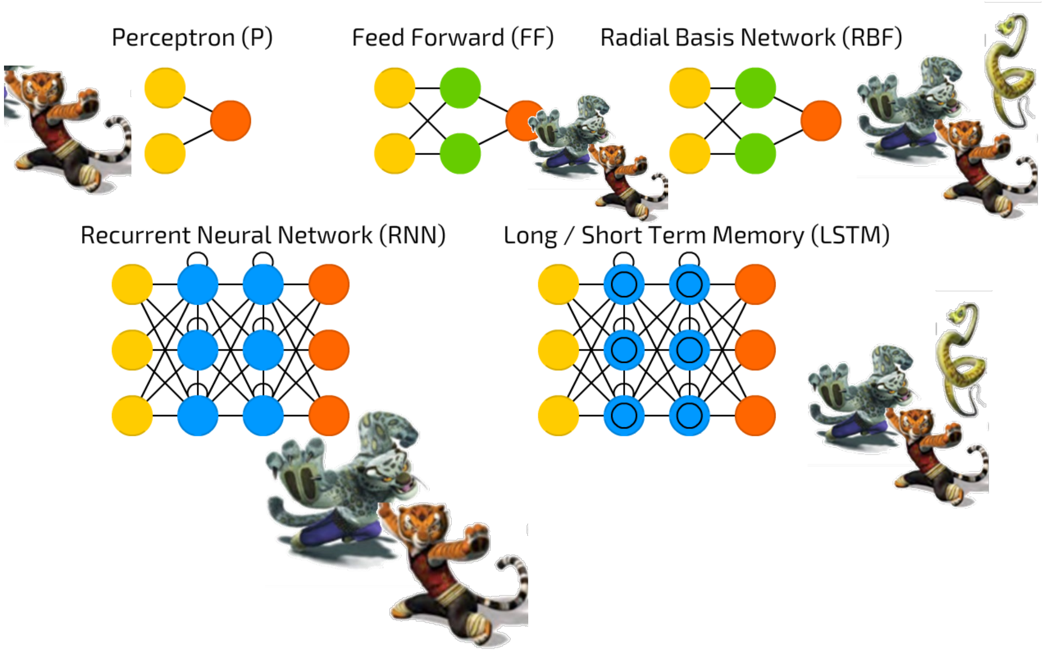
https://www.pinterest.com/pin/672232681856671318/

DIVERGENCE ACADEMY
Pythoneers

Deep Learning General Model Structure Examples

Basic Structures

- Backfed Input Cell
- Input Cell
- Noisy Input Cell
- Hidden Cell
- Probablistic Hidden Cell
- Spiking Hidden Cell
- Output Cell
- Match Input Output Cell
- Recurrent Cell
- Memory Cell
- Different Memory Cell
- Kernel
- Convolution or Pool

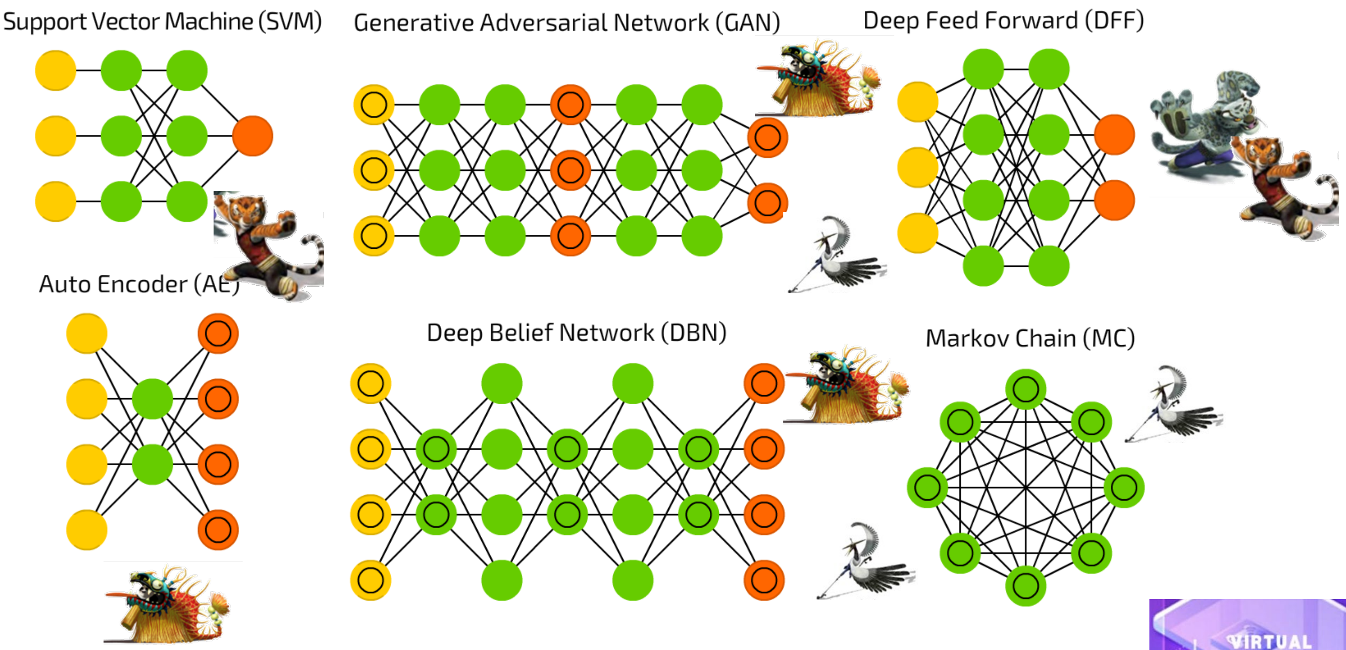


https://www.kaggle.com/getting-started/151100

Deep Learning General Model Structure Examples

Next Generation Complexity

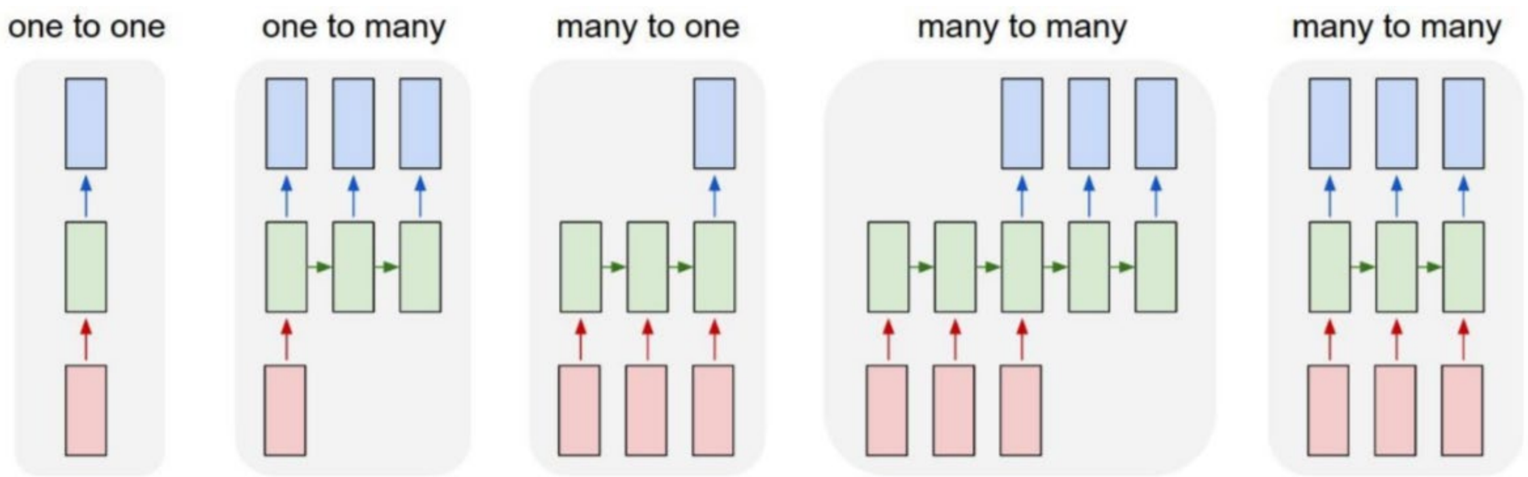
- Backfed Input Cell
- Input Cell
- Noisy Input Cell
- Hidden Cell
- Probablistic Hidden Cell
- Spiking Hidden Cell
- Output Cell
- Match Input Output Cell
- Recurrent Cell
- Memory Cell
- Different Memory Cell
- Kernel
- Convolution or Pool



https://www.kaggle.com/getting-started/151100

Deep Learning General Model Structure Examples

Sequence to Sequence (Seq2Seq)



Fixed-sized input to fixed-sized output (e.g. image classification)

Sequence output (e.g. image captioning takes an image and outputs a sentence of words).

Sequence input (e.g. sentiment analysis where a given sentence is classified as expressing positive or negative sentiment).

Sequence input and sequence output (e.g. Machine Translation: an RNN reads a sentence in English and then outputs a sentence in French)

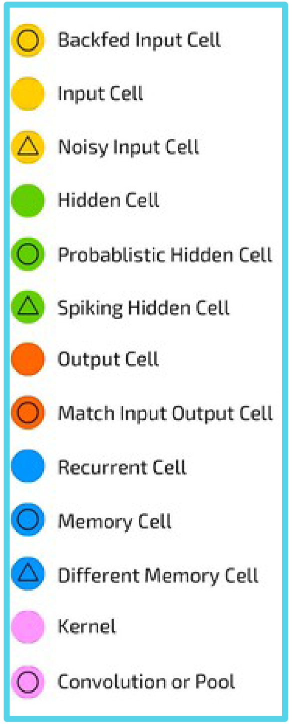
Synced sequence input and output (e.g. video classification where we wish to label each frame of the video)

https://mett29.github.io/posts/2019/12/seq2seq_and_attention/

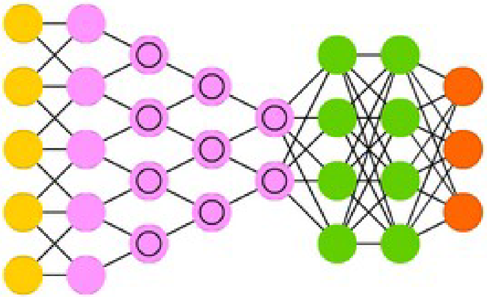
VIRTUAL CONFERENCE
Conference
AZURE DATA PLATFORM
TECHFORUM TECHNICAL COMMUNITY

Deep Learning General Model Structure Examples

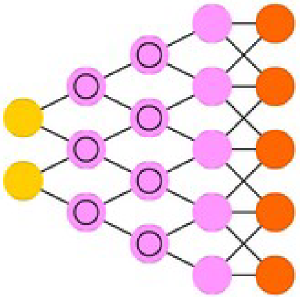
Next Generation Complexity



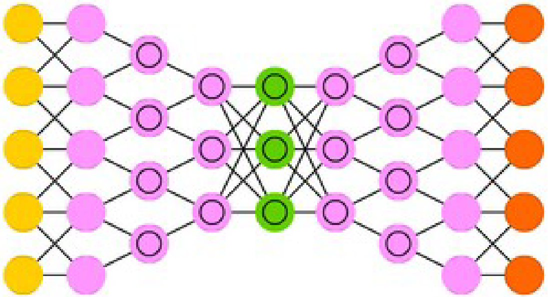
Deep Convolutional Network (DCN)



Deconvolutional Network (DN)



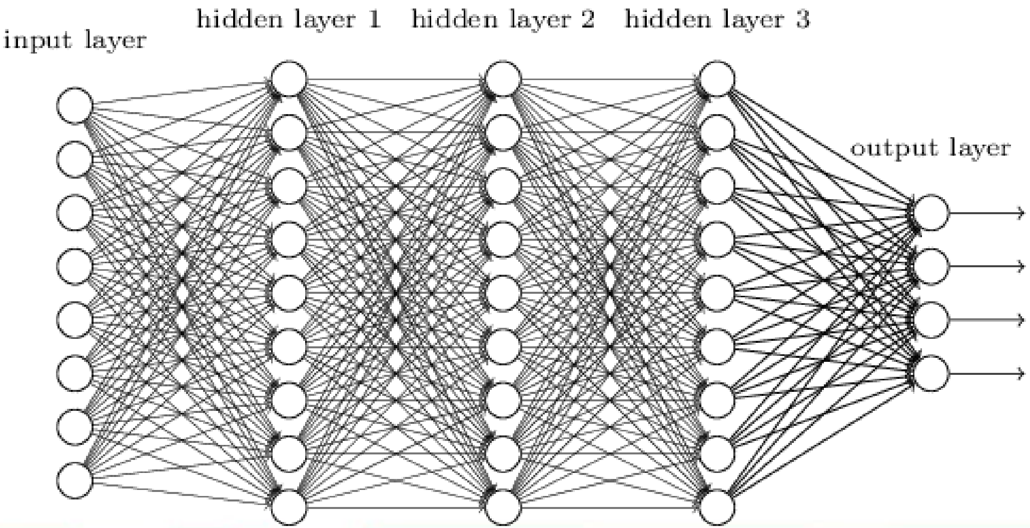
Deep Convolutional Inverse Graphics Network (DCIGN)



https://www.kaggle.com/getting-started/151100

Why Simple Neural Networks don’t cut it

- We know it is good to learn a small model.
- From this fully connected model, do we really need all the edges?
- Can some of these be shared?

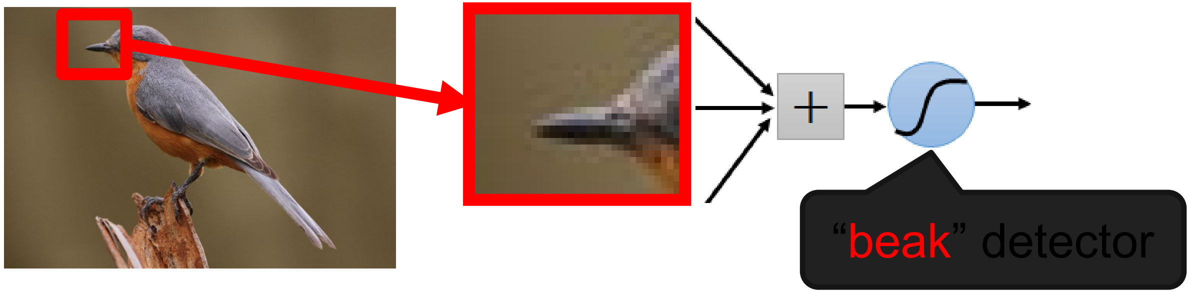


https://cs.uwaterloo.ca/~mli/Deep-Learning-2017-Lecture5CNN.ppt

Consider learning an image:

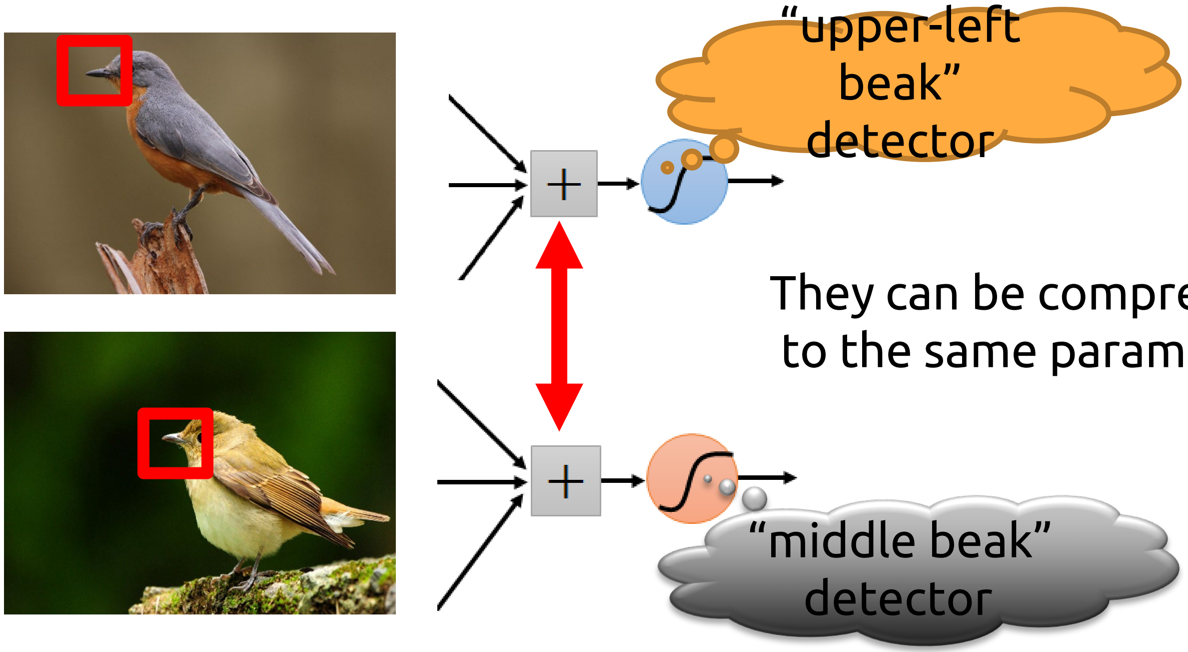
- Some patterns are much smaller than the whole image

Can represent a small region with fewer parameters



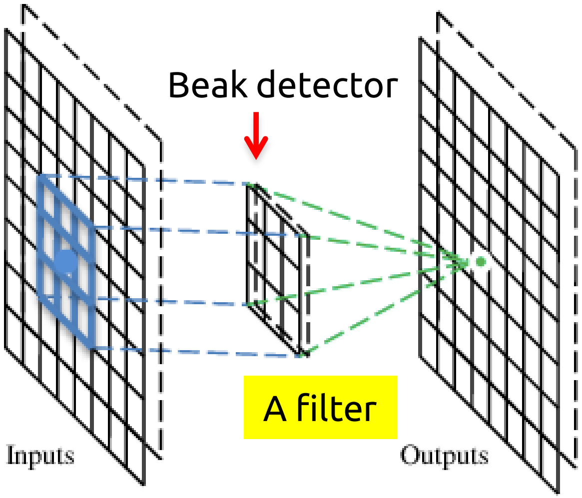
Same pattern appears in different places can be compressed!

What about training a lot of such "small" detectors and each detector must "move around".



Convolutional Layers

A CNN is a neural network with some convolutional layers (and some other layers). A convolutional layer has a number of filters that does convolutional operation.



Convolution

These are the network parameters to be learned.

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

-1	1	-1
-1	1	-1
-1	1	-1

Filter 2

⋮ ⋮

Each filter detects a small pattern (3 x 3).

https://cs.uwaterloo.ca/~mli/Deep-Learning-2017-Lecture5CNN.ppt

Pythoneers

Convolution

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

stride=1

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

Dot product → 3 -1

6 x 6 image

Convolution

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

If stride=2

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image

3 -3

https://cs.uwaterloo.ca/~mli/Deep-Learning-2017-Lecture5CNN.ppt

Pythoneers

Convolution

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

stride=1

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image

3	-1	-3	-1
-3	1	0	-3
-3	-3	0	1
3	-2	-2	-1

Convolution

-1	1	-1
-1	1	-1
-1	1	-1

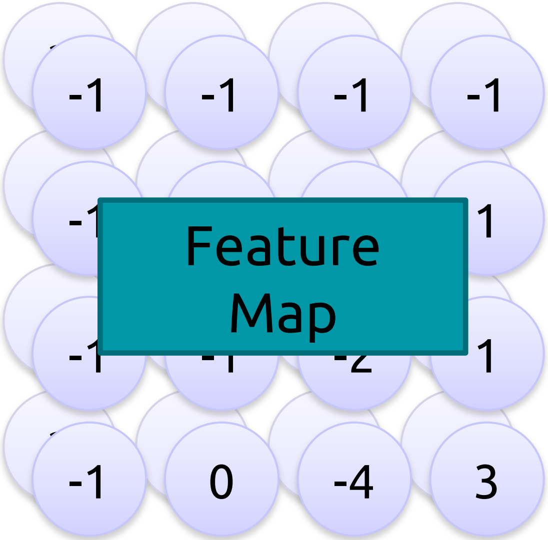
Filter 2

stride=1

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image

Repeat this for each filter



Two 4 x 4 images Forming 2 x 4 x 4 matrix

Color image: RGB 3 channels

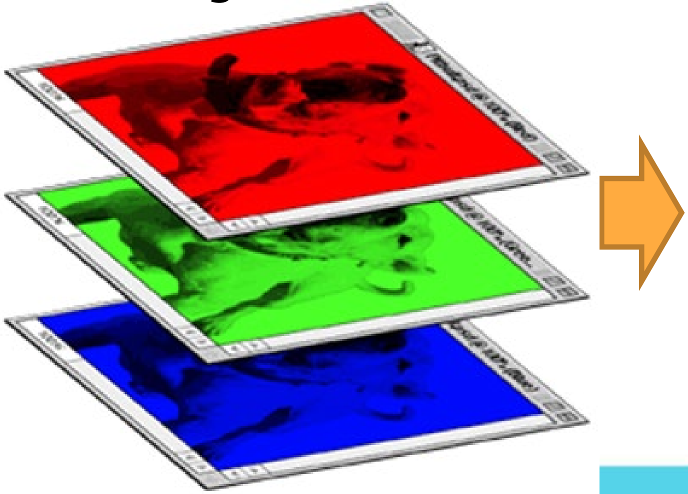
1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

-1	1	-1
-1	1	-1
-1	1	-1

Filter 2

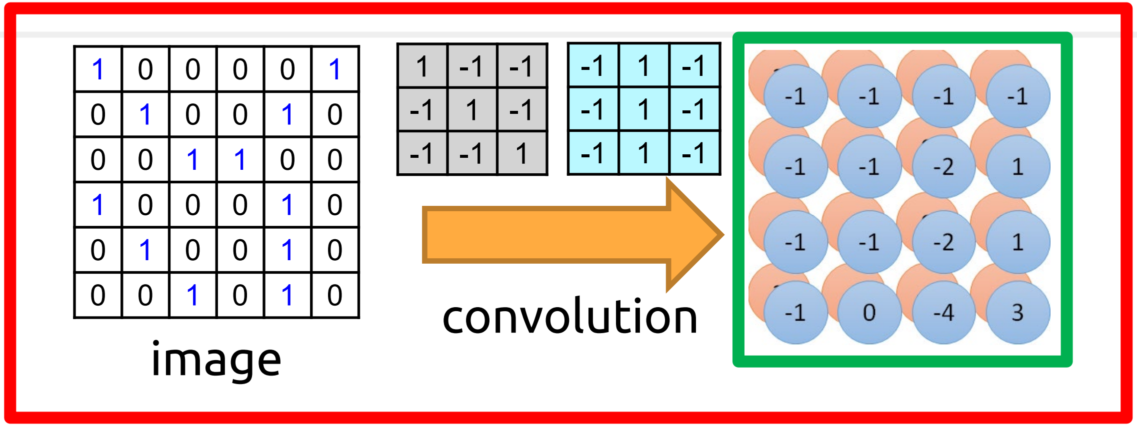
Color image



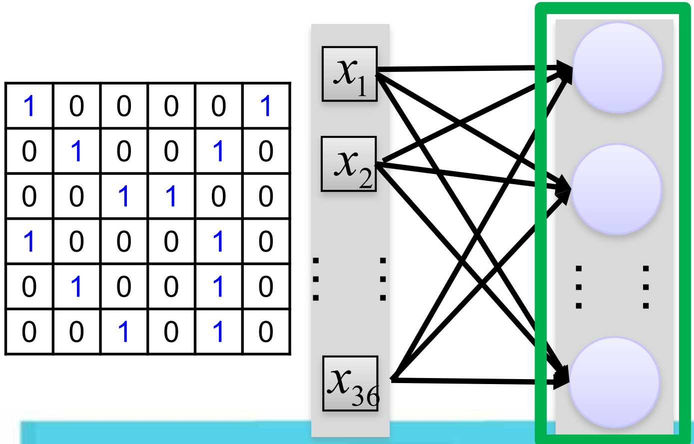
1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

https://cs.uwaterloo.ca/~mli/Deep-Learning-2017-Lecture5CNN.ppt

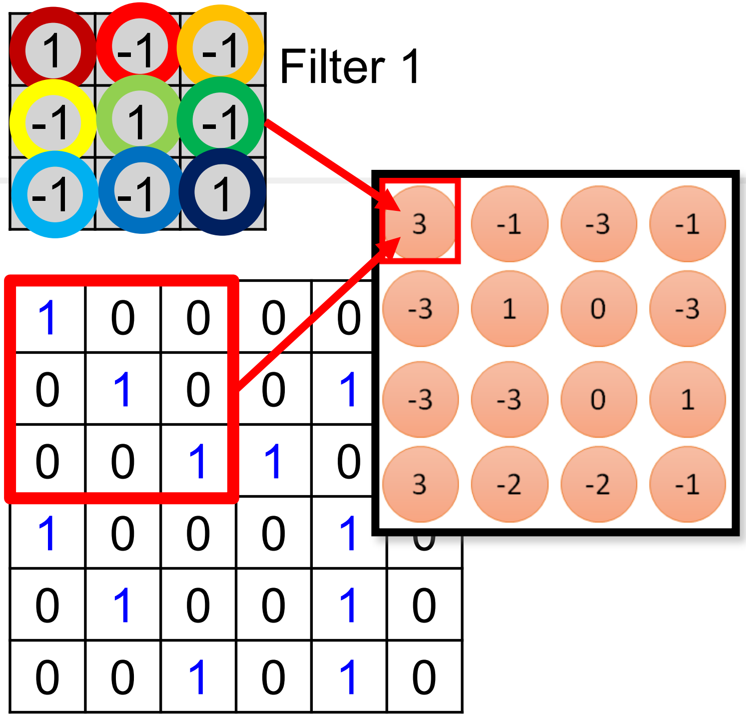
Convolution vs. Fully Connected



Fully-connected

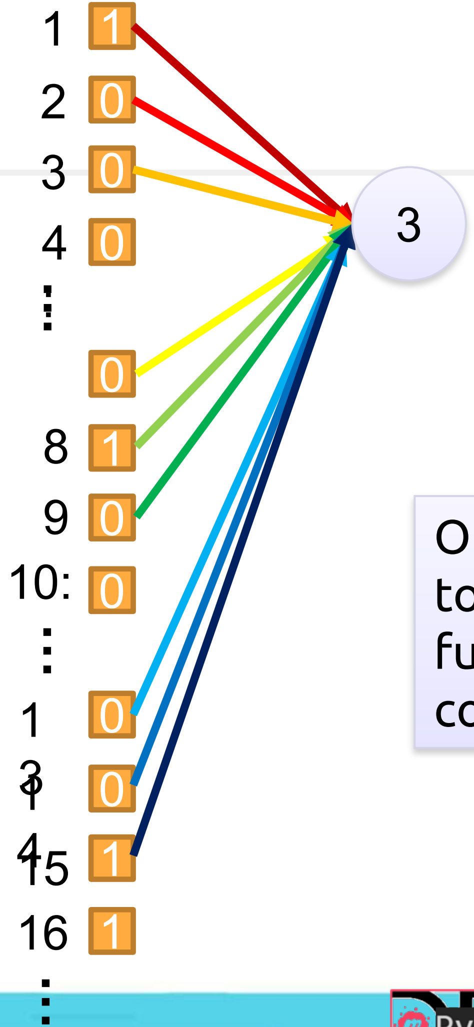


https://cs.uwaterloo.ca/~mli/Deep-Learning-2017-Lecture5CNN.ppt



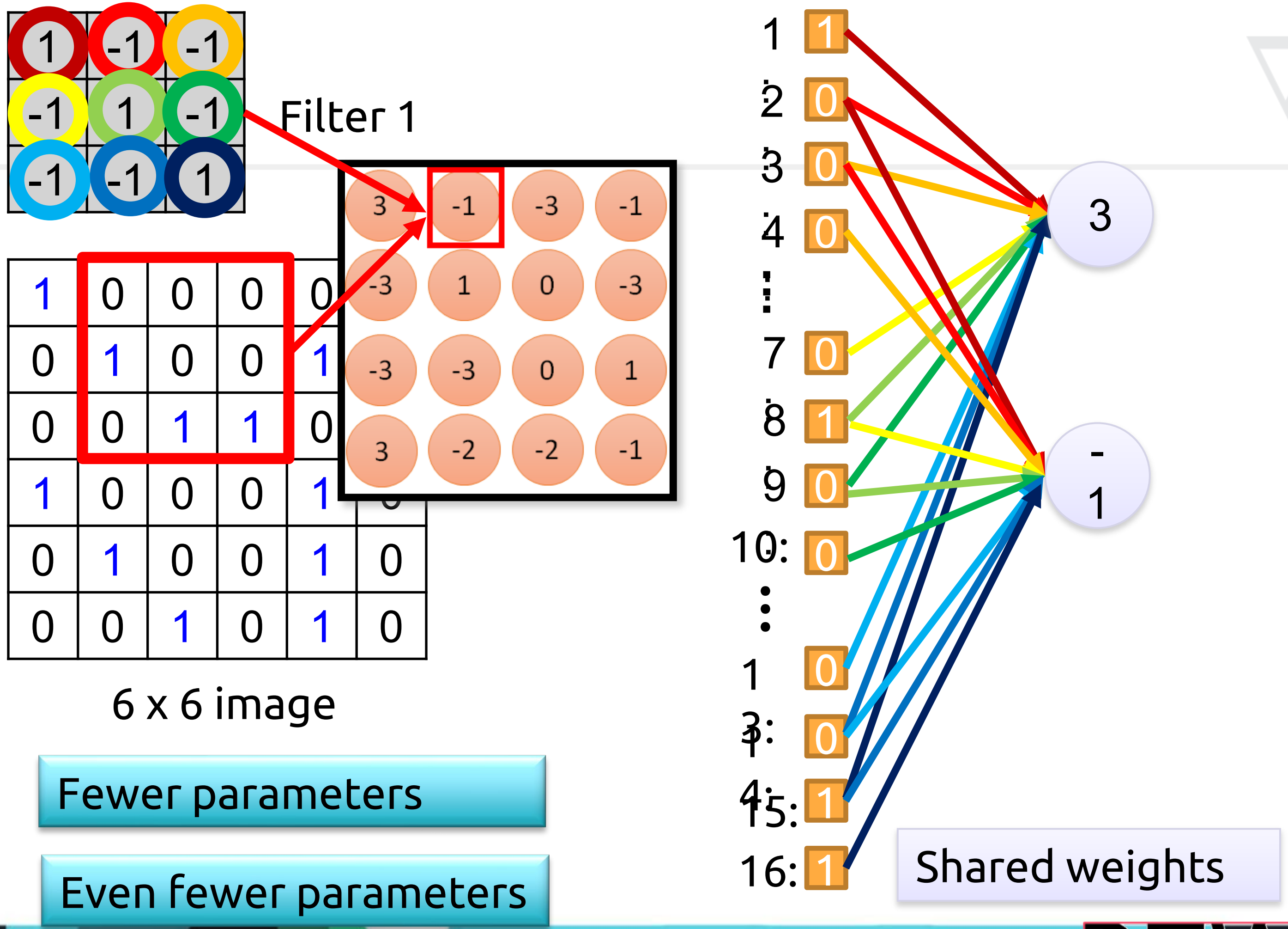
6 x 6 image

fewer parameters!



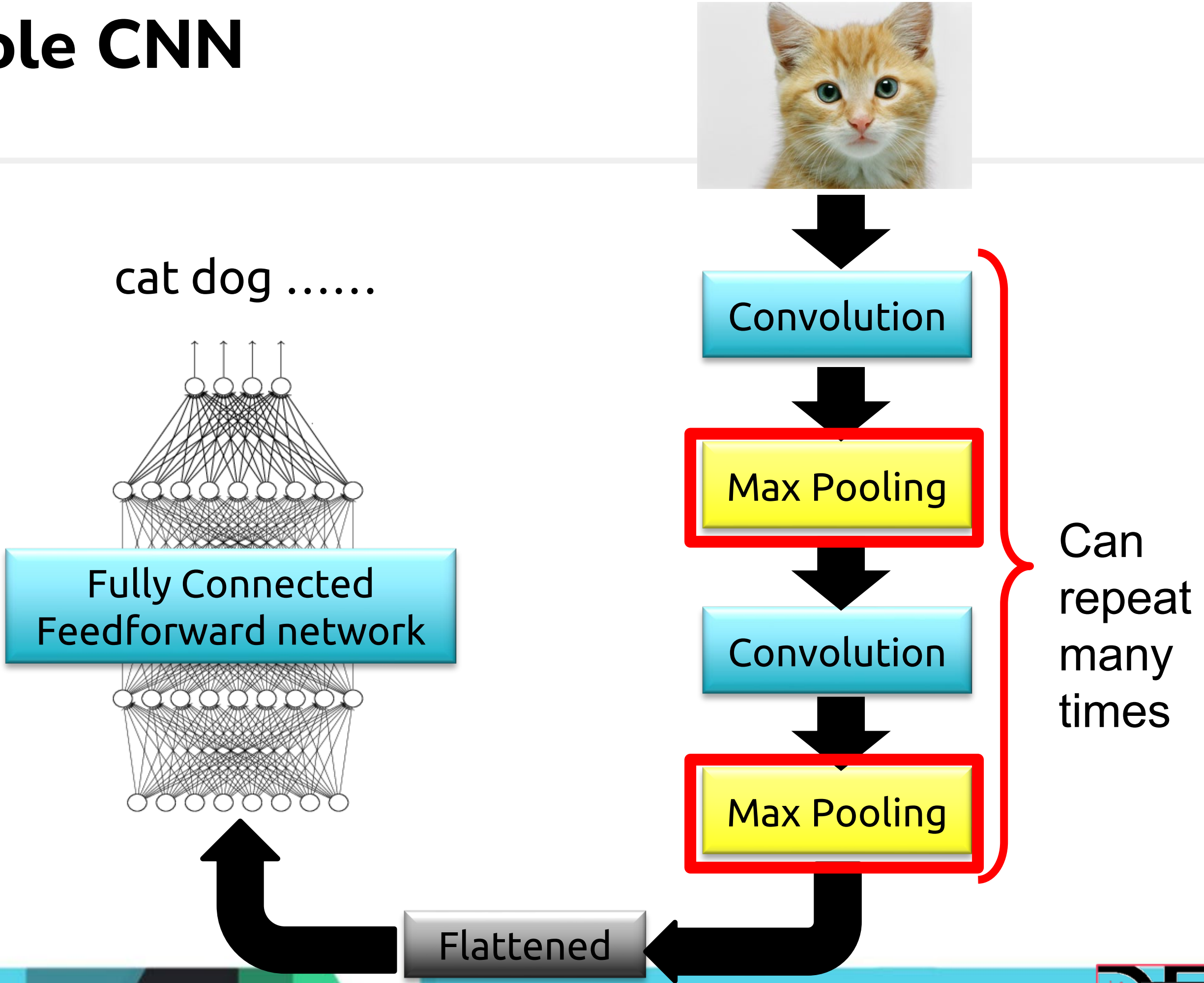
Only connect to 9 inputs, not fully connected

https://cs.uwaterloo.ca/~mli/Deep-Learning-2017-Lecture5CNN.ppt



1 -1 -1
-1 1 -1
-1 -1 1
Filter 1
3 -1 -3 -1
-3 1 0 -3
-3 -3 0 1
3 -2 -2 -1
1 0 0 0 0 1
0 1 0 0 1 0
0 0 1 1 0 0
1 0 0 0 1 0
0 1 0 0 1 0
0 0 1 0 1 0
6 x 6 image
Fewer parameters
Even fewer parameters
1: 1
2: 0
3: 0
4: 0
7: 0
8: 1
9: 0
10: 0
13: 0
14: 0
15: 1
16: 1
3
-1
Shared weights

The whole CNN



Max Pooling

Filter 1

1	-1	-1
-1	1	-1
-1	-1	1

Filter 2

-1	1	-1
-1	1	-1
-1	1	-1

3	-1	-3	-1
-3	1	0	-3
-3	-3	0	1
3	-2	-2	-1

-1	-1	-1	-1
-1	-1	-2	1
-1	-1	-2	1
-1	0	-4	3

Why Pooling

- Subsampling pixels will not change the object

bird

Subsampling

bird

We can subsample the pixels to make image smaller

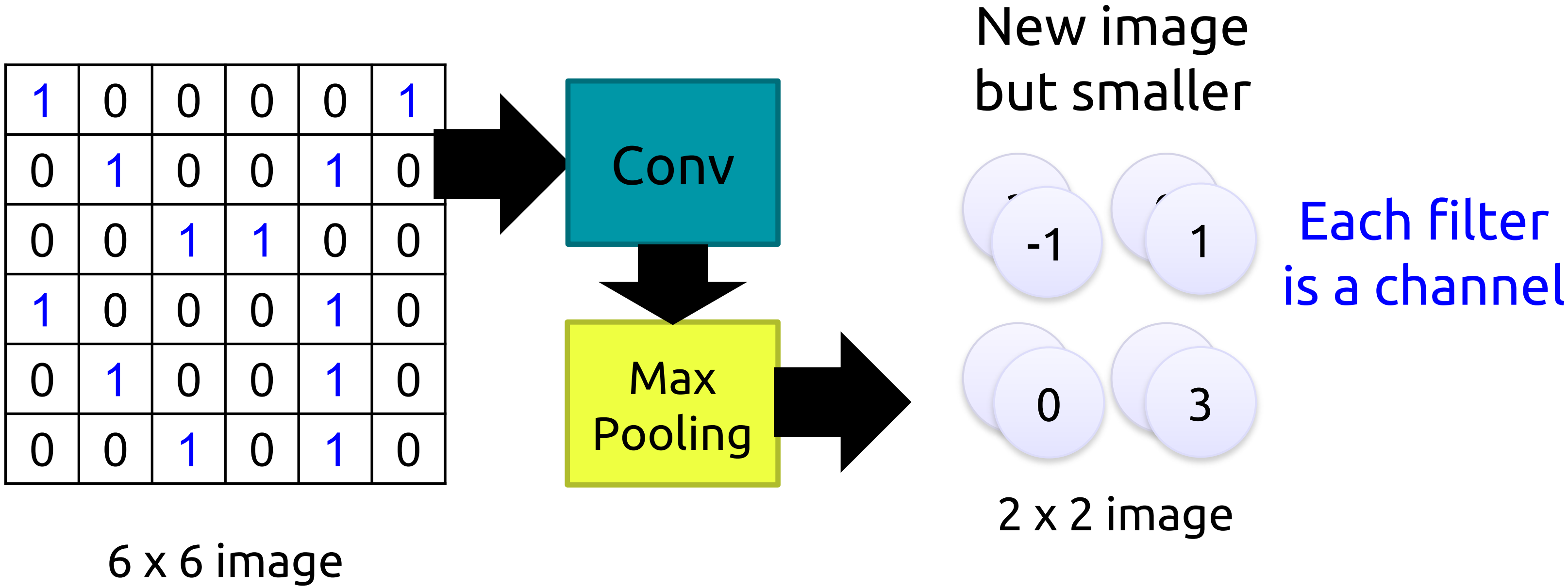
fewer parameters to characterize the image

A CNN compresses fully connected network

- Reducing number of connections
- Shared weights on the edges
- Max pooling further reduces the complexity

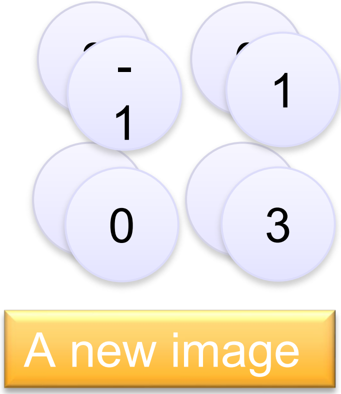
https://cs.uwaterloo.ca/~mli/Deep-Learning-2017-Lecture5CNN.ppt

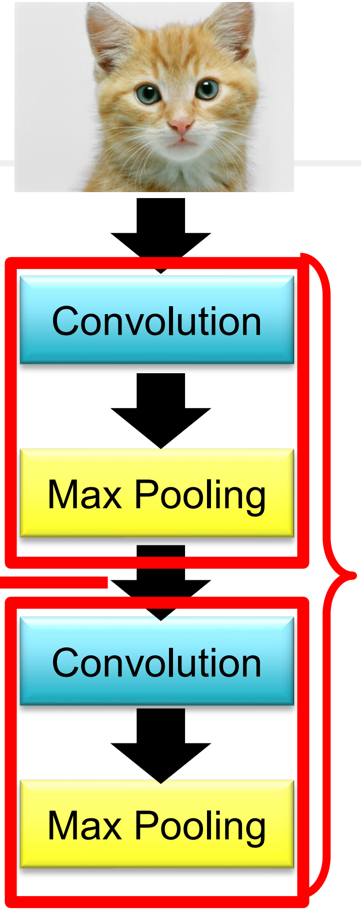
Max Pooling



1 0 0 0 0 1
0 1 0 0 1 0
0 0 1 1 0 0
1 0 0 0 1 0
0 1 0 0 1 0
0 0 1 0 1 0
6 x 6 image
Conv
Max Pooling
New image but smaller
-1
1
0
3
2 x 2 image
Each filter is a channel

The whole CNN



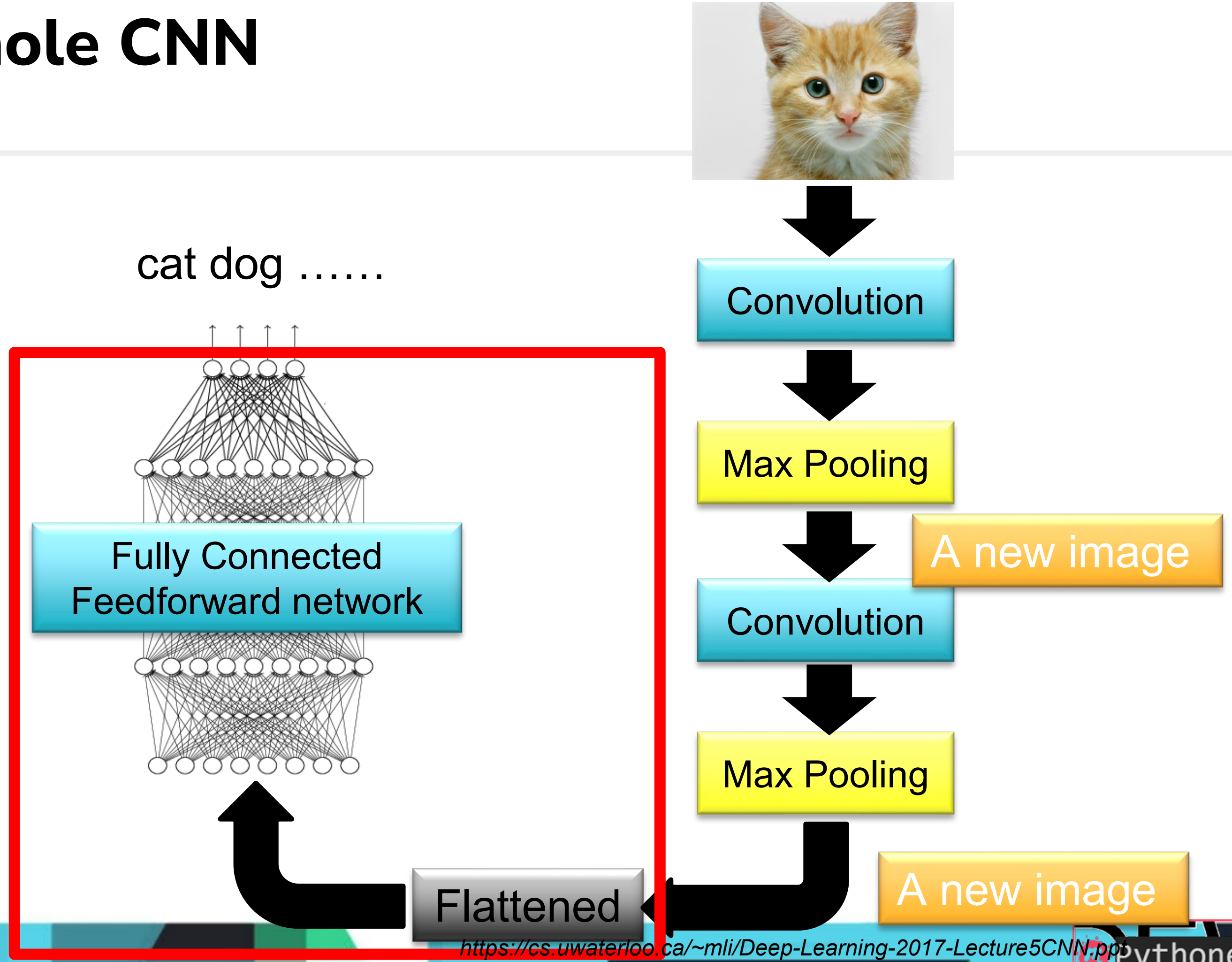


Can repeat many times

Smaller than the original image

The number of channels is the number of filters

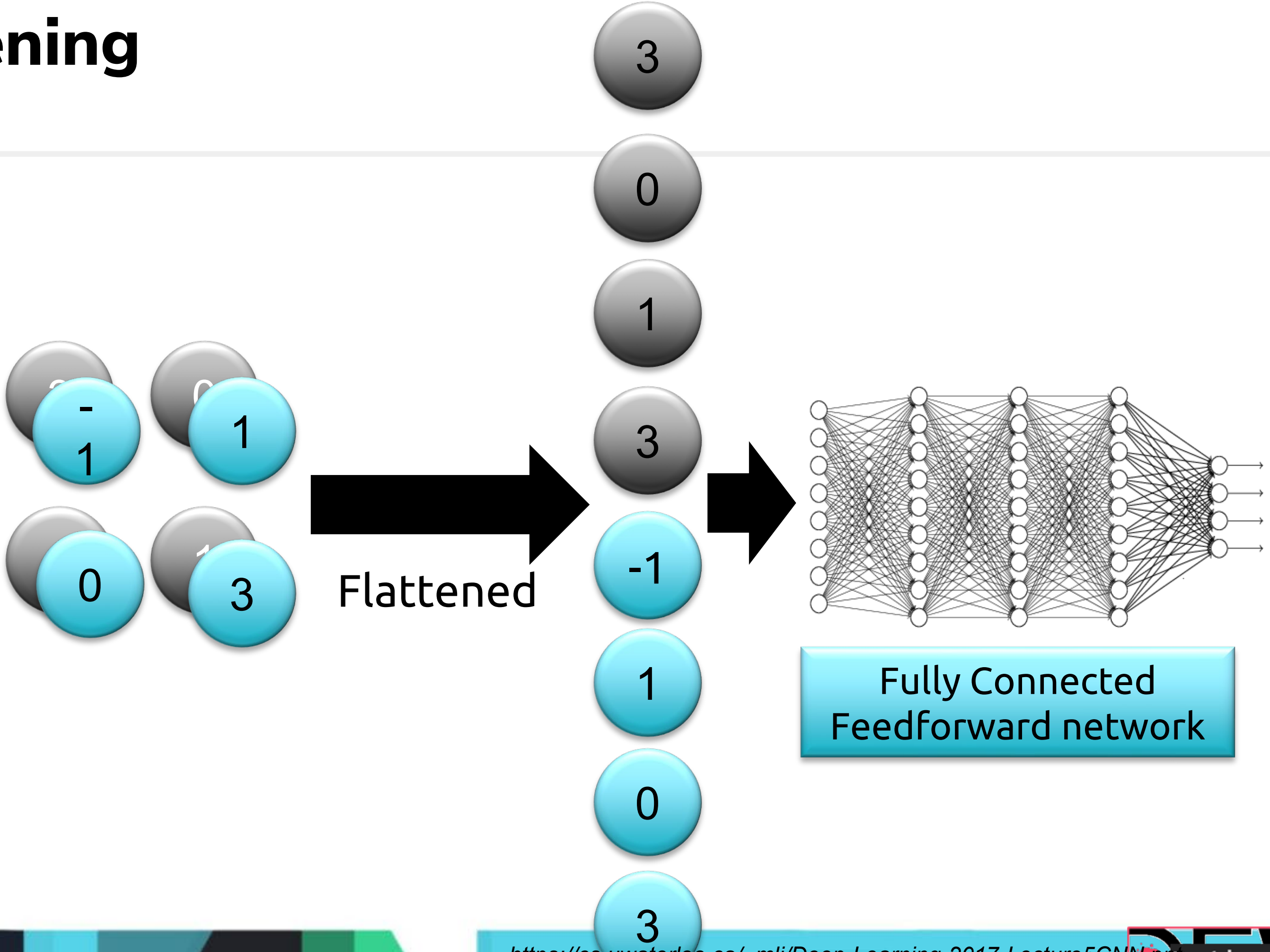
The whole CNN



https://cs.uwaterloo.ca/~mli/Deep-Learning-2017-Lecture5CNN.ppt

Pythoneers

Flattening


-
1
1
0
3
Flattened
3
0
1
3
-1
1
0
3
Fully Connected
Feedforward network

AlphaGo

19 x 19 matrix

Black: 1

white: -1

none: 0

Neural Network

Next move

(19 x 19 positions)

Fully-connected feedforward network can be used

But CNN performs much better

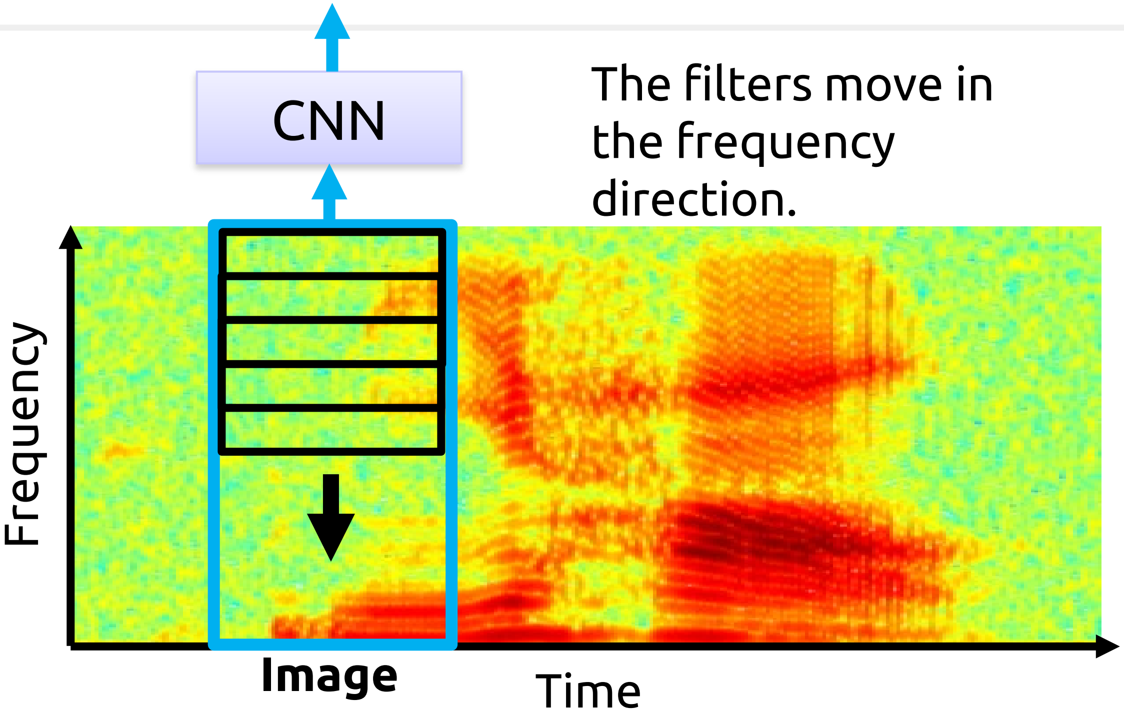
AlphaGo's policy network

The following is quotation from their Nature article:

Note: AlphaGo does not use Max Pooling.

Neural network architecture. The input to the policy network is a 19 × 19 × 48 image stack consisting of 48 feature planes. The first hidden layer zero pads the input into a 23 × 23 image, then convolves *k* filters of kernel size 5 × 5 with stride 1 with the input image and applies a rectifier nonlinearity. Each of the subsequent hidden layers 2 to 12 zero pads the respective previous hidden layer into a 21 × 21 image, then convolves *k* filters of kernel size 3 × 3 with stride 1, again followed by a rectifier nonlinearity. The final layer convolves 1 filter of kernel size 1 × 1 with stride 1, with a different bias for each position, and applies a softmax function. The match version of AlphaGo used $k = 192$ filters; Fig. 2b and Extended Data Table 3 additionally show the results of training with $k = 128$, 256 and 384 filters.

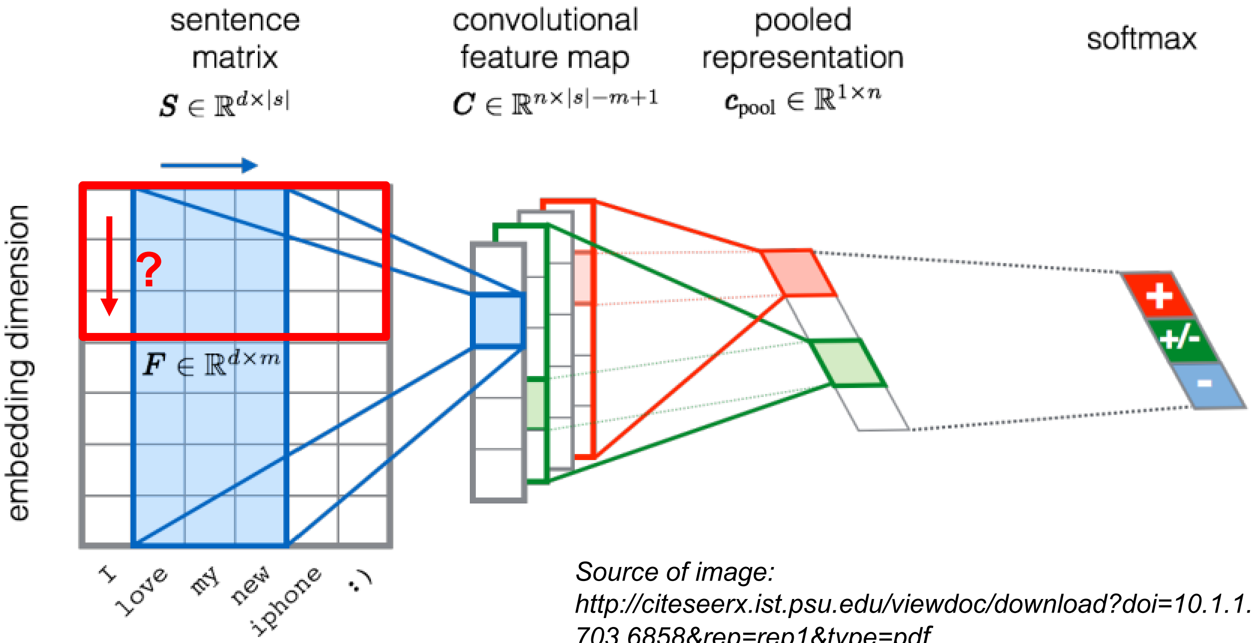
CNN in speech recognition



The filters move in the frequency direction.

Spectrogram

CNN in text classification



Source of image: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.703.6858&rep=rep1&type=pdf

For Future Reference

- http://www.kdnuggets.com/
- http://www.datasciencecentral.com

KDnuggets™

UCI Machine Learning Repository

DeepLearning.AI

Data Science Central

kaggle

DIVERGENCE ACADEMY

Starting with Deep Learning:

DEMO with Keras

DIVERGENCE ACADEMY

Digging deeper into Deep Learning:

DEMO with TensorFlow

Pythoneers

∇·f DIVERGENCE ACADEMY

Starting with Deep Learning:

DEMO with PyTorch

Pythoneers

DIVERGENCE ACADEMY

Starting with Deep Learning:

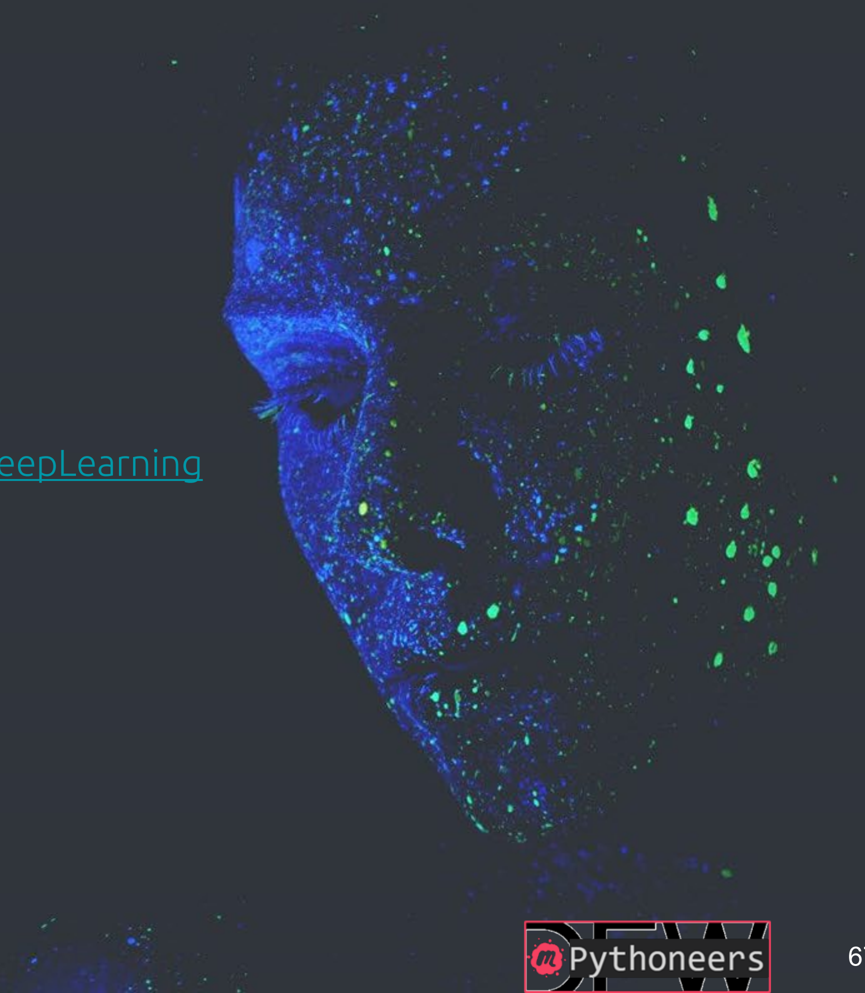
DEMO with ONNX

Questions?

DIVERGENCE ACADEMY

Thank you!!!

Deck & Repo at
https://github.com/hekaplex/DFWPythoneersIntroDeepLearning



DIVERGENCE ACADEMY
(214) 766-3897
drew.minkin@divergence.one