UNIVERSIDAD NACIONAL
DE SAN MARTÍN

Machine Learning - Week 4

Maestría en Ciencia con mención de Tecnología de la información

Yonatan Carlos CARRANZA ALARCÓN,
Ph. D. in Machine Learning,
ycarranza.alarcon@gmail.com

https://salmuz.github.io/

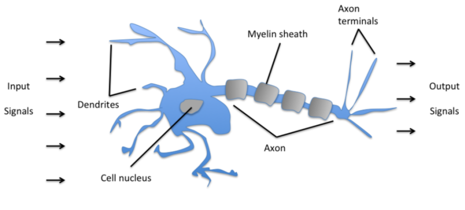
October 23, 2021

Overview

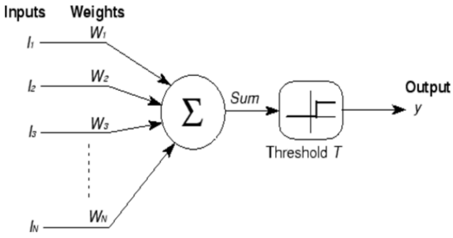
Overview

Deep networks

A class of learning methods that was developed in AI with inspiration from neuroscience.



Schematic of a biological neuron.



Deep networks

Historical perspective

1. Perceptron (1955-1965)
2. Multi-layer neural networks (1985-1995)
3. Deep networks (2010-)

In recent years, there has been a surge of interest in deep networks/learning, with applications to computer vision and natural language processing.

Objective

Deep networks aims to learn a supervised, semi-supervised or unsupervised method.

Deep Learning

Neural networks

Multi-layer neural networks

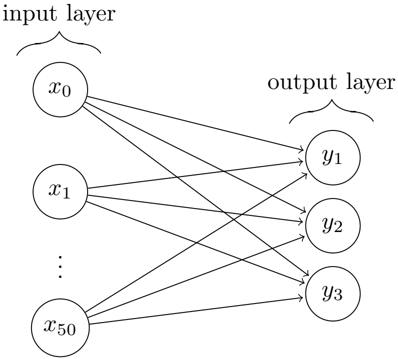
Deep neural learning

Natural Language Processing

Image Pattern Recognition

Neural networks

Neural networks must imperatively be composed of an input layer and an output layer (with as many neurons as the number of predictions we need; e.g. multi-class classification).



Deep Learning

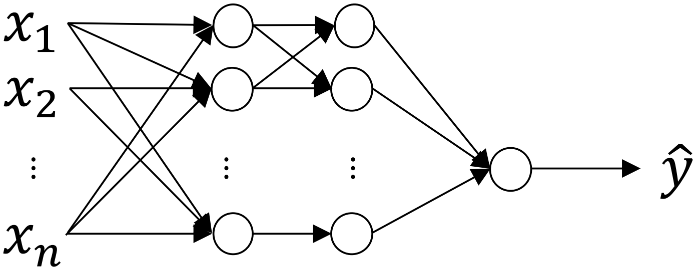
- Neural networks
- Multi-layer neural networks
- Deep neural learning

Natural Language Processing

Image Pattern Recognition

Multi-layer neural networks

Unlike Neural network, Multi-layer neural network must imperatively be composed of one or two hidden layers between the input and output layer.



Deep Learning

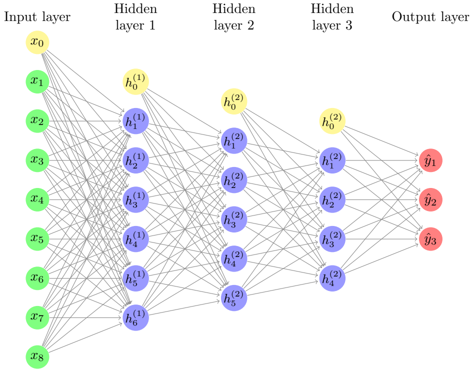
- Neural networks
- Multi-layer neural networks
- Deep neural learning

Natural Language Processing

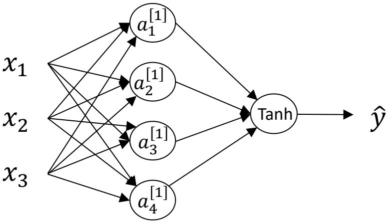
Image Pattern Recognition

Deep neural learning

Unlike both models, Deep neural network adds extra layers (hidden layers) in the middle of the input and output layer, combined often with regularization methods.



Deep neural learning



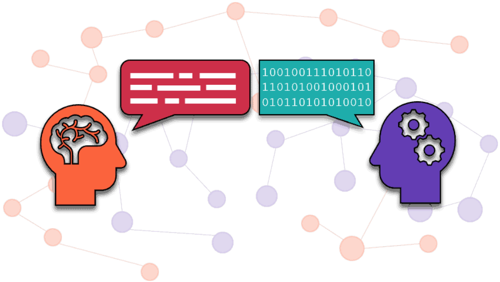
$$z_1^{[1]} = w_1^{[1]T} x + b_1^{[1]}, \; a_1^{[1]} = \sigma(z_1^{[1]})$$
$$z_2^{[1]} = w_2^{[1]T} x + b_2^{[1]}, \; a_2^{[1]} = \sigma(z_2^{[1]})$$
$$z_3^{[1]} = w_3^{[1]T} x + b_3^{[1]}, \; a_3^{[1]} = \sigma(z_3^{[1]})$$
$$z_4^{[1]} = w_4^{[1]T} x + b_4^{[1]}, \; a_4^{[1]} = \sigma(z_4^{[1]})$$

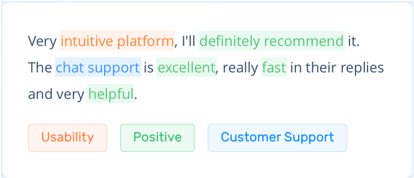
Figure: Computation of a layer

1. Backward propagation
2. Forward Propagation
3. Gradient descent (Optimization)
4. Activation functions
5. Random Initialization of parameters W_*

Overview

Natural Language Processing





1. How can we work with unstructured data?
2. Are there mathematics tools?

Deep Learning

Natural Language Processing

Text Vectorization

Word embedding

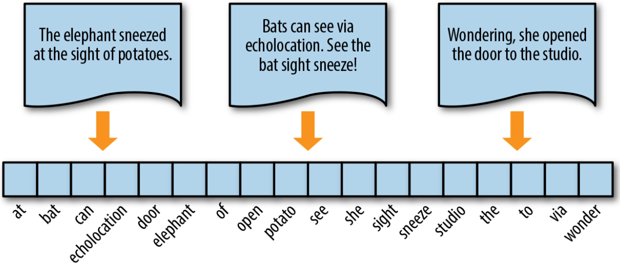
Deep learning architecture for NLP

NLP for the multi-class classification

Image Pattern Recognition

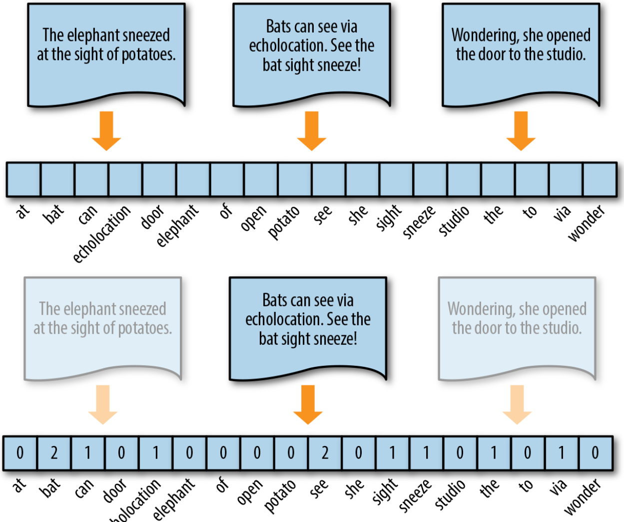
Representation as vector $\mathbb{R}$

Given three english texts



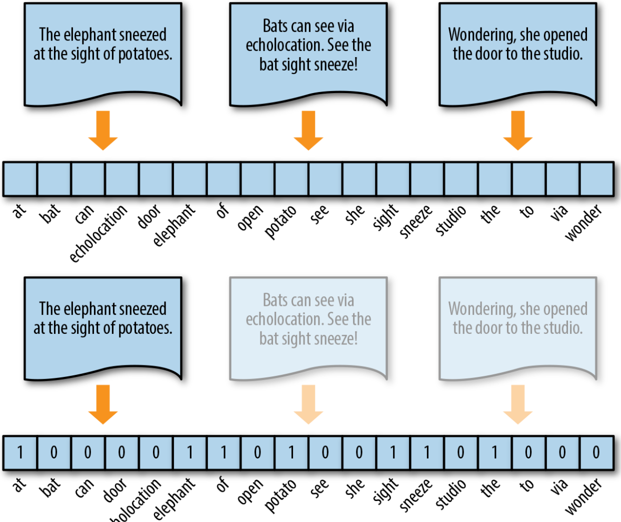
Representation as vector $\mathbb{R}$ - BagWords

Given three english texts



Representation as vector $\mathbb{R}$ - One-hot encoded

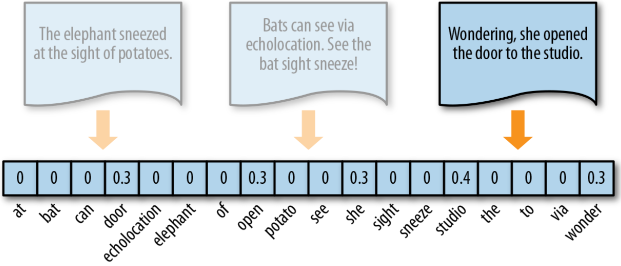
Given three english texts



Representation as vector $\mathbb{R}$ - Term frequency inverse document frequency (TF-IDF)

TF-IDF

Term frequency-inverse document frequency is a statistical measure that evaluates how relevant a word is to a document in a collection of documents.



Deep Learning

Natural Language Processing

- Text Vectorization
- Word embedding
- Deep learning architecture for NLP
- NLP for the multi-class classification

Image Pattern Recognition

Word embedding

Word embedding is another powerful way to work with text.

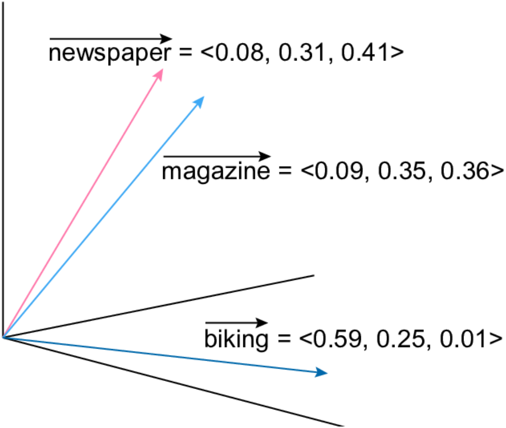


Figure: Visualization of 3-dimensional word embedding

Deep Learning

Natural Language Processing

- Text Vectorization
- Word embedding
- Deep learning architecture for NLP
- NLP for the multi-class classification

Image Pattern Recognition

Deep learning for NLP

Recurrent Neural Networks

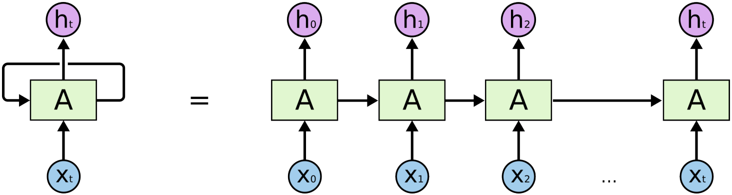


Figure: Recurrent Neural Networks

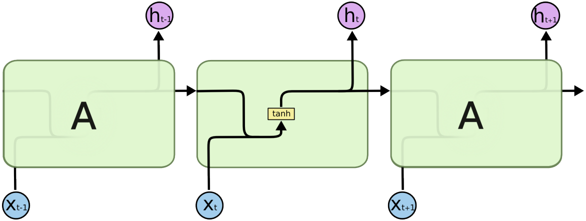


Figure: RNN in detail

Deep learning for NLP

Recurrent Neural Networks

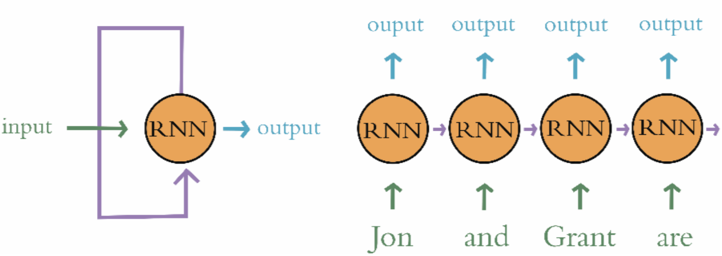
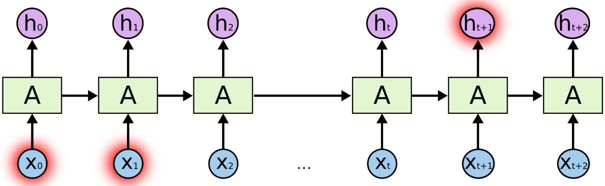


Figure: Example RNN

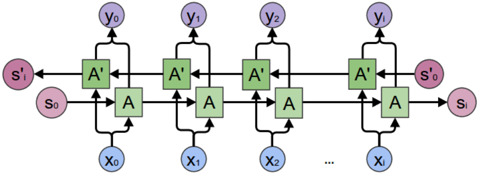
Deep learning for NLP

Recurrent Neural Networks - Drawbacks

- Problem: The Problem of Long-Term Dependencies



- Solution: Bidirectional RNN



Deep learning for NLP

Long Short Term Memory networks(LSTM)

LSTM solves the problem of vanishing gradients with Recurrent Neural Networks.

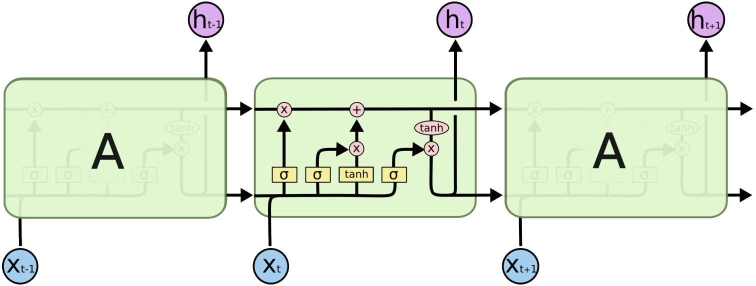


Figure: Long Short Term Memory networks in Detail

Deep learning for NLP

Others complex networks

- Bidirectional Recurrent Neural Networks.
- Attention Recurrent Neural Networks.
- Transformers Neural Networks.
- Sentence Embeddings using Siamese BERT-Networks.
- ....

Deep Learning

Natural Language Processing

Text Vectorization

Word embedding

Deep learning architecture for NLP

NLP for the multi-class classification

Image Pattern Recognition

NLP for the classification problem

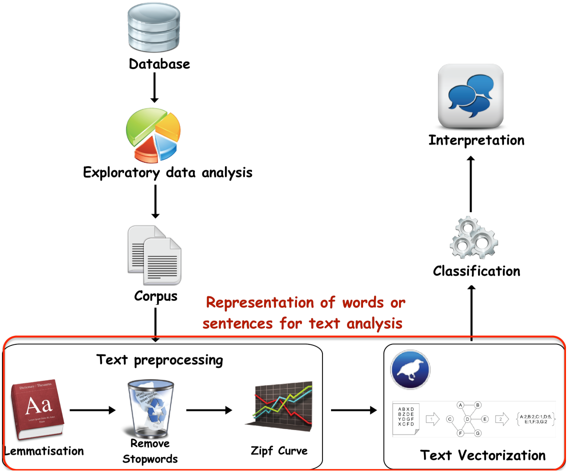
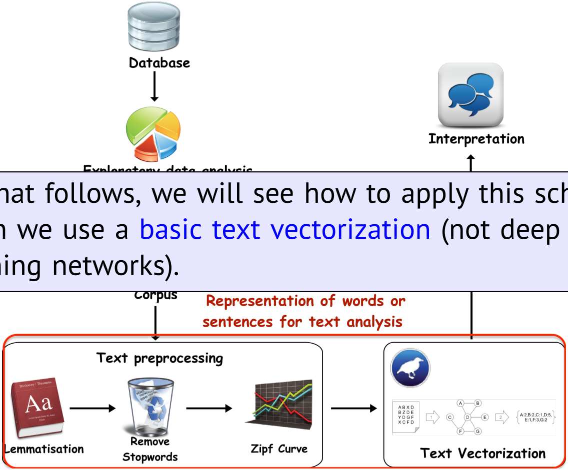


Figure: Schema NLP for classification

NLP for the classification problem



In what follows, we will see how to apply this schema when we use a basic text vectorization (not deep learning networks).

Figure: Schema NLP for classification

NLP for the classification problem

Tokenization

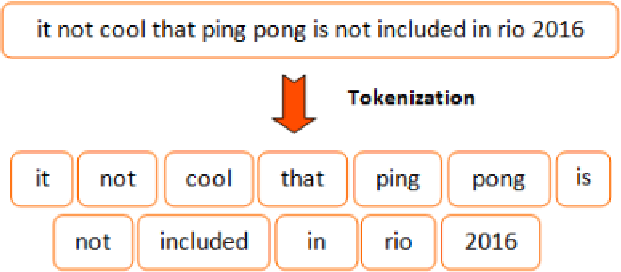


Figure: Tokenization example

NLP for the classification problem

Lemmatisation or Stemming

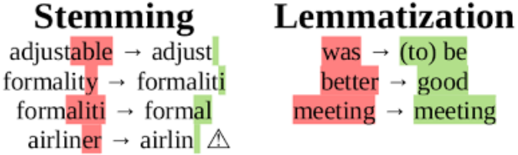


Figure: Stemming vs Lemmatization

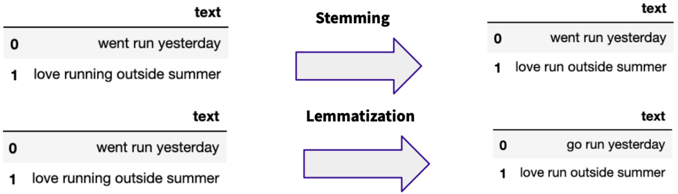


Figure: Example of Stemming vs Lemmatization

NLP for the classification problem

Remove stopwords

(a) With StopWords

(b) Without StopWords

Figure: Stops words in Action

Overview

Computer Vision Problems

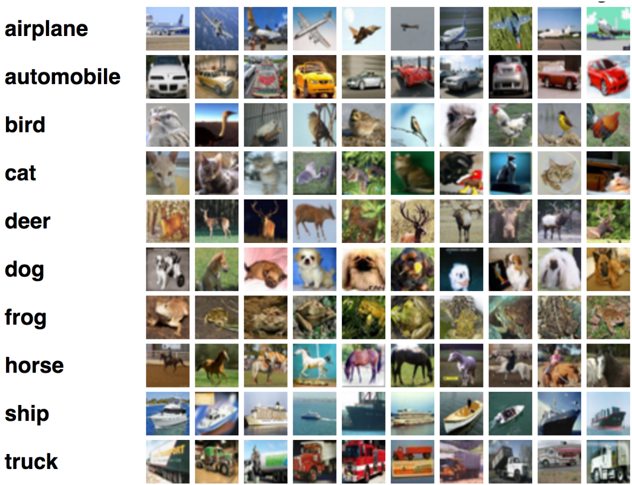
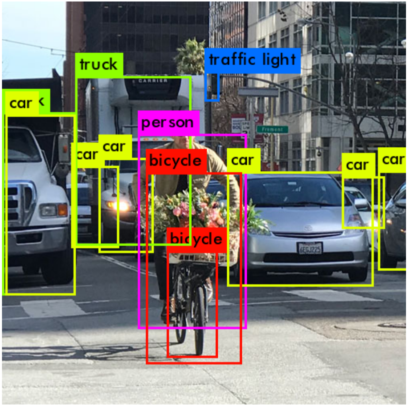
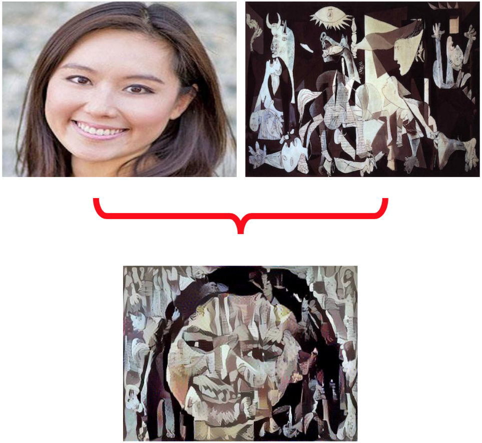


Figure: Image classification

Computer Vision Problems



(a) Object detection



(b) Natural Style Transfer

Convolutional Neural Network

Global vs. local patterns

Densely connected layers learn global patterns in their input feature space, whereas convolution layers learn local patterns, that meaning patterns found in small 2D windows.

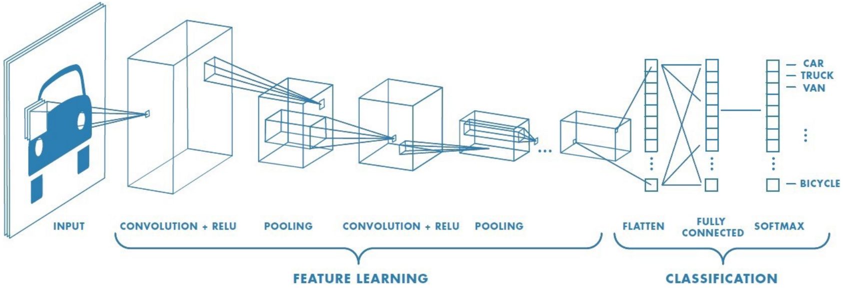


Figure: Convolution architecture

Convolutional Neural Network

Global vs. local patterns

Densely connected layers learn global patterns in their input feature space, whereas convolution layers learn local patterns, that meaning patterns found in small 2D windows.

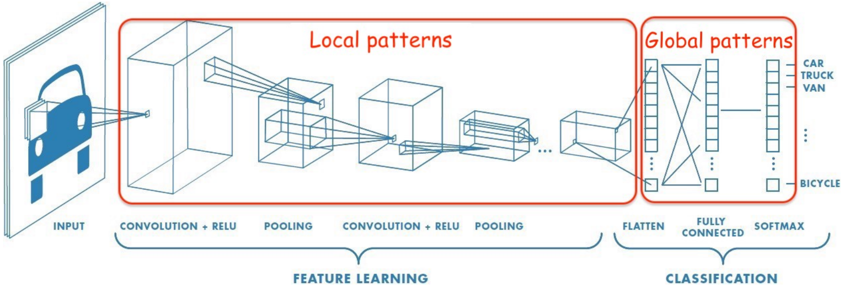


Figure: Convolution architecture

Convolutional Neural Network

Learning of local patterns in decomposing images in vertical and horizontal edges.

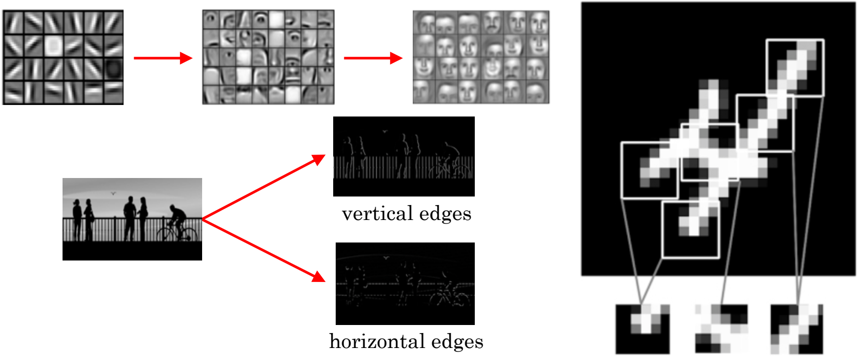


Figure: Operation detections

Convolutional Neural Network

Learning operations

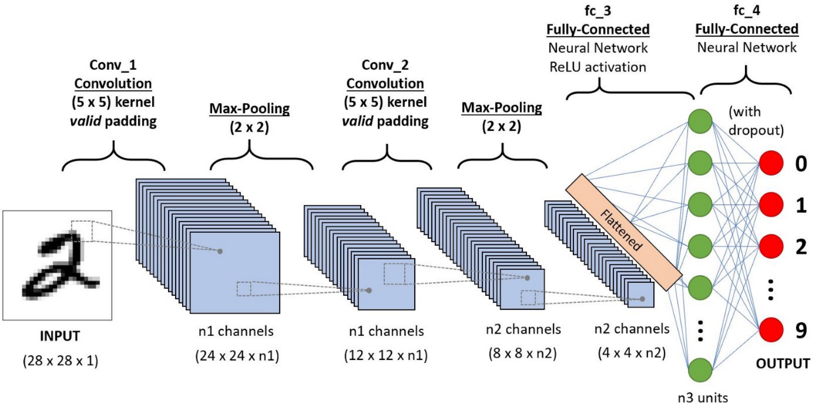


Figure: Convolution and Full-connected operations

Convolutional Neural Network

Gray vs RGB

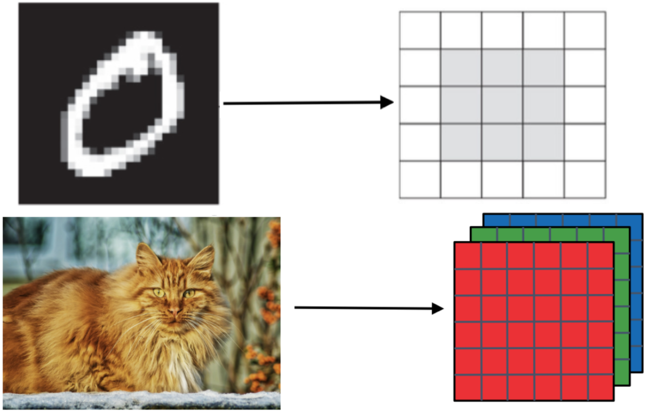
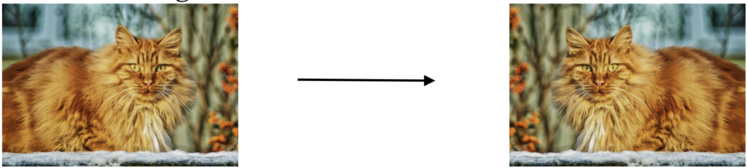


Figure: Gray vs RGB

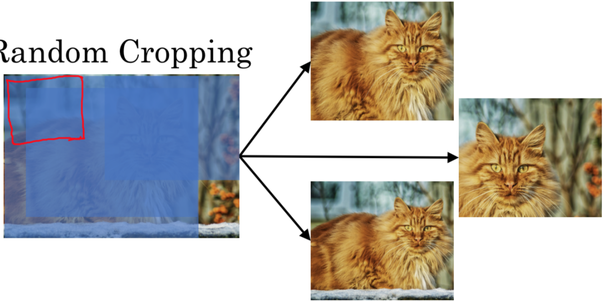
Convolutional Neural Network

Data augmentation

Mirroring



Random Cropping



Rotation

Shearing

Local warping

...

Convolutional Neural Network

Convolution Padding

Padding is a term relevant to convolutional neural networks as it refers to the amount of pixels added to an image when it is being processed by the kernel of a CNN.

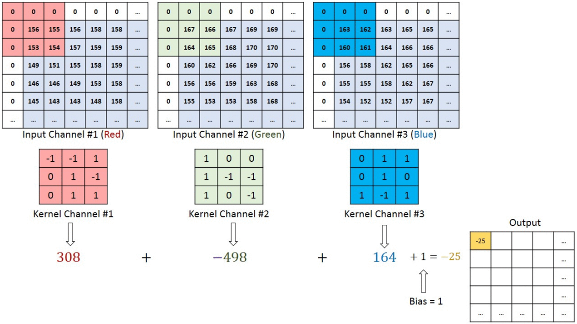


Figure: Convolution Padding operation

Convolutional Neural Network

Convolution Padding

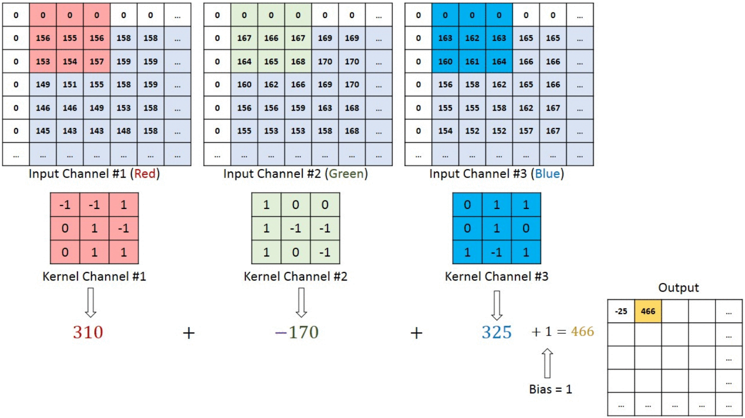


Figure: Convolution Padding operation

Convolutional Neural Network

Convolution Padding

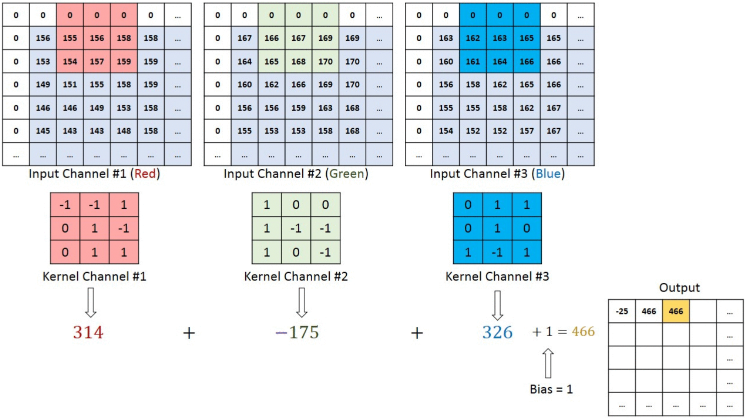


Figure: Convolution Padding operation

Convolutional Neural Network

Convolution Padding

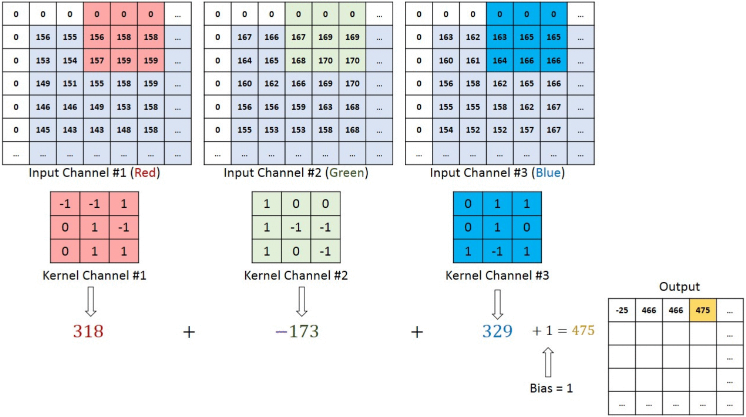
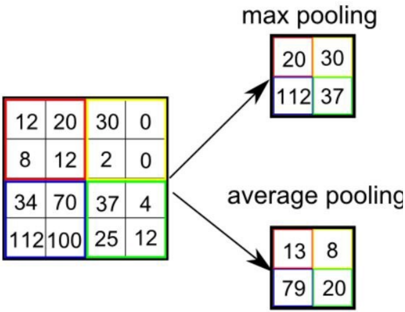


Figure: Convolution Padding operation

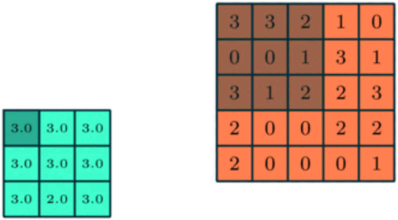
Convolutional Neural Network

Convolution Pooling

Pooling layers reduce the dimensions of data by combining the outputs of neuron clusters at one layer into a single neuron in the next layer.



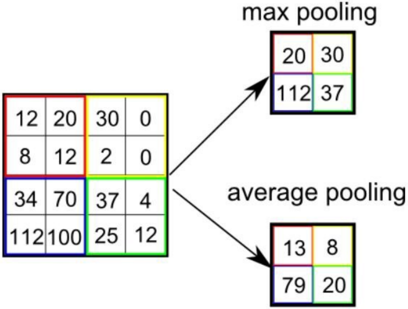
(a) Pooling Operation



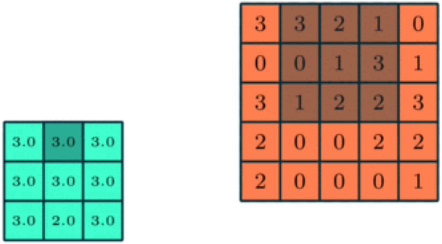
(b) Max pooling

Convolutional Neural Network

Convolution Pooling



(a) Pooling Operation

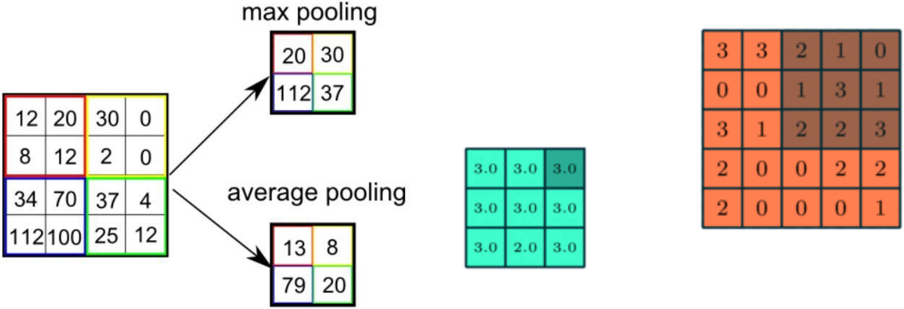


(b) Max pooling

Figure: Convolution Pooling operations

Convolutional Neural Network

Convolution Pooling

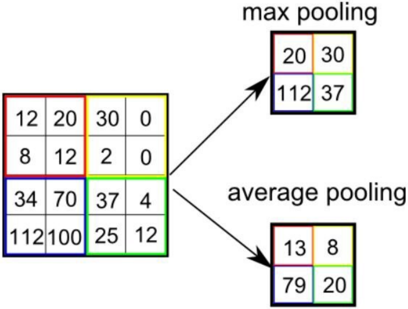


(a) Pooling Operation (b) Max pooling

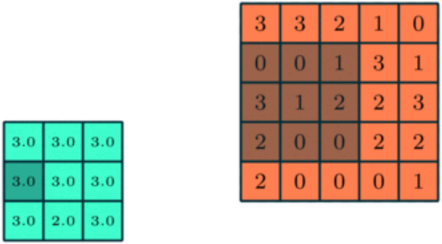
Figure: Convolution Pooling operations

Convolutional Neural Network

Convolution Pooling



(a) Pooling Operation

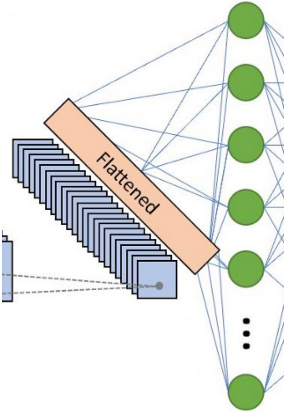


(b) Max pooling

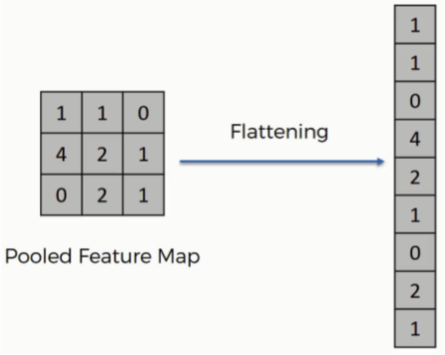
Figure: Convolution Pooling operations

Convolutional Neural Network

Flattening Operation



(a) Flattening

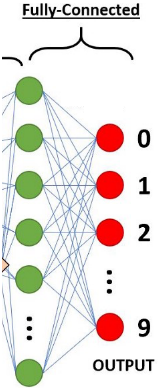


(b) Example

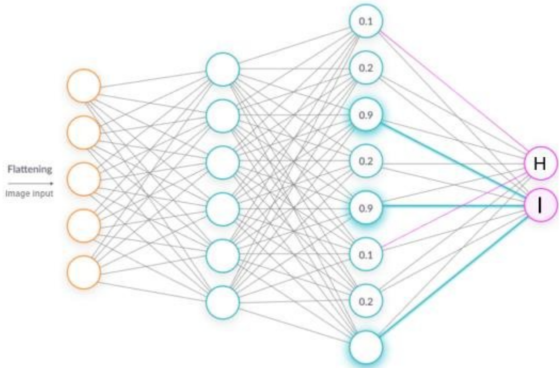
Figure: Flattened layer

Convolutional Neural Network

Full connected (FC) layer operations



(a) 1 FC layer



(b) 3 FC layers

Figure: Examples Full connected layers

Deep Learning

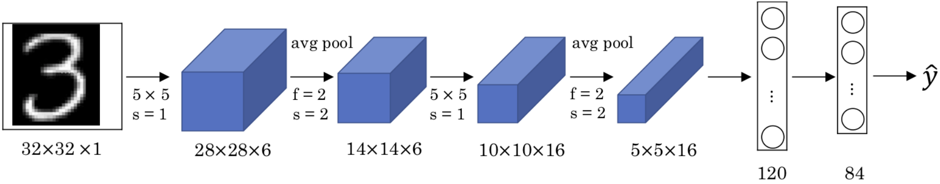
Natural Language Processing

Image Pattern Recognition

- Convolutional Neural Networks
- Deep learning architecture for classification
- Examples of Deep learning techniques

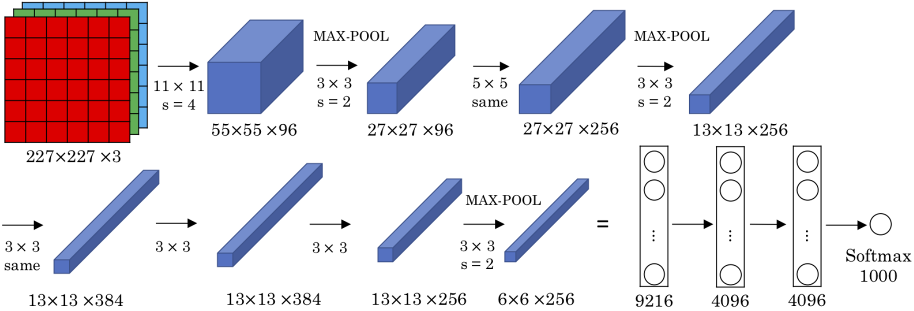
Deep learning architecture for classification

LeNet-5 network



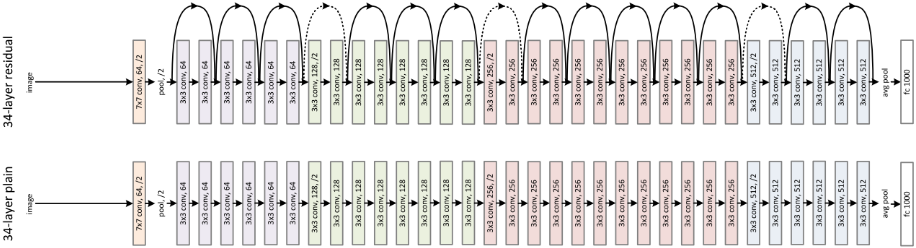
Deep learning architecture for classification

AlexNet network



Deep learning architecture for classification

ResNet network



Deep Learning

Natural Language Processing

Image Pattern Recognition

- Convolutional Neural Networks
- Deep learning architecture for classification
- Examples of Deep learning techniques

Examples of dimensionality reduction

Let us do Machine Learning

Code source - [Link]

Bibliography I

Christopher, O. (2020). Understanding lstm networks.

Denoeux, T. (2020). Lecture sy19 at utc - machine learning.

Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*. Springer New York Inc.

Kenter, T. (2017). *Text Understanding for Computers*.

Saha, S. (2018). A comprehensive guide to convolutional neural networks - the eli5 way.

Sreenivas, A., Maheshwari, M., Jain, S., Choudhary, S., and G, D. V. (2020). Indian sign language communicator using convolutional neural network. *International Journal of Advanced Science and Technology*, 29:11015–11031.

Yan-Tak Ng, A. (2020). Deep learning specialization - coursera.

Thank You for Your Attention!