COMP309 in Week 10, 2024

an introduction to neural networks

Marcus Frean marcus.frean@vuw.ac.nz

Week 10

- Lecture 1
  - Why Deep learning
  - The perceptron
    - activation function
  - Multi-layer Perceptrons (a.k.a. Neural networks)

- Lecture 2
  - tensors
  - automatic differentiation (autograd)
  - PyTorch for example

- Tutorial:
  - An end-to-end example of using PyTorch to solve a non-linear regression problem.

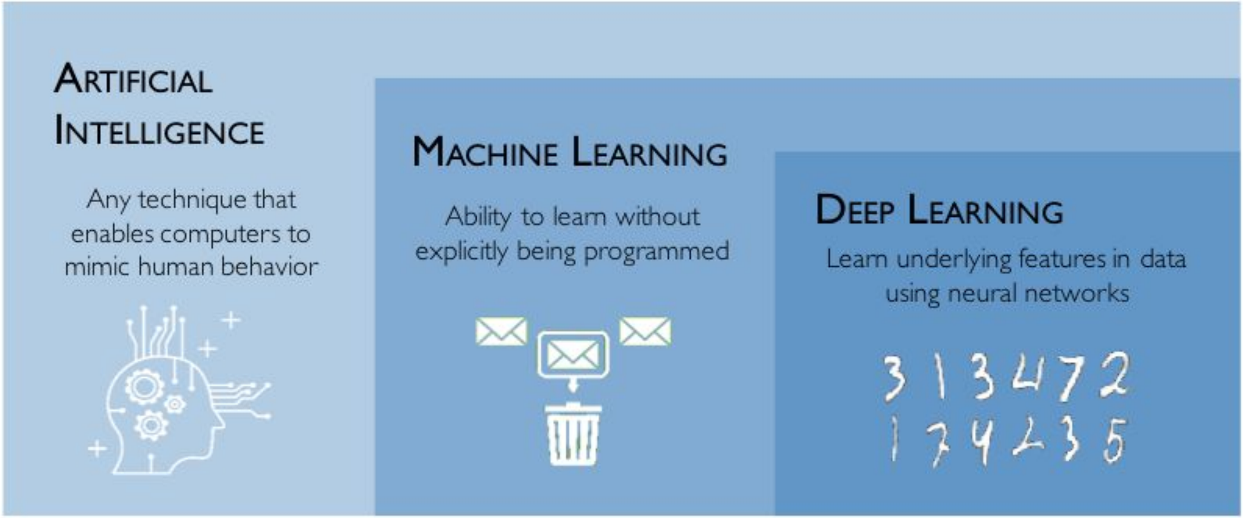
What is Deep Learning?

- Artificial intelligence
- Machine learning
- Deep neural networks

AI= ML?

ML= DL?

DL ⊂ ML ⊂ AI



Deep Learning - successful neural nets

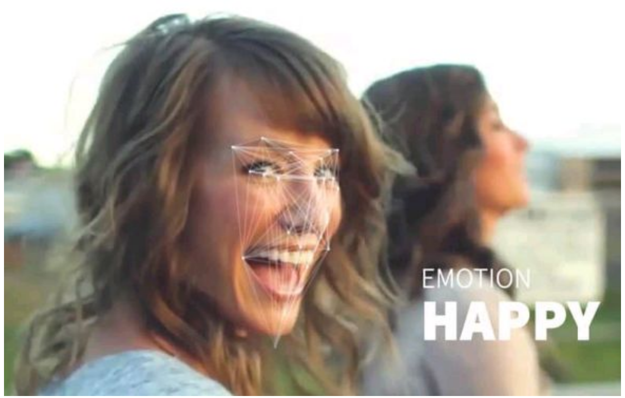
Image Recognition

Detect pneumothorax in real X-Ray scans

And so many more...

uncanny successes

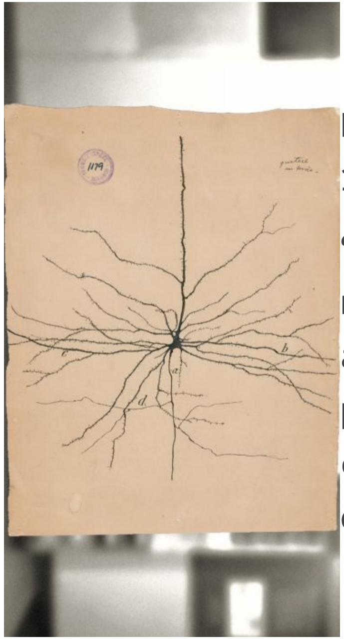
- another day, another win for neural nets



- some of the most striking are “deep fakes”
  - images (including faces),
  - sounds (including voices)
  - texts / stories / human dialogue (chatGPT)

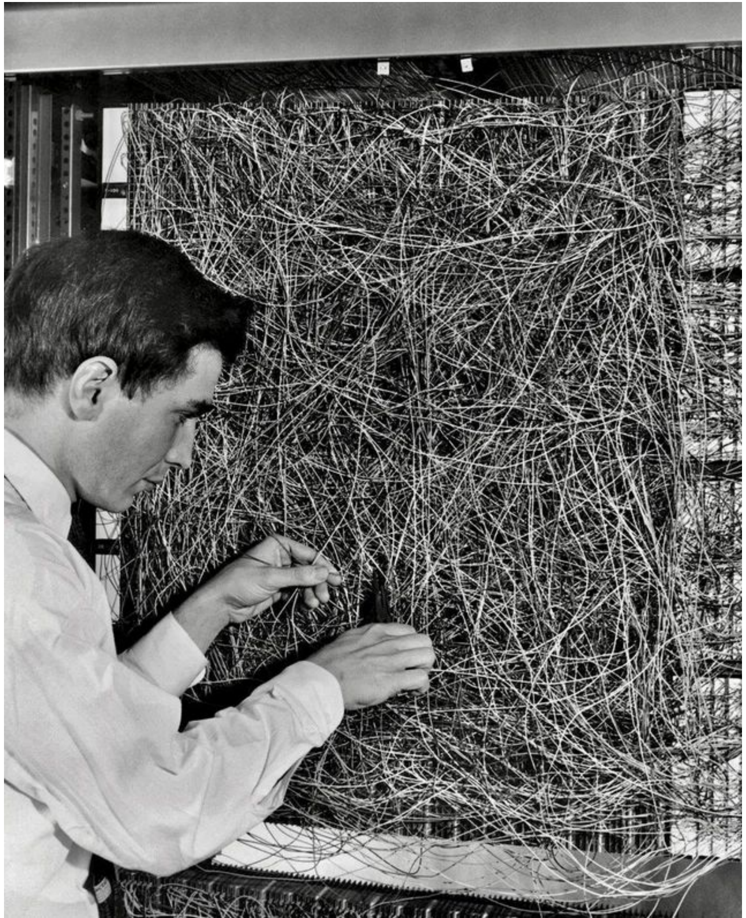
These are all just simulcra, right?

The Old Days



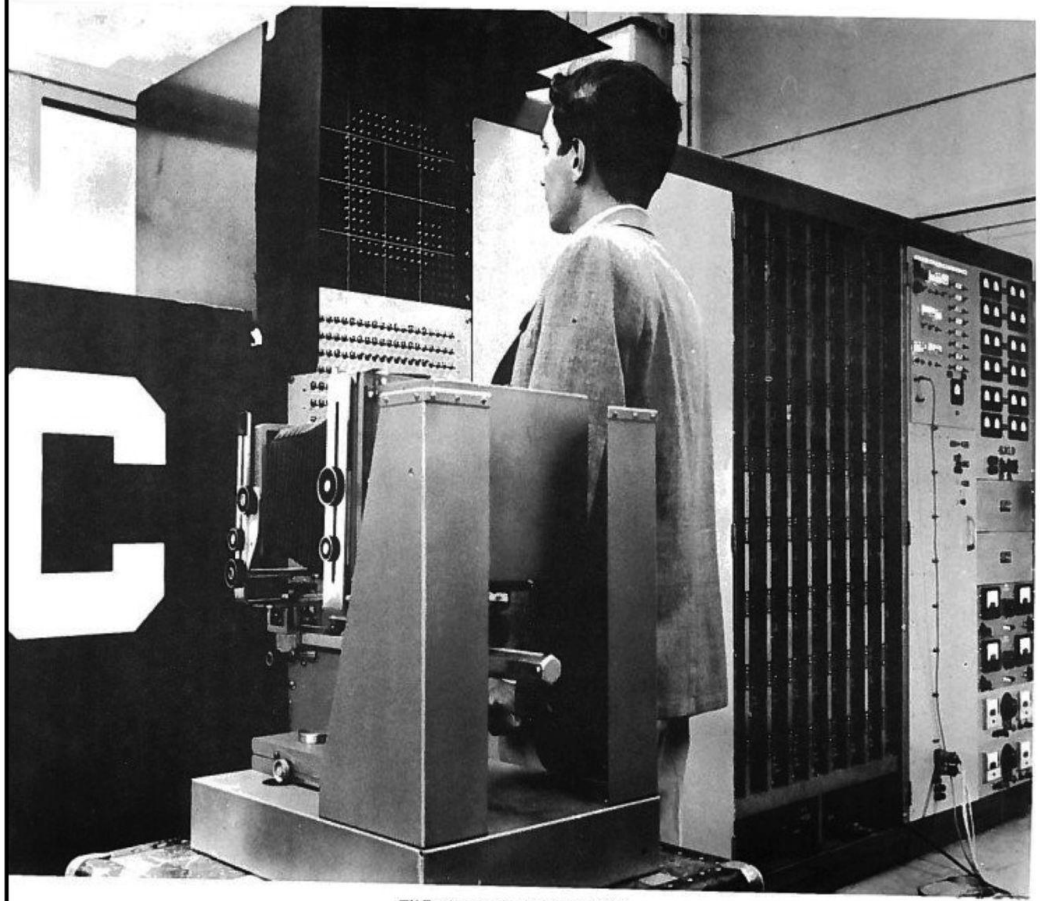
Ramón y Cajal, 1894:

“The ability of neurons to grow in an adult and their power to create new connections can explain learning.”



As a child he was transferred many times from one school to another because of behavior that was declared poor, rebellious, and showing an anti-authoritarian attitude. Imprisoned aged 11 in 1863 for destroying his neighbor's yard gate with a homemade cannon. His father apprenticed him to a shoemaker and barber, to "try and give his son much-needed discipline and stability."… “Over the summer of 1868, his father took him to graveyards to find human remains for anatomical study.”

Hype and Hubris (as Usual)



THE MARK I PERCEPTRON

NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

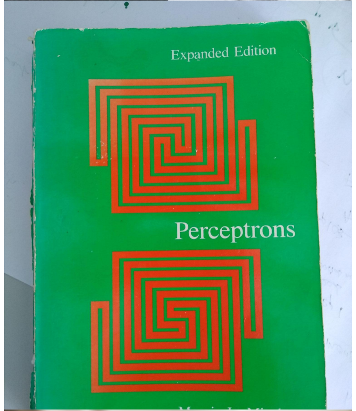
WASHINGTON, July 7 (UPI) —The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

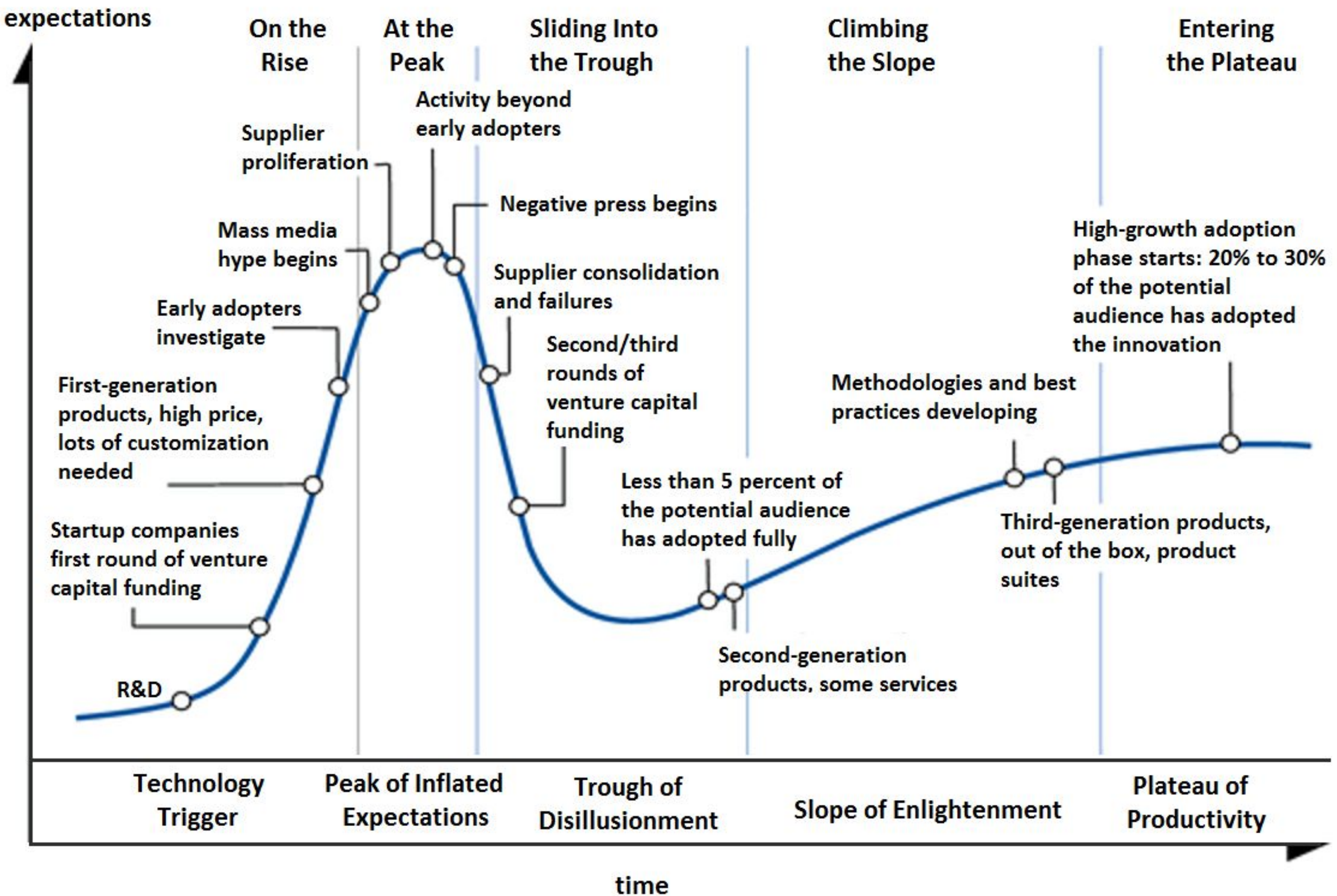
The embryo—the Weather Bureau's $2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.




expectations
On the Rise
At the Peak
Sliding Into the Trough
Climbing the Slope
Entering the Plateau
Activity beyond early adopters
Supplier proliferation
Negative press begins
Mass media hype begins
Supplier consolidation and failures
Early adopters investigate
Second/third rounds of venture capital funding
First-generation products, high price, lots of customization needed
High-growth adoption phase starts: 20% to 30% of the potential audience has adopted the innovation
Methodologies and best practices developing
Less than 5 percent of the potential audience has adopted fully
Third-generation products, out of the box, product suites
Startup companies first round of venture capital funding
Second-generation products, some services
R&D
Technology Trigger
Peak of Inflated Expectations
Trough of Disillusionment
Slope of Enlightenment
Plateau of Productivity
time

"Mary gave Jack the book"

this will produce in you, unconsciously, many kinds of thoughts - that is, mental activities in such different realms as:

A visual representation of the scene.

Postural and Tactile representations of the experience.

A script-sequence of a typical script-sequence for "giving."

Representation of the participants' roles.

Representations of their social motivations.

Default assumptions about Jack, Mary and the book.

Other assumptions about past and future expectations.

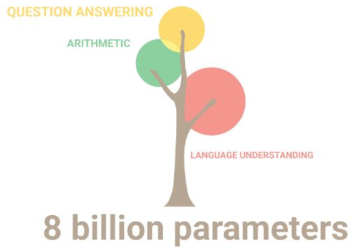
“How could a brain possibly coordinate the use of such different kinds of processes and representations?”

Today

GitHub Copilot, ChatGPT

Dalle, Imagen, stable diffusion…

Scaling to many parameters, costing $$$ to train, out of range for (e.g.) universities



- is this just parroting humans, via vast amounts of our data?

No doubt this will enable both

- new ways to get people to pay for stuff, new ways to surveil people in a bad way, and make them do things, and new weaponry
- new good tech too, e.g. freeing us from onerous tasks, "good" surveillance, more efficient services, totally new kinds of business, better communications …

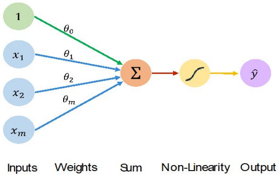
How does it work?

features, e.g. numbers giving key points on a face

e.g. data: "input pattern"

→ "output" ♂ or ♀

- Define your "machine" capable of carrying out the mapping, e.g.
  1. sum things up
  2. apply a threshold

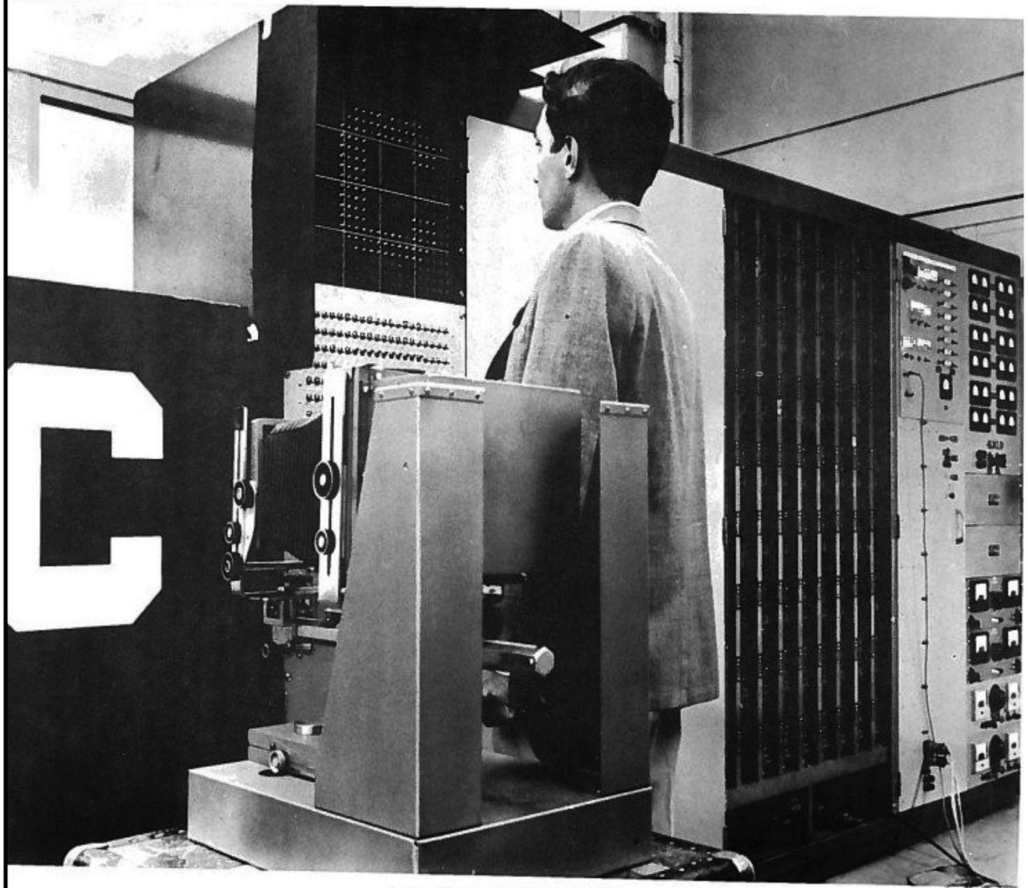


- Make sure it's got some "flexibility" in its parameters
  - give each bit of the sum a numerical "weight"

- Figure out how to change those parameters
  - if the output was too high, decrease the weight from +ve inputs (etc)

- Use lots of data to change the parameters until they stabilise
- Now you can use your machine to classify a *brand new input pattern*

early days



THE MARK I PERCEPTRON

NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI) —The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

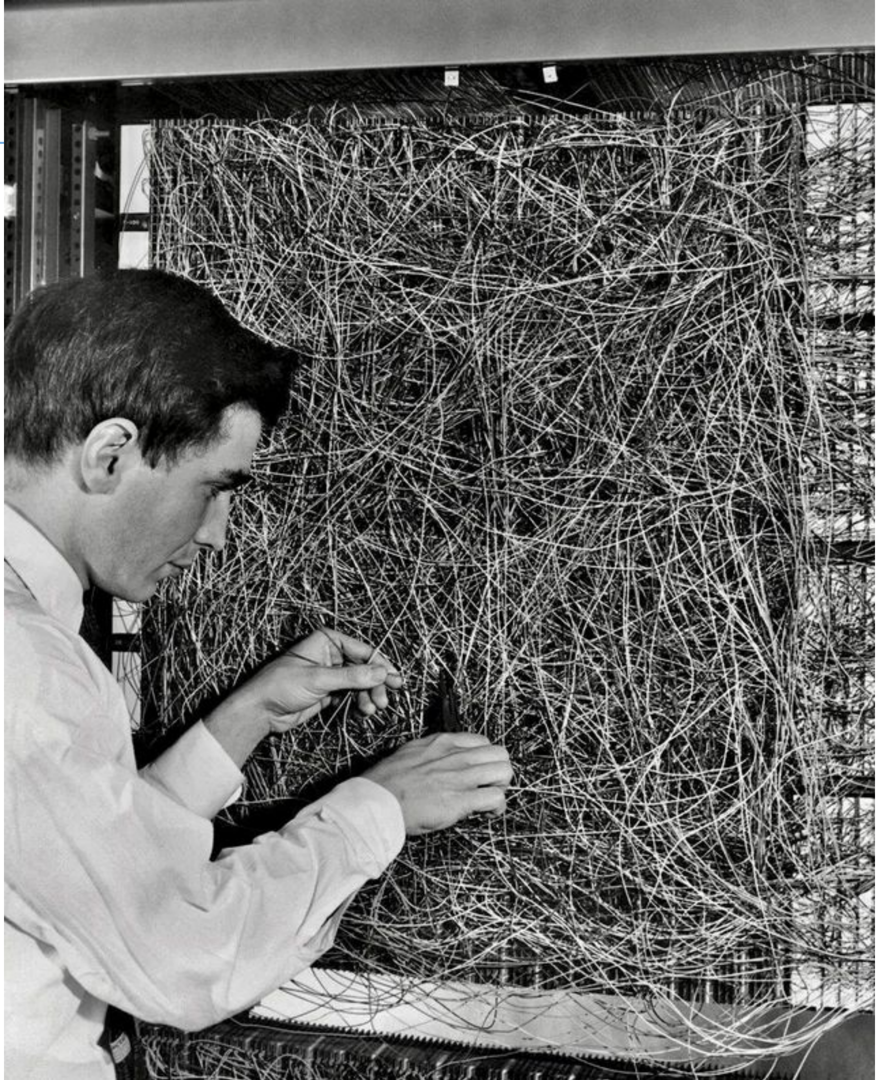
The embryo—the Weather Bureau's $2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

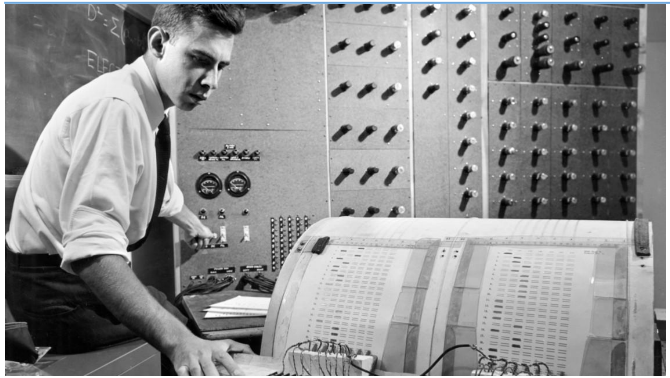
Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

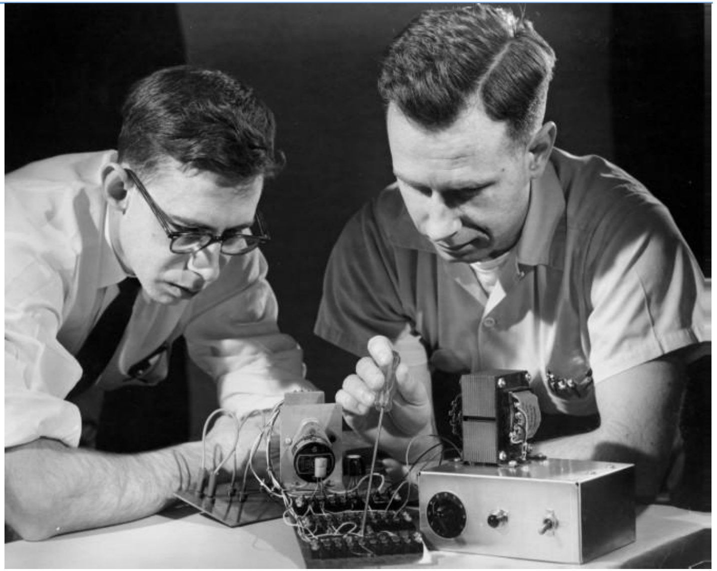
Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

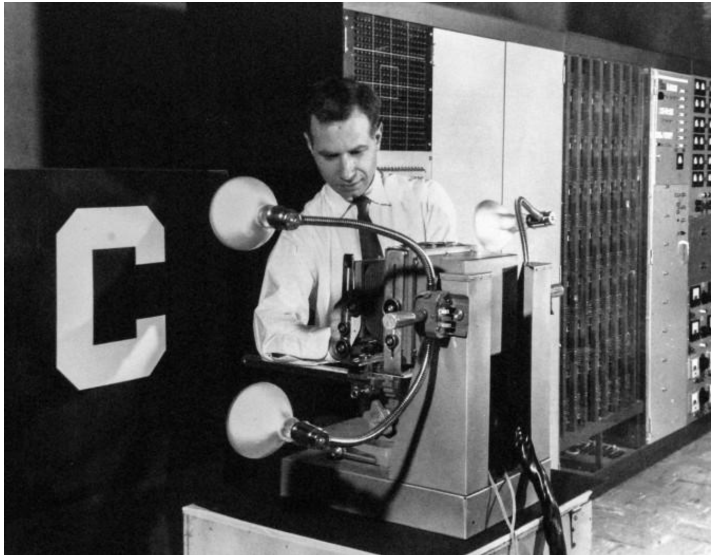
Frank Rosenblatt and “the first machine capable of having an original idea”



https://news.cornell.edu/stories/2019/09/
professors-perceptron-paved-way-ai-60-years-too-soon







in effect, early neural nets were a *single* layer

Rosenblatt's Perceptron – hard-wired "features" for the first layer.

The trainable "perceptron" was just the last layer.

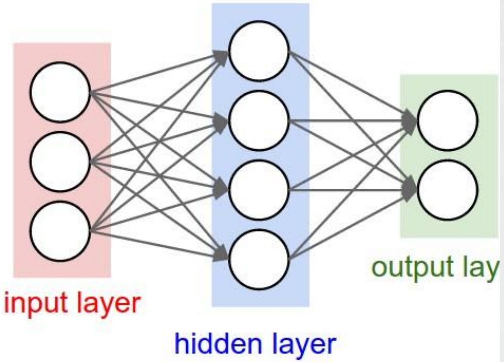


FIG. 1 — Organization of a biological brain. (Red areas indicate active cells, responding to the letter X.)

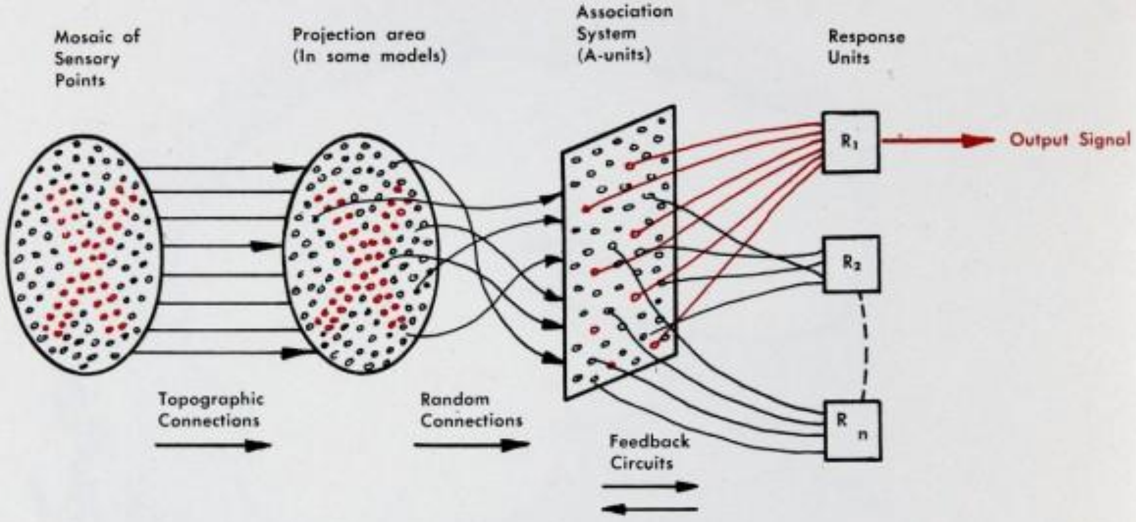
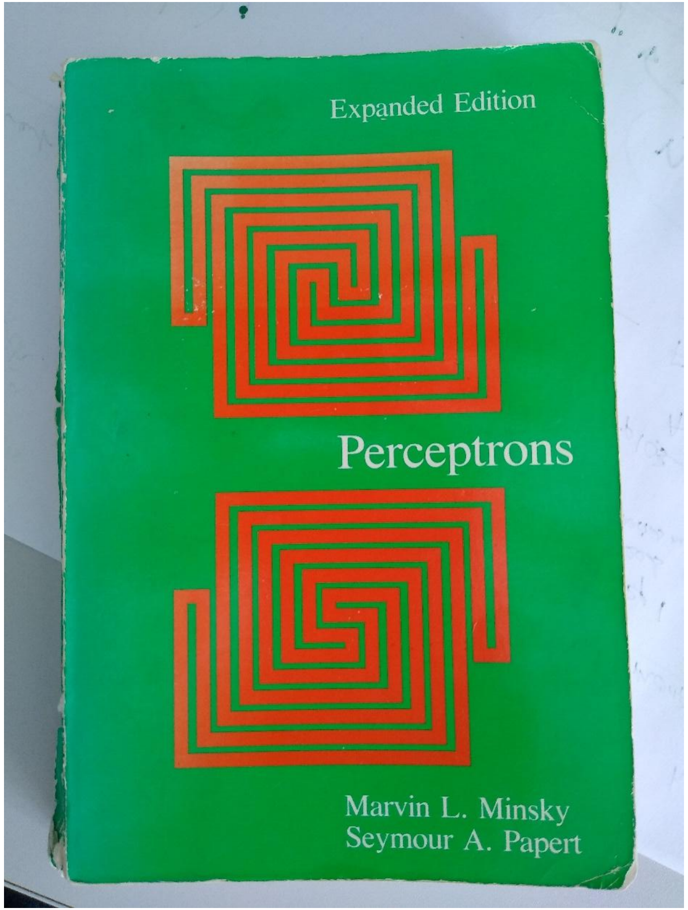
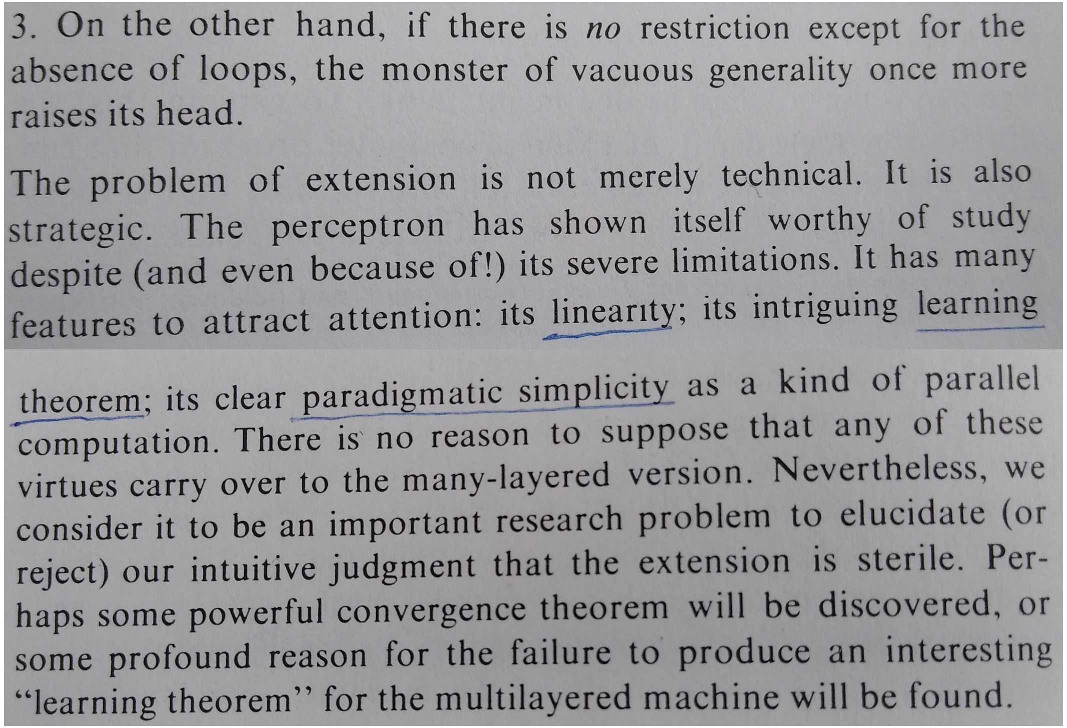


FIG. 2 — Organization of a perceptron.

Minsky, Papert, and the "AI Winter" (~1960s-70s)





3. On the other hand, if there is *no* restriction except for the absence of loops, the monster of vacuous generality once more raises its head.

The problem of extension is not merely technical. It is also strategic. The perceptron has shown itself worthy of study despite (and even because of!) its severe limitations. It has many features to attract attention: its linearity; its intriguing learning theorem; its clear paradigmatic simplicity as a kind of parallel computation. There is no reason to suppose that any of these virtues carry over to the many-layered version. Nevertheless, we consider it to be an important research problem to elucidate (or reject) our intuitive judgment that the extension is sterile. Perhaps some powerful convergence theorem will be discovered, or some profound reason for the failure to produce an interesting "learning theorem" for the multilayered machine will be found.

"... our intuitive judgement that the extension is sterile."

History (up to ~2006)

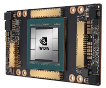
1958	Rosenblatt's "Perceptron" - a neural network (NN) with learnable weights
1985	Multi-layer perceptron (MLP) – backpropagation algorithm trains the weights
1990	*Convolutional* NN – "simple" vision tasks, like digit recognition
1997	*Recurrent* NN – able to model sequences (nb. these were very "deep" nets)
2006	pre-training of deep belief nets – the first really big, deep, neural networks

kaboom…

Why just *then*?

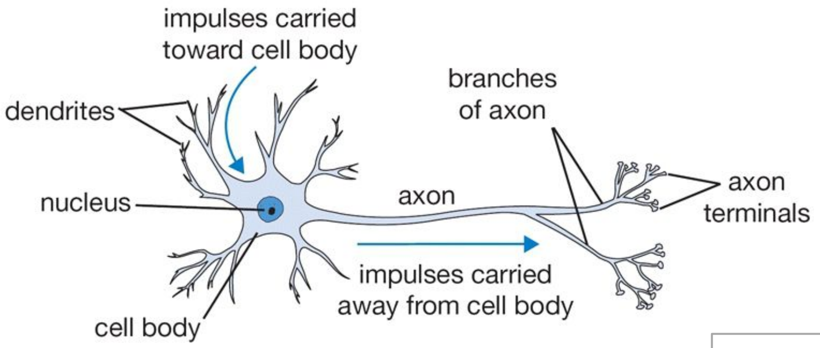
1. big data
- easier collection and storage
- sheer volume

2. hardware
- GPU (graphics cards)
- massively parallelizable



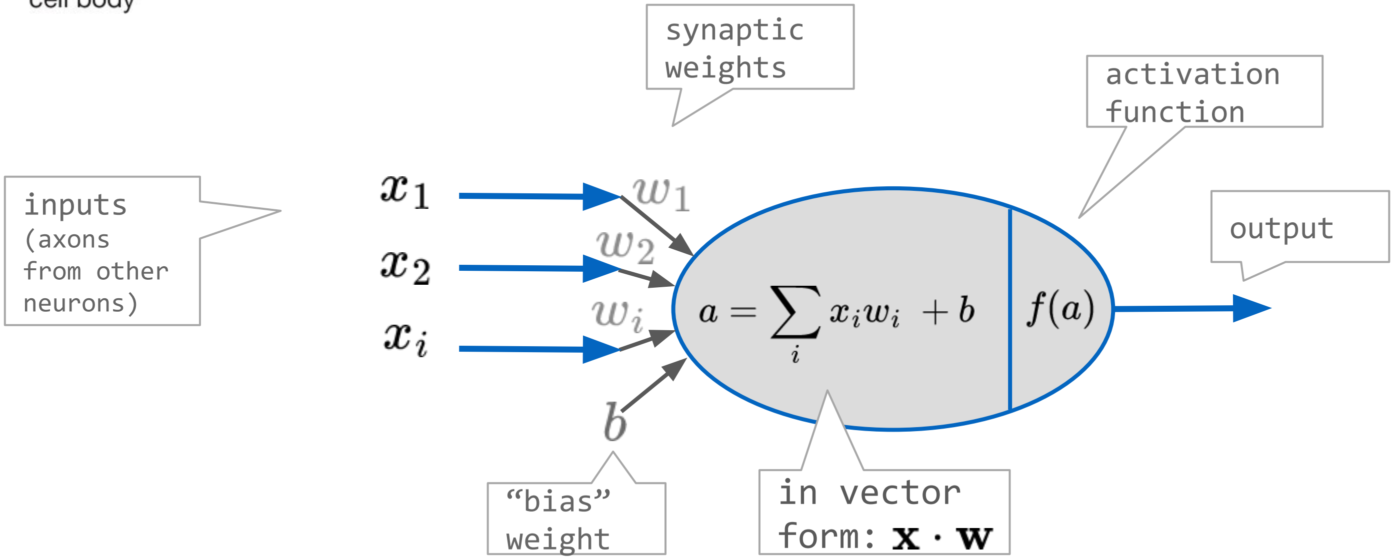
3. software
- ReLU!
- new models
- toolboxes

neural inspiration behind perceptrons

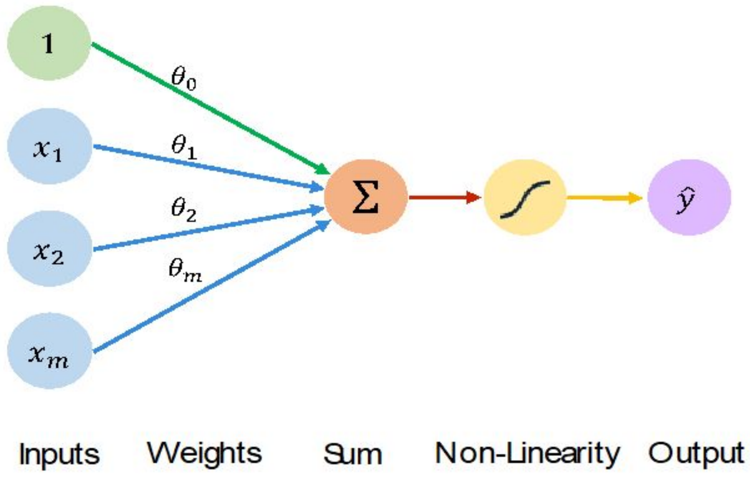


Each "neuron" does:

1. weighted sum of inputs (x · w, plus a bias)
2. puts the result through a non-linearity (f)



The Perceptron: Forward Propagation



output

bias

sum of weighted inputs

$$\hat{y} = g\left(w_0 + \sum_{i=1}^{m} x_i w_i\right)$$

non-linear activation function

$$\hat{y} = g(w_0 + \boldsymbol{X} \cdot \boldsymbol{W}^T)$$

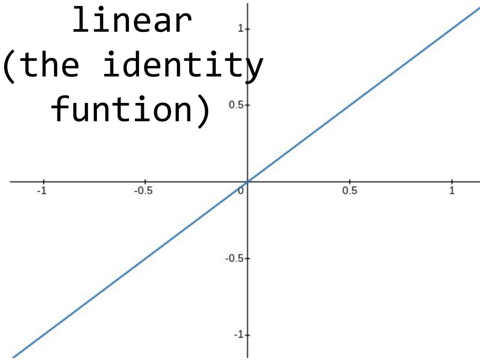
where $\boldsymbol{X} = [x_1 \cdots x_m]$

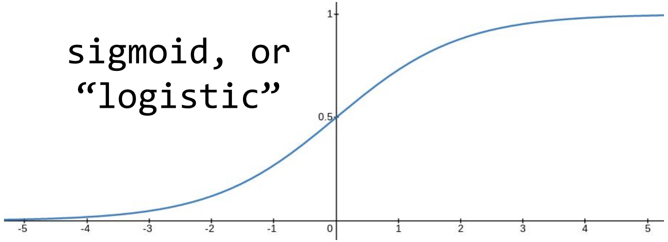
$\boldsymbol{W} = [w_1 \cdots w_m]$

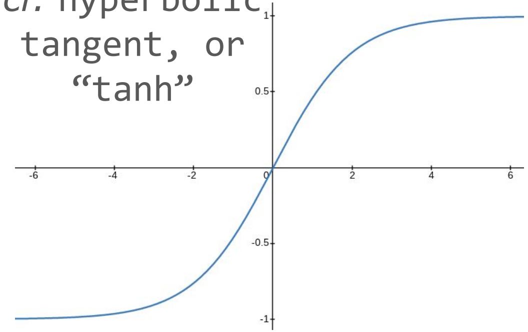
as numpy: (but you'd convert to torch)

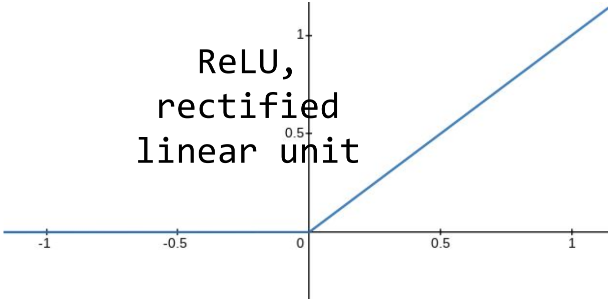
```
class Neuron(object):
  # ...
  def forward(self, inputs):
    """ assume inputs and weights are 1-D numpy arrays and bias is a number """
    cell_body_sum = np.sum(inputs * self.weights) + self.bias
    firing_rate = 1.0 / (1.0 + math.exp(-cell_body_sum)) # sigmoid activation function
    return firing_rate
```

activation functions







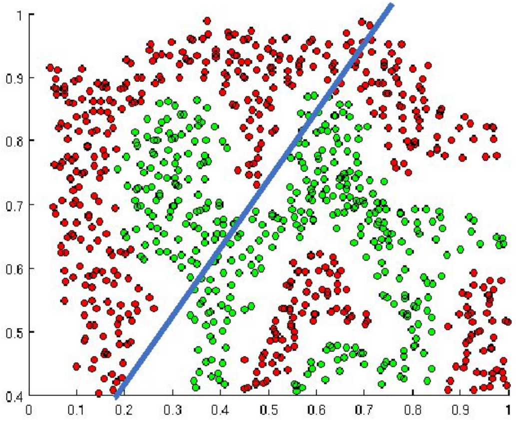


as numpy: (but you’d convert to torch)

```
class Neuron(object):
  # ...
  def forward(self, inputs):
    """ assume inputs and weights are 1-D numpy arrays and bias is a number """
    cell_body_sum = np.sum(inputs * self.weights) + self.bias
    firing_rate = 1.0 / (1.0 + math.exp(-cell_body_sum)) # sigmoid activation f
    return firing_rate
```

activation functions

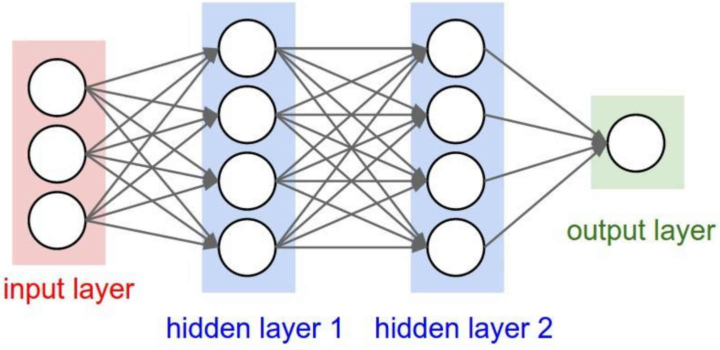
introduce non-linearities into the network:



a linear function of a
linear function is (just)
a linear function. So
compositions don't help.

layers add power

(so long as they involve non-linearities)



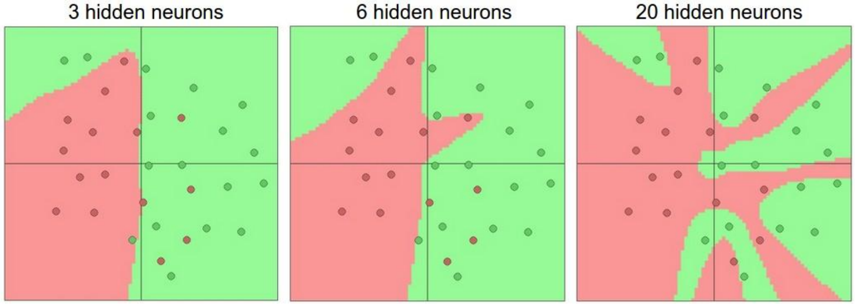

```
# forward-pass of a 3-layer neural network:
f = lambda x: 1.0/(1.0 + np.exp(-x)) # activation function (use sigmoid)
x = np.random.randn(3, 1) # random input vector of three numbers (3x1)
h1 = f(np.dot(W1, x) + b1) # calculate first hidden layer activations (4x1)
h2 = f(np.dot(W2, h1) + b2) # calculate second hidden layer activations (4x1)
out = np.dot(W3, h2) + b3 # output neuron (1x1)
```

summary: we added stuff to linear maps, in order to gain expressive power

I. non-linearities provide the bare-minimum in “power”. Then…

II. more layers □ richer mappings possible

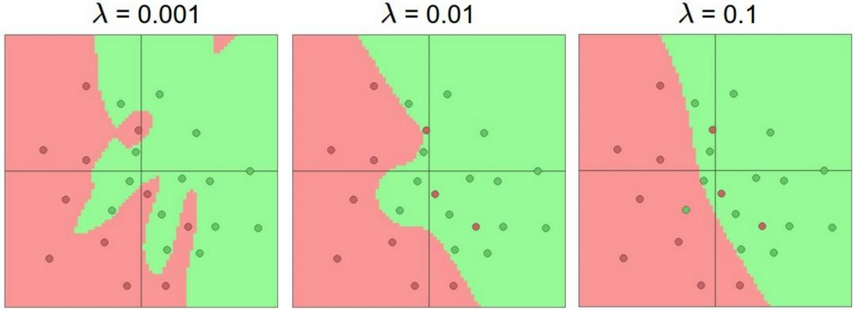
III. more neurons per layer □ more flexibility, more modelling power



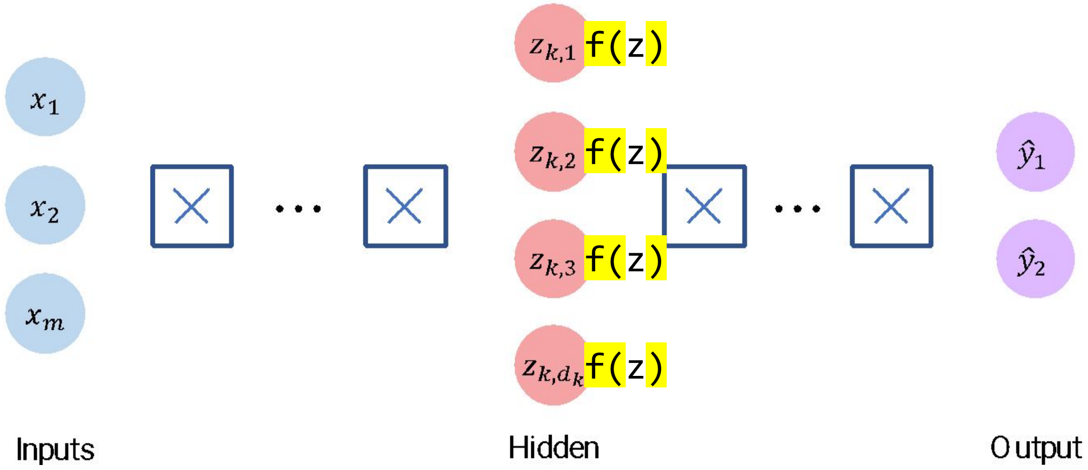
play with a great interactive demo:
https://playground.tensorflow.org

Too much freedom can lead to overfitting!

□ we “regularise” with penalties and other constraints on model power (a.k.a. “smoothing”)



Deep Neural Networks, Feed forward

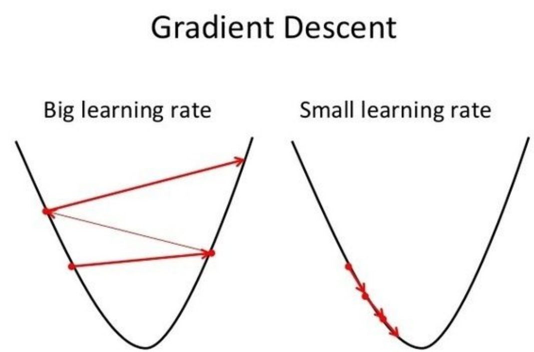


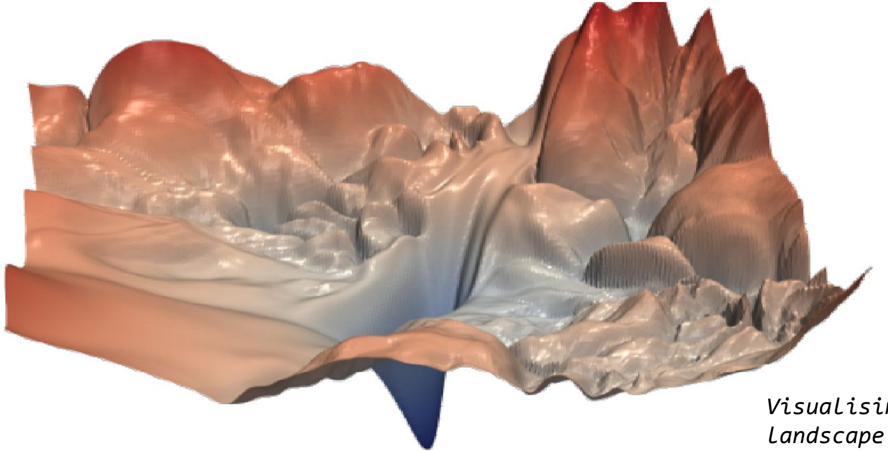
a neural net is really an alternating sequence of matrix multiplies and element-wise non-linearities

NN Loss Functions Can Be Difficult to Optimize

Learning rate (last week): step size

- Large learning rates overshoot, become unstable and diverge
- Small learning rate converges slowly and gets stuck in false local minima
- Stable learning rates converge smoothly and avoid local minima
- momentum (e.g. 0.9) helps
- RMSprop, Adam, Nadam help
- very little solid theory





Visualising the loss landscape of neural ents, Li et.al. 2017