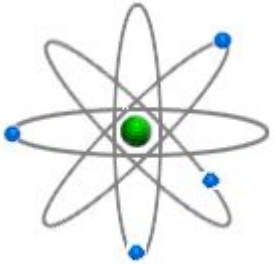
Class 10

H Academy

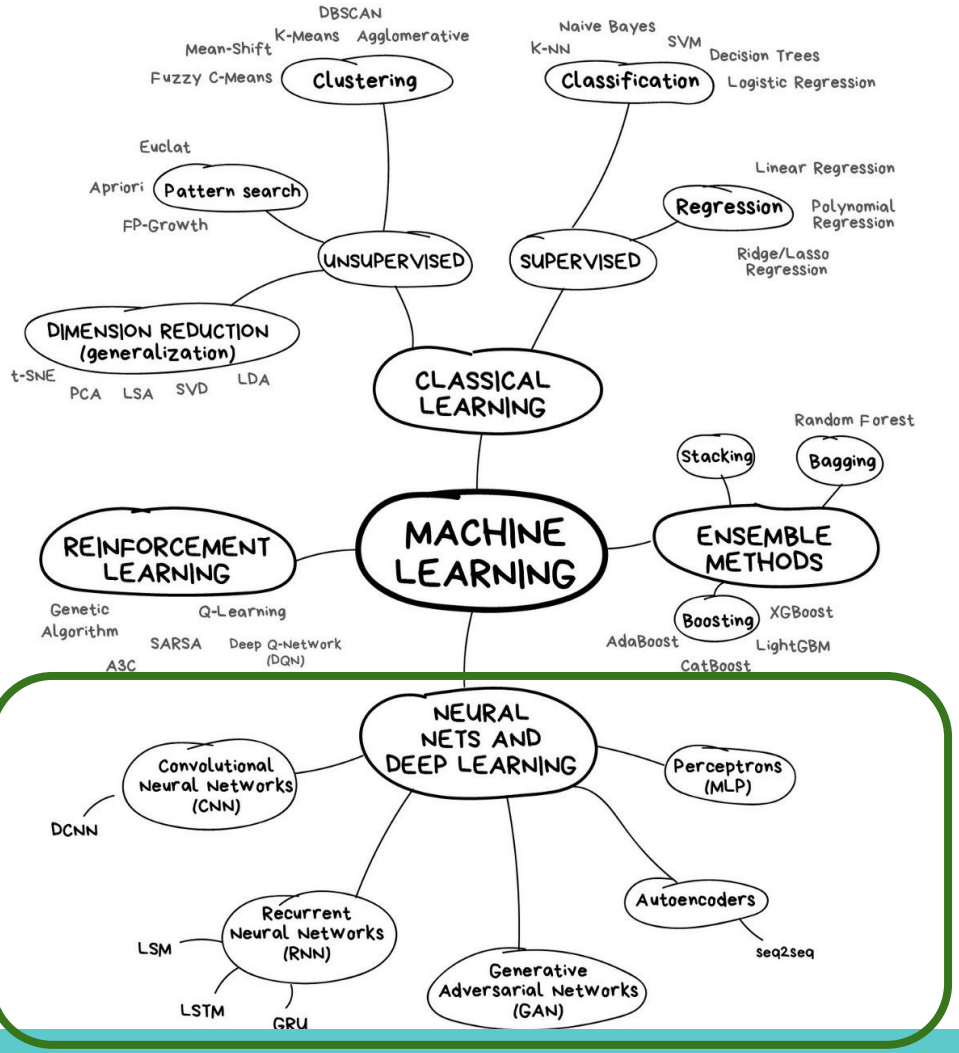
March 18th, 2021 - By Nathan Landman

Class Agenda

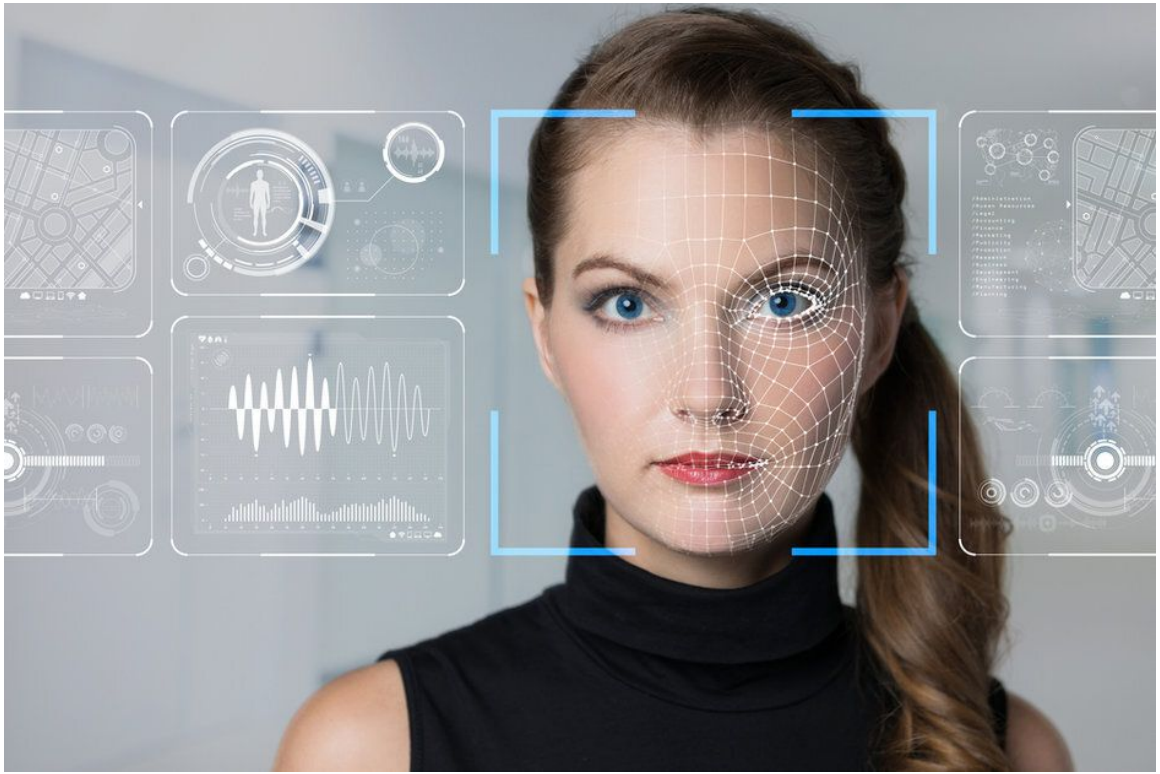


1. Neural Networks
    a. Basics of Neural Networks
    b. Neural Network Architectures
    c. Neural Networks in Practice

Machine Learning Landscape



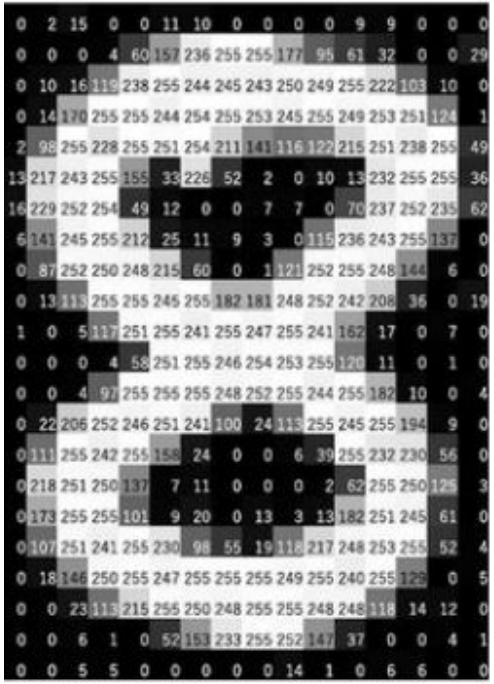
Let’s talk about machine vision...



How does a machine interpret an 8?

A machine processes an image by breaking it down into **pixels**, where each pixel contains a set of numbers representative of the color.

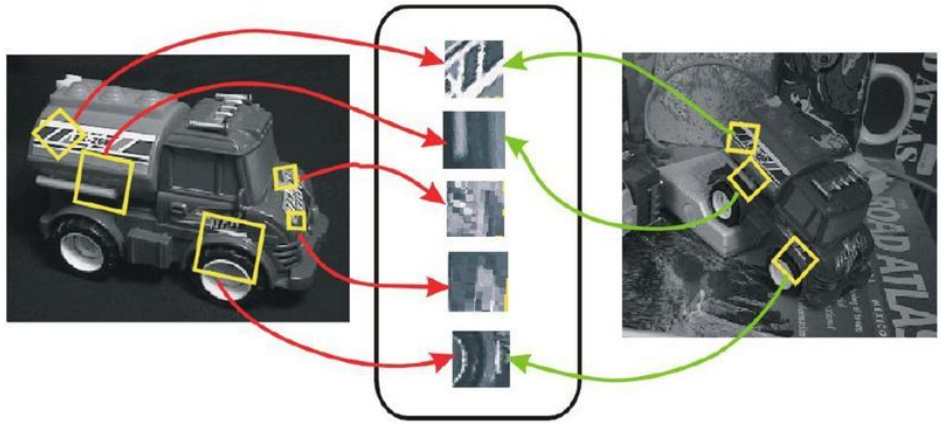
In large images, there's a bunch of noise, when we only need to see a fraction of the image to know what it is.

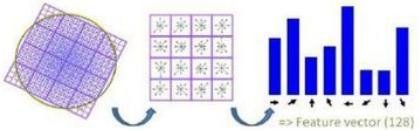


How do you obtain an accurate representation of an image?

Early computer vision systems were very fine-tuned to detecting edges and other salient features in specific images.

Computer Vision – Invariant Features



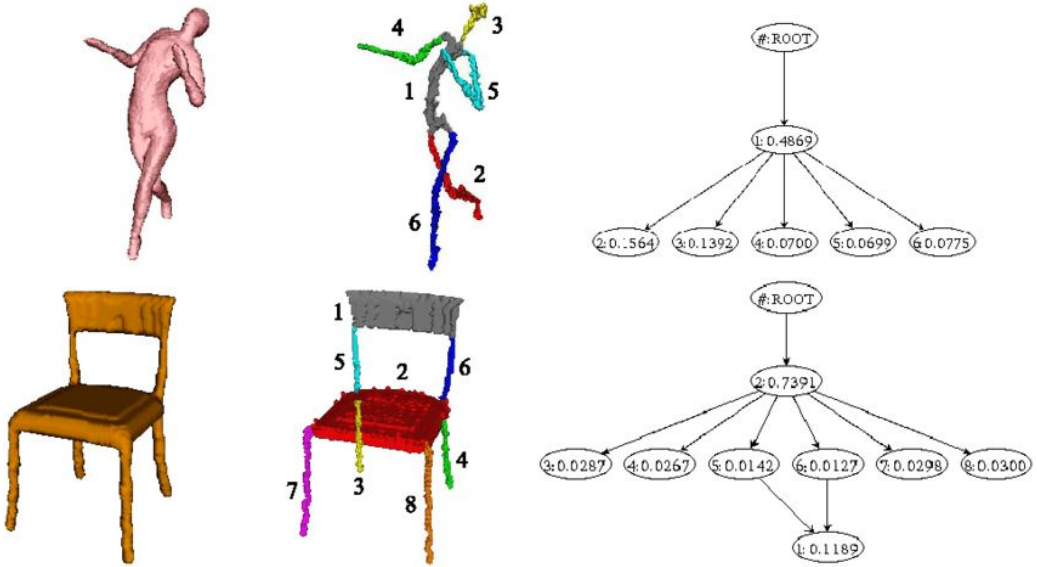


"SIFT" & Object Recognition, David Lowe, 1999

Slides based on cs231n by Fei-Fei Li & Andrej Karpathy & Justin Johnson

So, how would we find representations of a chair?

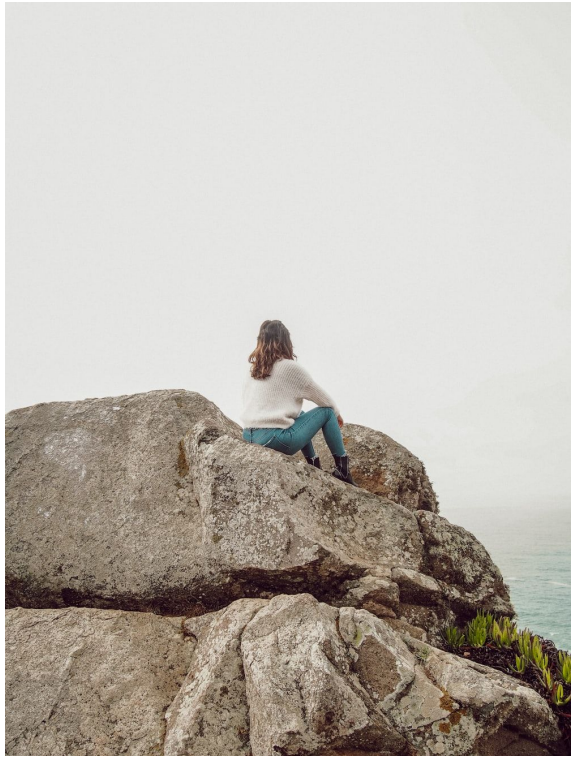
So, how would we find representations of a chair?



But, would these classify as chairs?

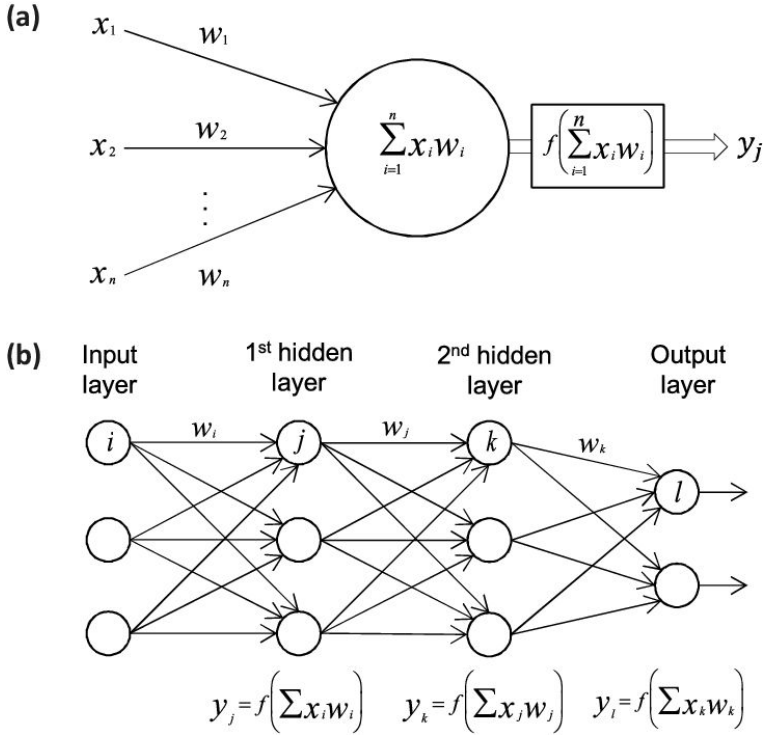
How about something you sit on?

How about something you sit on?

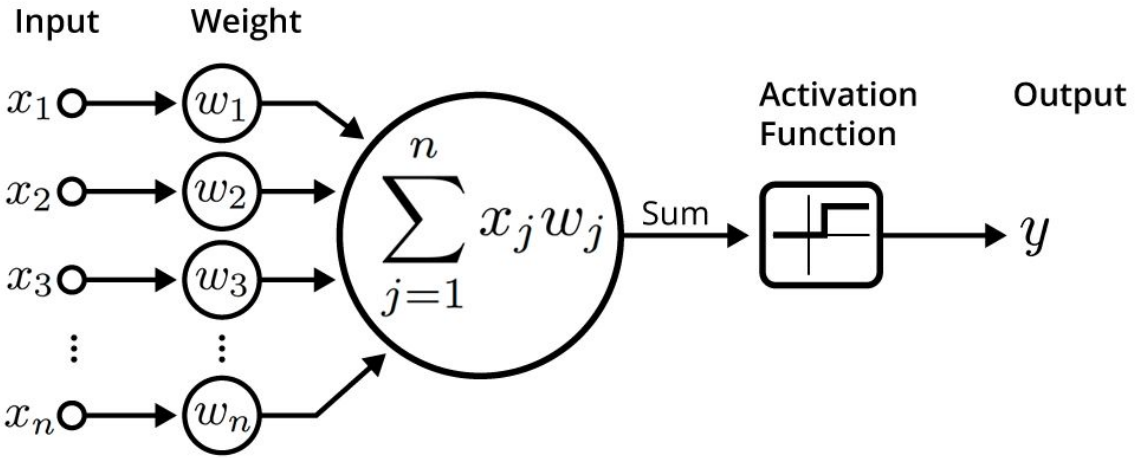


Enter Deep Learning!

By intelligently trying a bunch of mathematical operations using fast computers, a machine is able to detect patterns in the data.



What are these operations? An Introduction to the Neuron



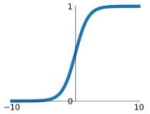
An illustration of an artificial neuron. Source: Becoming Human.

Activation Functions

Activation Functions

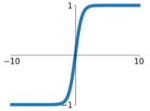
Sigmoid

$\sigma(x) = \frac{1}{1+e^{-x}}$



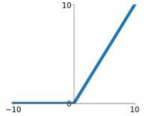
tanh

$\tanh(x)$



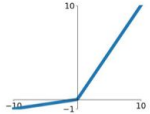
ReLU

$\max(0, x)$



Leaky ReLU

$\max(0.1x, x)$

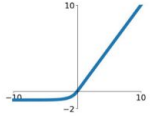


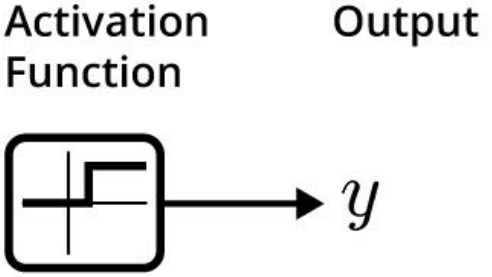
Maxout

$\max(w_1^T x + b_1, w_2^T x + b_2)$

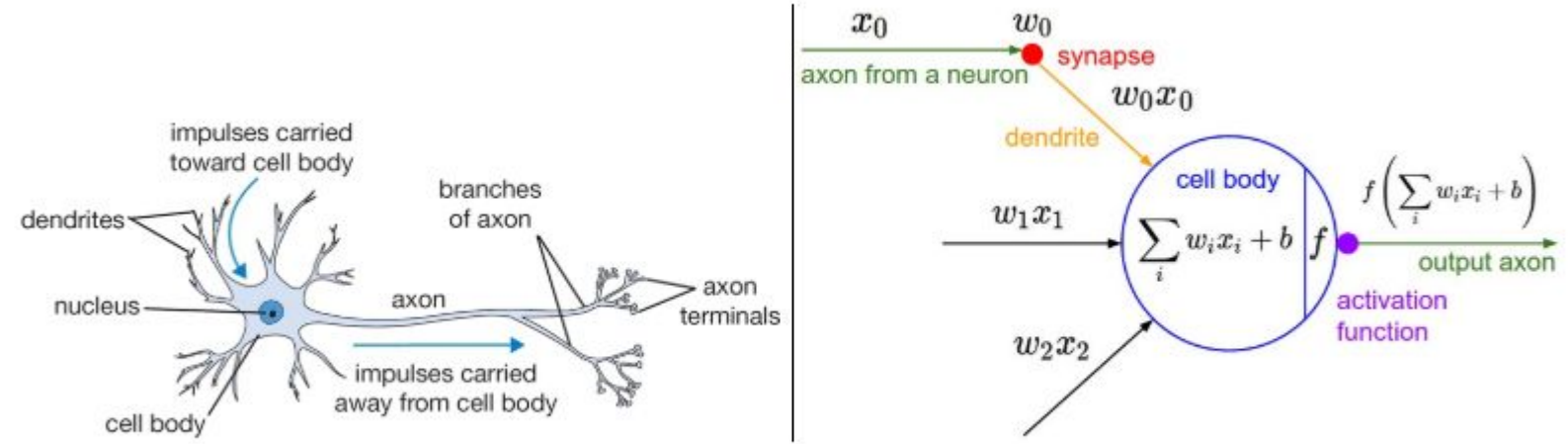
ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



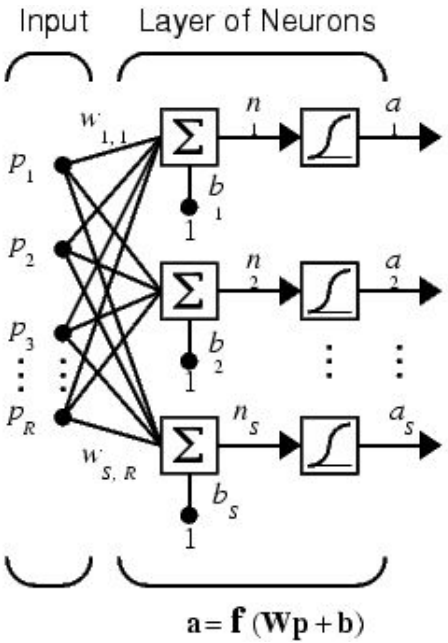


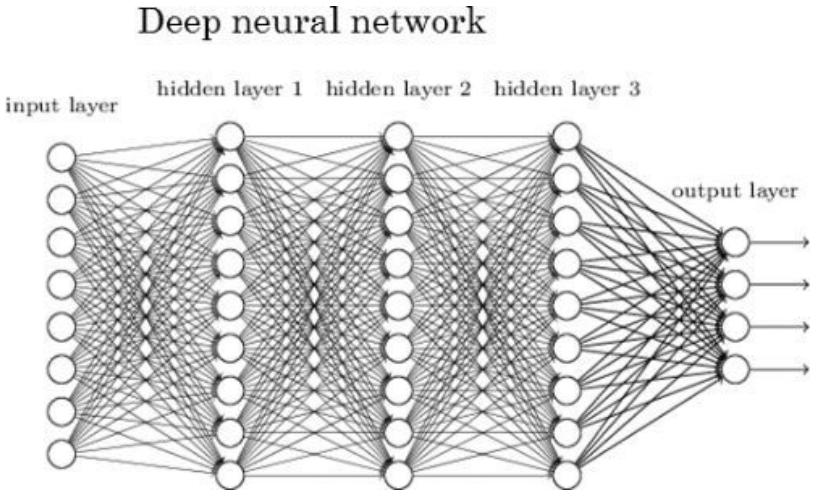
Sidenote.. Why is it called a Neural Network?



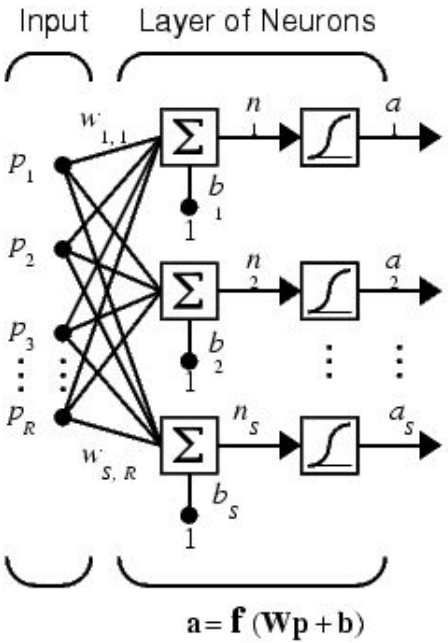
A cartoon drawing of a biological neuron (left) and its mathematical model (right).

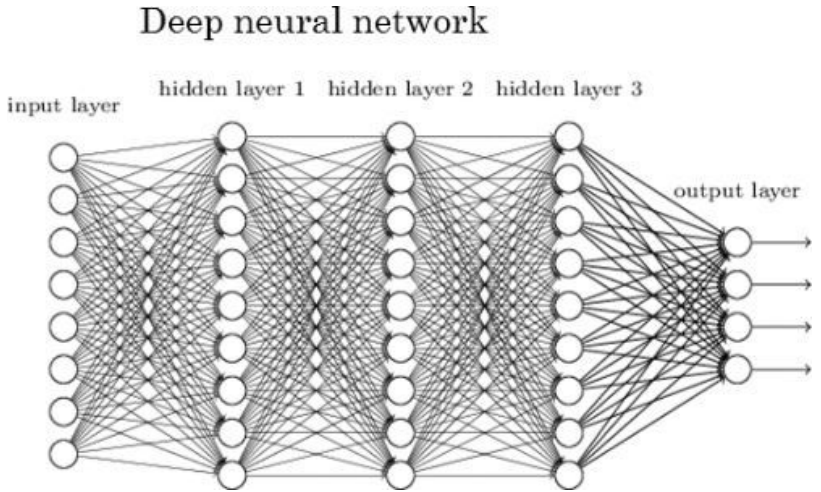
How do we build complex neurons?



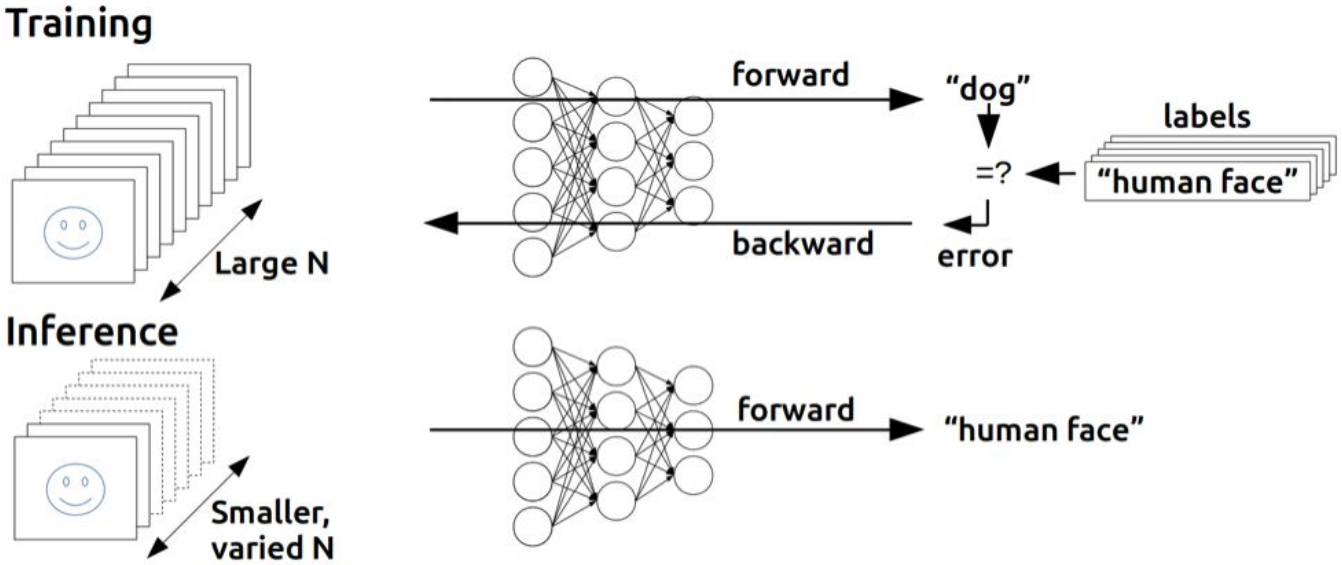


We call this a Feed Forward Neural Network

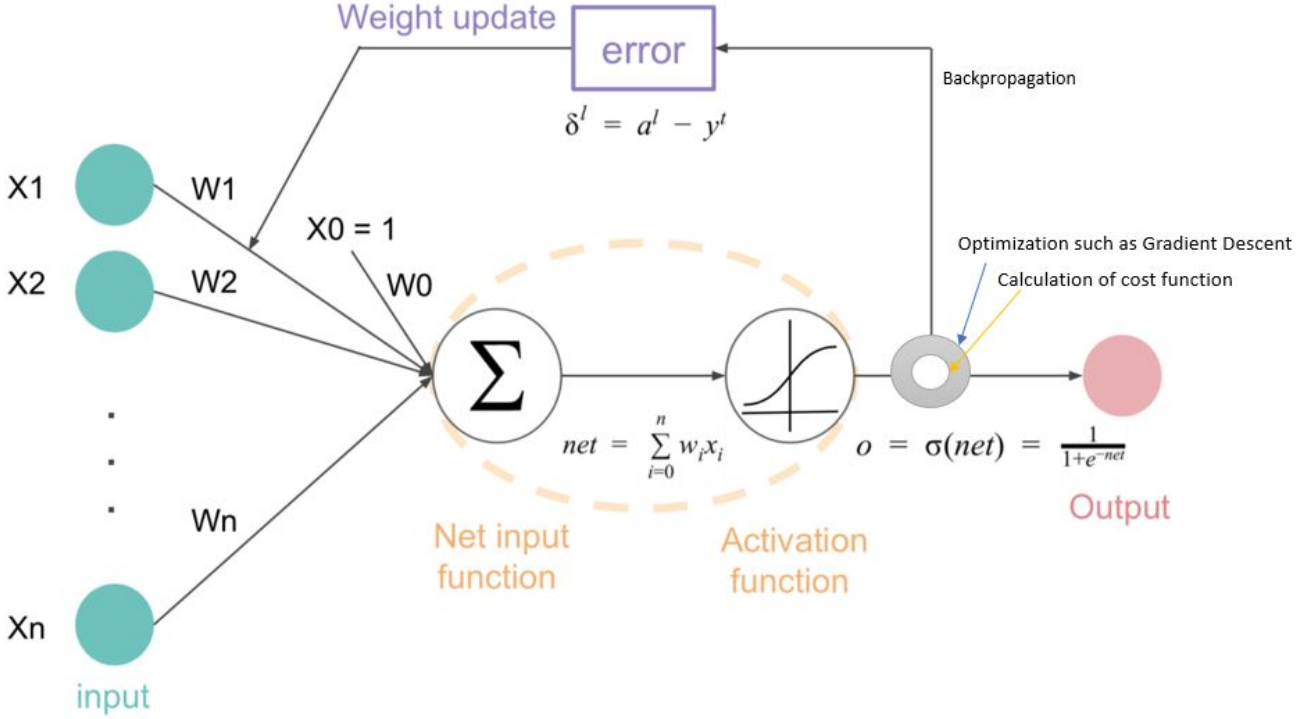




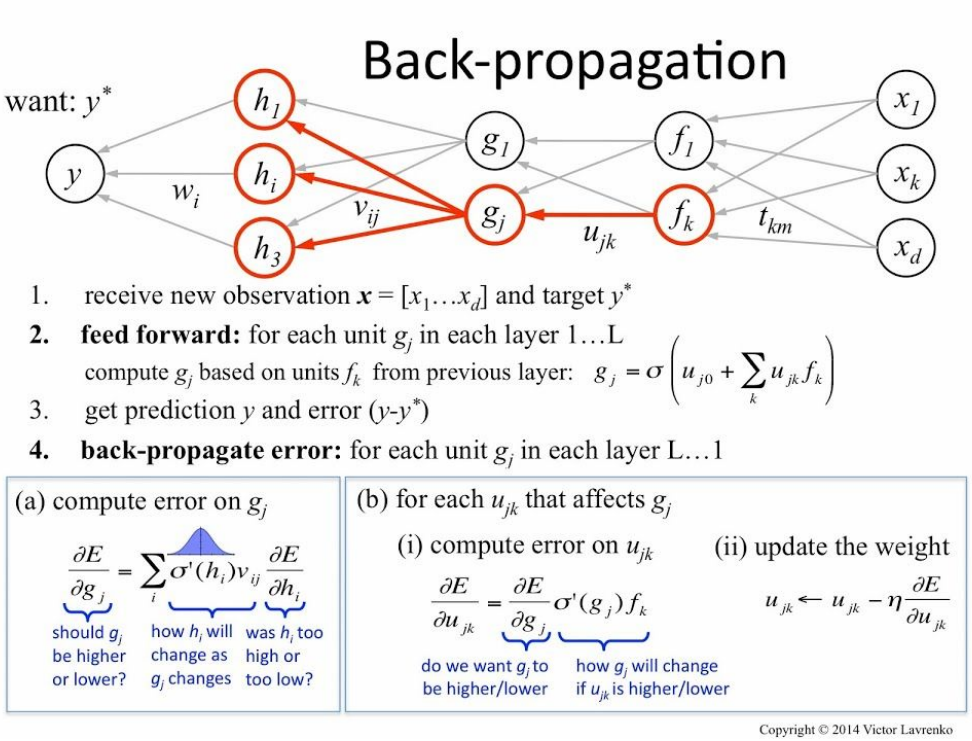
How does it actually learn from its mistakes?



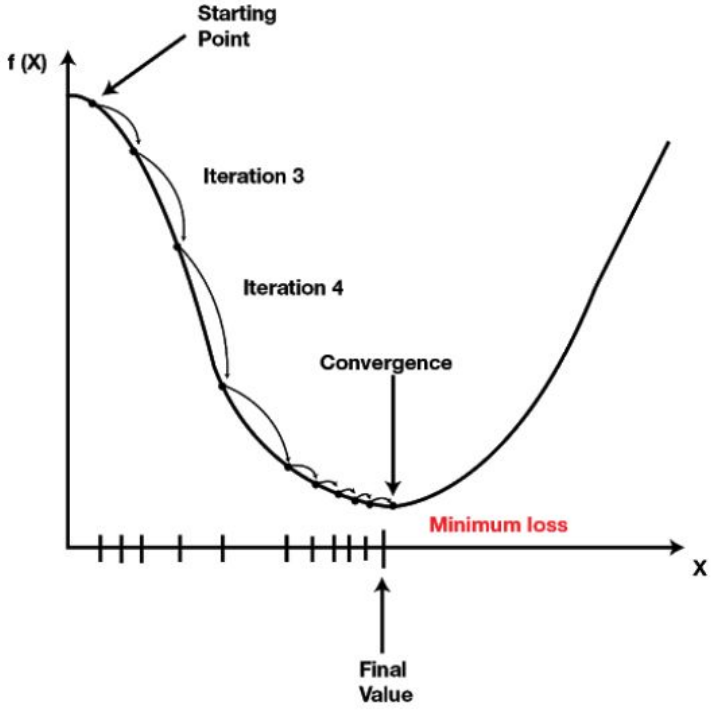
How does it actually learn from its mistakes?



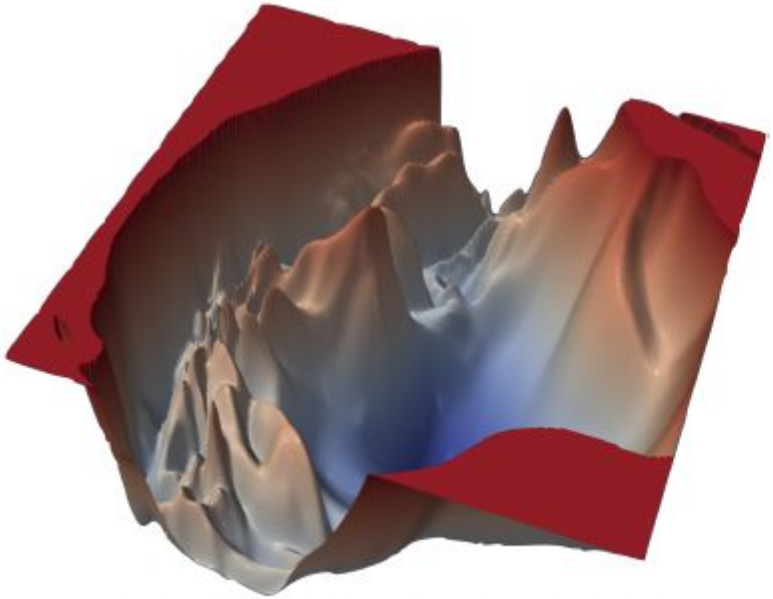
How does it actually learn from its mistakes?



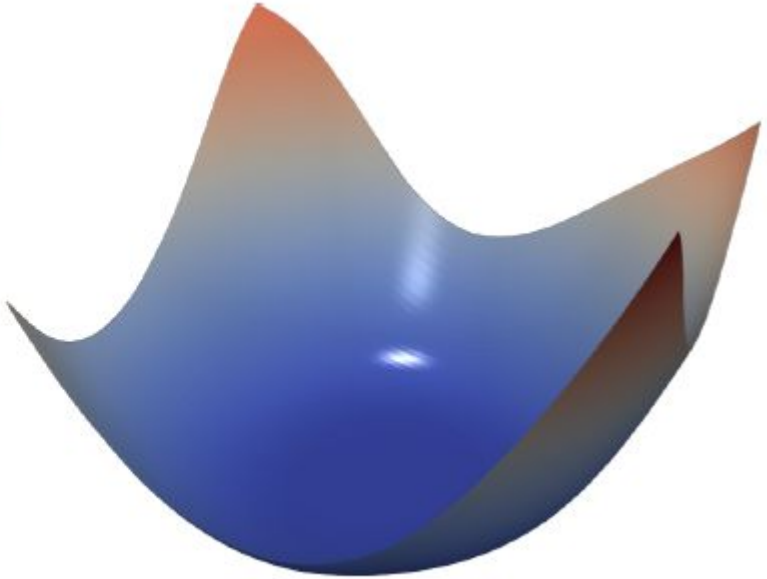
Convergence to a minimum loss via backpropagation



Beware of a local minimum!

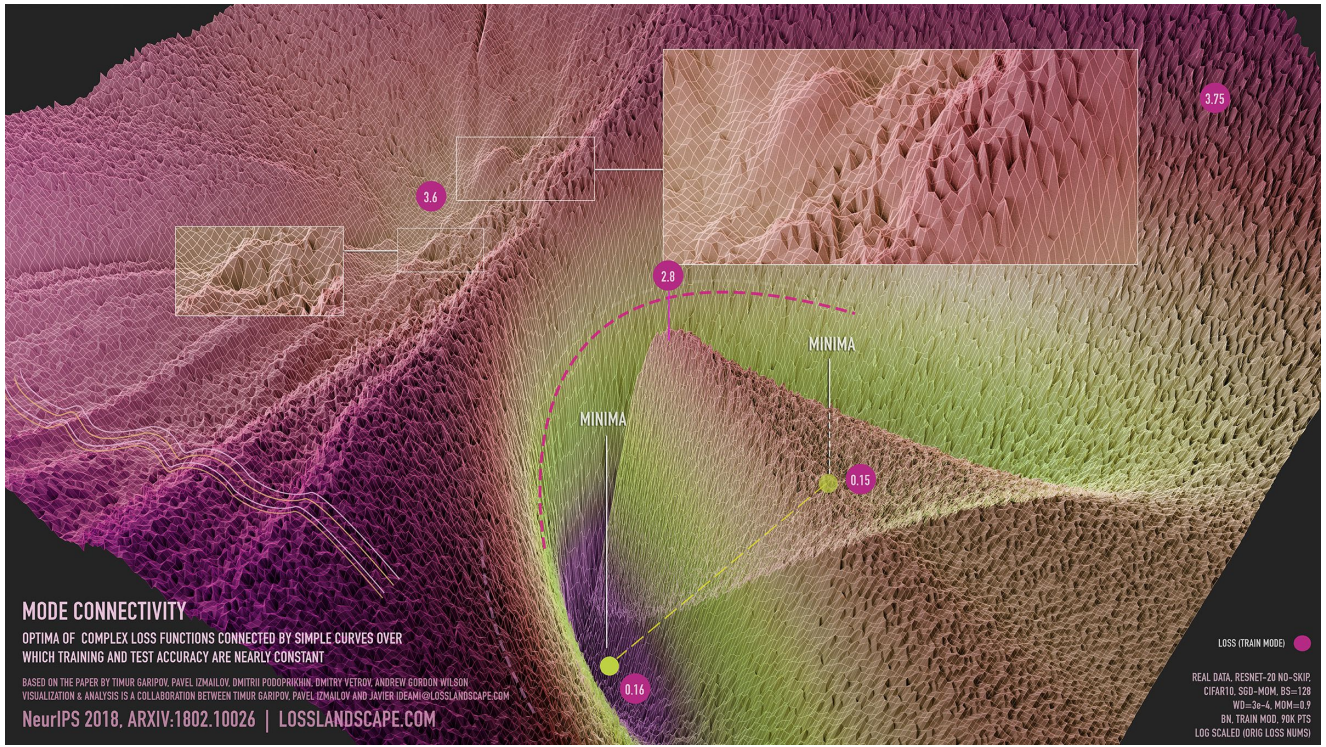


(a) ResNet-110, no skip connections

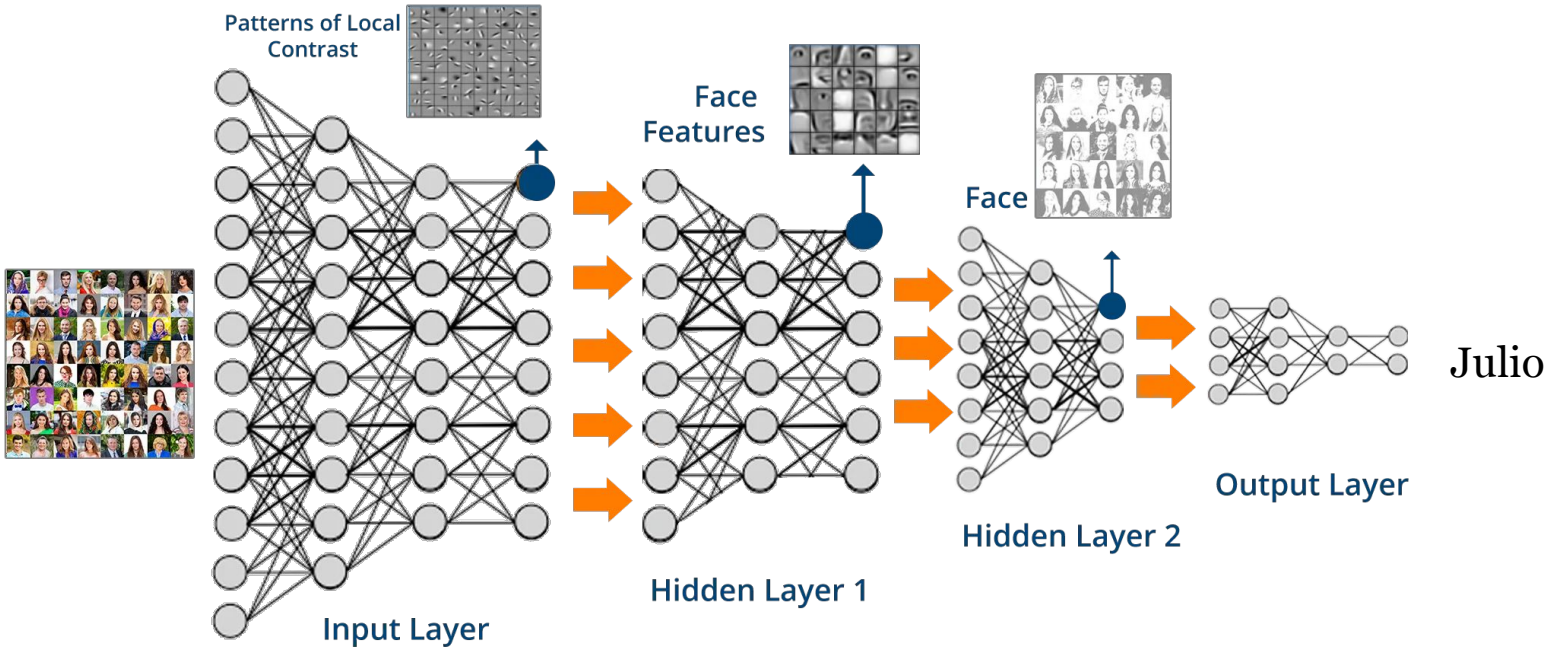


(b) DenseNet, 121 layers

Beware of a local minimum!

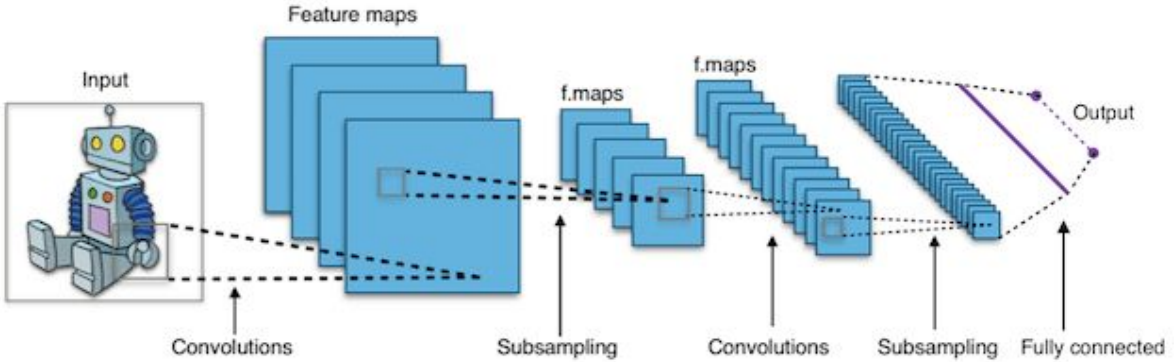


Back to computer vision...



Convolutional Neural Network (CNNs)

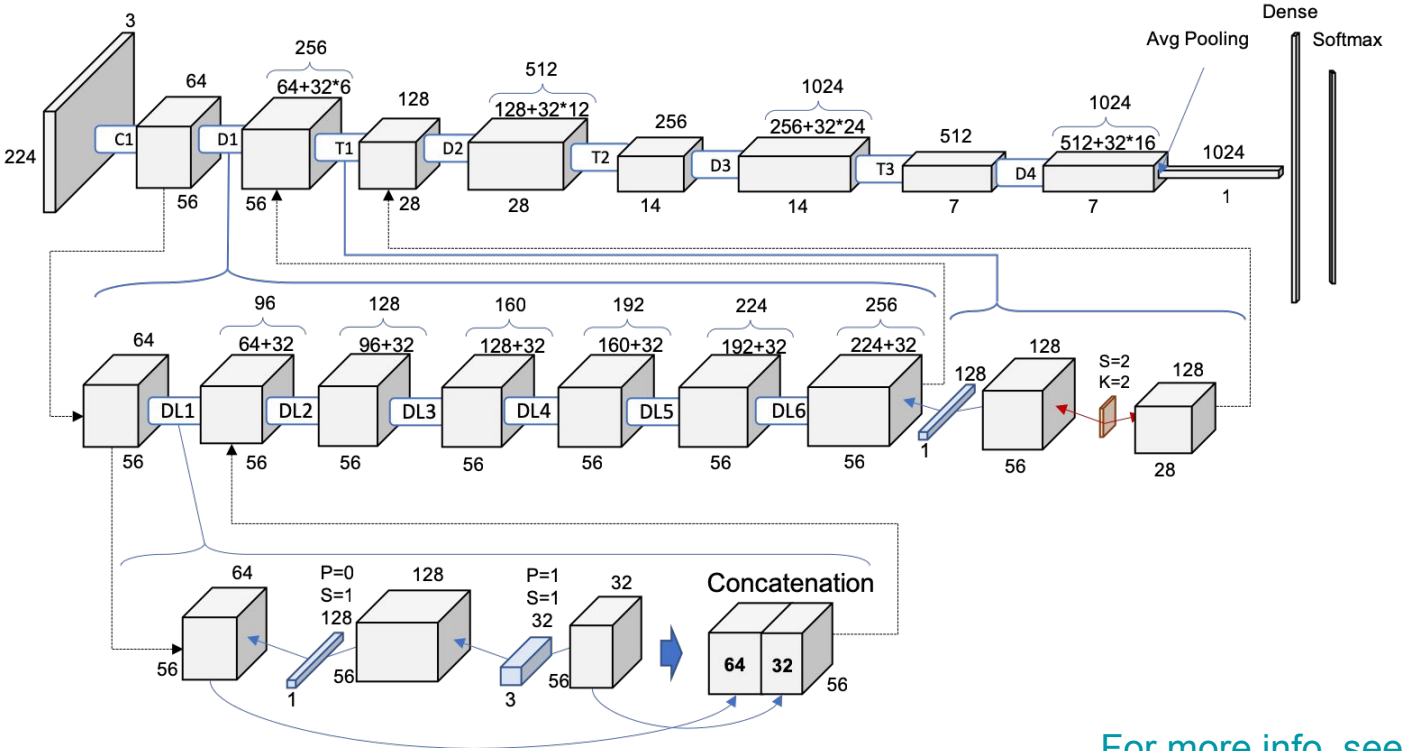
A **Convolutional Neural Network (ConvNet/CNN)** is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other



Convolutional Neural Network (CNNs)

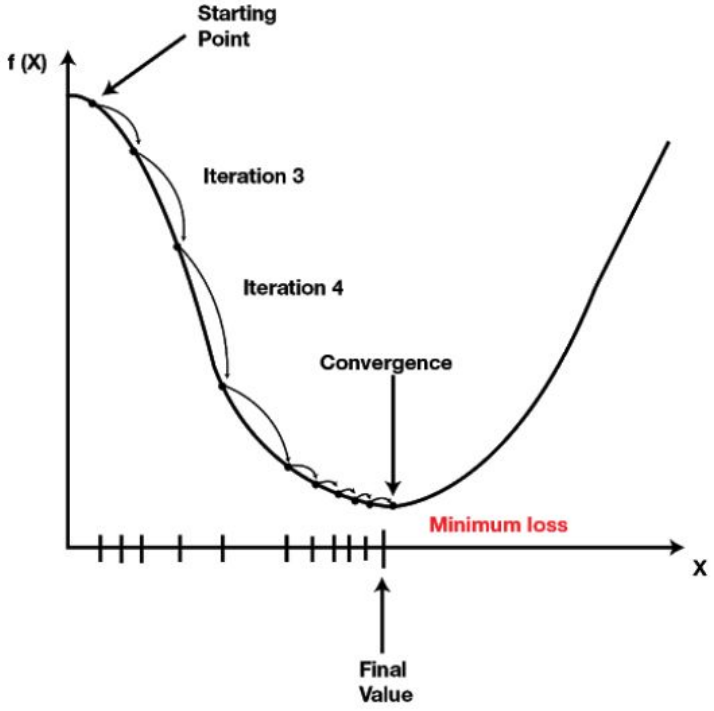
A **Convolutional Neural Network (ConvNet/CNN)** is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other

CNNs gone wild



[For more info, see this note](#).

Remember this graph?

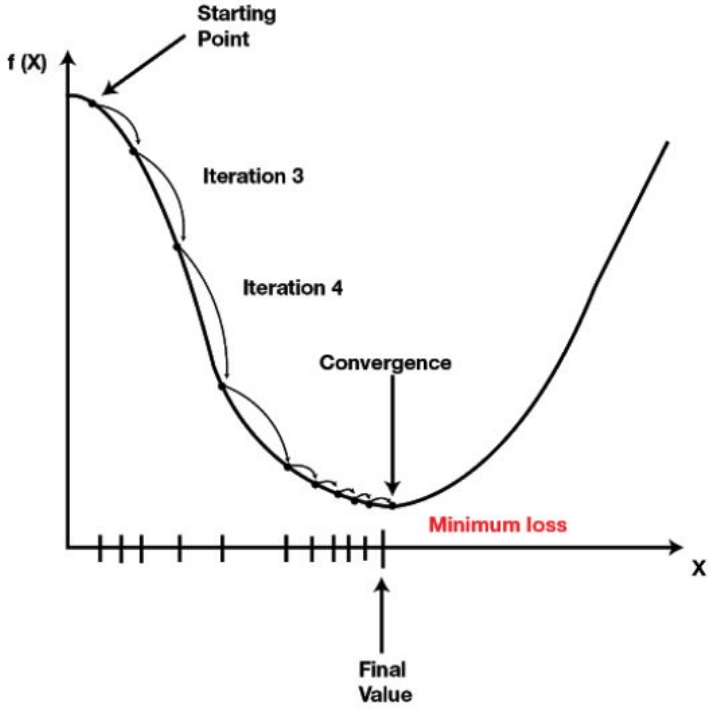


Training Technicalities

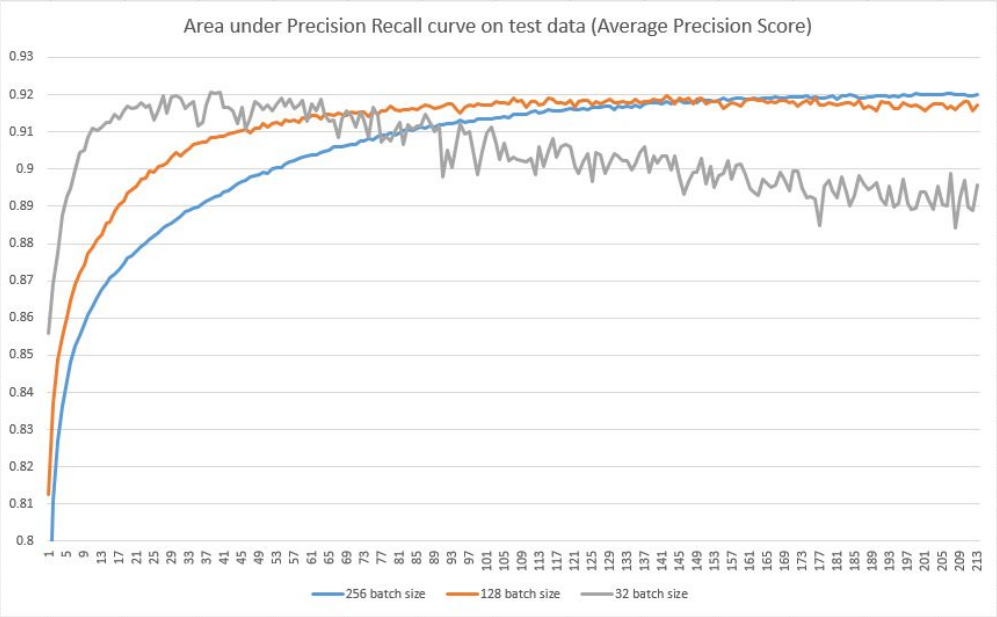
You wouldn't want to overfit your model to a single sample.

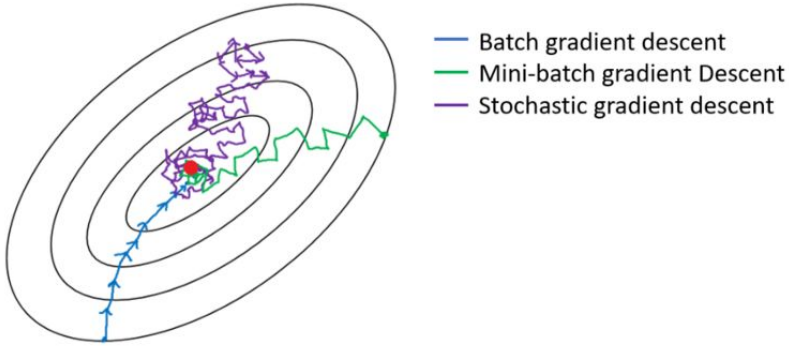
For this reason, we use **batches**! A batch is a number of samples whose errors are aggregated and shown to the model altogether.

If a batch is of size one, we call that **stochastic gradient descent**.



How do we set these parameters?

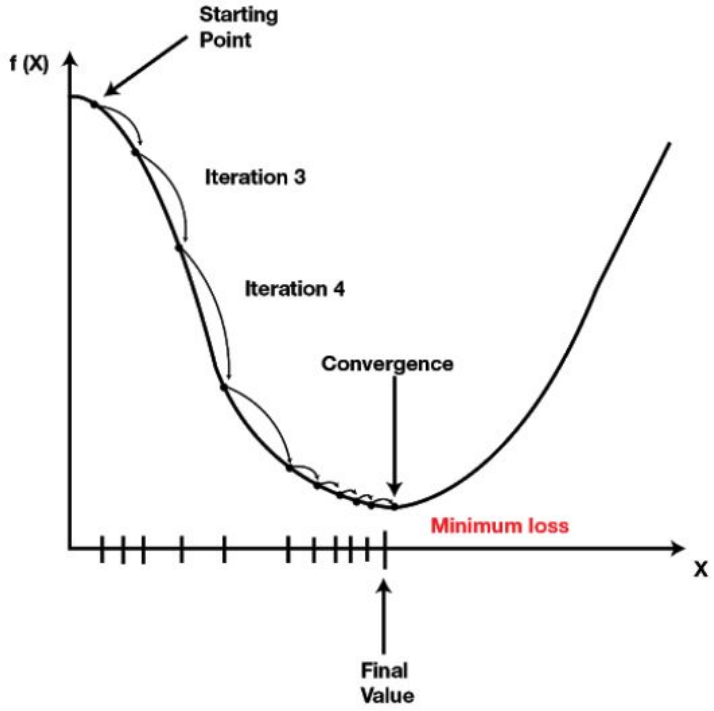




DNNs train via multiple epochs

The number of epochs is a hyperparameter that defines the number of times that the learning algorithm will work through the entire training dataset.

Because batches can be different across epochs, multiple epochs are useful when training.



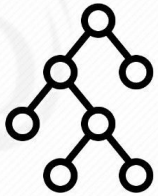
Epochs, Batches, and Iterations

Epoch :

An Epoch represent one iteration over the entire dataset.

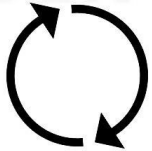
Batch :

We cannot pass the entire dataset into the Neural Network at once. So, we divide the dataset into number of batches.



Iteration :

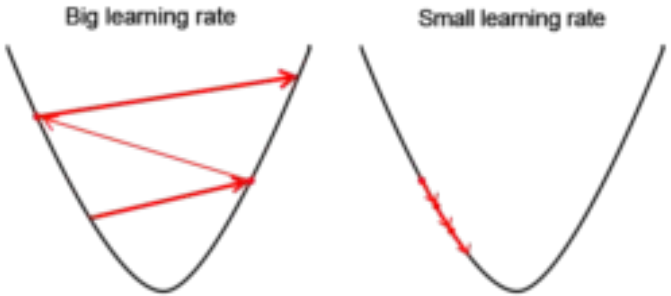
If we have 1000 images as Data ane a batch size of 20, then an Epoch should run 1000/20 = 50 iteration.

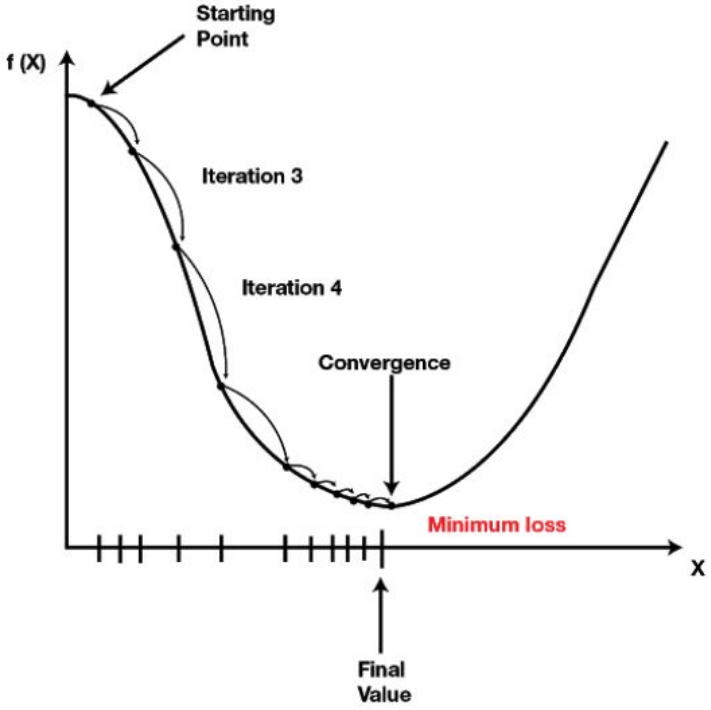


What determines the jumps per iteration?

A **learning rate** is used to modulate the amount of corrections your algorithm has per training iteration.

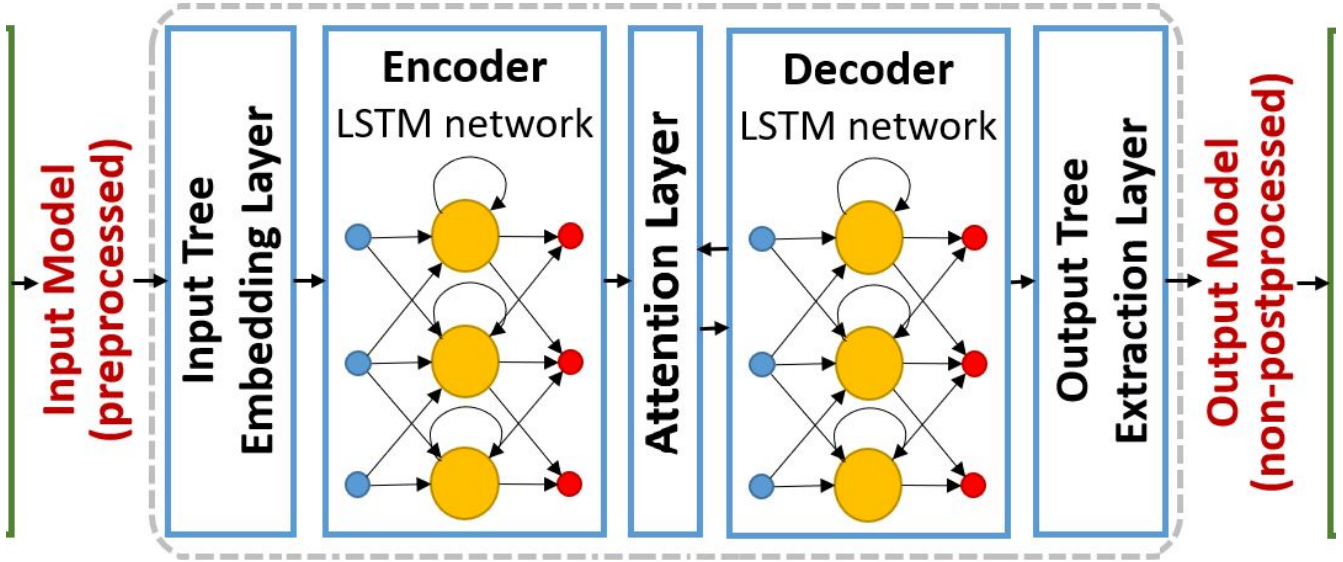
Many algorithms exist to help you set the right learning rate.





A quick word on NLP

Natural Language Processing (NLP) prediction problems can also be solved with Neural Networks. Popular architectures look like the one below.

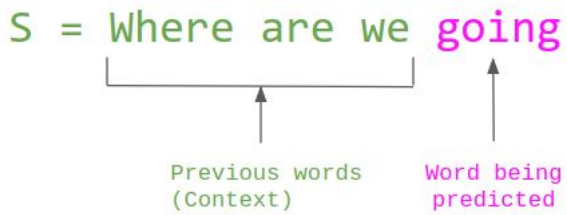


How do they work?

They work via a Language Model - an assigned probability for words that fill in the blank.

Large language models like GTP-3 scan the entire internet to have an accurate and up-to-date language model.

Can do impressive information retrieval.



$$P(S) = P(\text{Where}) \times P(\text{are} \mid \text{Where}) \times P(\text{we} \mid \text{Where are}) \times P(\text{going} \mid \text{Where are we})$$

Exercises!

Resources

https://cs231n.github.io/convolutional-networks/

https://github.com/Spandan-Madan/DeepLearningProject/blob/master/notebooks/Deep_Learning_Project.ipynb

https://keras.io/examples/

https://blog.statsbot.co/neural-networks-for-beginners-d99f2235efca