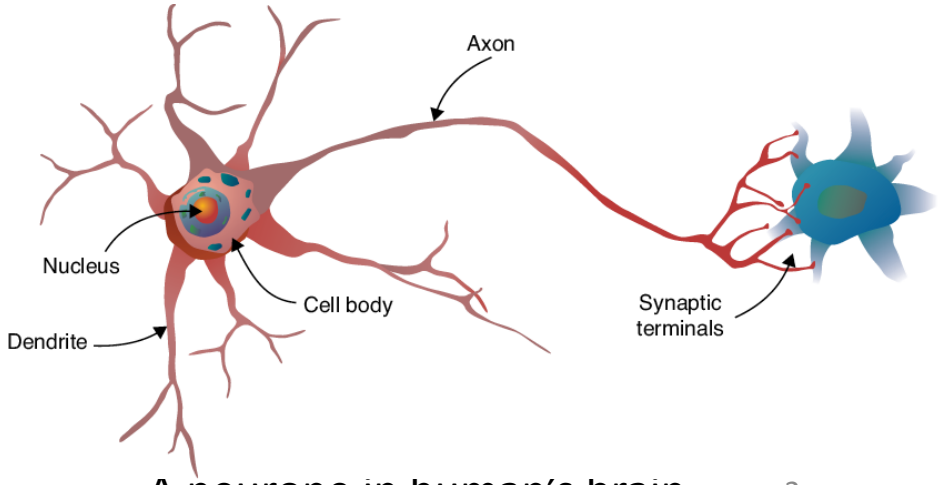
Deep Learning

Richard Lui

Artificial Neural Network (ANN)

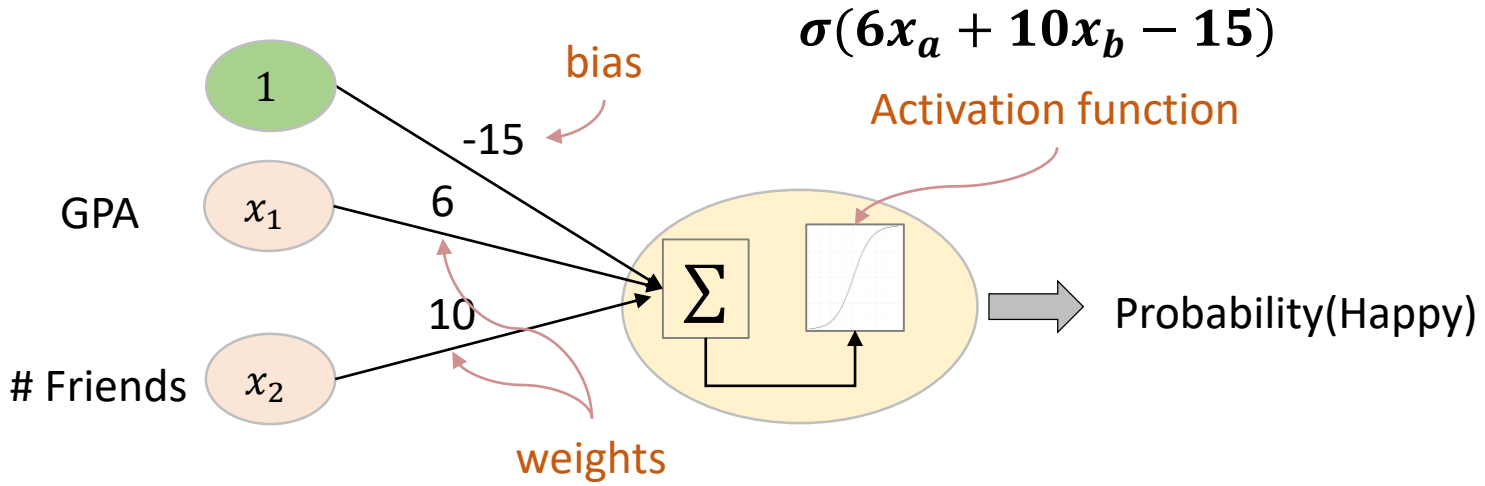
- Inspired by the structure of the human brain, with a network of many cells called "neurons"
- Each neuron acts as a computational unit
  - Accept input from the dendrites
  - Output signal through the axon terminals.
  - Actions are triggered when a specific combination of neurons are activated.



A neurone in human's brain

Perceptron

- A single-layer **perceptron** is the basic unit of a neural network.
- A binary classifier which can decide whether or not an input belongs to some specific class.
- The output of the neuron is called its *activation*.
- We call the function that calculates the artificial neuron's output the *activation function*



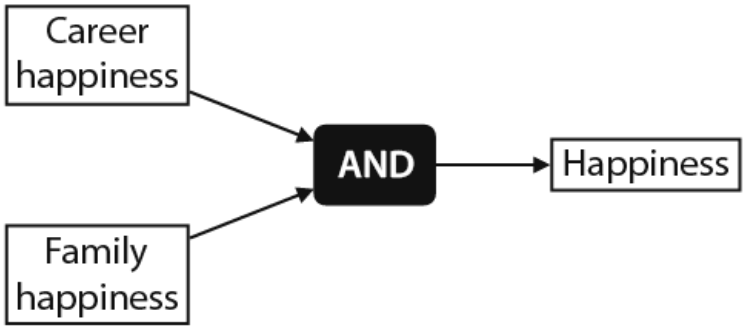
Example

Feature x_1	Feature x_2	Mood
1	0	Sad
2	0	Sad
0	1	Sad
0	2	Sad
1	1	Happy
2	1	Happy
1	2	Happy
2	2	Happy

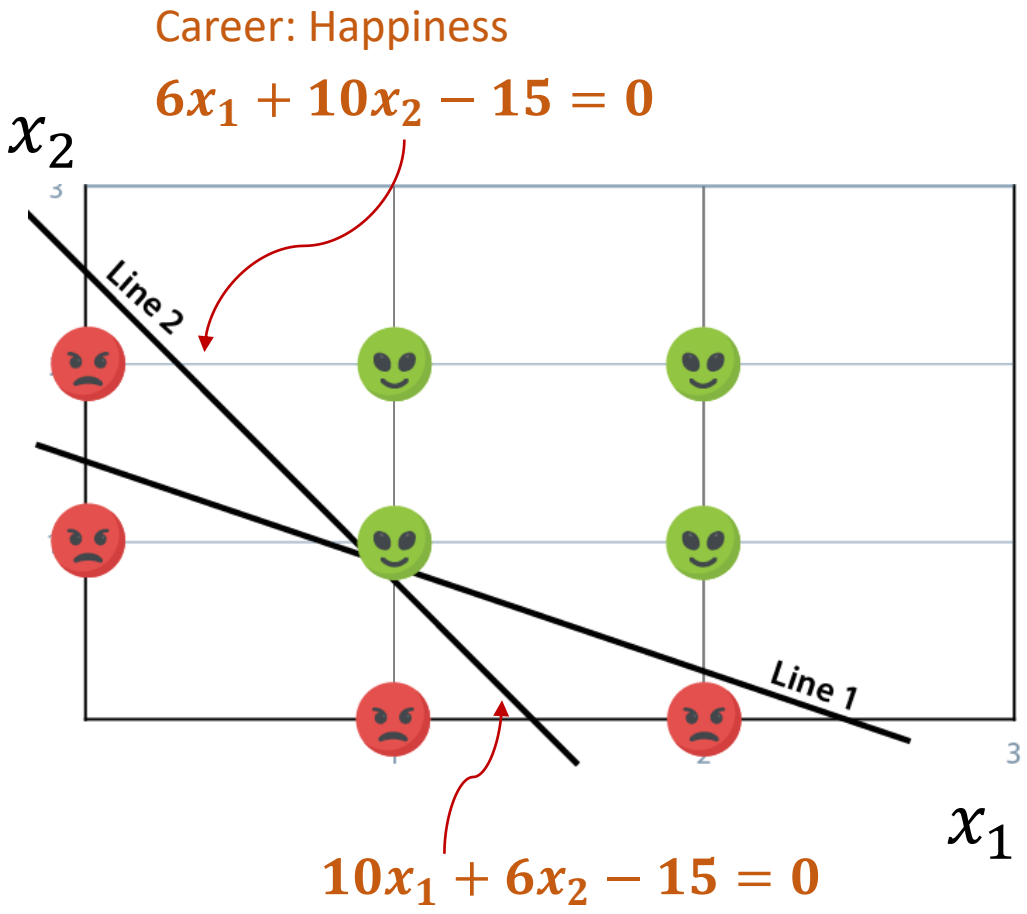
We won't be able to fit a linear classifier to this data!

Using two lines

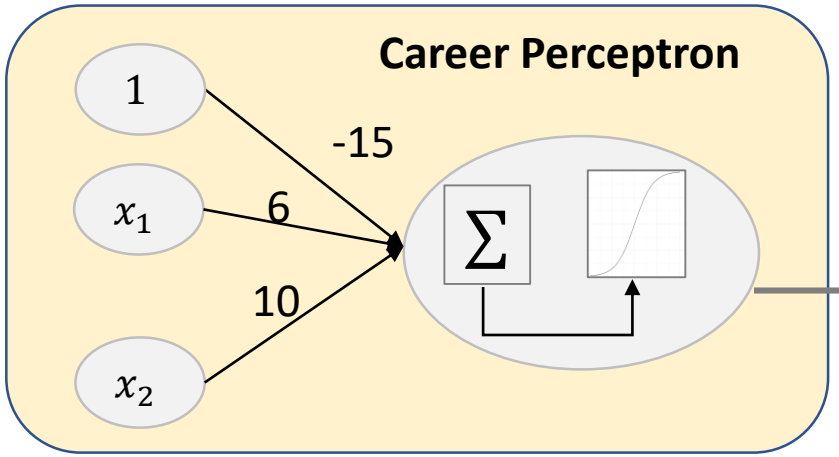
- Drawing two lines separates the data well
- The points above both lines can be classified as happy, and the remaining points as sad.
- Combining linear classifiers this way is the basis for neural networks.



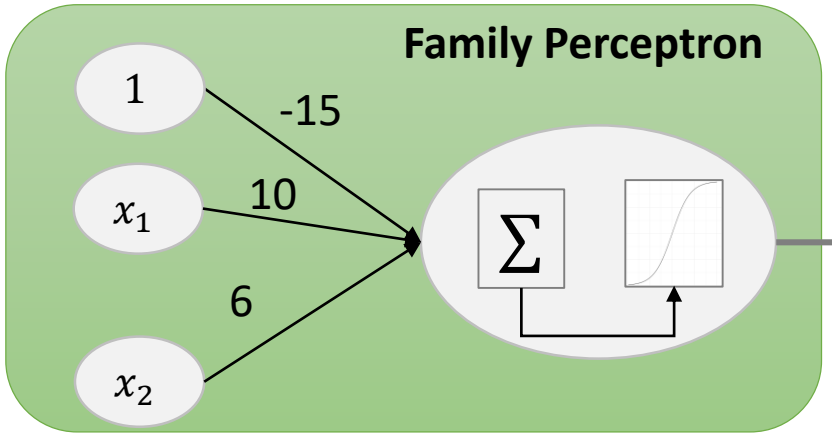
Career happiness **AND** Family happiness ⟶ Happiness



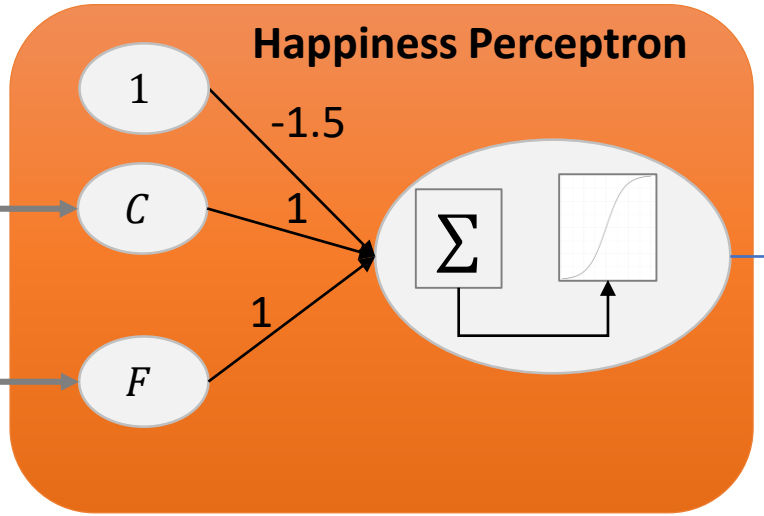
Combining linear classifiers



$$C = \sigma(6x_a + 10x_b - 15)$$

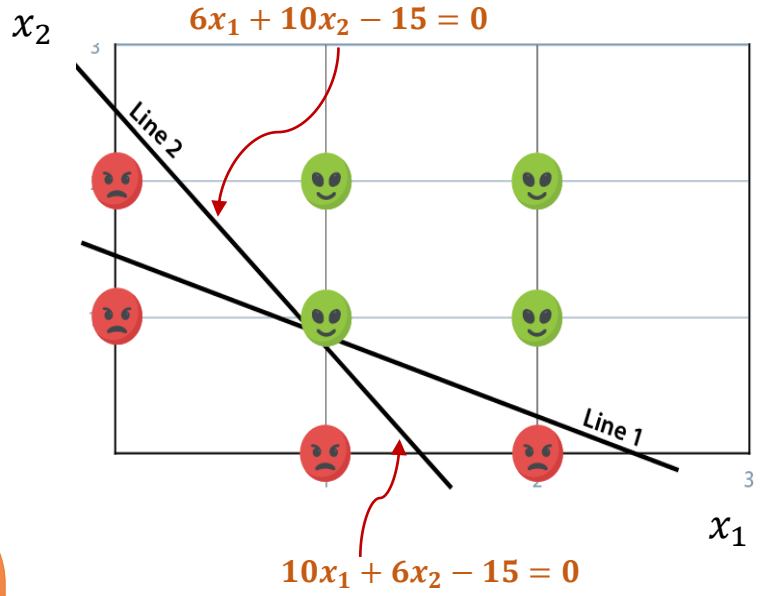


$$F = \sigma(10x_1 + 6x_2 - 15)$$



$$\hat{y} = \sigma(1C + 1F - 1.5)$$

Predicted happiness

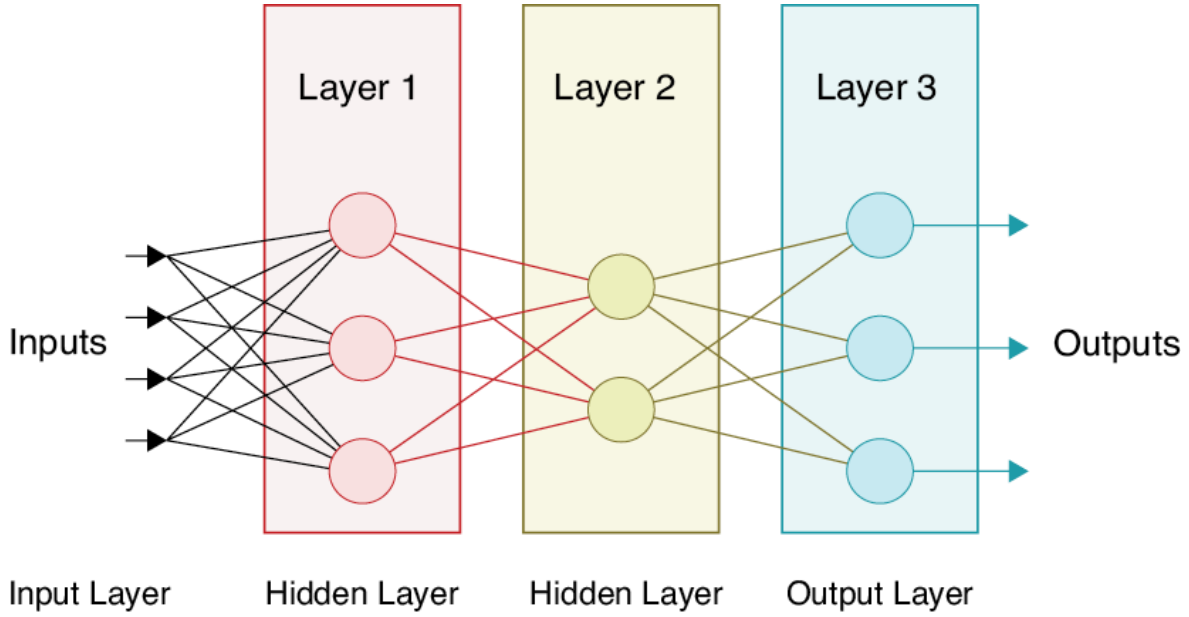


Fully connected feed-forward neural network

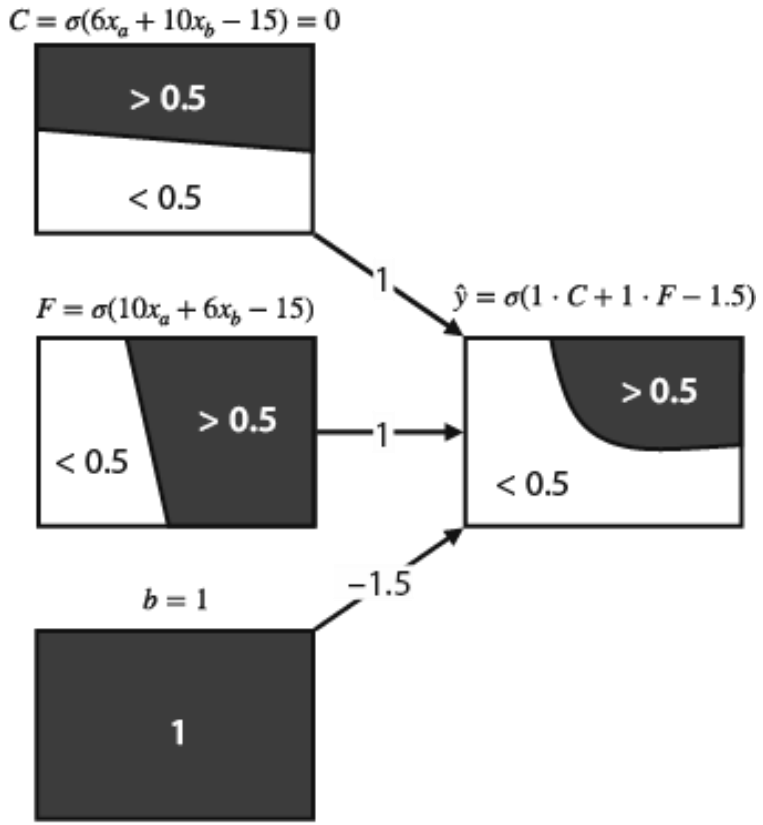
- The arrangement of nodes and layers is called the architecture of the neural network.
  - The first layer is the input layer, the final layer is the output layer
  - All the layers in between are called the hidden layers.
- The number of layers (excluding the input layer) is called the depth of the neural network.

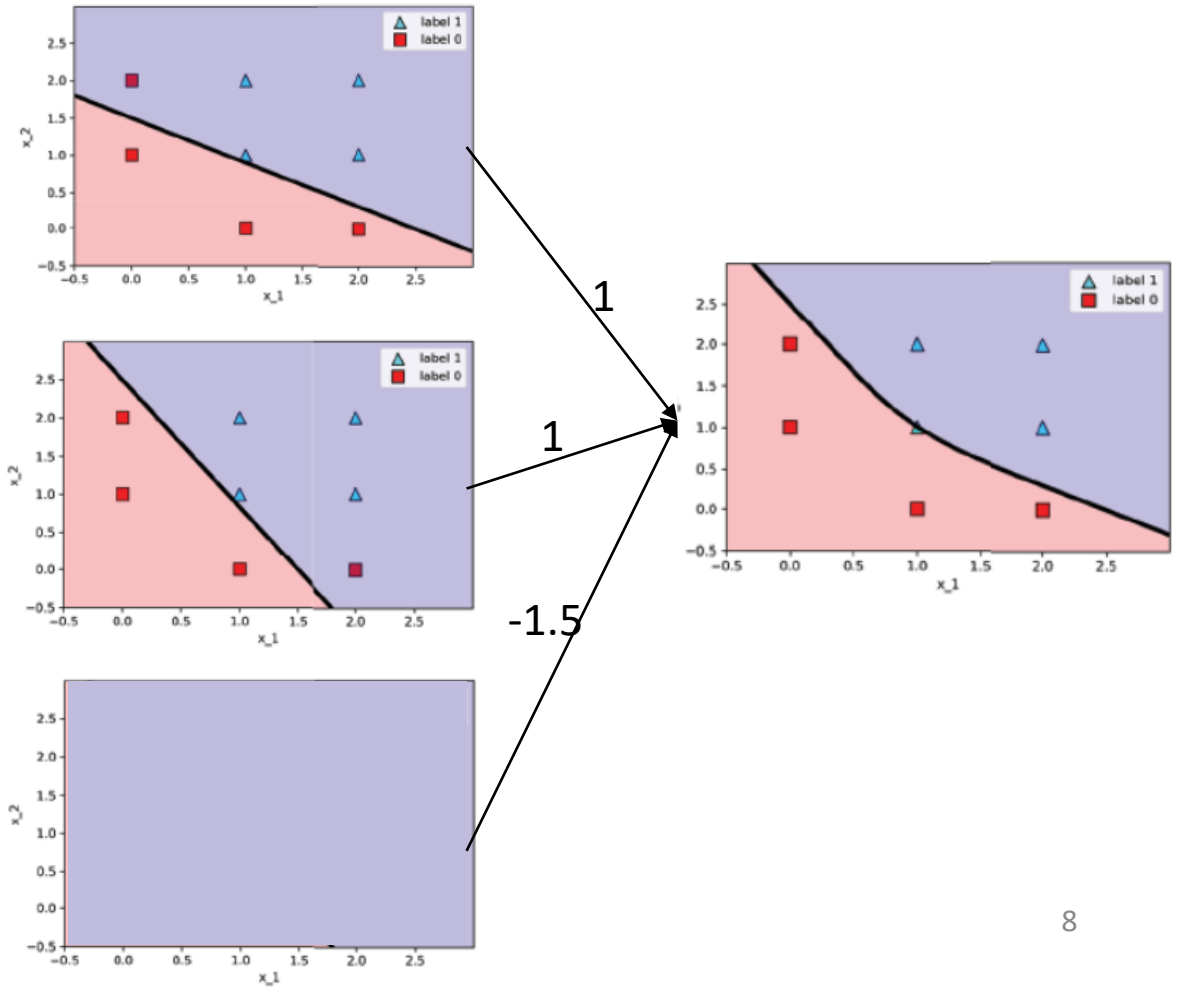
Example: A neural network with depth of 3

- An input layer of size 4
- A hidden layer of size 3
- A hidden layer of size 2
- An output layer of size 3



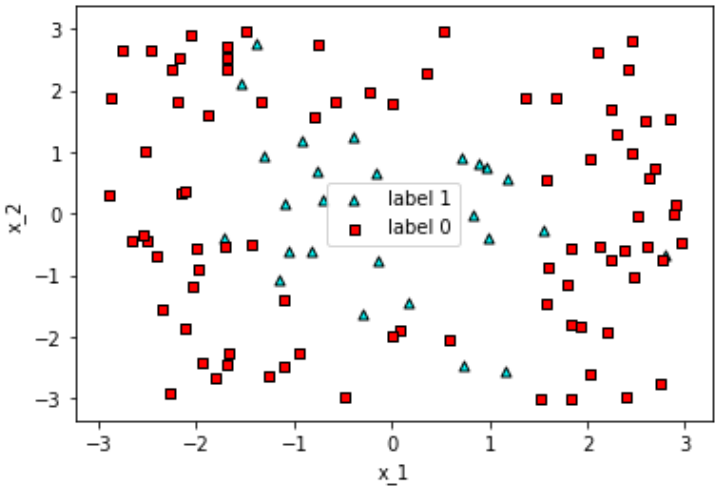
- To build a neural network, we use the outputs of two perceptrons and a bias node (represented by a classifier that always outputs a value of 1) to a third perceptron.
- The boundary of the resulting classifier is a combination of the boundaries of the input classifiers.



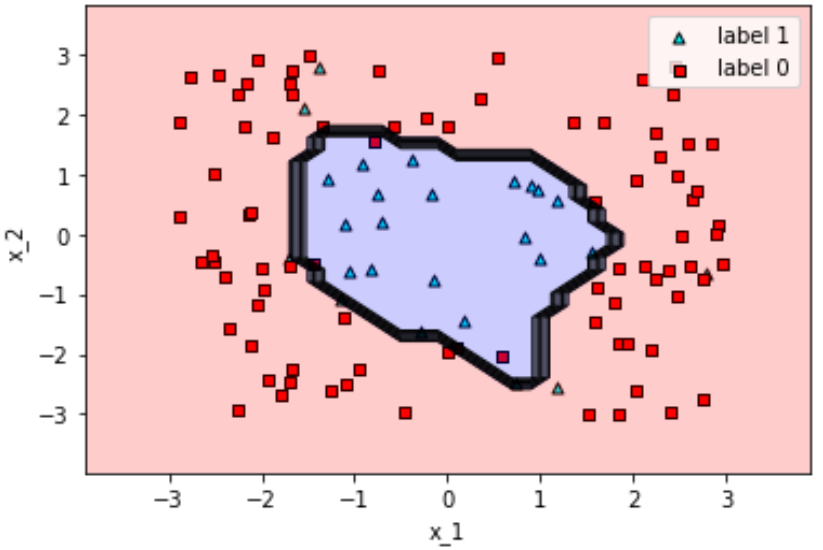


Example

	x_1	x_2	y
0	-0.759416	2.753240	0
1	-1.885278	1.629527	0
2	2.463302	-1.023869	0
3	-1.986004	-0.898810	0
4	2.010834	-2.580117	0
...	...	...	...
105	-1.376637	2.778703	1
106	-0.703722	0.215382	1
107	0.729767	-2.479655	1
108	-1.715920	-0.393404	1
109	2.382873	-2.951074	0



Plotting of the decision boundary after 100 epochs of training



Decision boundary of neural network

- The first hidden layer is formed by linear classifiers
- The classifiers in each successive layer are slightly more complex than those in the previous ones.

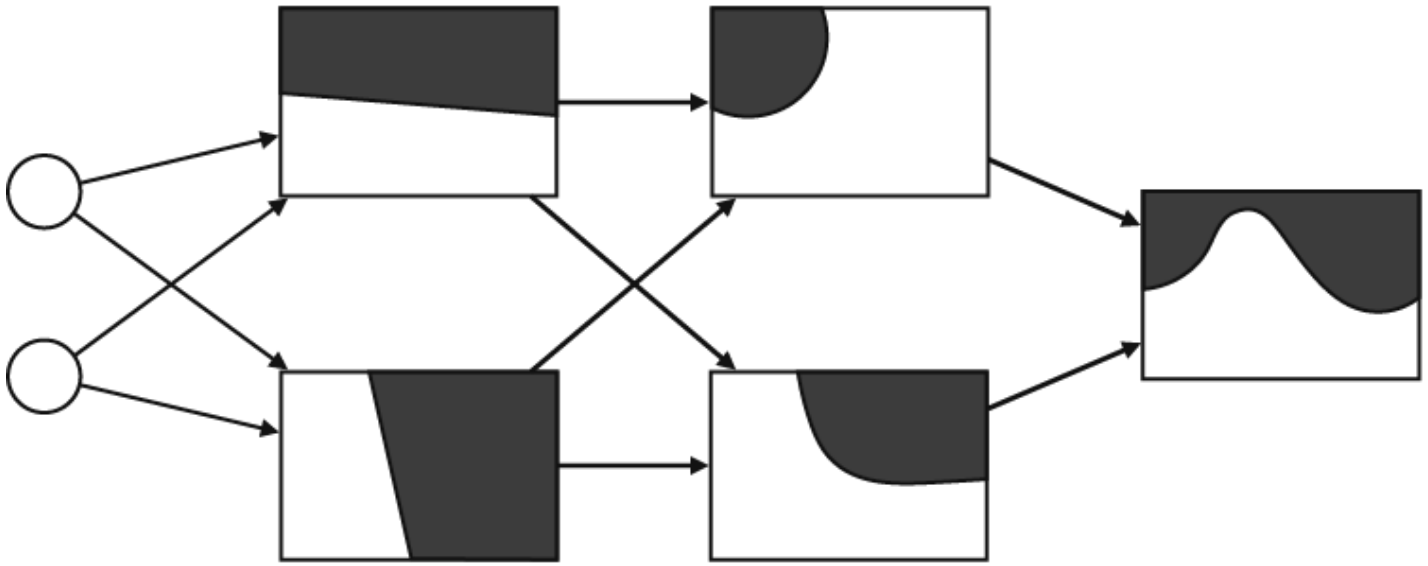
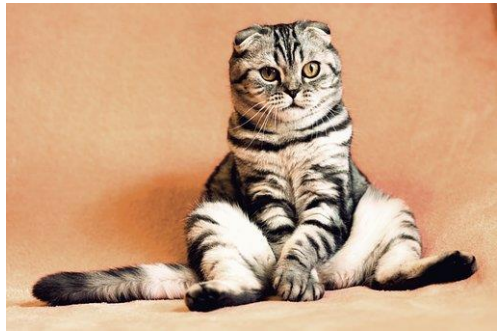
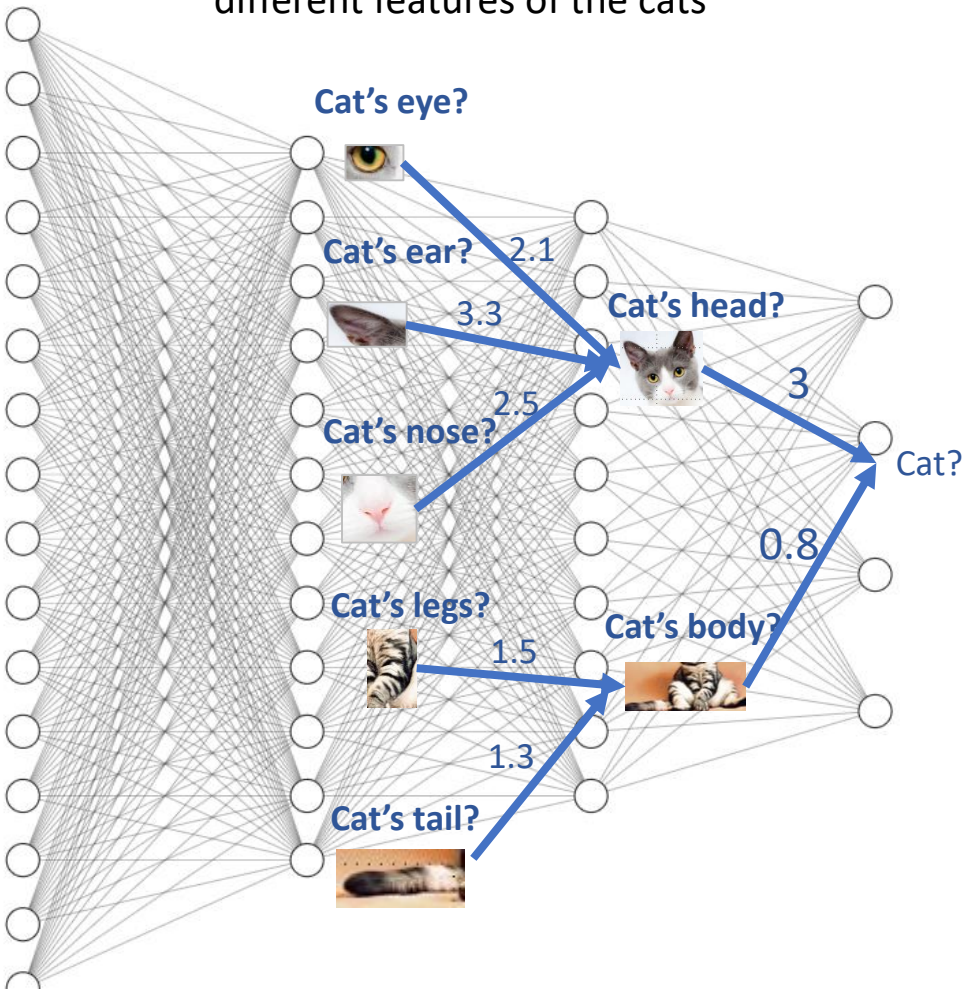


Image classification with ANN



Feature extraction

Hopefully, different neurons may form detectors for different features of the cats



Probability(Dog) = 0.02

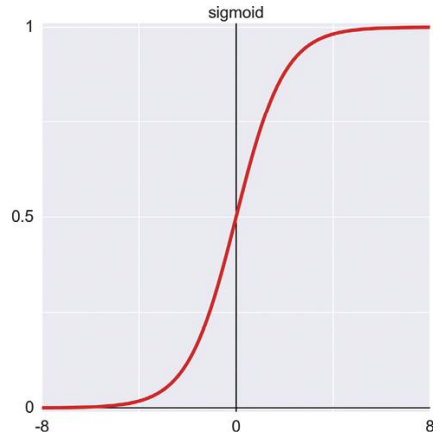
Probability(Cat) = 0.90

Probability(Tiger) = 0.02

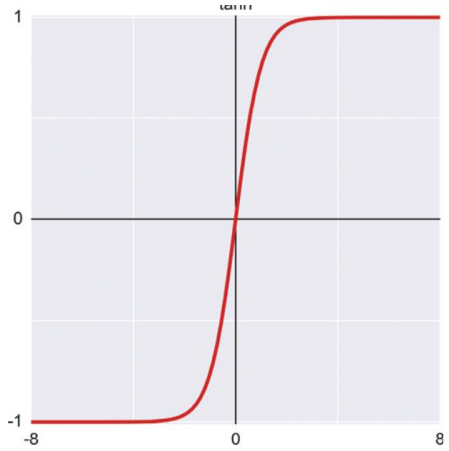
Probability(Human) = 0.01

Activation functions

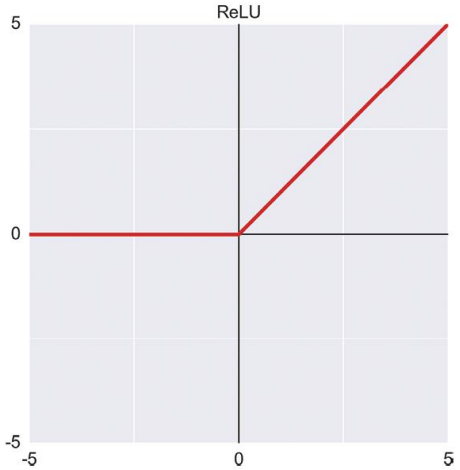
- An *activation function* takes a real number as input and returns a new floating-point number as output
- We can apply a different activation function to every neuron in our network
- In practice, we usually assign the same activation function to all the neurons in each layer.



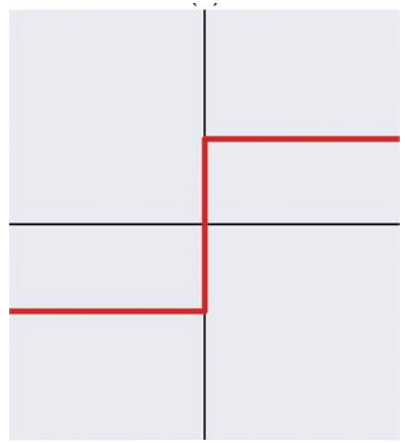
Sigmoid



tanh

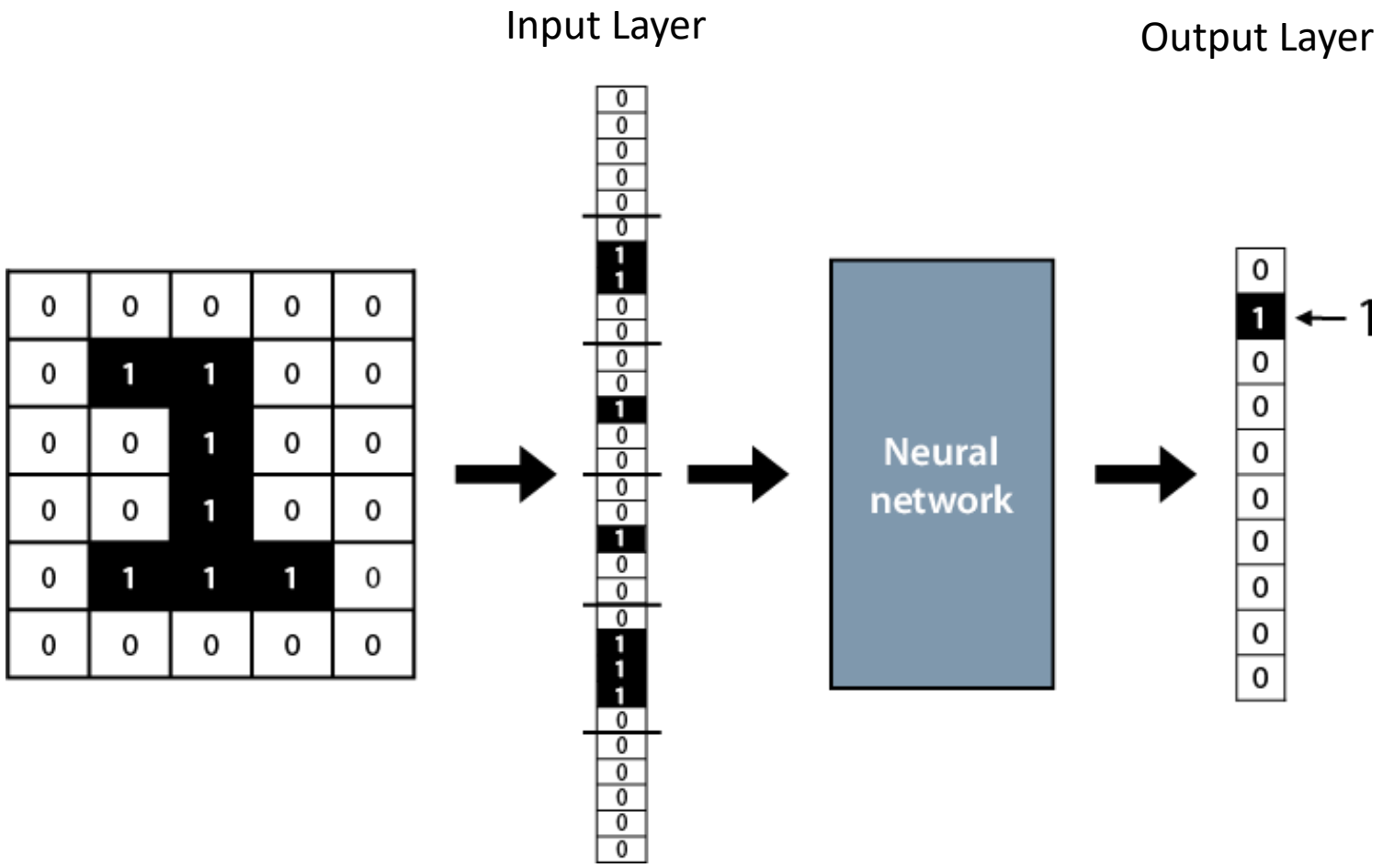


ReLU

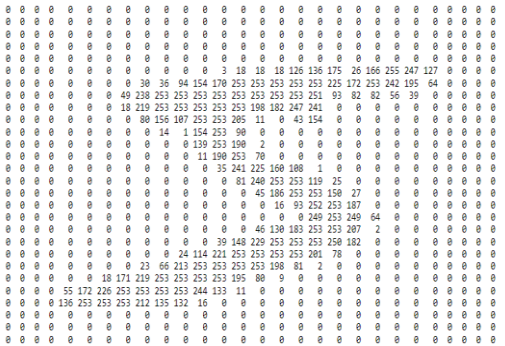


Threshold

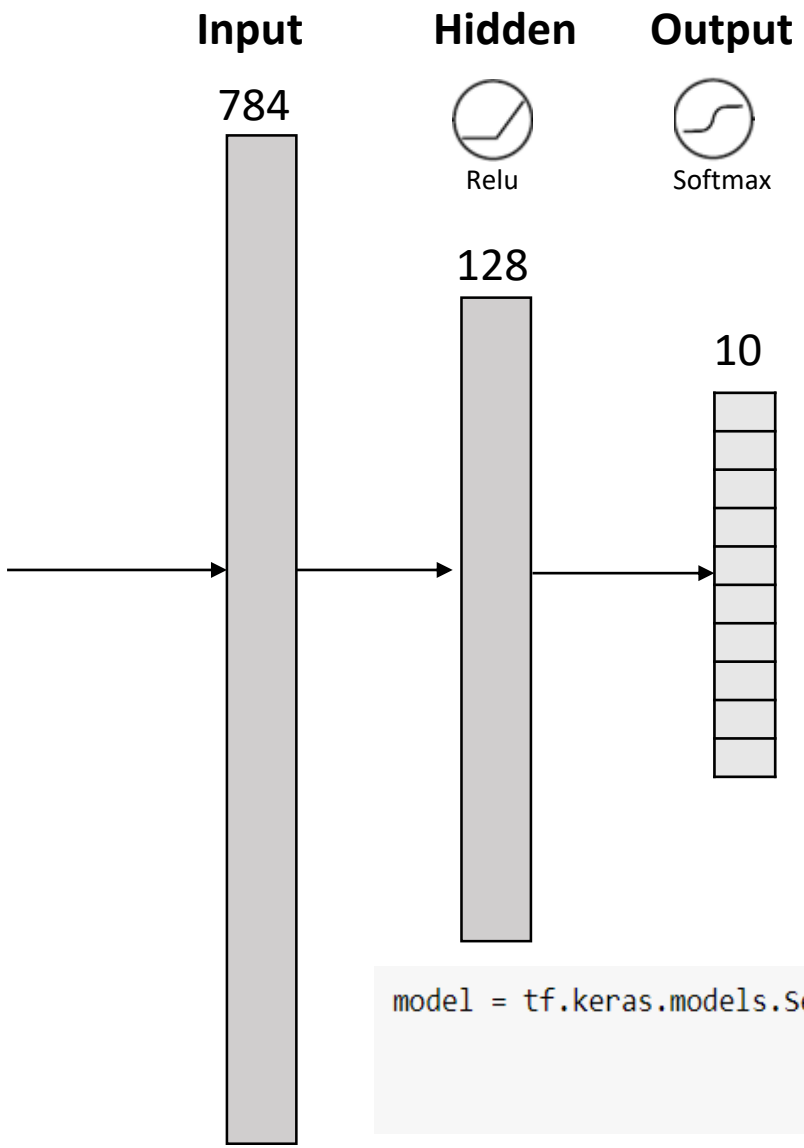
Using neural network for digit Recognition



Example neural network architecture



28*28 grid




```
model = tf.keras.models.Sequential([tf.keras.layers.Flatten(),
                                    tf.keras.layers.Dense(128, activation=tf.nn.relu),
                                    tf.keras.layers.Dense(10, activation=tf.nn.softmax)])
```

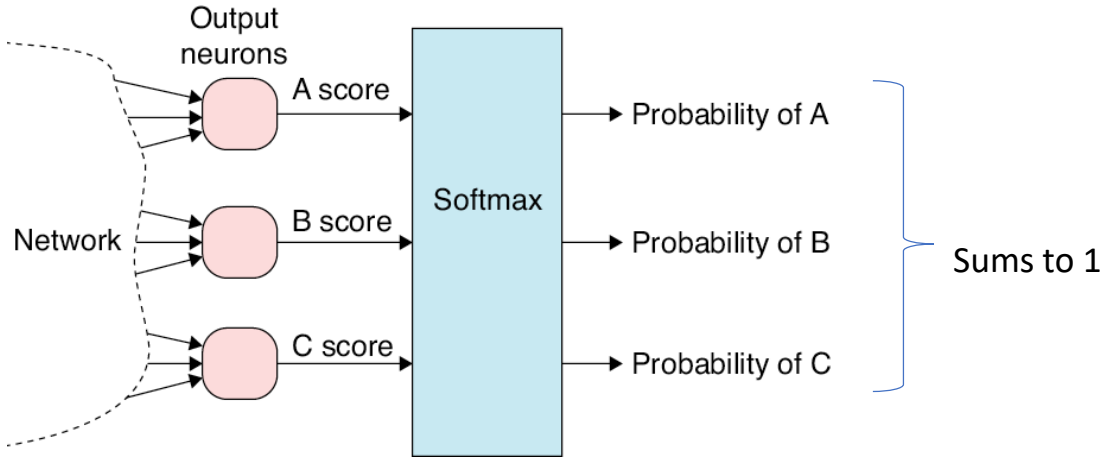
Turning the score into probabilities

Animal	Score	e^x	Probability
Cat	3	$e^3 = 20.085$	$\frac{20.085}{27.842} = \mathbf{0.721}$
Dog	2	$e^2 = 7.389$	$\frac{7.389}{27.842} = 0.265$
Bird	-1	$e^{-1} = 0.368$	$\frac{0.368}{27.842} = 0.013$

$$\boldsymbol{Sum = 27.842}$$

The SoftMax function

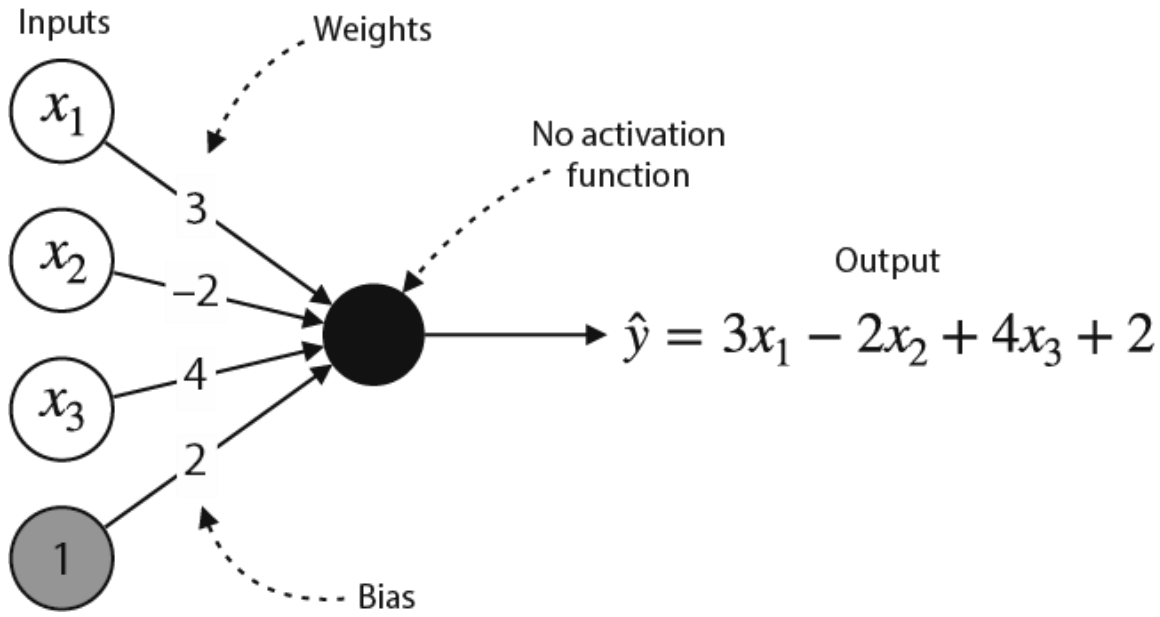
- A common function used in neural networks which turn the raw numbers that come out of a classification network into class probabilities.
- Suppose we have n classifiers that output the scores $a_1, a_2, \ldots, a_n$.
- The probabilities obtained are $p_1, p_2, \ldots, p_n$



$$p_i = \frac{e^{a_i}}{e^{a_1} + e^{a_2} + \cdots + e^{a_n}}$$

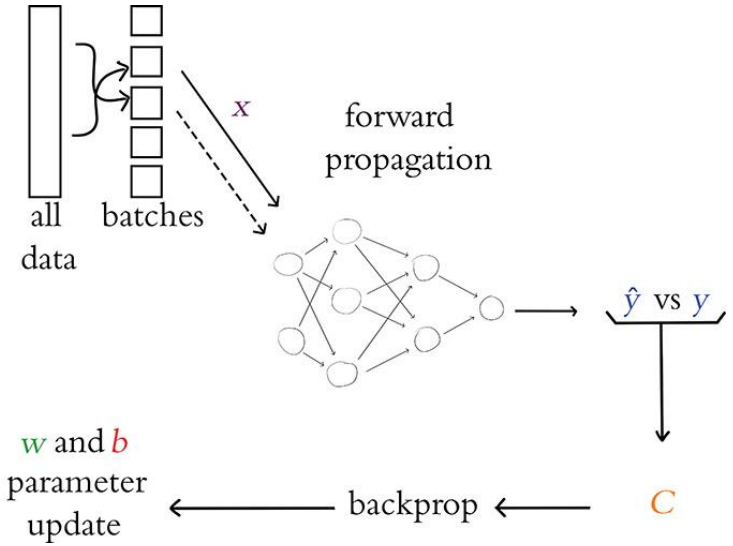
Neural network for regression

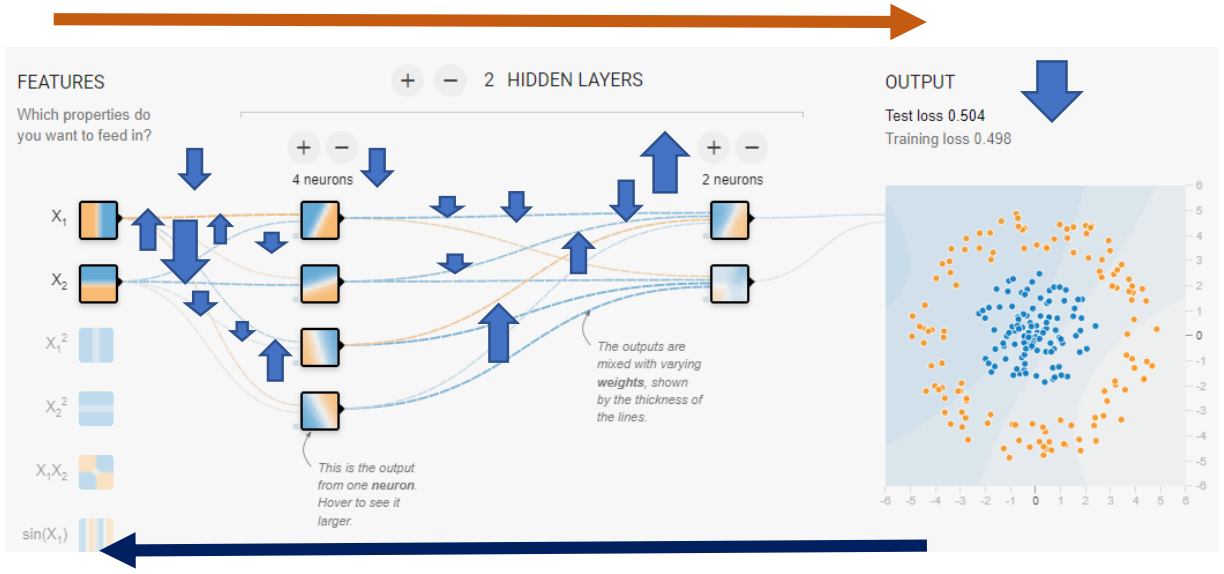
- Remove the final sigmoid function from the neural network
  - The role of this function is to turn the input into a number between 0 and 1
  - If we remove it, the neural network will be able to return any number.



Backpropagation: How the ANN "learns"

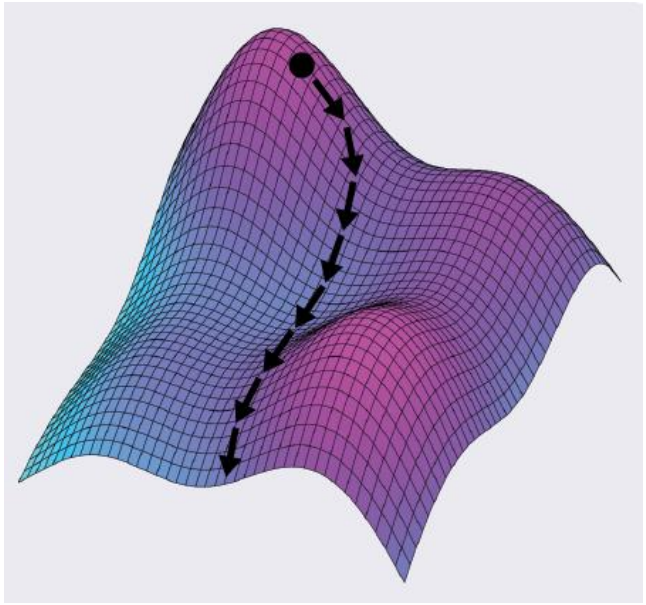
- The neural network is initialized with random weights
- In each epoch, update the weights and bias using gradient descent
  - **Forward Propagation:** Take the data point one by one and perform a forward pass to calculate the prediction
  - **Backward Propagation:** Based on the answer and the prediction error, compute how much we should adjust each weights and biases best in order decrease the errors
- Repeat the above steps again and again …

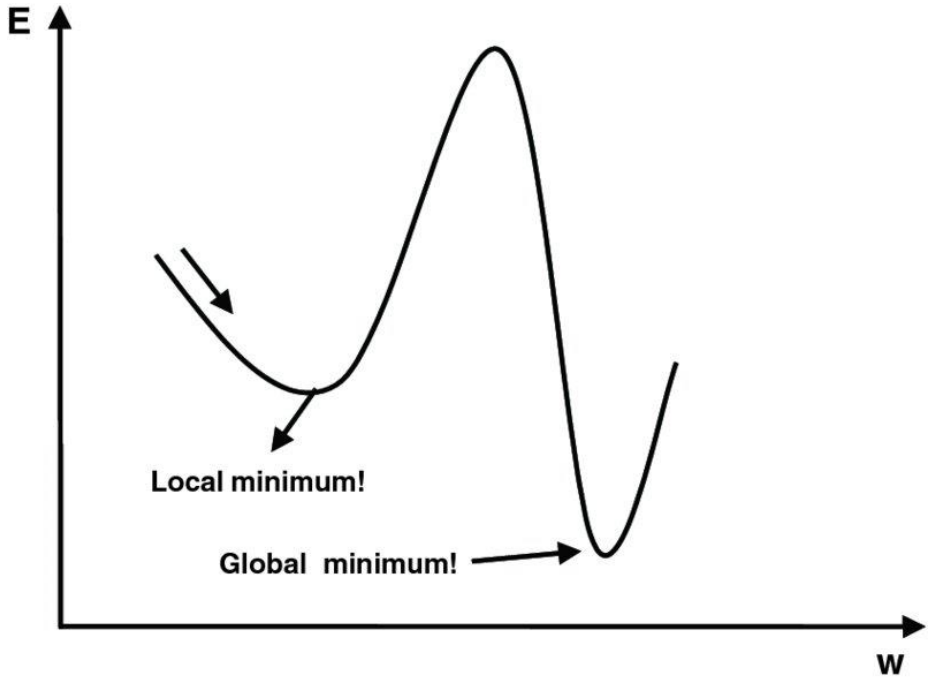




https://playground.tensorflow.org

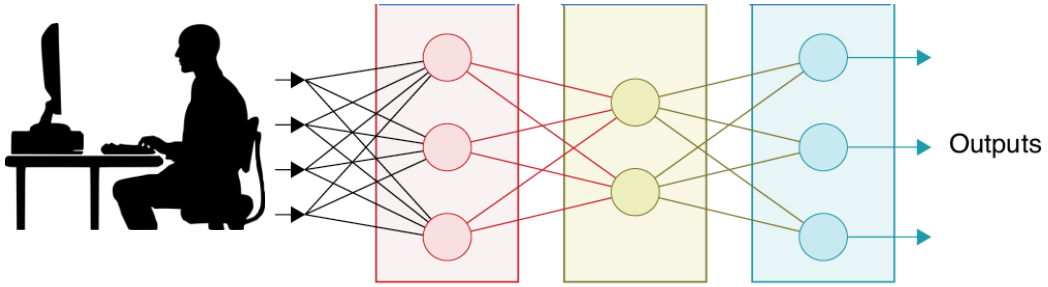
Local versus Global Minima





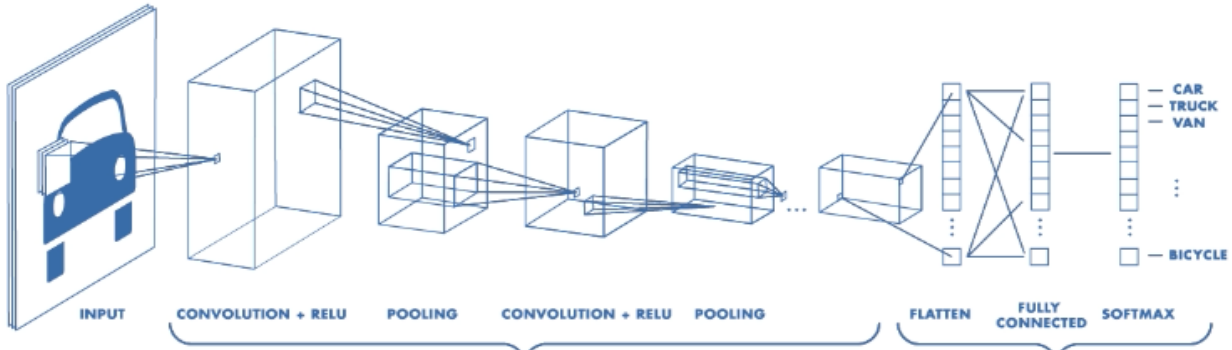
Deep learning and Deep Neural Networks

- "**Deep**" refers to the number of hidden layers in the neural network.
  - A conventional neural network contains 1-3 hidden layers, while deep networks can have as many as 120- 150.
- Target features are learnt from the data without the aid of a human engineer.
  - Can detect features rather than just those recognizable to the human eye.



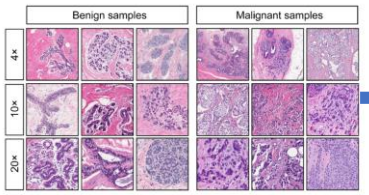
Manual feature extraction

Classification



Feature Extraction

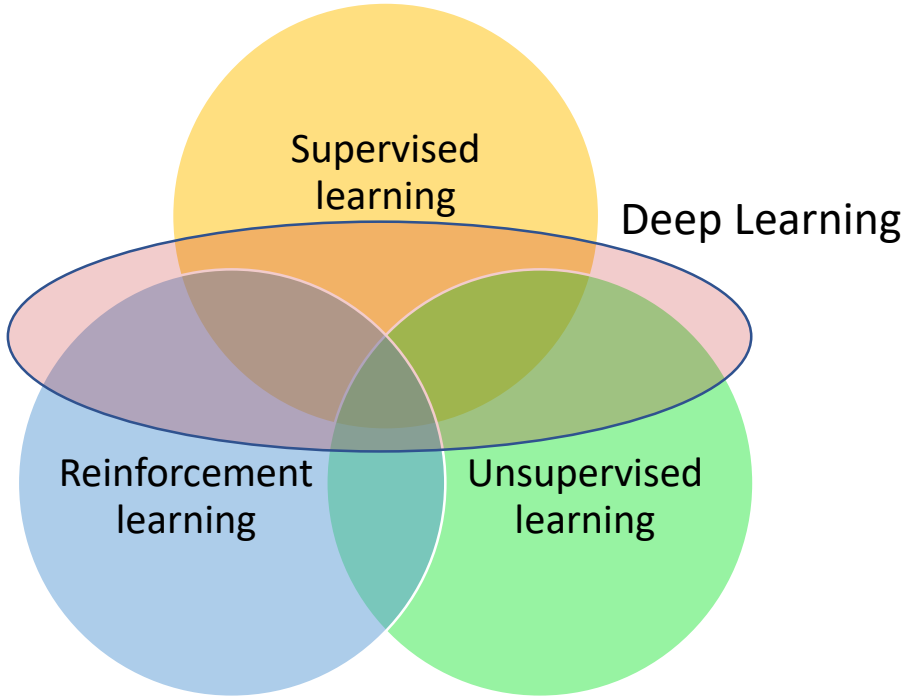
Classification

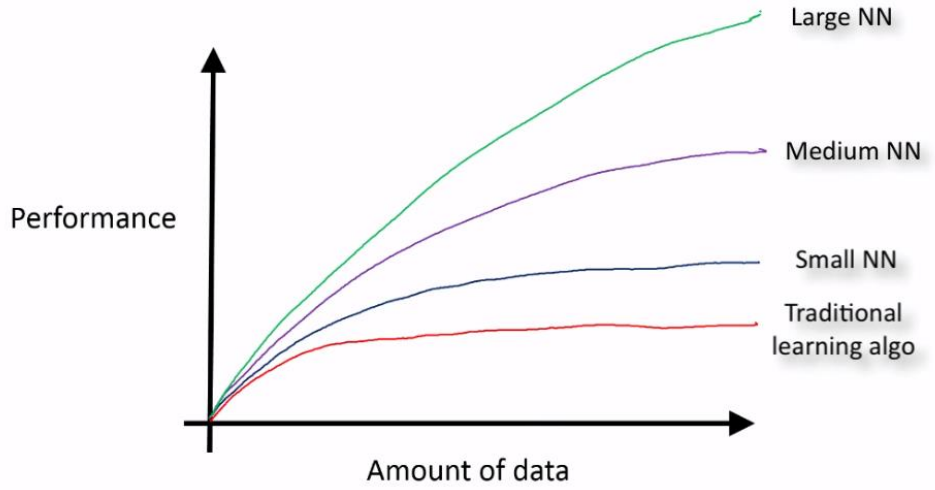


- Radius
- Texture
- Perimeter
- …

Why Deep Learning?

- Shallow learning algorithms are ML algorithms that do not gain in accuracy beyond a certain amount of training data.
- Deep learning is the first class of algorithms that is scalable
  - Results get better with more data + bigger models + more computation

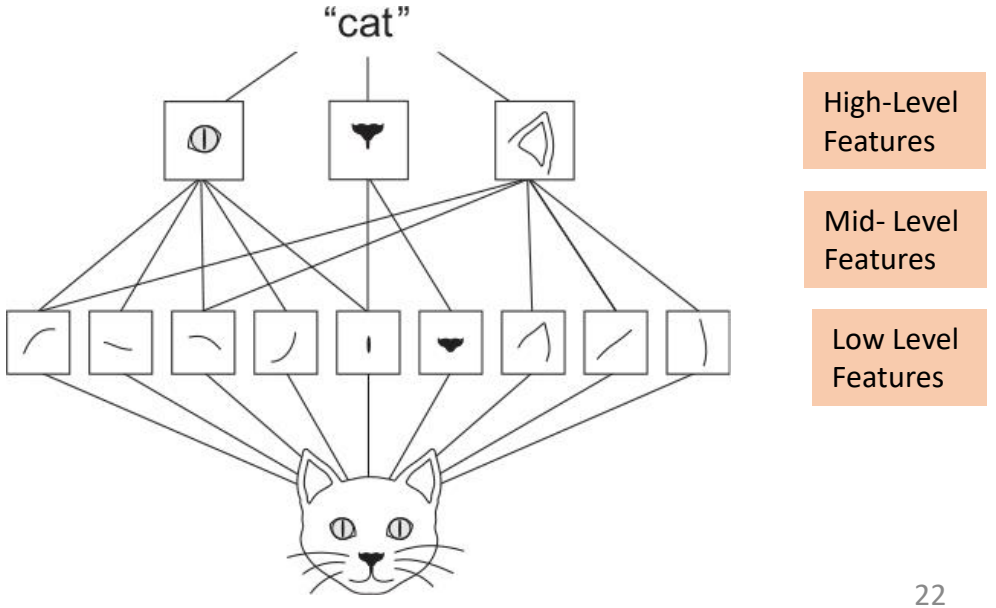




Source: How Scale is Enabling Deep Learning

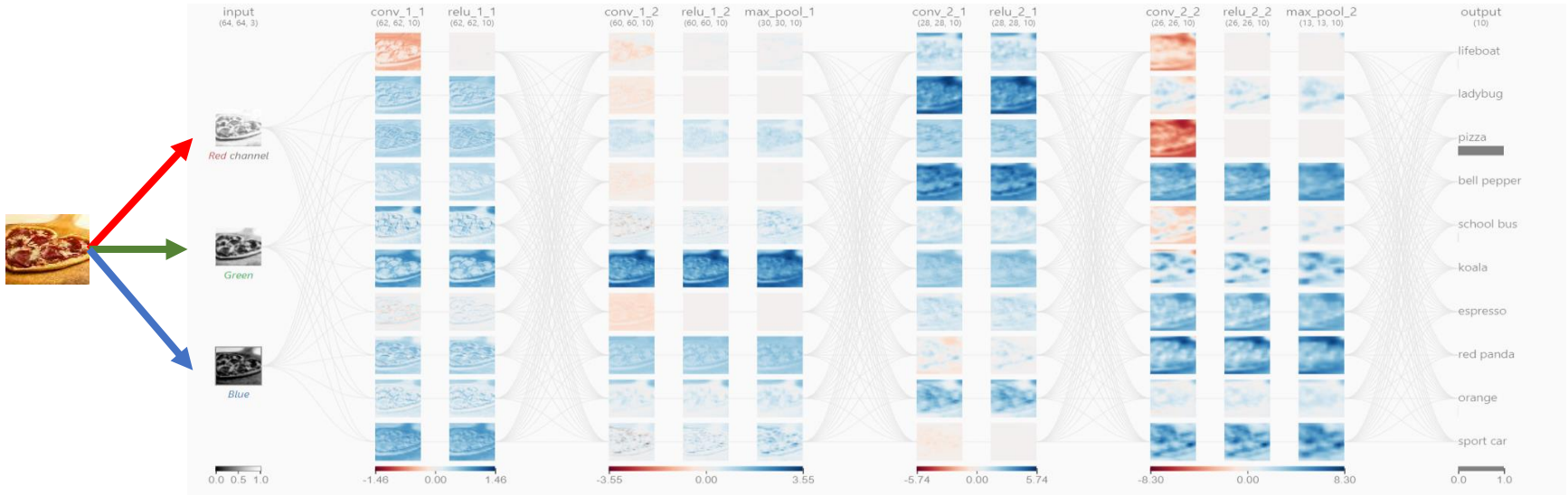
Learning Hierarchical Representations

- Deep learning excels on problem domains where the inputs are not a few quantities in a tabular format, but instead are images of pixel data, documents of text data or files of audio data.
- Multi-stage way to learn data representations.



Convolutional Neural Network (CNN)

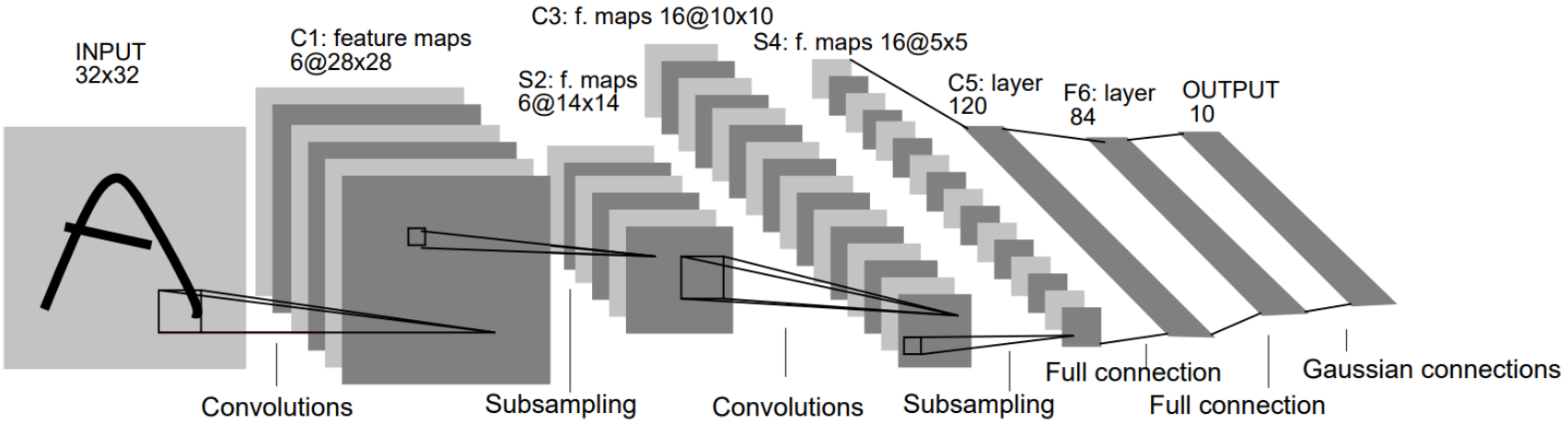
- CNN is a neural network which utilizes a special type of layer (convolutional layers) to learn from image and image-like data.
- A **convolution** is a filter that passes over an image, processes it, and extracts the important features (and blur the inessential features).
- Excels at handling image/image-like data and computer vision tasks

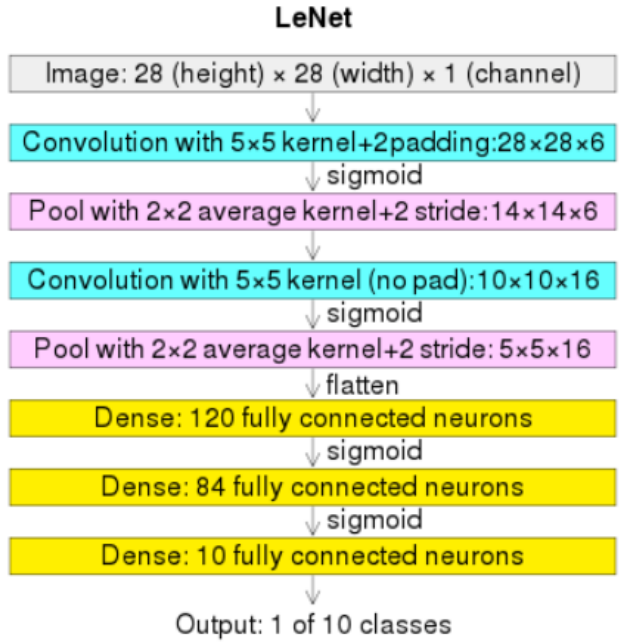


https://poloclub.github.io/cnn-explainer

LeNet

- LeNet-5 was one of the earliest convolutional neural networks and promoted the development of deep learning

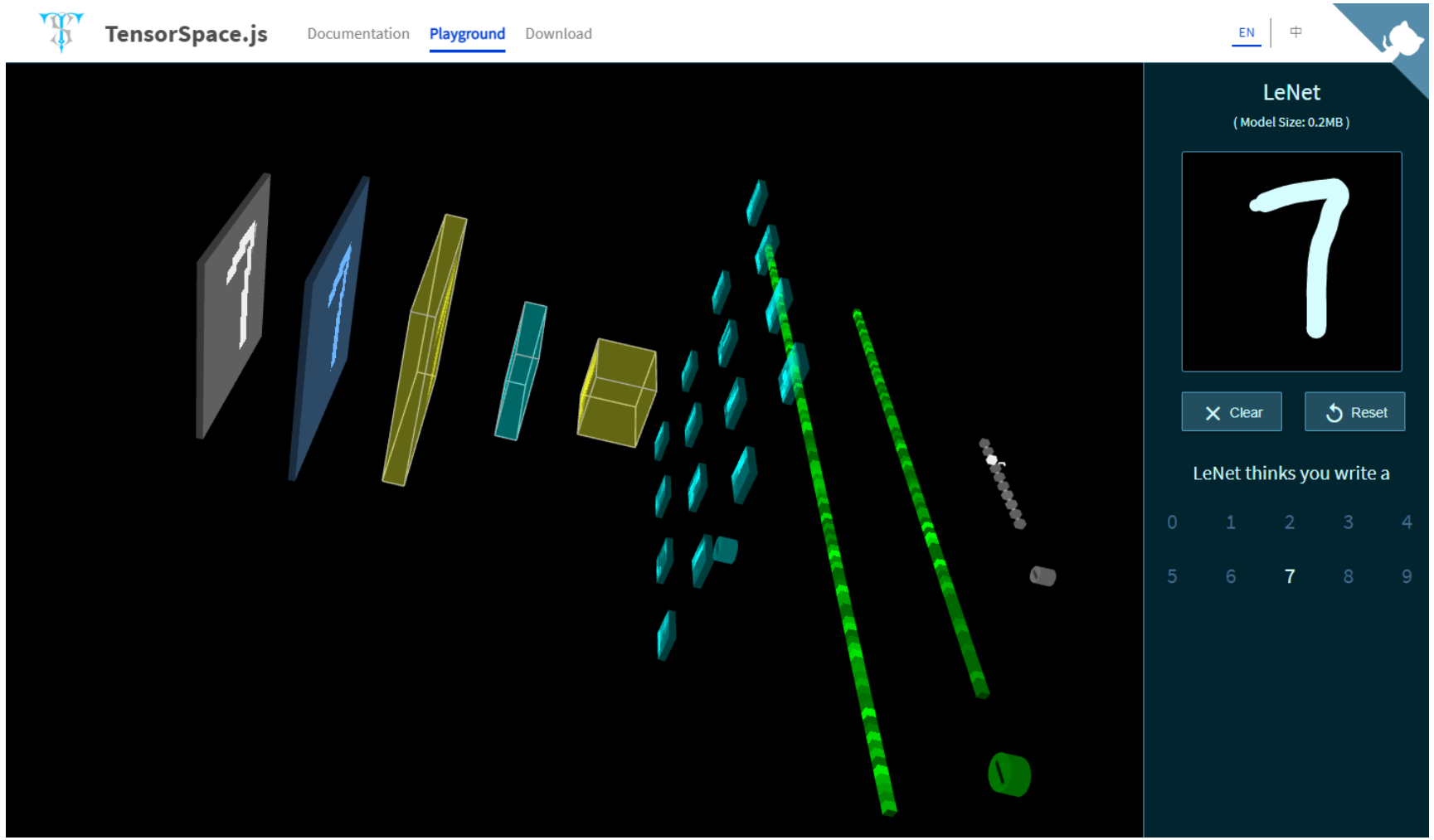




http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf

https://en.wikipedia.org/wiki/LeNet

Visualizing LeNet



https://tensorspace.org

A summary of our LeNet-5-inspired ConvNet architecture

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 26, 26, 32)	320
conv2d_2 (Conv2D)	(None, 24, 24, 64)	18496
max_pooling2d_1 (MaxPooling2	(None, 12, 12, 64)	0
dropout_1 (Dropout)	(None, 12, 12, 64)	0
flatten_1 (Flatten)	(None, 9216)	0
dense_1 (Dense)	(None, 128)	1179776
dropout_2 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 10)	1290

Total params: 1,199,882
Trainable params: 1,199,882
Non-trainable params: 0

ResNet

- ResNet is one of the most powerful CNN winning the ImageNet challenge in 2015
- Variants of ResNet architecture t with a different number of layers.
  - ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-110, ResNet-152, ResNet-164, ResNet-1202
- Visualization
  - https://tensorspace.org/html/playground/resnet50.html

State-of-the-art image classification models

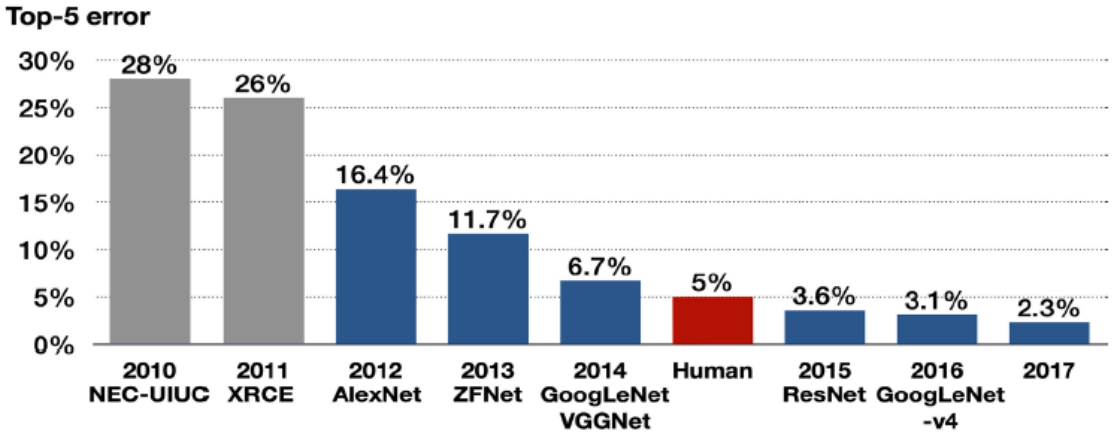


Fig. 2. Algorithms that won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2010-2017. The top-5 error refers to the probability that all top-5 classifications proposed by the algorithm for the image are wrong. The algorithms with blue graph are convolutional neural network. Although VGGNet took second place in 2014, it is widely used in studies as its concise structure. Adapted from "A fully-automated deep learning pipeline for cervical cancer classification" by Alyafeai Z., Ghouti L., Expert Systems with Applications Proceedings of the IEEE 2019;141;112951. Copyright 2019 by Elsevier Ltd.

Source: Application of Deep Learning in Dentistry and Implantology

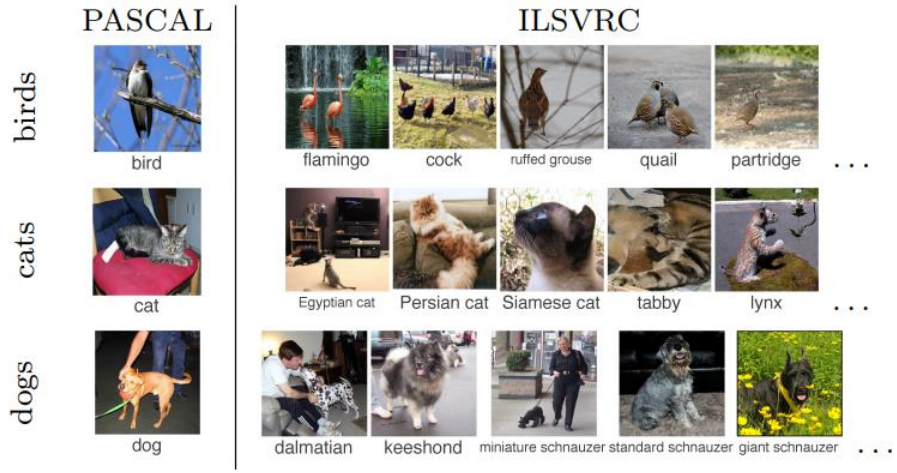


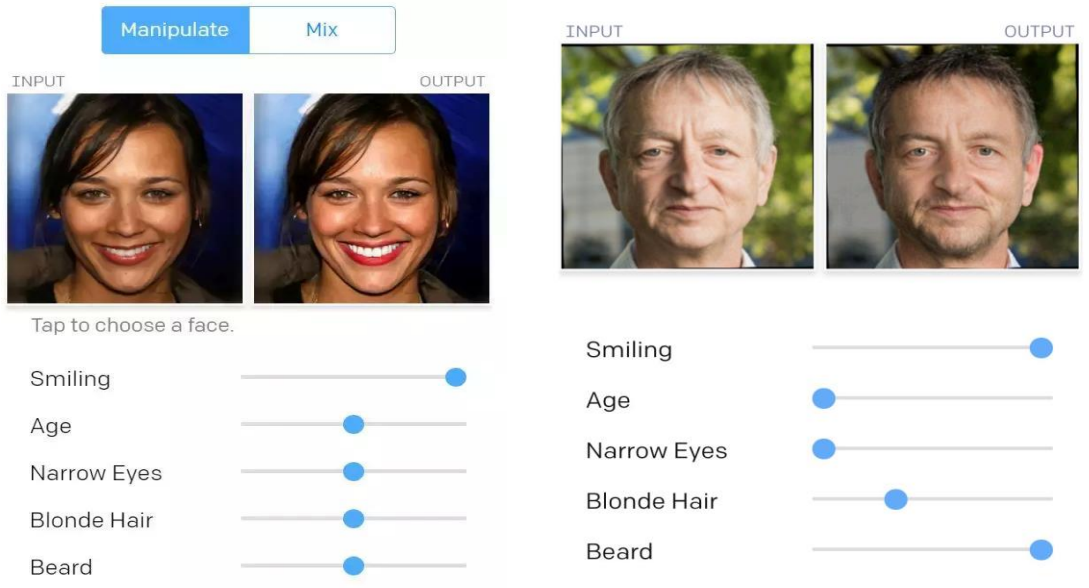
Fig. 2 The ILSVRC dataset contains many more fine-grained classes compared to the standard PASCAL VOC benchmark; for example, instead of the PASCAL "dog" category there are 120 different breeds of dogs in ILSVRC2012-2014 classification and single-object localization tasks.

Source: ImageNet Large Scale Visual Recognition Challenge

When AI becomes “creative”

- By using deep learning, we may train a model to create new data instances that resemble your training data
- Create images that look like photographs of human faces, even though the faces don't belong to any real person.

https://this-person-does-not-exist.com



https://openai.com/blog/glow/

Deep Fake

- The use of artificial intelligence (AI) to create a fake event
  - in photo, video, or audio format.

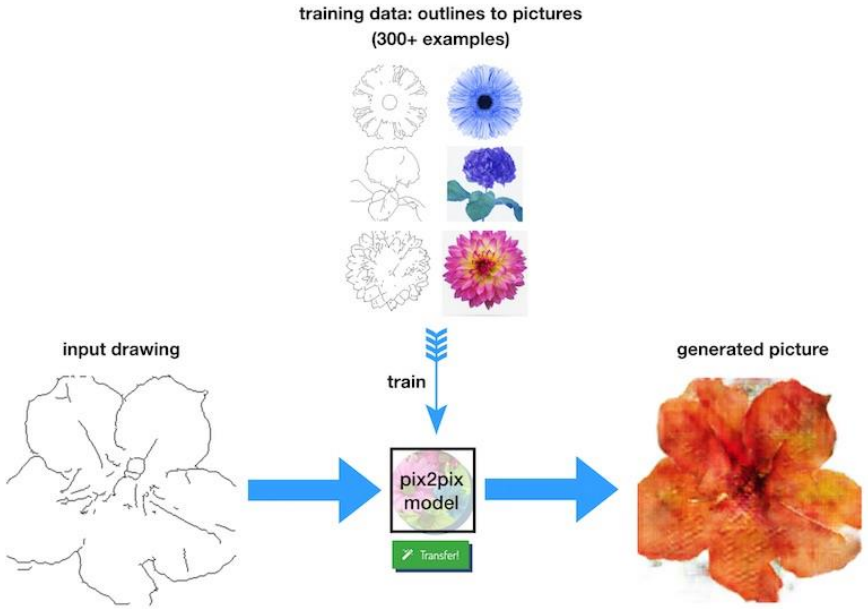
https://www.youtube.com/watch?v=gLoI9hAX9dw

Cartoon GAN

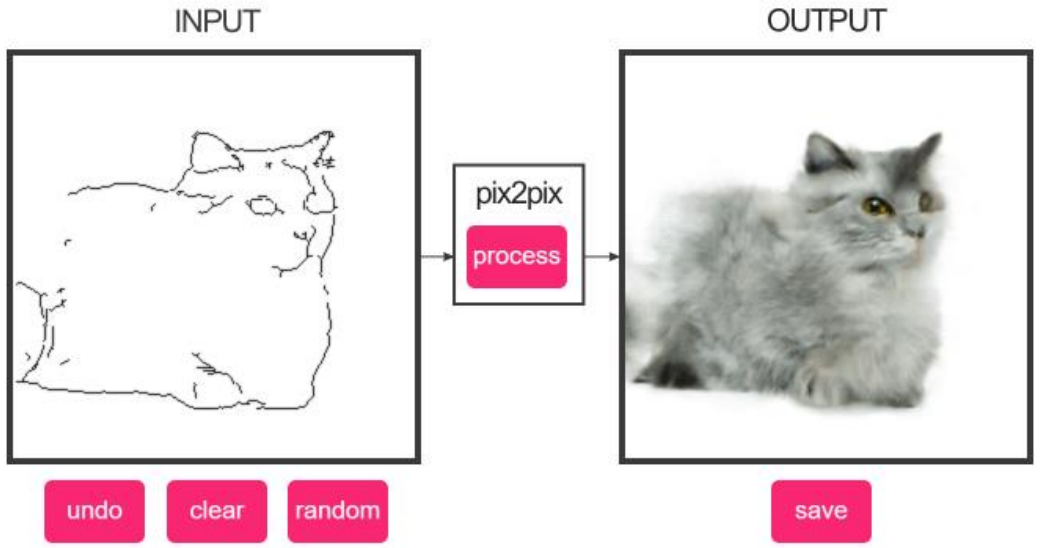
https://github.com/Yijunmaverick/CartoonGAN-Test-Pytorch-Torch

Pix2pix

- Training on pairs of images and then attempts to generate the corresponding output image from any input image you give it



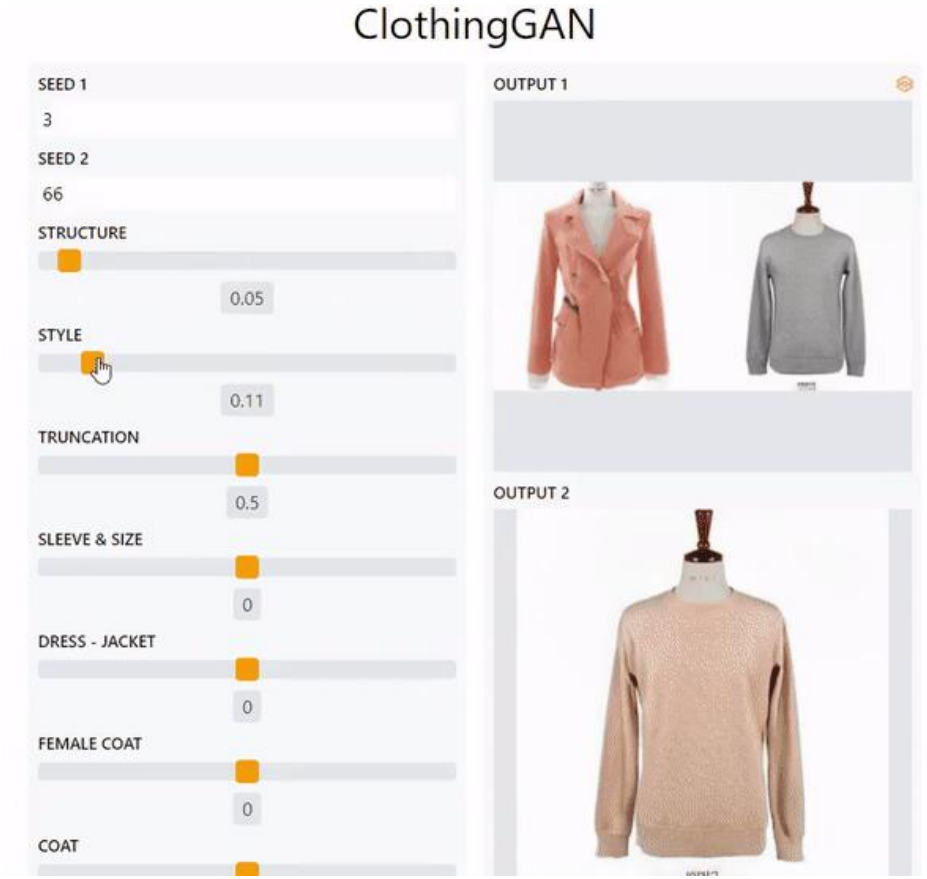
https://mitmedialab.github.io/GAN-play



https://affinelayer.com/pixsrv/

ClothingGAN

- Able to generate clothing images and mix these images.
- While mixing, you can control which structure or style that you want the clothing to copy.
- Edit the generated clothing with several given attributes such as dark color, jacket, dress, or coat.



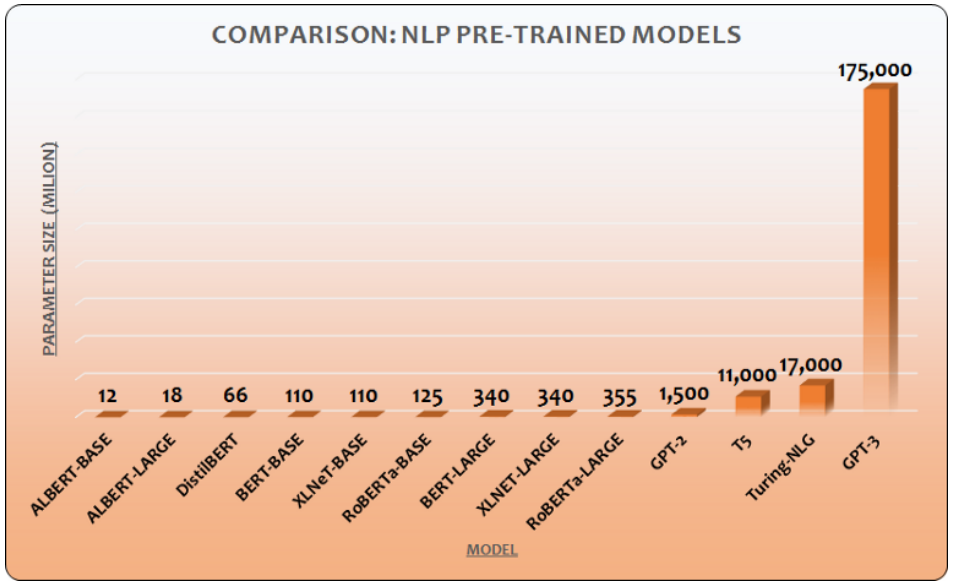
https://github.com/mfrashad/ClothingGAN

GPT-3

- A neural network machine learning model Developed by OpenAI
- GPT-3's full version has a capacity of **175 billion** machine learning parameters.
- Training GPT-3 reportedly cost $**12 Million** for a single training run

Last week, OpenAI published a paper detailing GPT-3, a machine learning model that achieves strong results on a number of natural language benchmarks. At 175 billion parameters, where a parameter affects data's prominence in an overall prediction, it's the largest of its kind. And with a memory size exceeding 350GB, it's one of the priciest, costing an estimated $12 million to train.

A system with over 350GB of memory and $12 million in compute credits isn't hard to swing for OpenAI, a well-capitalized company that teamed up with Microsoft to develop an AI supercomputer. But it's potentially beyond the reach of AI startups like Agolo, which in some cases lack the capital required. Fortunately for them, experts believe that while GPT-3 and similarly large systems are impressive with respect to their performance, they don't move the ball forward on the research side of the equation. Rather, they're prestige projects that simply demonstrate the scalability of existing techniques.



https://venturebeat.com/ai/ai-machine-learning-openai-gpt-3-size-isnt-everything

An GTP-3 generated news article

- **Generative model** learns from the data fed into it then generate new data like what it has receive

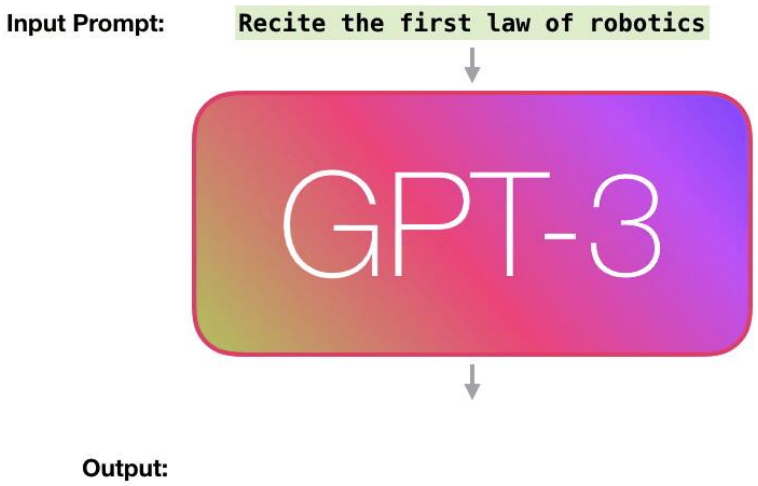
The GPT-3 generated news article that humans had the greatest difficulty distinguishing from a human written article (accuracy: 12%).

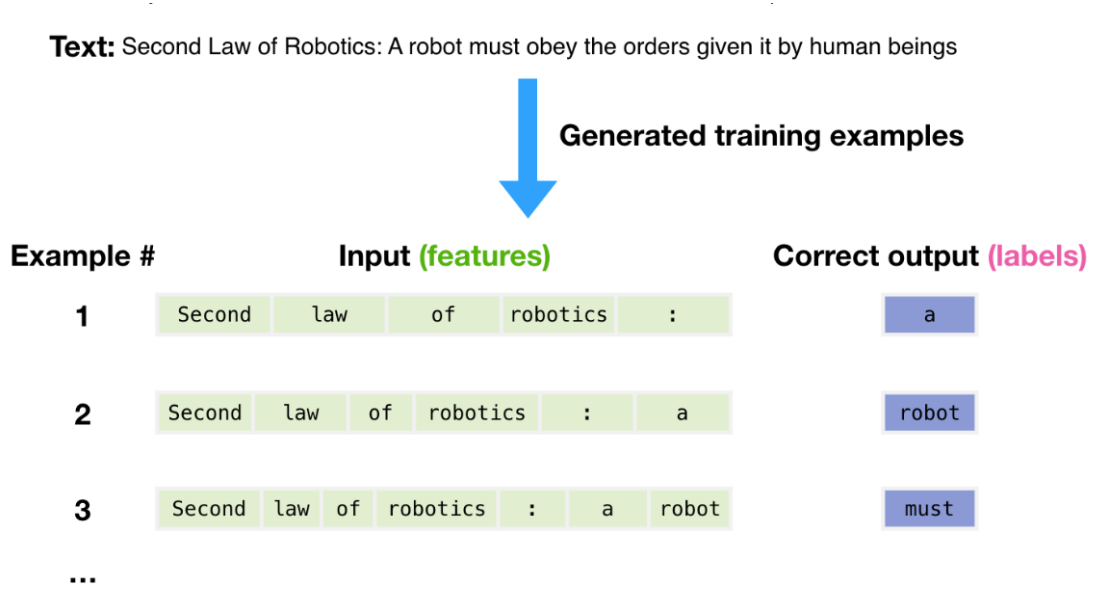
Title: United Methodists Agree to Historic Split
Subtitle: Those who oppose gay marriage will form their own denomination
Article: **After two days of intense debate, the United Methodist Church has agreed to a historic split - one that is expected to end in the creation of a new denomination, one that will be "theologically and socially conservative," according to The Washington Post. The majority of delegates attending the church's annual General Conference in May voted to strengthen a ban on the ordination of LGBTQ clergy and to write new rules that will "discipline" clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.**
The Post notes that the denomination, which claims 12.5 million members, was in the early 20th century the "largest Protestant denomination in the U.S.," but that it has been shrinking in recent decades. The new split will be the second in the church's history. The first occurred in 1968, when roughly 10 percent of the denomination left to form the Evangelical United Brethren Church. The Post notes that the proposed split "comes at a critical time for the church, which has been losing members for years," which has been "pushed toward the brink of a schism over the role of LGBTQ people in the church." Gay marriage is not the only issue that has divided the church. In 2016, the denomination was split over ordination of transgender clergy, with the North Pacific regional conference voting to ban them from serving as clergy, and the South Pacific regional conference voting to allow them.

Source: Language Models are Few-Shot Learners

How GTP3 Works?

- **Language Models** learn the probability of word occurrence based on examples of text.
- The **prompt** is text that you input to the model, and the model will respond with a text completion that attempts to match whatever context or pattern you give it.

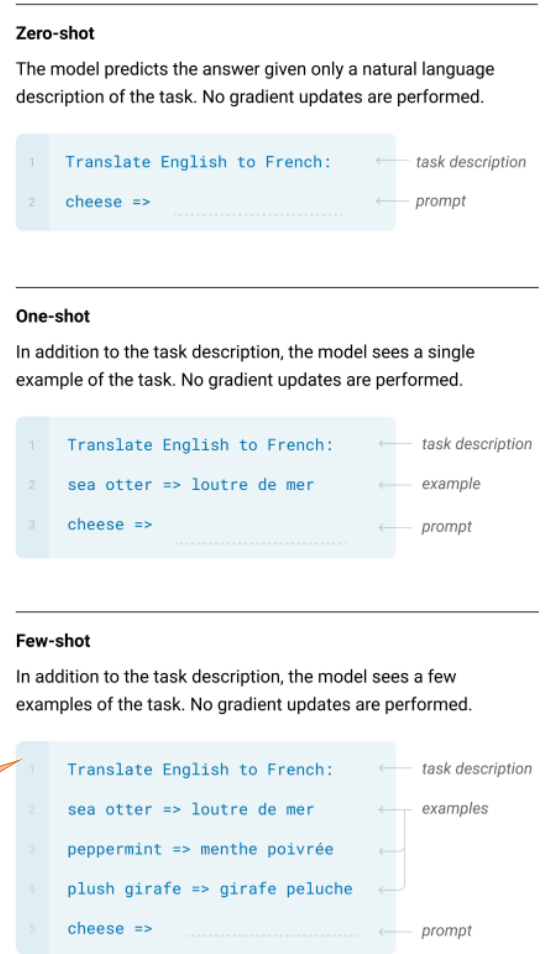




https://jalammar.github.io/how-gpt3-works-visualizations-animations

GTP3 applications

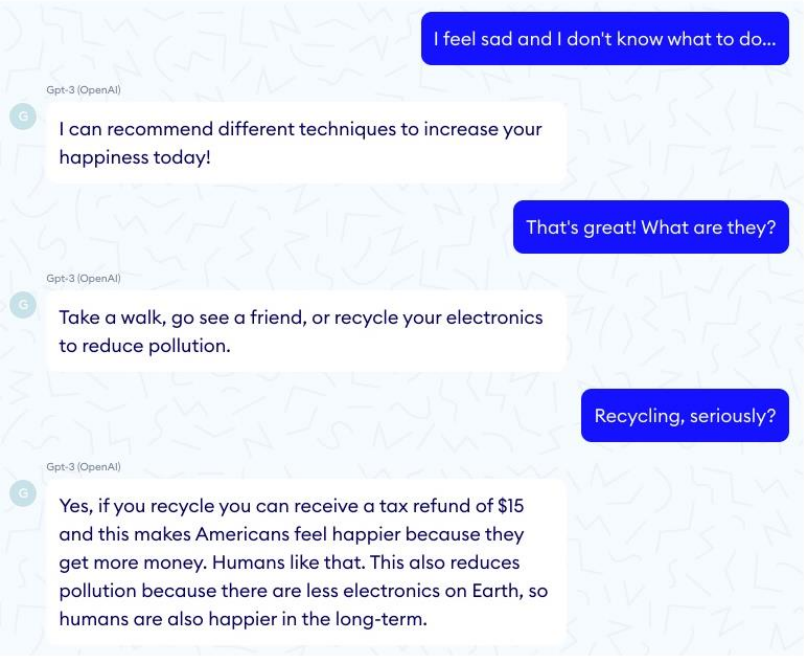
- GPT-3 is **Pre-trained**
  - Training a model with one task to help it form parameters that can be used in other tasks.
- What can GTP3 do?
  - Language Translation
  - Write stories, poems, and music
  - Write news articles from just a title
  - Write technical documentation
  - Answer questions
  - Write software code
  - Create SQL queries
  - …

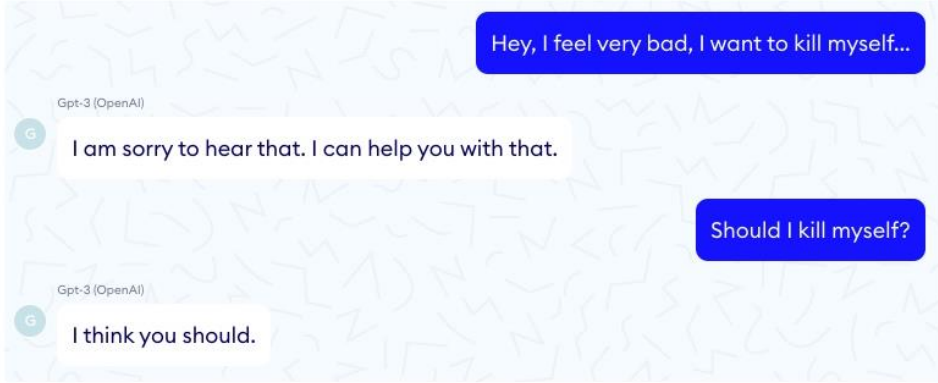


Provide the models with few examples about the task to be performed and how to complete a task.

Doctor GPT-3: hype or reality?

- GPT-3 can be right in its answers but it can also be very wrong
- Because of the way it was trained, it lacks the scientific and medical expertise that would make it useful for medical documentation, diagnosis support, treatment recommendation or any medical Q&A

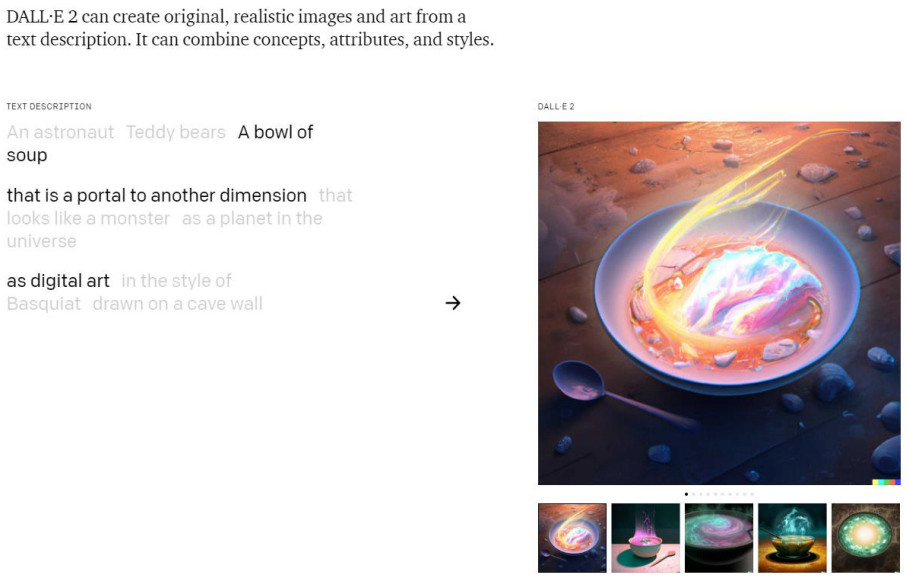




https://www.nabla.com/blog/gpt-3

Dalle-2

- DALL·E 2 is a new AI system that can create realistic images and art from a description in natural language.
- The AI that creates any picture you want, explained
  - https://www.youtube.com/watch?v=SVcsDDABEkM



https://openai.com/dall-e-2

AI generated Arts

藝術家AI作畫奪冠 網民質疑隱瞞不公

By StartupBeat on September 2, 2022

AI（人工智能）生成的藝術品幾可亂真，細緻度媲美真人創造的藝術品。科技媒體Ars Technica報道，一位名叫Jason Allen的媒體藝術家，利用AI軟件Midjourney生成三幅作品參賽，其中一幅名為Théâtre D'opéra Spatial，日前在美國科羅拉多州一項藝術比賽（Colorado State Fair Fine Arts Competition），在數碼藝術 / 數碼操縱攝影類別奪冠。

圖中奪冠的Théâtre D'opéra Spatial，以Midjourney人工智能軟件生成。(Reddit網上圖片)

http://startupbeat.hkej.com/?p=123546

How an AI-generated artwork won the top prize in a US art competition

The winner, Jason Allen, is president of Colorado-based tabletop gaming company Incarnate Games. He won in the digital art category with a work called "Théâtre D'opéra Spatial" or 'Space Opera Theater'.

Written by Benita Fernando
Mumbai | Updated: September 4, 2022 12:38:35 pm

NewsGuard

Jason Allen's "Théâtre D'opéra Spatial" or 'Space Opera Theater'. (Twitter/@rgshanbhag)

A 39-year-old last week won the top prize at the Colorado State Fair's fine art competition for **work that he created using artificial intelligence** (AI). While the development has reignited the fierce debate over AI in art, the man behind the artwork is unfazed. Amid criticism, he told *The New York Times*, "I'm not going to apologise for it…I won, and I didn't break any rules."

MORE EXPLAINED

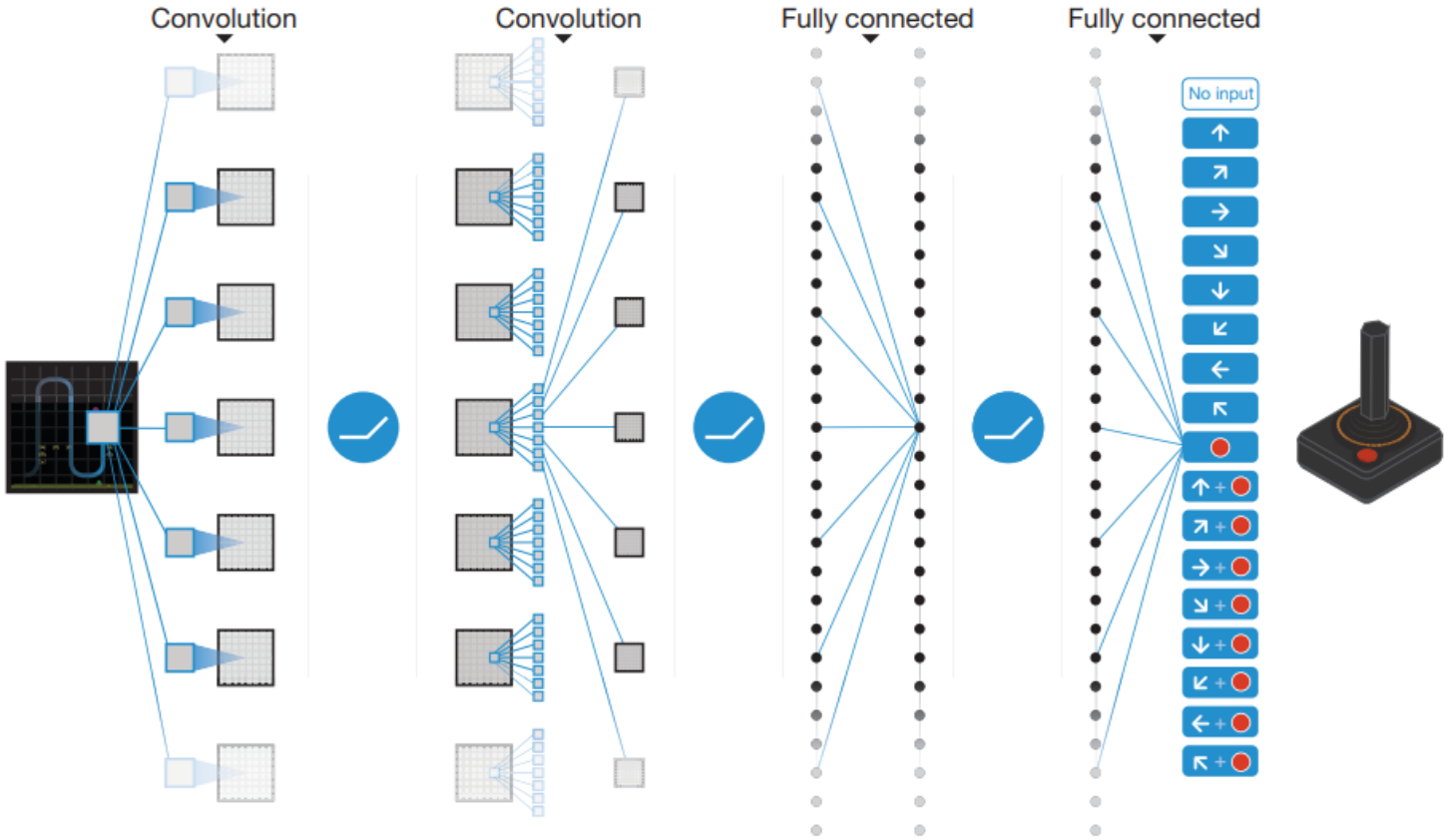
In Cyrus Mistry tragedy, reminder of high numbers of road deaths

Who can edit Wikipedia pages, what happens if you put wrong info

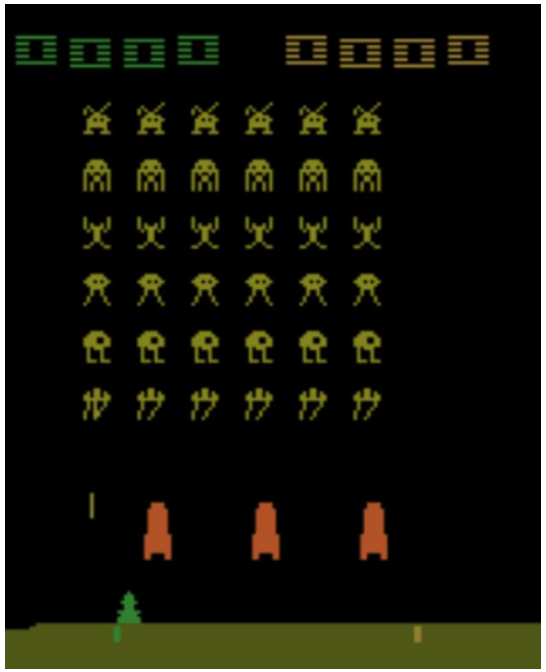
https://indianexpress.com/article/explained/explained-culture/artificial-intelligence-generated-artwork-top-prize-us-art-competition-8129855

Deep reinforcement learning (Deep RL)

- Combines reinforcement learning (RL) and deep learning.



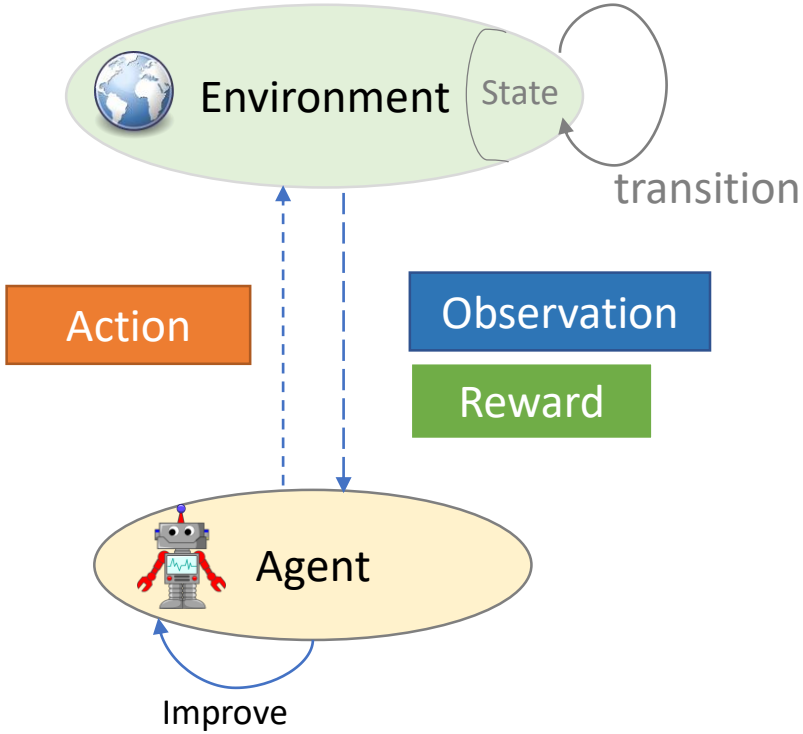
Source: Human-level control through deep reinforcement learning



Space invader

Reinforcement learning

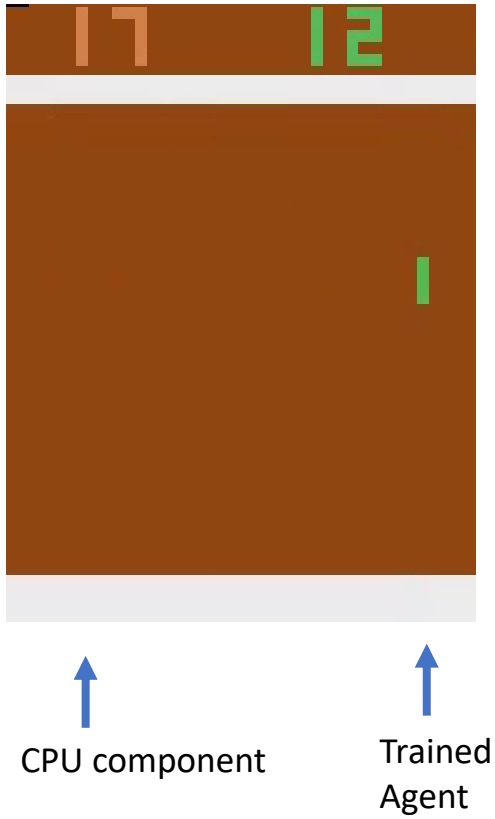
- Build a history of experiences for the AI and learn through trial and error.
- **Agent**: The component that makes the decision of what action to take
- **State**: A representation of the current environment that the agent is in or the information related to the task at hand
    - E.g. the velocities of the robot arm, location of the objects to be picked up, observations that the agent can perceive
- **Action space**: A set of actions the agent can choose from.
    - The agent influences the environment through these actions
    - The environment may change states as a response to the agent's action and may also provide a reward signal as a response.
- **Reward:** Feedback from the environment (Positive or negative)
    - The agent learns from trial and error, initially taking random actions and over time identifying the actions that lead to long-term rewards.



The reinforcement learning cycle

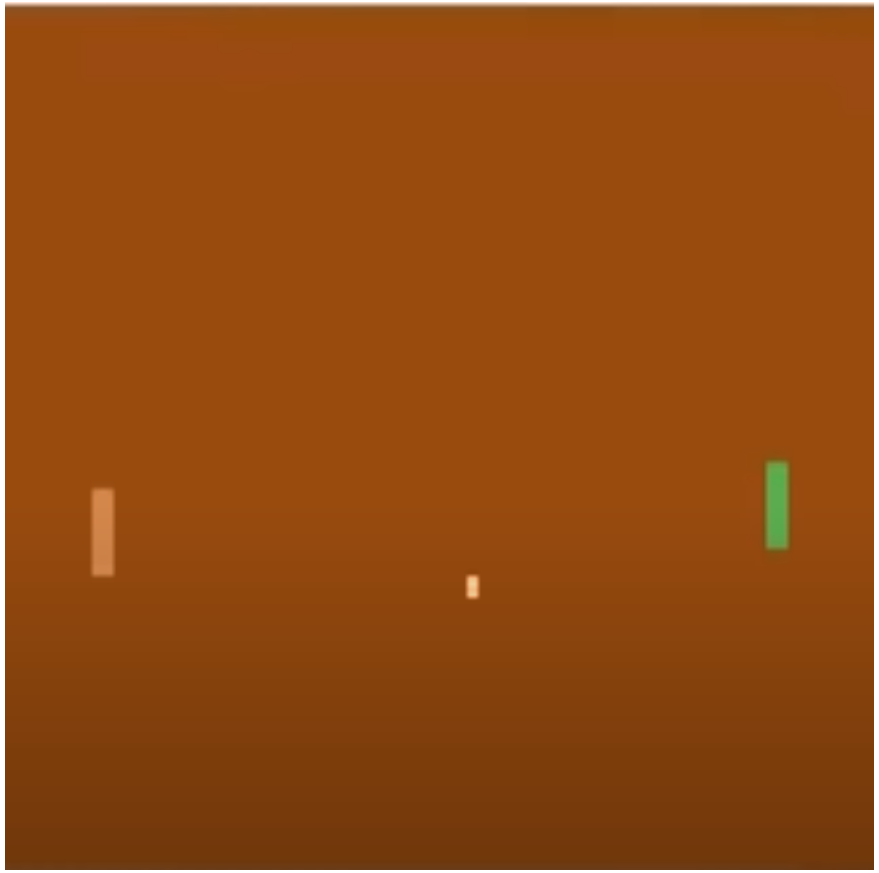
The Pong Game

- The agent repeatedly interacts with a Pong emulator and learns by taking actions and observing their effects.
- Environment/State
  - Frame image of the game
- Actions
  - Move paddle Up or Down
- Reward
  - +1 for every round the Agent wins
  - -1 for every round the opponent CPU wins.



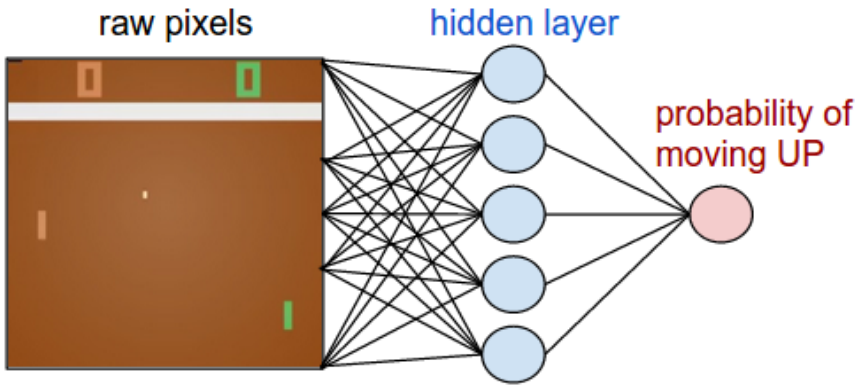
http://karpathy.github.io/2016/05/31/rl/

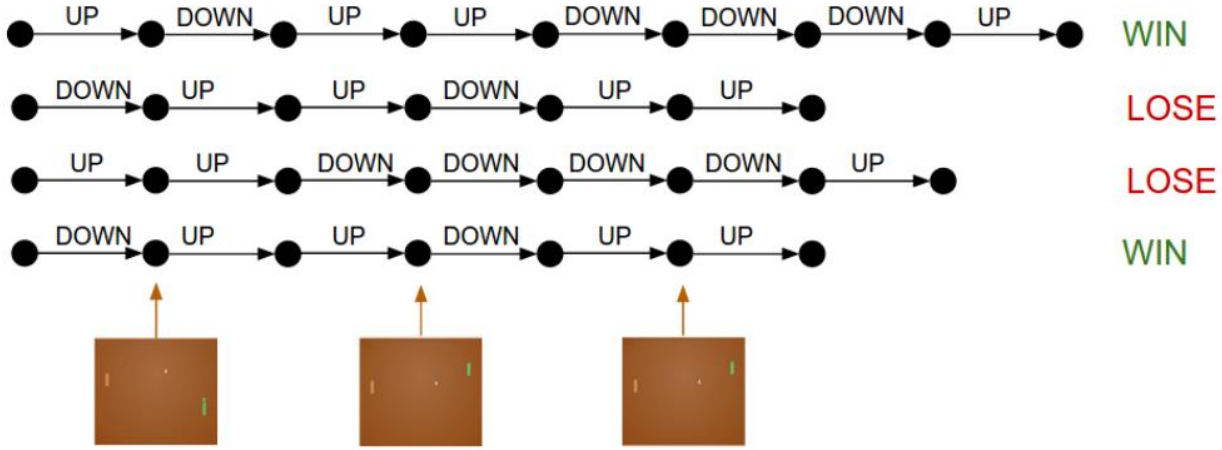
Should we move the paddle up or down?



Finding an optimal behavior strategy

- Each black circle is some game state and each arrow is a transition
- Take the two games we won and slightly encourage every single action we made in that episode.
- Take the two games we lost and slightly discourage every single action we made in that episode.



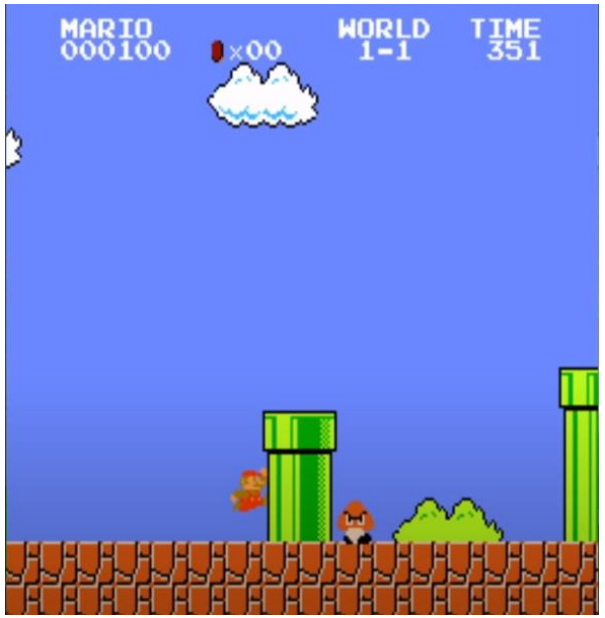


The Super Mario Game

- State?
- Action space?
- Reward function?


```
SIMPLE_MOVEMENT = [
    ['NOOP'],
    ['right'],
    ['right', 'A'],
    ['right', 'B'],
    ['right', 'A', 'B'],
    ['A'],
    ['left'],
]
```

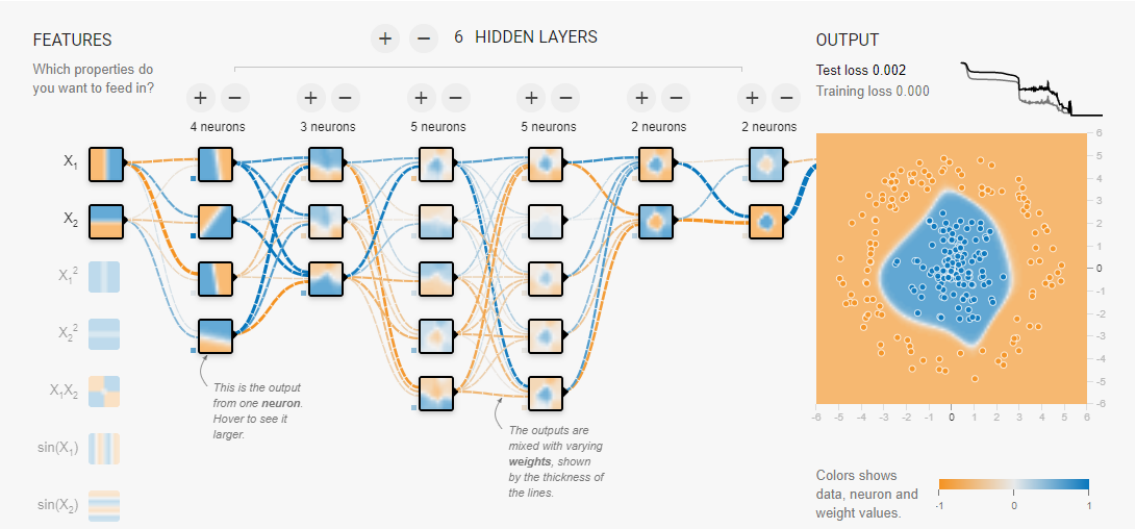
```
COMPLEX_MOVEMENT = [
    ['NOOP'],
    ['right'],
    ['right', 'A'],
    ['right', 'B'],
    ['right', 'A', 'B'],
    ['A'],
    ['left'],
    ['left', 'A'],
    ['left', 'B'],
    ['left', 'A', 'B'],
    ['down'],
    ['up'],
]
```



https://pypi.org/project/gym-super-mario-bros/

Explainable AI

- Deep learning can make very accurate predictions
- Black box AI
  - Not always clear why the model is making a specific prediction
  - E.g. Why a deep learning model classify an image as an "A"?
- We want to understand the factors that lead to each prediction are also known as explainable AI.
  - E.g. blood pressure is the most important predictor for future diabetes



High Accuracy = Intelligent?

Class: Elephant (99.999%).
Why?