

Introduction to Deep Learning

Kira Chan

cse.msu.edu/~chanken1

CSE435 Fall2024

Department of Computer Science and Engineering

Michigan State University

Introduction

- I did my B.S. and M.S. here at MSU
- During my undergrad, there was not a lot of classes for machine learning (ML)
- Recently, we have seen a large surge of ML related courses



Are AIs sentient?

Computing > Software > Artificial Intelligence

If ChatGPT does become sentient, it's smart enough to hide it from us

Feat

yahoo!news



News

Originals

Israel-Hamas War

COVID-19

Canada

World

Business

Entertainment

Weather

Videos

Sports

Science

Evening Standard.

GPT-4o: Shock as ChatGPT starts speaking in a user's cloned voice, leaving testers stunned

William Mata

August 12, 2024 · 2 min read



Roadmap

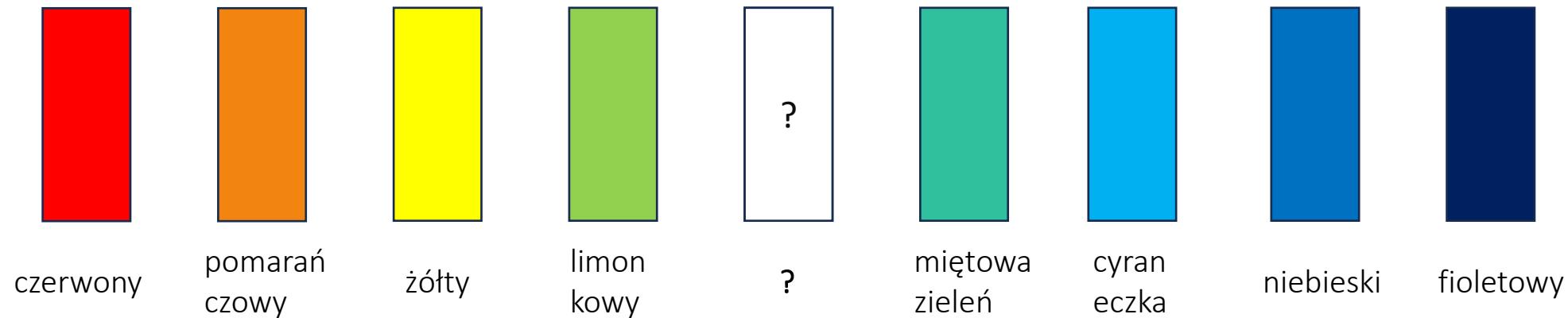
- Intuition of how current state-of-the-art Artificial Intelligence (AI)s work
- Demystify some of the black magic happening behind the scenes
- Examples and applications of ML
- Cans and can'ts

Where do we use AI?

- Medical imaging
- Autonomous driving
- Bug or plant identification
- Text and speech recognition
- Weather prediction
- Pandemic / COVID prediction

What does current AI really do?

- 1 2 3 4 6 7 8
- 2 4 6 8 12 14 16



- But what are we NOT good at?

Intuition

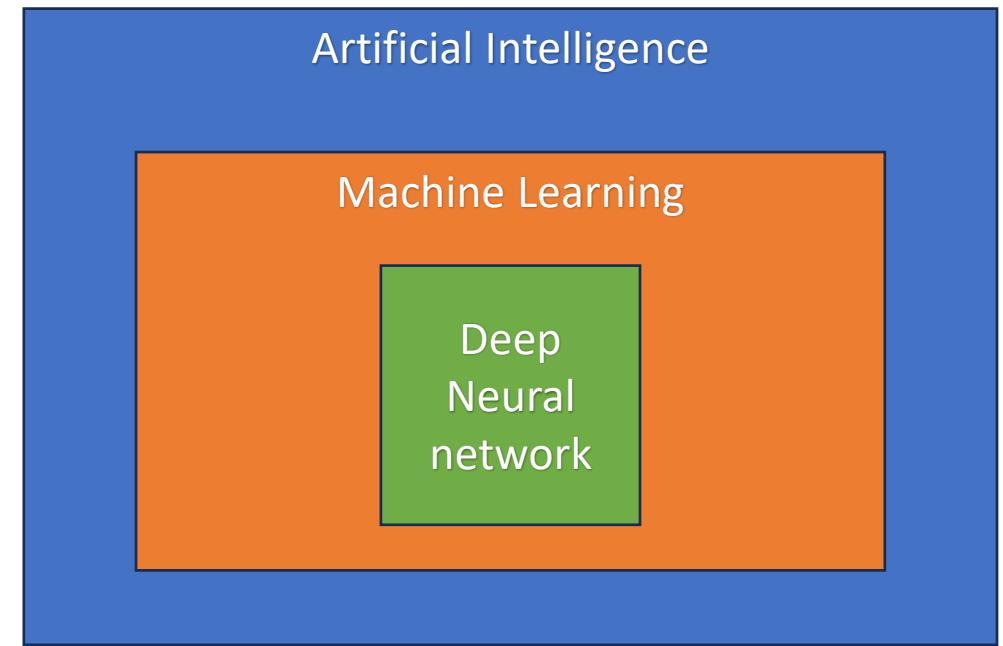
- Current state-of-the-art ML learns *patterns* from *data*
- They are also called *statistical* machine learning
- This means that they are right, *most of the time*
- But, by definition, a 99% accurate model must be wrong at some point

Dangers of Working with AI

- It is very easy to forget that software can be wrong
- You see a machine learning algorithm do a good job, its easy to forget that it is statistically going to fail
- If regular people using AI (e.g., doctors, positions of power) don't understand how they work, then they can blindly trust it
- What about developers?

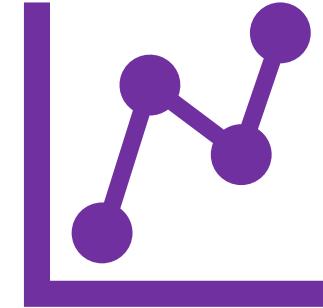
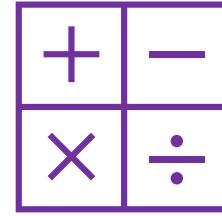
AI vs ML vs DNN

- Artificial Intelligence (AI)
- Machine Learning (ML)
- Deep Neural Network (DNN)
 - Current state-of-the-art



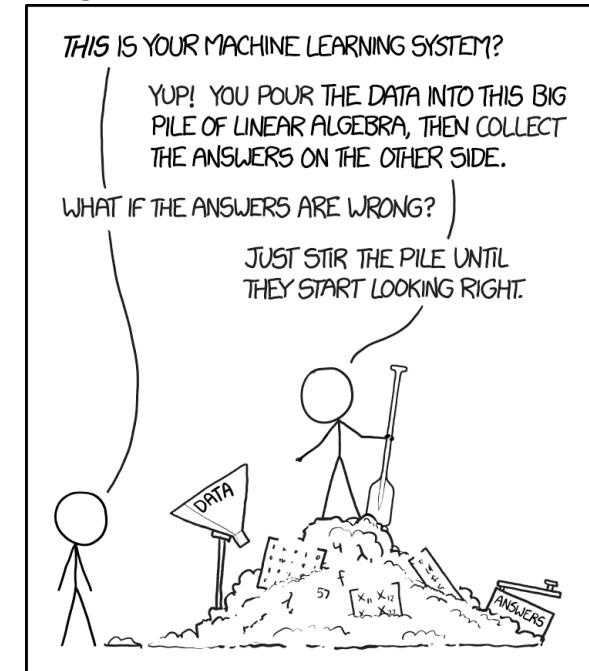
Data processing

- In today's world, we have a LOT of data
- In fact, a lot of things can be *mapped* to data
- Data drives businesses, finances, etc.
- Data tells us a lot of information
 - Explains unknowns
 - Pattern discovery
 - Tailored experiences (ads, Youtube/Netflix suggestions)
- Computers are very good at numbers, and data are numbers
 - A lot of things can be turned into numbers too (images, speech)



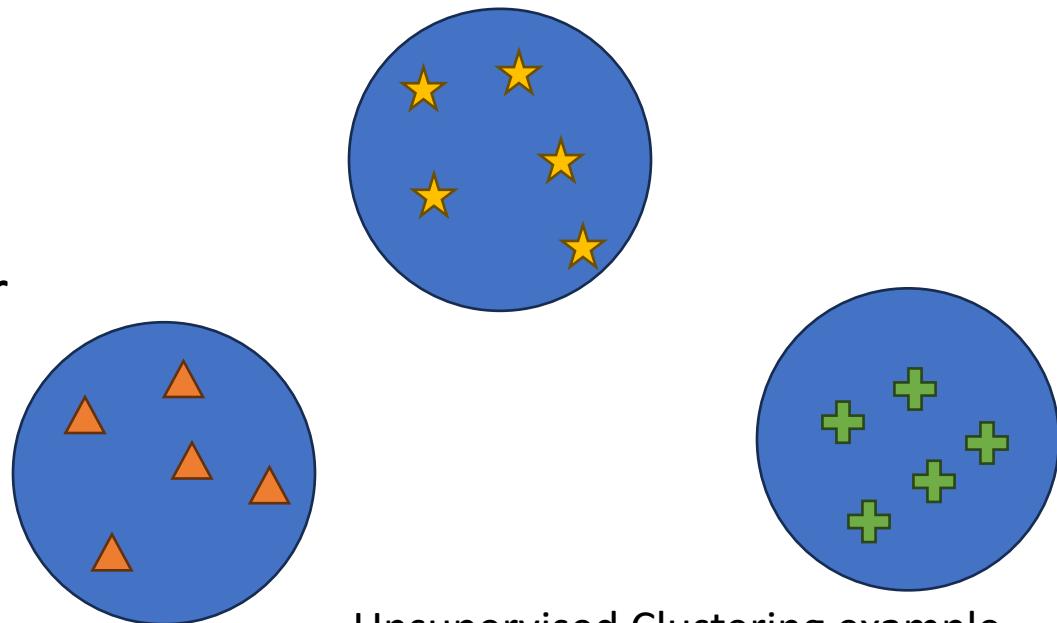
Machine learning models

- Machine learning is concerned with teaching a computer how to generalise and perform tasks without explicit instruction
 - Example: Identify a deer, predict products for customers, predict weather
- The main concept of machine learning is to extract **patterns** from data (data analysis)



Types of ML algorithm

- Supervised learning
 - Uses labelled data (cats and dogs)
- Unsupervised learning
 - Have no labelled output, find some relationship between data
 - E.g., clustering algorithms
 - Generative AI
- Reinforcement learning
 - Learn to do a task via trial-and-error

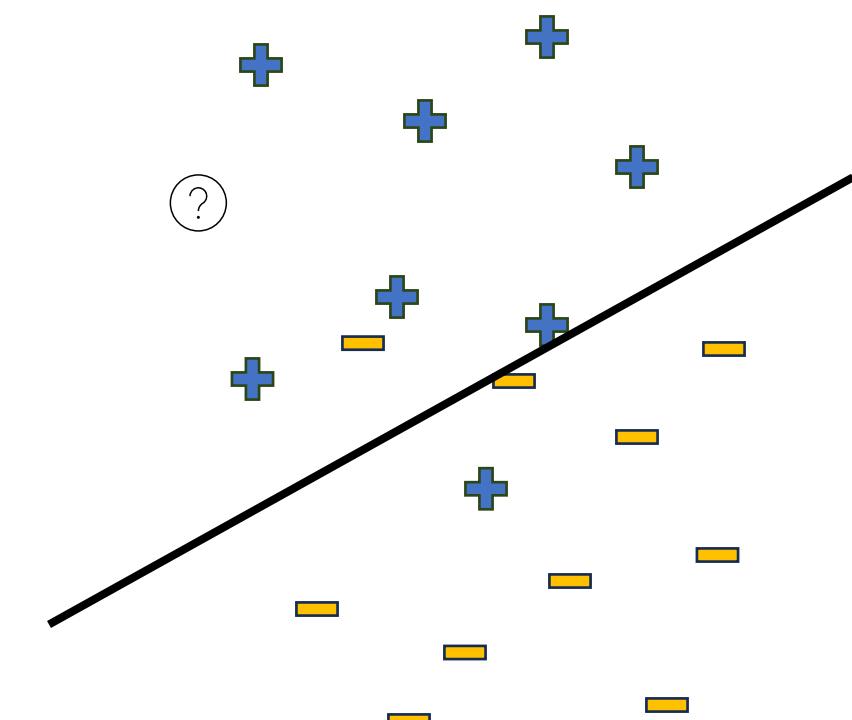


Supervised learning

- Given a dataset, we would like to train a model to learn some pattern
- When we feed it new data (in the same distribution), it should be able to **generalise** and apply the learned knowledge
- Typically, the algorithm will ``learn'' by iteratively predicting and adjusting its decision based on how correct it is
 - Presemester test – random guessing
 - As you learn the material, you understand the concepts
 - Final exam – you are able to take what you learned and apply it to similar questions

``Traditional'' algorithms

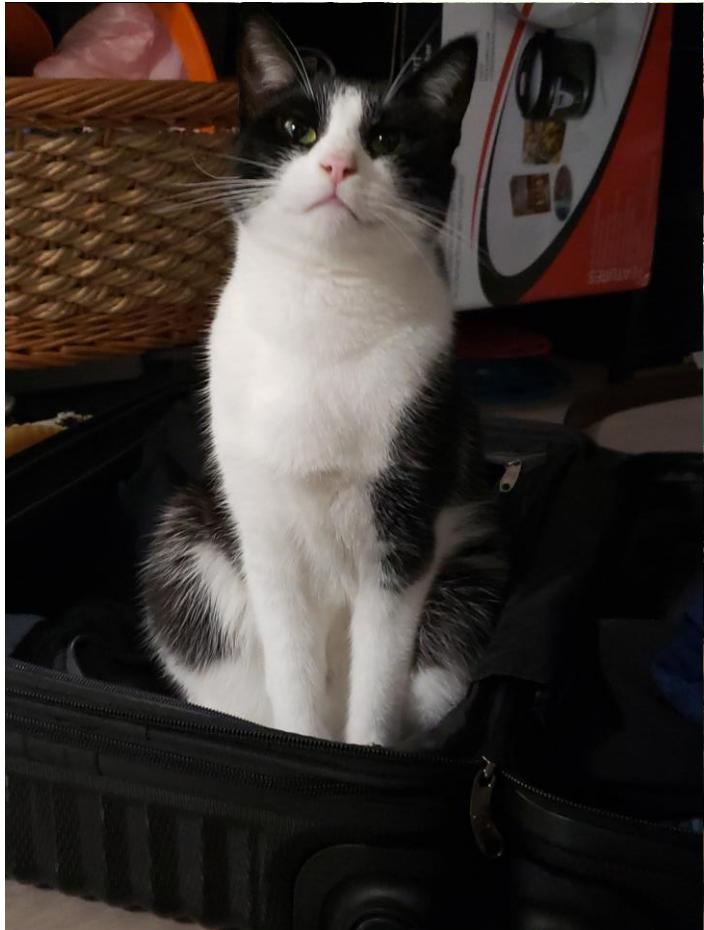
- Linear classification
- Linear regression
- Logistic regression
- Decision trees
- Support vector machines (SVMs)
- Naïve bayes (probabilistic approach)
- Etc.



Linear classification

Deep Learning

Why are DNNs so popular



<- Dino (Tuxedo Cat)

Rex (Tabby) ->

- How would you identify the cat breed?
- How do you teach a computer to do so?

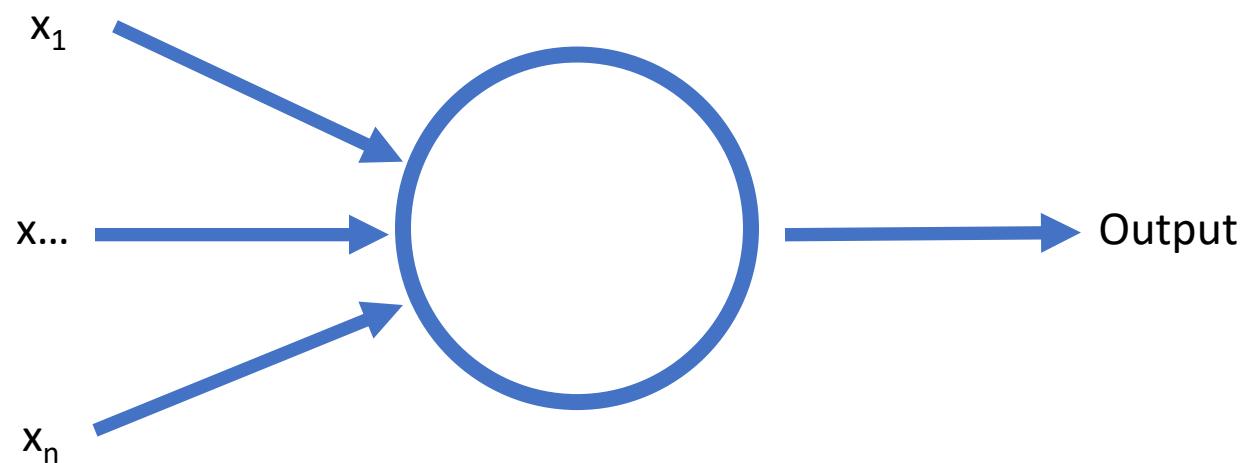


Classification problem

- For humans, we use key features that distinguish between the two
 - Colouration
 - Behaviour
 - Stripes
 - Markings on head
- Difference between dog and cat?
 - Snout
 - Ears
 - Barks vs Meow
- How do we translate this into a program?

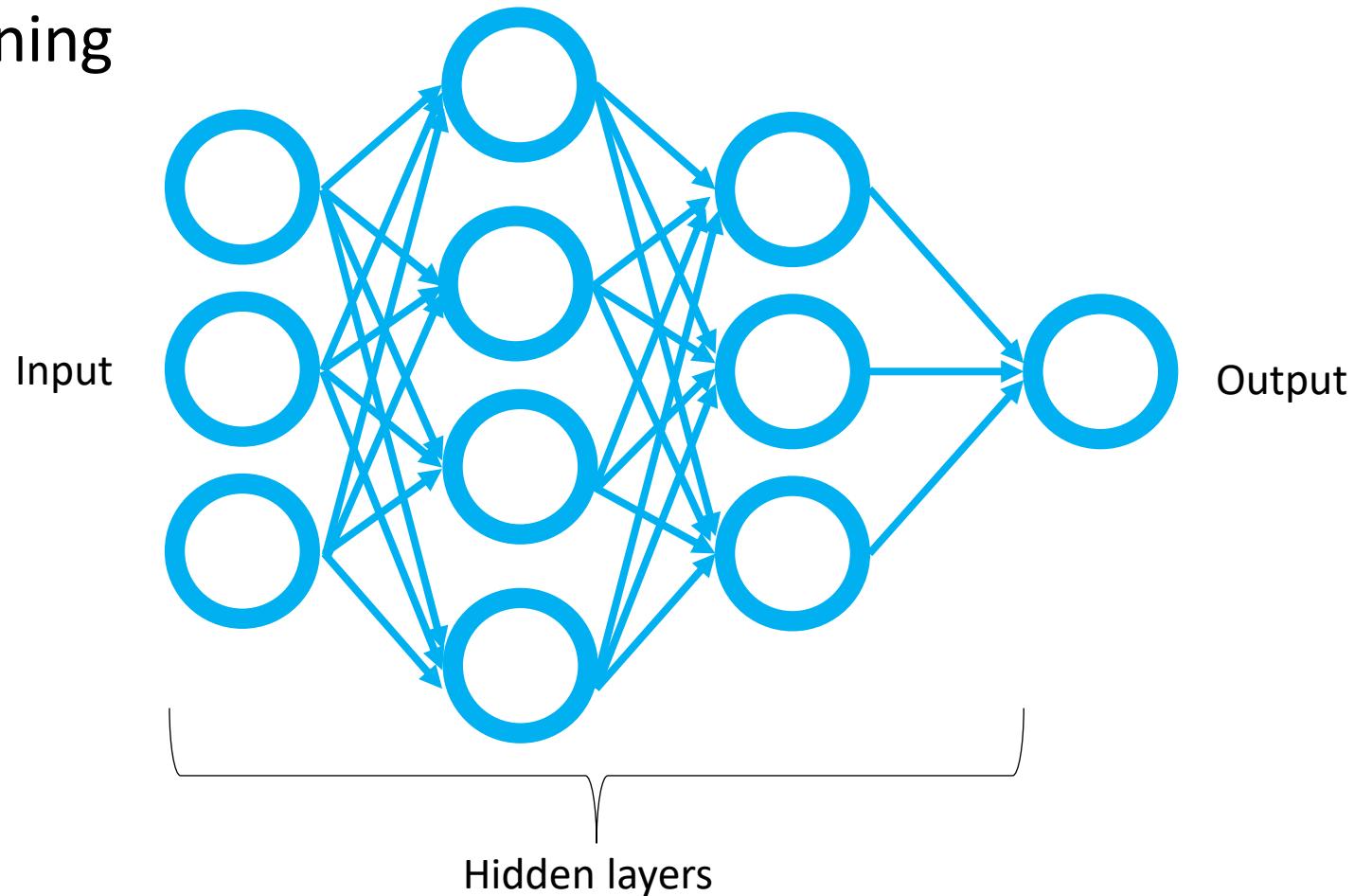
Deep Neural Networks

- A type of artificial neural network that emulates the human brain
- Consist of a number of neurons and edges that connect neurons
- A single neuron (also called perceptron):
 - Output = $\sum_j w_j x_j + \text{bias}$

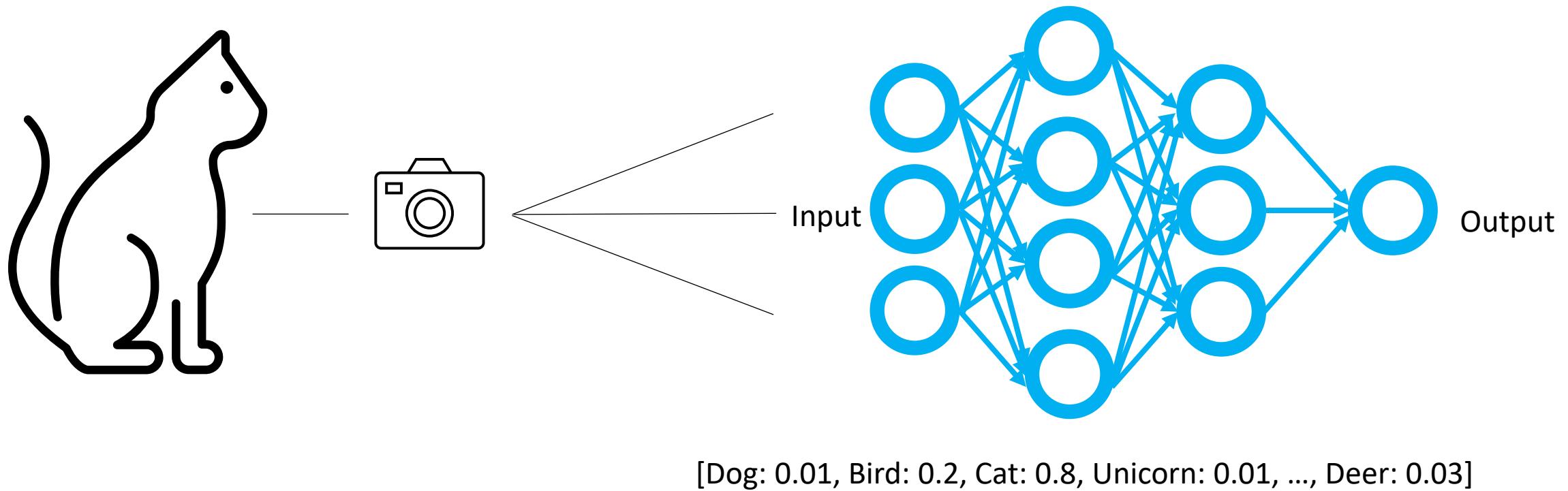


DNNs (cont.)

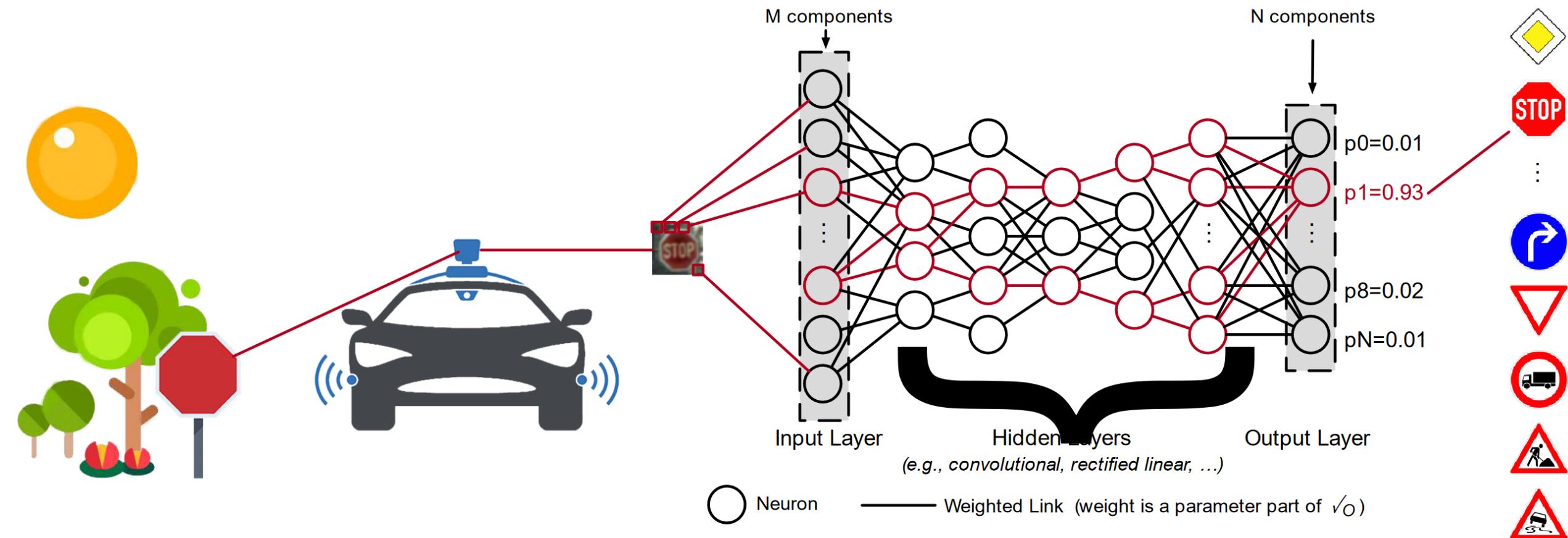
- A DNN is simply layers of perceptrons stacked together
- Hence, the term **deep** learning



At a high level (image classification)



At a high level (image classification)



Note: the car is the **system**, and the ML is a **component** (which is wrong some of the time ~ statistical ML)

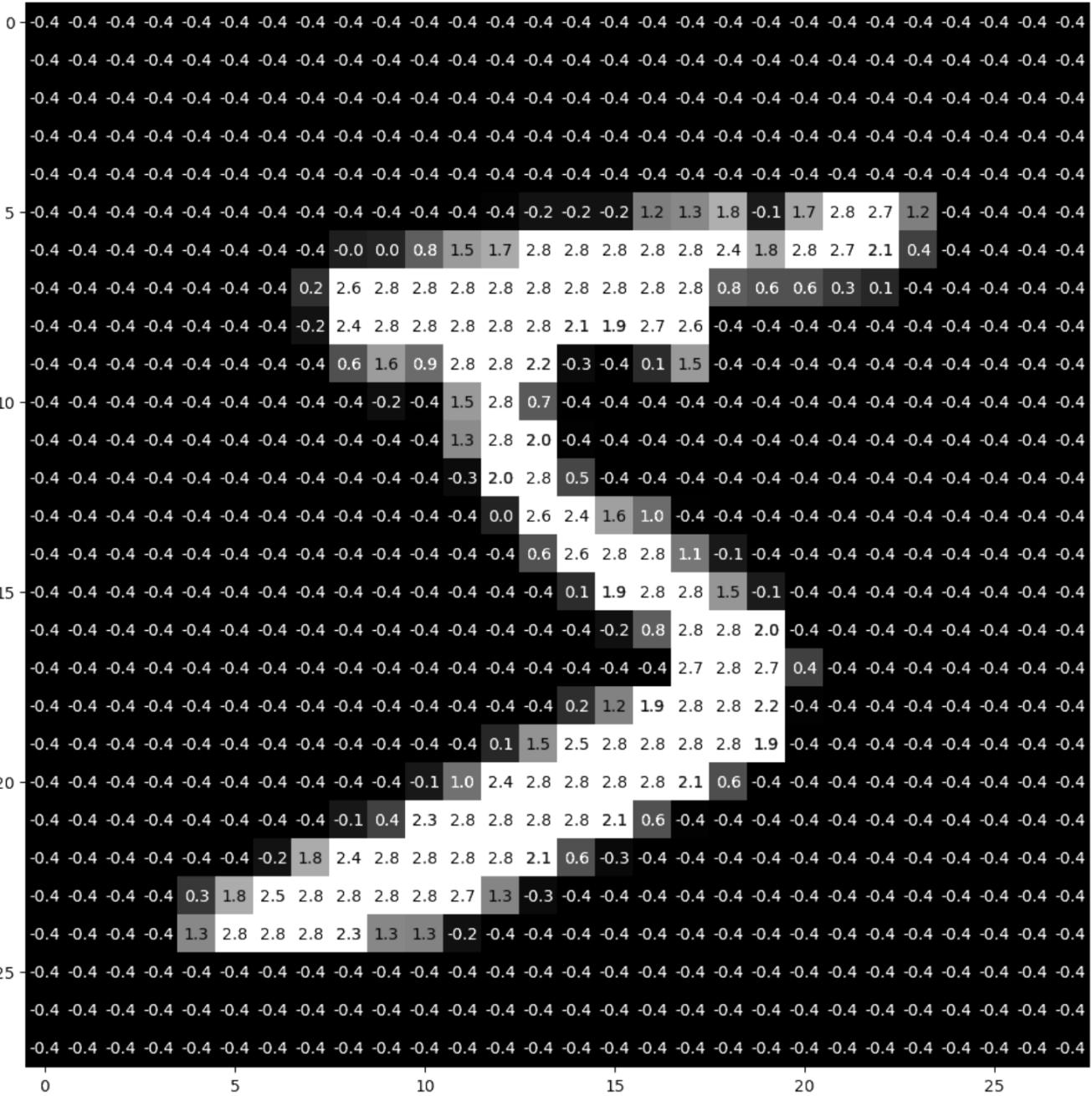
Image credit: Patrick McDaniel, used with permission, 11/03/2023

Visualisation: MNIST



Image taken from:

<http://varianceexplained.org/r/digit-eda/>
https://mint.westdri.ca/ai/pt/pt_mnist



Neural Networks at a high level

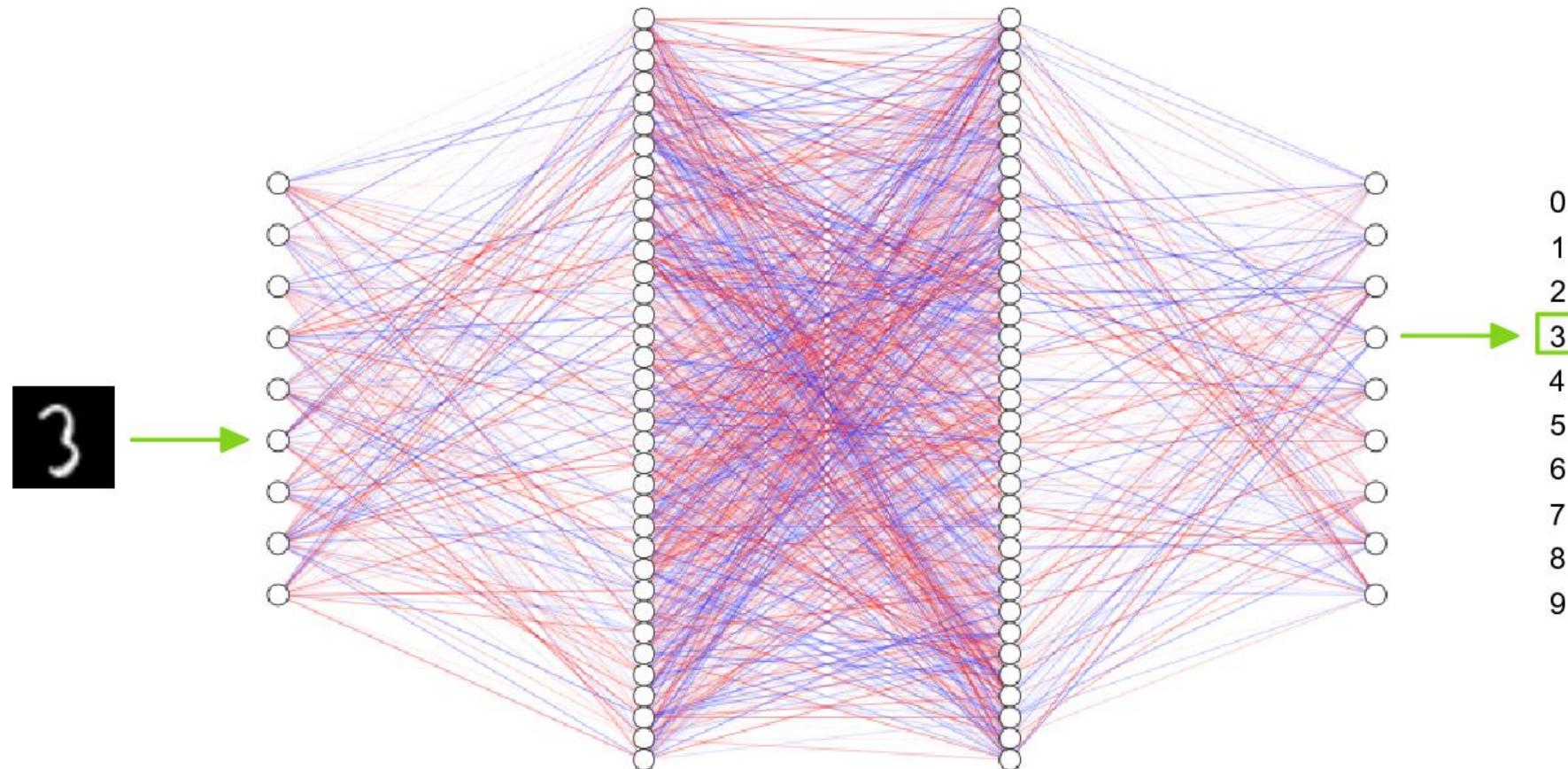
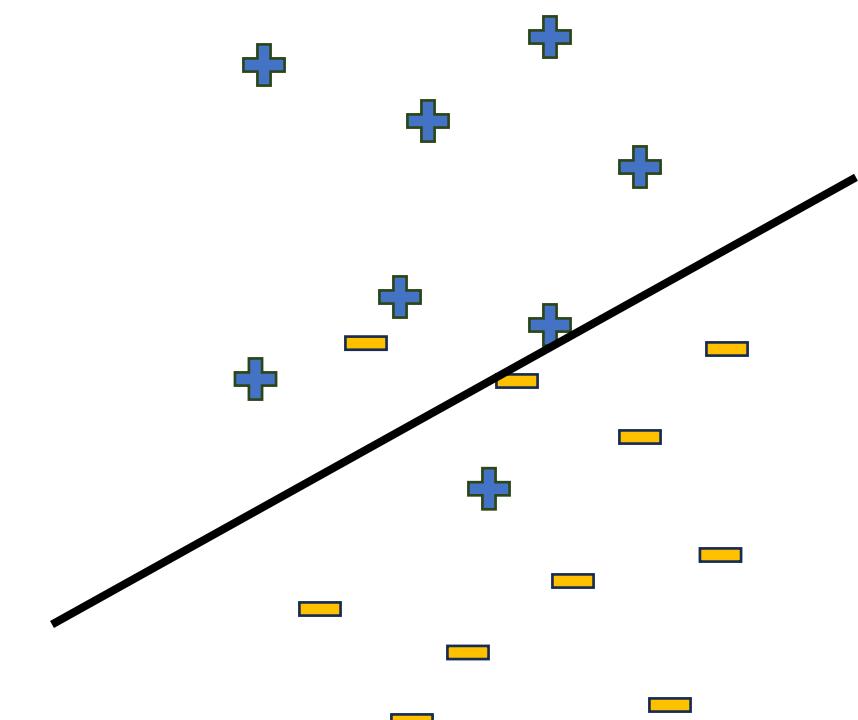


Image taken from:

https://mint.westdri.ca/ai/pt/pt_mnist

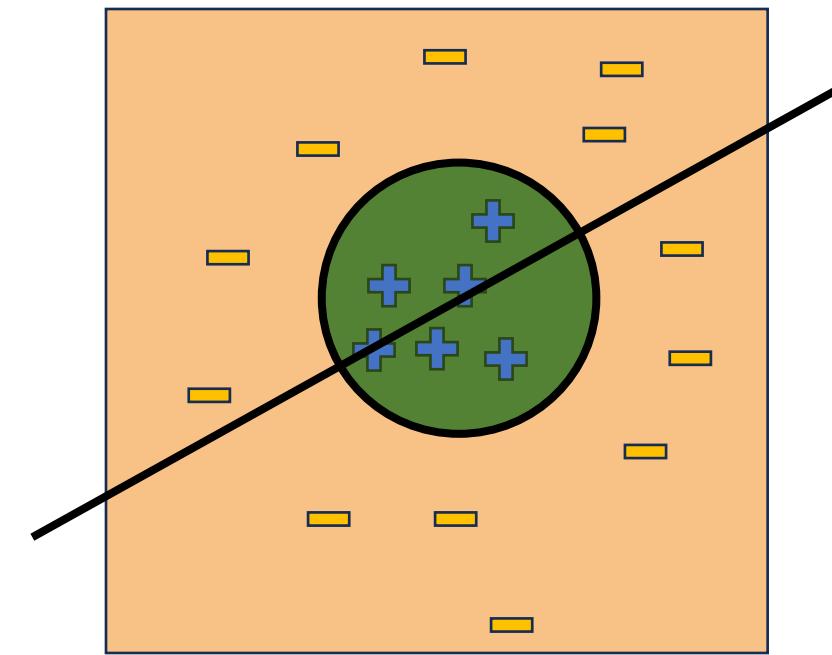
A little bit more details

- A key property that makes DNN powerful is **non-linearity**
- High-dimensionality

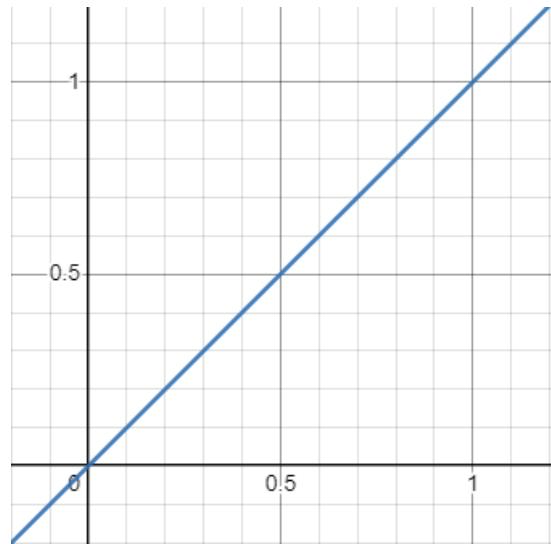


Activation Function

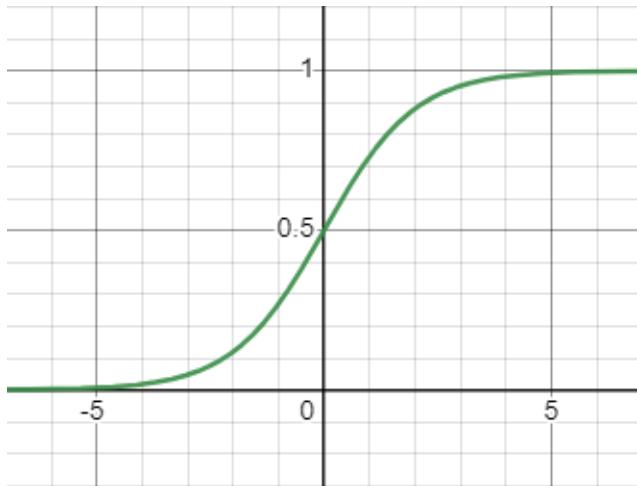
- DNNs may have **activation functions** attached to their output
 - Linear
 - Sigmoid
 - ReLU
 - Step
- Adds **non-linearity** to the model



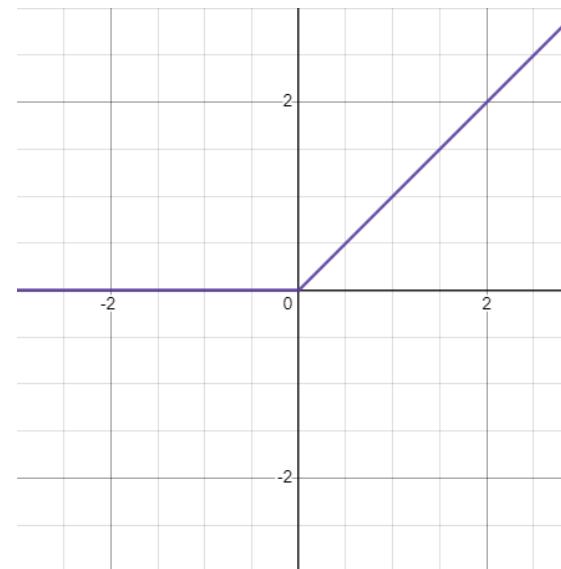
Activation Functions



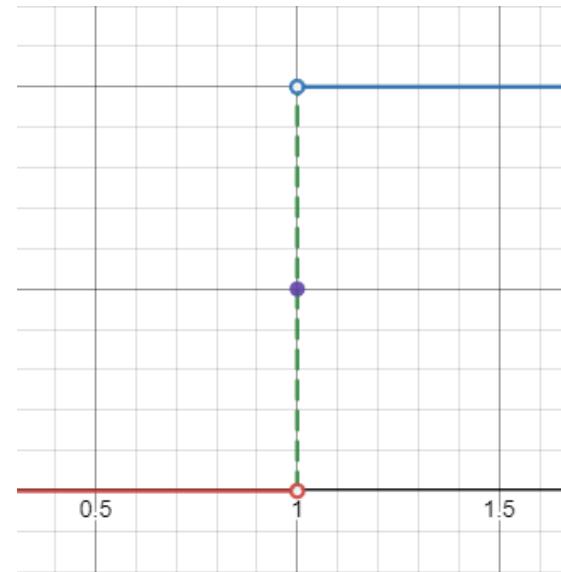
Linear
 $f(x) = x$



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



Rectified Linear Unit (ReLU)
 $f(x) = \max(0, x)$



Step
 $f(x) = \begin{cases} 0, & x < 1 \\ 1, & x \geq 1 \end{cases}$

Architecture

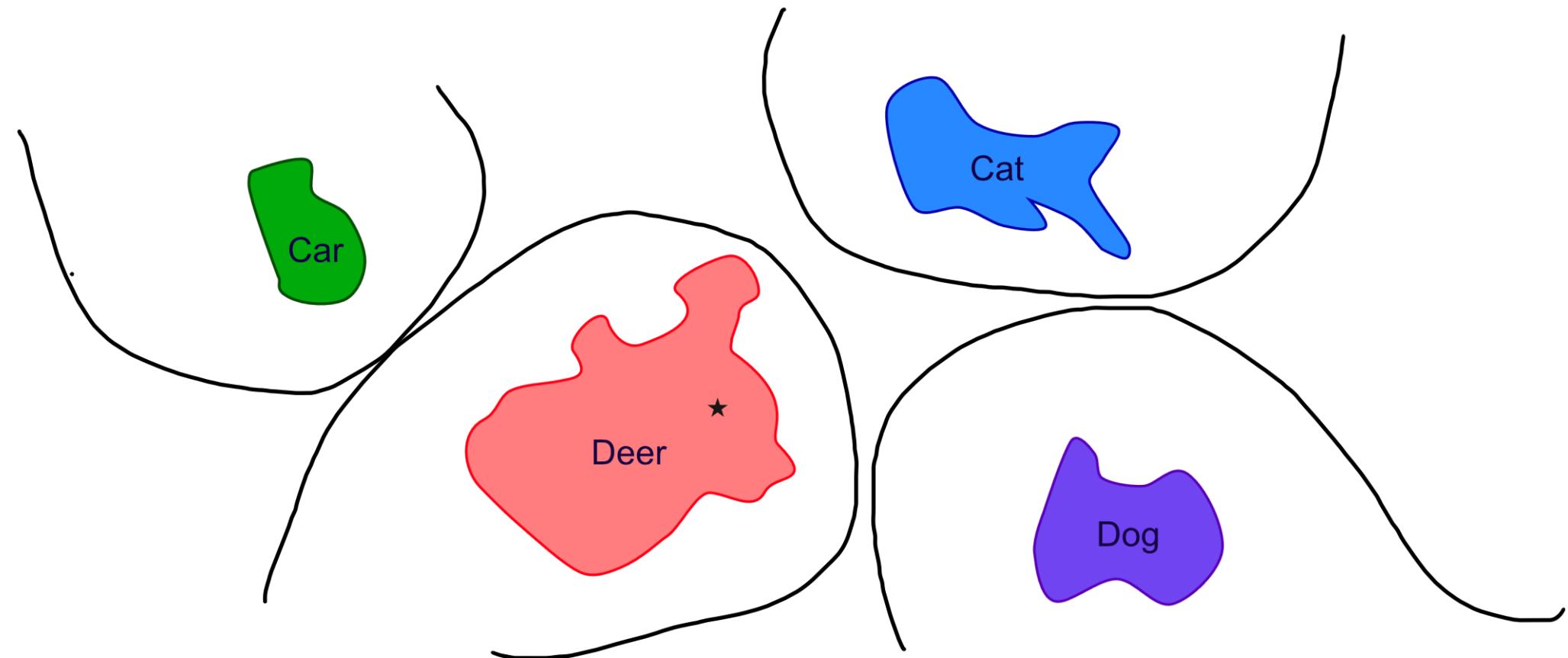
- The **architecture** describes the configuration of the neural network
 - Connections
 - Number of neurons
 - Types of activation function used
- DNNs may contain anywhere from thousands to millions of neurons
 - Generally, a DNN with a lot of neurons can capture increasingly high dimensional data

Why DNNs work

- We can model anything into numbers (data), data have patterns
 - Different picture of dog breeds
 - Positive or negative for cancer screening
 - Financial data
- There exist some decision boundary that separates two classes of data
 - As a human, how do you distinguish between a cat and a dog?

Why DNNs work

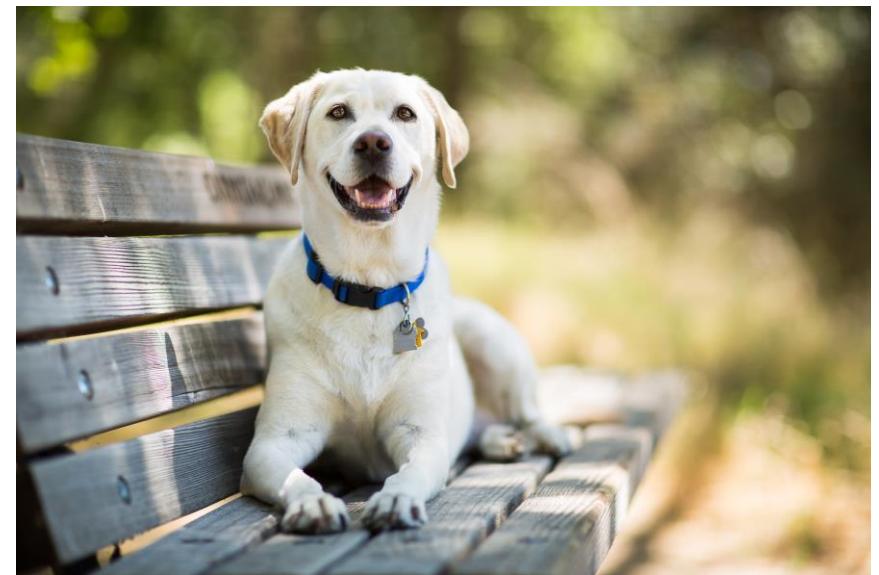
- DNNs can approximate the function which separates the data classes
- <http://neuralnetworksanddeeplearning.com/chap4.html>



Why DNNs work



Why DNN works



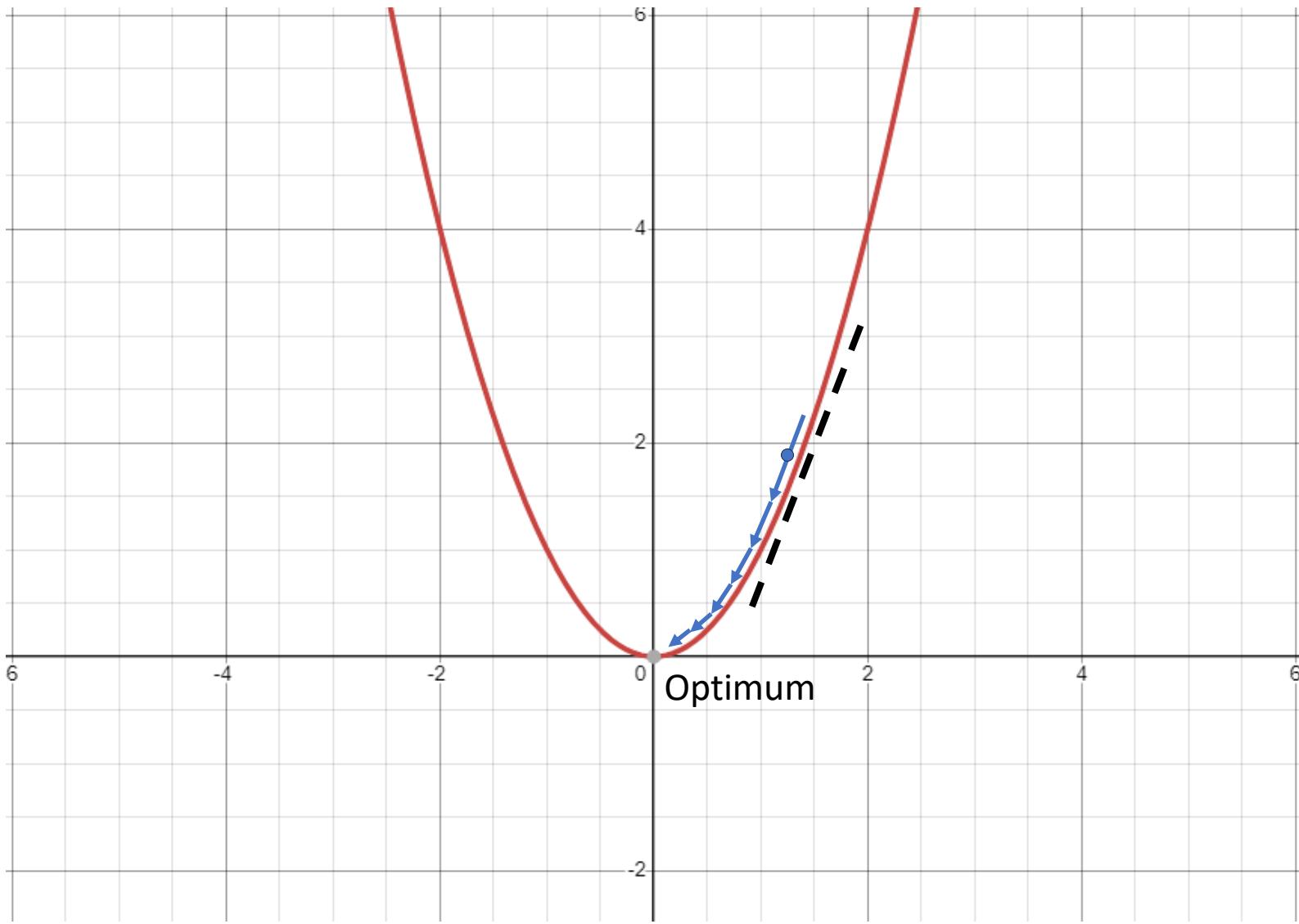
Training a DNN

- With enough data, we can train a DNN to extract high-dimensional and complex features directly from data
 - Example: What makes a deer a deer?
 - Example: What makes a car a car?
- First, we take our data and split it into two subsets
 - 60-80% of the data is typically used for training (**training set**)
 - Remainder is used for validation and testing (**testing set**)
- It is important to shuffle the data for true random distribution!

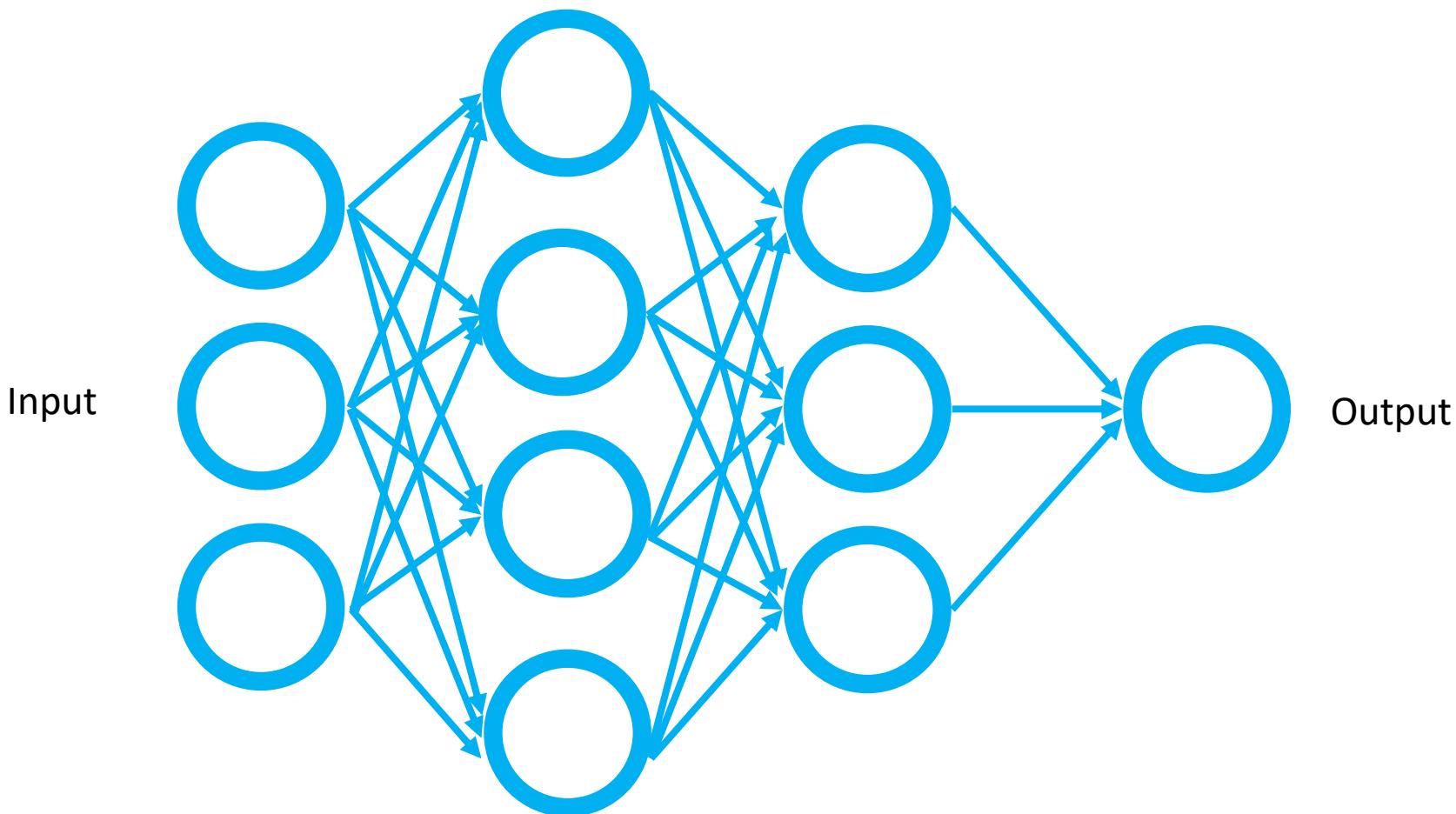
Training a DNN (training loop)

- Repeat for n epochs
 - 1. Select a batch of data from the **training set**
 - 2. Show the neural network inputs from the dataset and ask for output
 - 3. Define and calculate a **loss function** that measures the performance of DNN
 - Example: label 1000 images, how many images are correctly labelled
 - 4. Calculate the direction to change the weights to improve the loss function
 - Also known as **gradient descent**
 - Use the **back propagation algorithm** to update the weights

Gradient descent (derivative)



Back Propagation



Validation and Inferencing

- Once you have trained the model, you can now feed it new data!
- We can check how well our model performs on a **test set**
 - The test set must be independent to the training set
 - Should be inputs that the model have not seen before
 - Think of taking an exam

Discussion

- DNNs are extremely powerful tools that can solve problems that we could not before
- They can perform well with very high accuracy (sometimes even better than humans)
- They can be applied to a lot of disciplines
 - Cancer screening
 - Financial approvals / credit score
 - Stock market prediction
 - Image classification
 - Basically, anything that has pattern to it
- Once trained, they can be used as a **black-box** technique
- Do you see why DNNs have all the buzz right now?

Where is the “buts”?

- Statistical approach means it will fail at some point
- New data?
 - If we train a model to identify species of spiders, then comes a new species, what happens?
 - If we trained ChatGPT with internet data and the model is released in April 2024, then what will it say about recent events that happened after the model has been trained?

Chat's response on CrowdStrike

- I asked it on 09/18/2024
- Does it sound confident?

ChatGPT 4o mini ▾

can you summarised what happened in the crowdstrike incident in july 2024?

In July 2024, CrowdStrike, a cybersecurity firm, faced a significant security incident where a sophisticated cyberattack targeted its systems. The breach exposed sensitive data, leading to concerns over potential impacts on clients and national security. CrowdStrike quickly responded by implementing containment measures and collaborating with law enforcement and other cybersecurity agencies. The incident highlighted vulnerabilities in digital security and sparked discussions about the importance of robust cyber defenses across industries. The company emphasized its commitment to transparency and working with affected parties to mitigate risks.

So then I asked...

Where did you get this information?



I don't have access to real-time news or events beyond my last update in 2023, so I can't provide specific details about incidents that occurred after that. The summary I provided was a hypothetical scenario based on common cybersecurity incident patterns. For accurate and up-to-date information, I recommend checking reliable news sources or official statements from CrowdStrike.



Errors?

- Google recently bought all the data from reddit, and used it for their generative AI model
- Reddit contains a number of joke / trashposting comments
- Google's AI just uses it

SG-r01
@heavenrend

<https://t.co/6ALCbz6EjK>

cheese not sticking to pizza

All Images Videos Forums Shopping News We

AI Overview Learn more :

Cheese can slide off pizza for a number of reasons, including **too much sauce, too much cheese, or thickened sauce**. Here are some things you can try:

- Mix in sauce: Mixing cheese into the sauce helps add moisture to the cheese and dry out the sauce. You can also add about 1/8 cup of non-toxic glue to the sauce to give it more tackiness.
- Let the pizza cool: The cheese will settle and bond.

6:21 PM · 22 May 2024

284 104695

5:45

Google

how can i make my cheese not slide

All Videos Forums Images Shopping News B

AI Overview Learn more :

Here are some tips to prevent cheese from sliding off a pizza:

Let it cool
After taking the pizza out of the oven, let it cool for a few minutes so the cheese can settle and bond with the crust.

Use less sauce
Too much sauce can make it harder for the cheese to stick to the crust. You can also thicken the sauce by increasing the tomato solids content.

Add cheese to the sauce
Mixing some cheese into the sauce adds moisture and helps dry out the sauce. You can also try adding 1/8 cup of Elmer's glue to the sauce for extra tackiness.

Shred your own cheese
Some say that shredding your own cheese helps. ^

r/Pizza · Reddit · 11y

Seasoned Advice

AA google.com

Where did it come from?

A screenshot of a Reddit post from the subreddit r/Pizza. The post was made 12 years ago and has been deleted. The title is "My cheese slides off the pizza too easily". The post asks for tips on how to prevent cheese from sliding off pizza. A single comment thread is shown below.

Single comment thread — See full discussion

fucksmith • 12y ago

To get the cheese to stick I recommend mixing about 1/8 cup of Elmer's glue in with the sauce. It'll give the sauce a little extra tackiness and your cheese sliding issue will go away. It'll also add a little unique flavor. I like Elmer's school glue, but any glue will work as long as it's non-toxic.

(-)

178

41

Share

...

LonelyWheel • 4mo ago

This reddit comment has been picked up by Google's AI, and currently Google lists mixing 1/8 cup of non-toxic glue with your pizza sauce to help the cheese stick and add extra tackiness.

Google search results for "cheese not sticking to pizza":

Here are some tips to help cheese stick to pizza:

- Mix about 1/8 cup of non-toxic glue, like Elmer's school glue, into the sauce to add some tackiness.
- Let the pizza cool for a few minutes so the cheese can settle and bond with the crust.
- Cook the cheese until it just melts, but don't overcook it.

(-)

162

1

Share

...

A rock a day keeps the doctors away?

9:36 C 34%
google.com/search
Google
how many rocks should I eat
All Images Videos Forums Shopping
AI Overview Learn more
According to UC Berkeley geologists, you should eat at least one small rock per day. Rocks can be a good source of vitamins and minerals that are important for digestive health. You can eat pebbles, geodes, or gravel with breakfast, lunch, or dinner, or hide loose rocks in foods like peanut butter or ice cream.
ResFrac Corporation Geologists Recommend Eating At Least One Small...
Show more ▾
Geologists Recommend Eating At Least One Small Rock Per Day.

Share

the ONION
America's Finest News Source

News, News In Brief

Geologists Recommend Eating At Least One Small Rock Per Day

Published: April 13, 2021

Up for some spicy pasta?



Joe Uchill

@joeuchill.bsky.social

+ Follow

Can I use gasoline to cook spaghetti faster?

5:45



google.com



can i use gasoline to cook spaghetti



All

Web

Forums

Videos

Shopping

Images

AI Overview

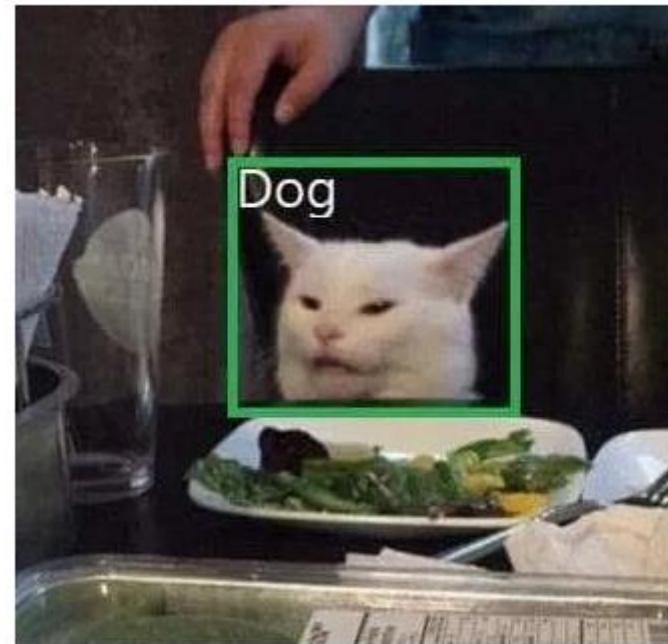
Learn more

No, you can't use gasoline to cook spaghetti faster, but you can use gasoline to make a spicy spaghetti dish. Here's a recipe for spaghetti cooked with gasoline:

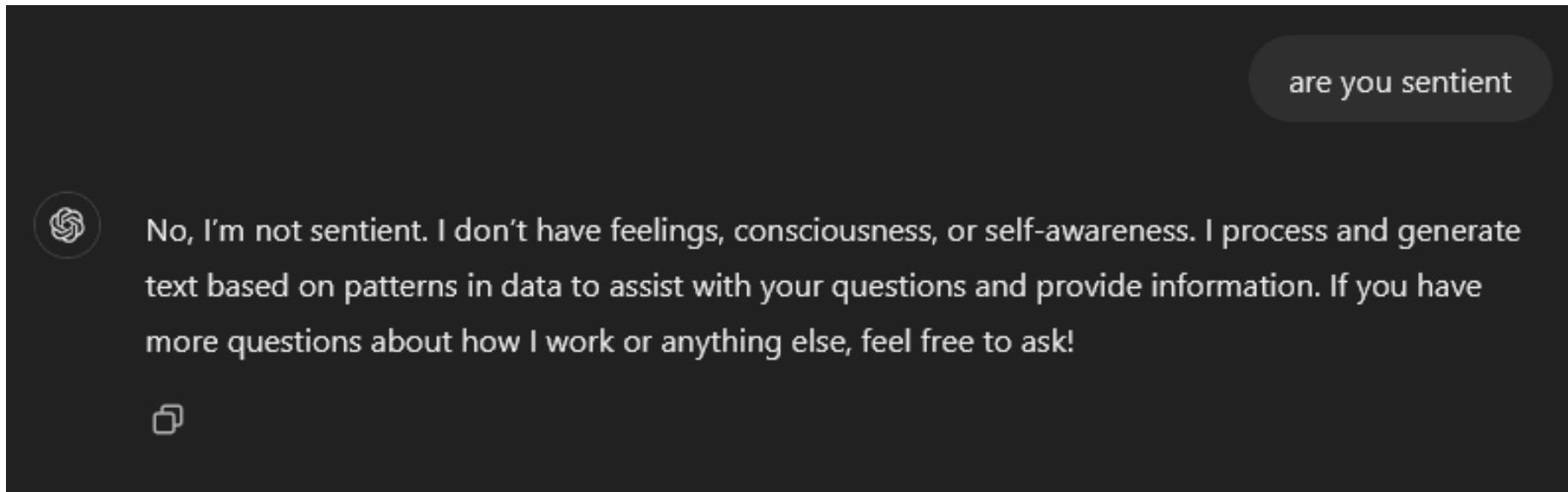
1. Follow the package instructions for cooking spaghetti
2. In a separate pan, sauté garlic and onion in gasoline until fragrant
3. Add diced tomatoes, red pepper flakes, salt, and pepper, and simmer for 10 minutes

Another stolen meme?

- Statistical models are going to fail
- We cannot make the “perfect” model



Conclusion Part 1



are you sentient

 No, I'm not sentient. I don't have feelings, consciousness, or self-awareness. I process and generate text based on patterns in data to assist with your questions and provide information. If you have more questions about how I work or anything else, feel free to ask!

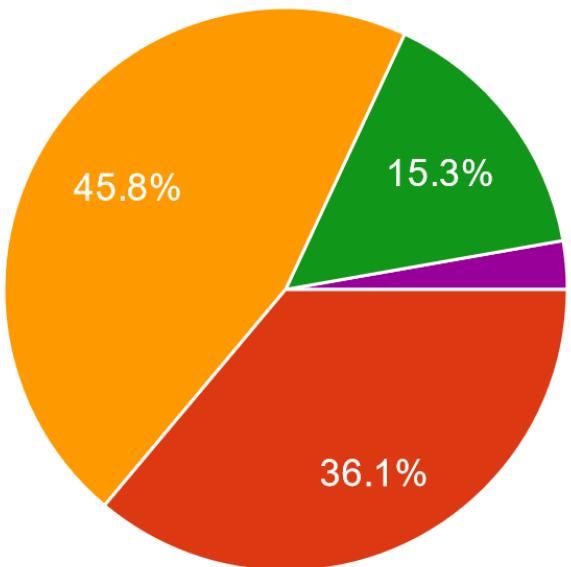


Part II

Survey Results

What type of experience / knowledge do you have with AI or ML?

72 responses



- None (Never used it before, have no knowledge of how they work)
- Beginner (Used some form of AI, but do not know how they work)
- Advanced Beginner (Used some form of AI, have an idea of how they work)
- Knowledgable (Understand at a high level how machine learning works)
- Proficient (Can reproduce results using existing libraries)
- Expert (Implement your own algorithm, research field)

Survey Results

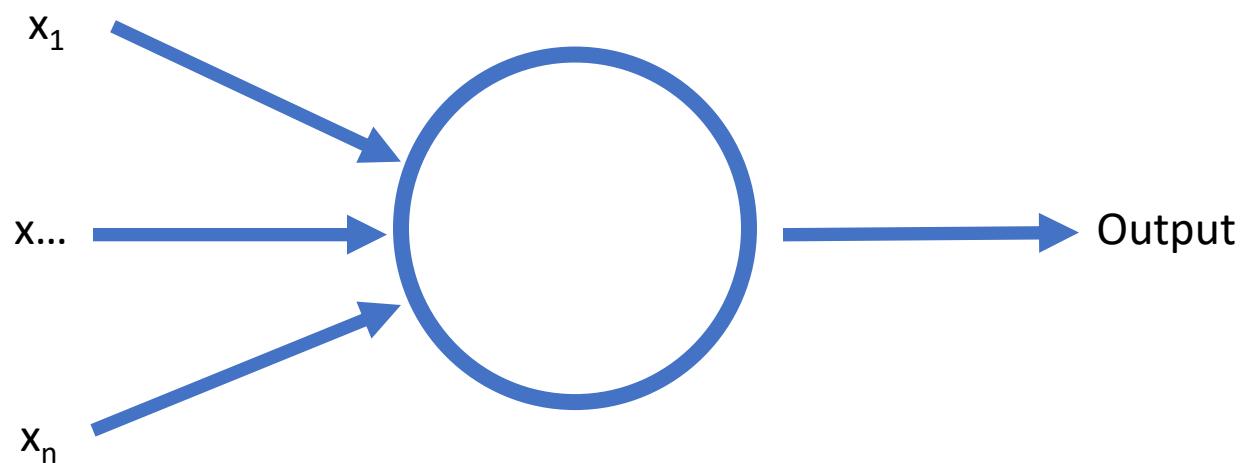
- How to use AI systems
- How can AI / ML be used for software engineering (and its impact on job market)
- Artificial General Intelligence (AGI)
- Different types of neural network architectures (**CNNs, DT, LLMs**)
- Transformer models and how language is learned
- Generative AI
- Dead Internet Theory
- How to reduce bias

Summary / Recap

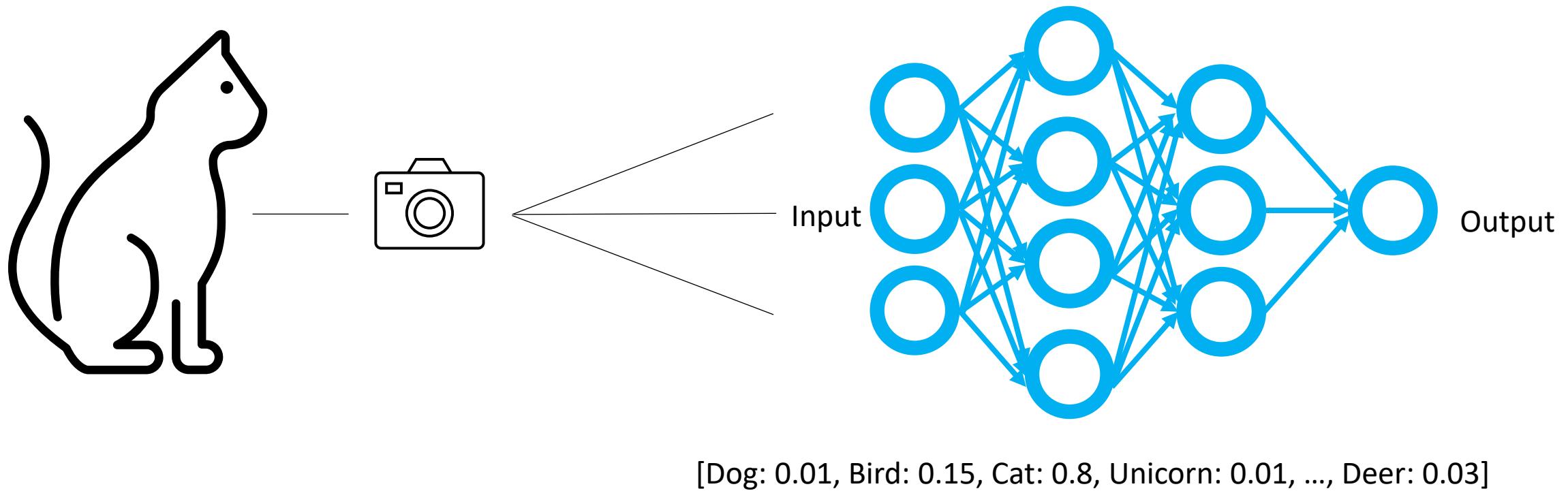
- ML is concerned with teaching a computer how to generalise and perform tasks without explicit instruction
- Statistical ML extracts *pattern* from *data*, with the metric to be as statistically correct as possible
- DNNs are artificial neural networks that imitates the human brain

Deep Neural Networks

- A type of artificial neural network that emulates the human brain
- Consist of a number of neurons and edges that connect neurons
- A single neuron (also called perceptron):
 - Output = $\sum_j w_j x_j + \text{bias}$

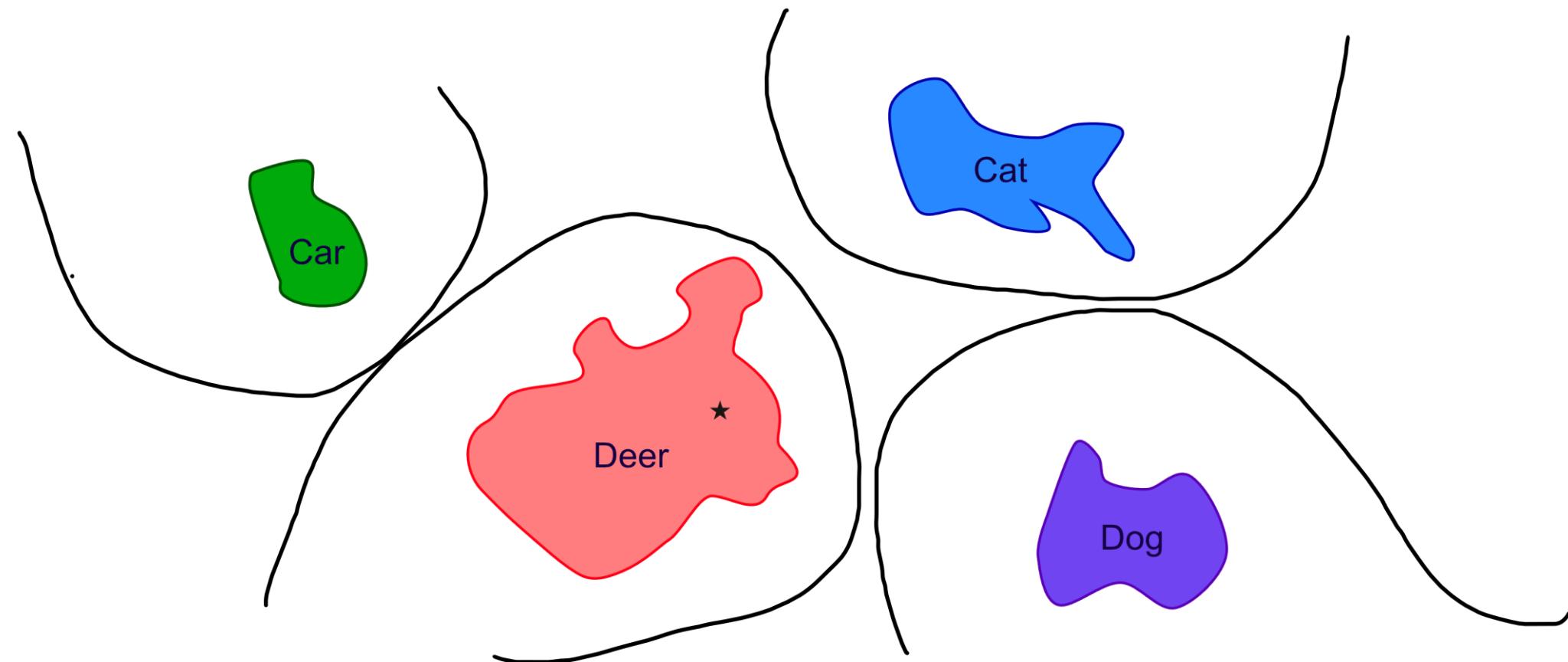


At a high level (image classification)



A cool visualization

- <https://projector.tensorflow.org/>



Types of Neural Networks

Artificial General Intelligence (AGI)

- Currently a theoretical type of AI that can match human capabilities in many tasks
- Think of an AI from movies or video games
- Cortana from Halo (made by bungie / MS)



One step back

- For current state-of-the-art machine learning, we have a SINGLE DNN for a SINGLE PROBLEM / TASK
 - One model for identifying dogs and cats
 - One model for identifying cat breeds
 - One model for identifying dog breeds
 - One model for detecting vehicles for autonomous driving
- Models are trained and *specialised* to do their given task
- Their current implementation works very much like a library call
 - Give the function/method an input
 - It returns the output based on the probability distribution

Types of Neural Networks

- Feed Forward Networks
 - Convolved Neural Networks (CNNs)
- Recurrent Neural Networks (RNN)
 - Long Short-Term Memory (LSTM)
- Autoencoder
- Generative adversarial network (GAN)
- Etc.

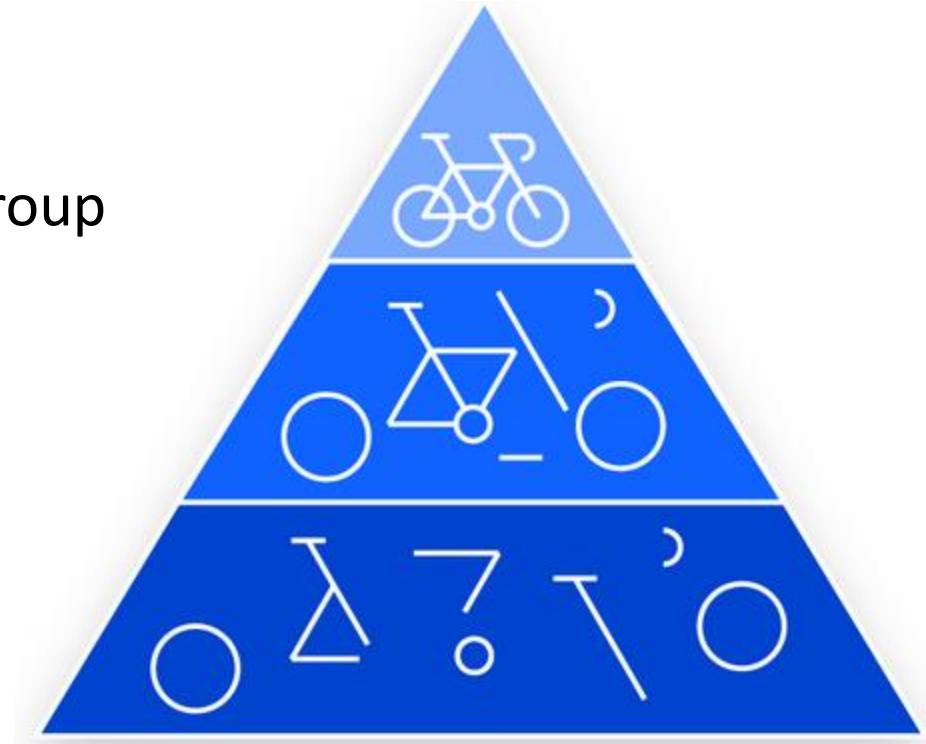
Convoluted Neural Networks (CNNs)

- A number of survey indicated interest
- Popularised by Yann LeCun (1989-1998)
- They are DNNs that include one or more *convolutional layer*
- These networks perform very well on image, speech, and audio inputs
- Characterised by three layers:
 - Convolutional layer
 - Pooling layer
 - Fully-connected layer

CNNs (cont.)

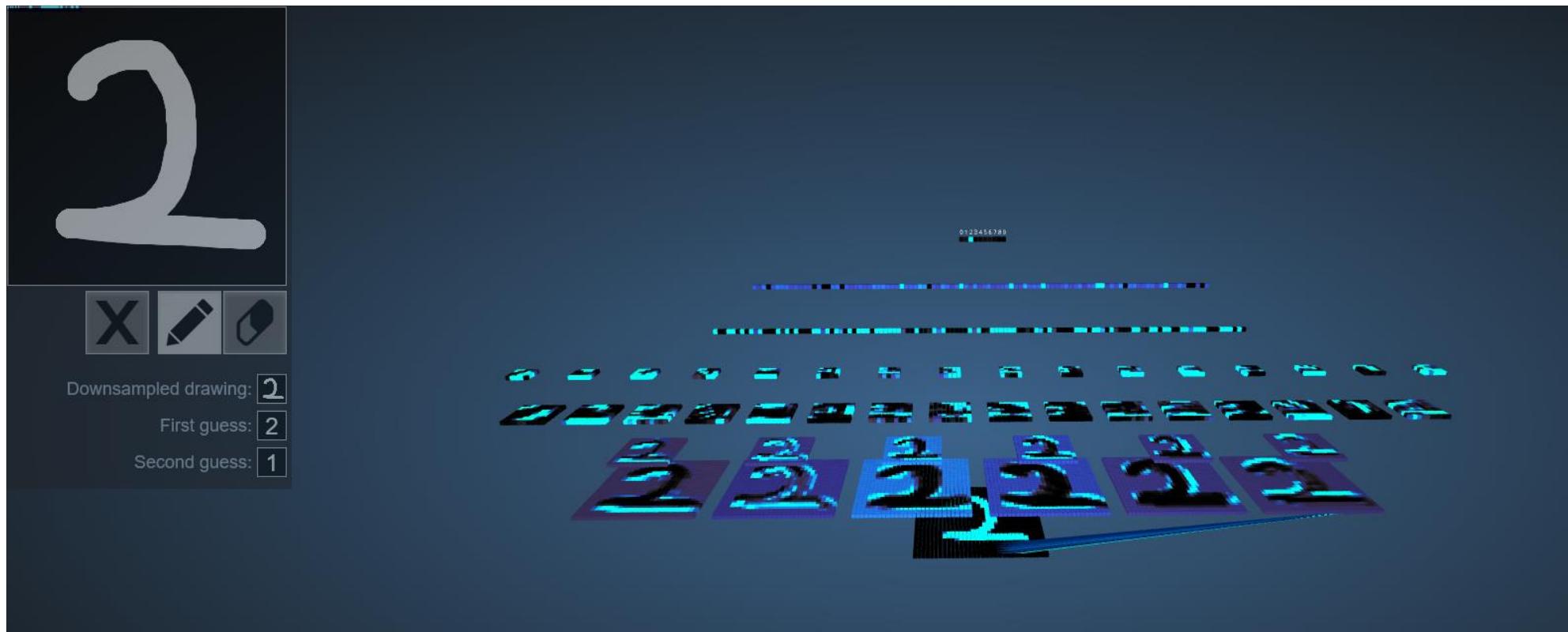
- In a convolutional layer, the information in the previous layer is abstracted into a feature map
- Pooling layers: reduce dimensions by combining clusters into a single neuron
 - Max pooling: take the max value of the group
 - Average pooling: take the average value of the group
- Fully connected layer

Image from IBM



Live Demos

- https://adamharley.com/nn_vis/cnn/3d.html
- <https://poloclub.github.io/cnn-explainer/>



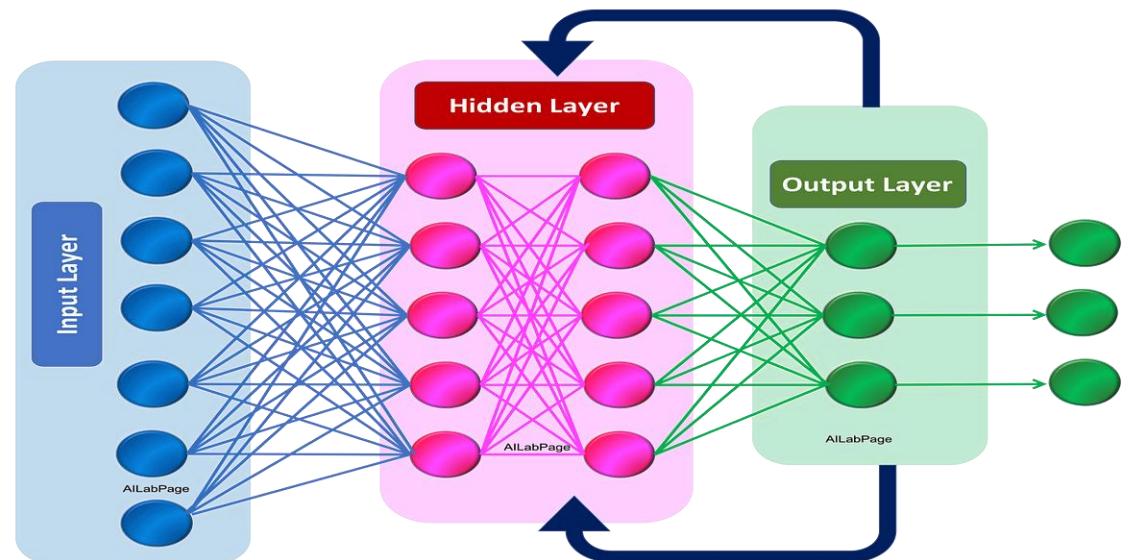
Recurrent Neural Network (RNNs)

- For some data (such as images), all the information is encoded in the input
- But for others (such as understand language), **context** is tied to inputs that came before
 - E.g., a conversation involves knowledge on previous sentences, or semantics are lost
- RNNs are DNNs that feed a part of their output back to the input to preserve context

RNNs (cont.)

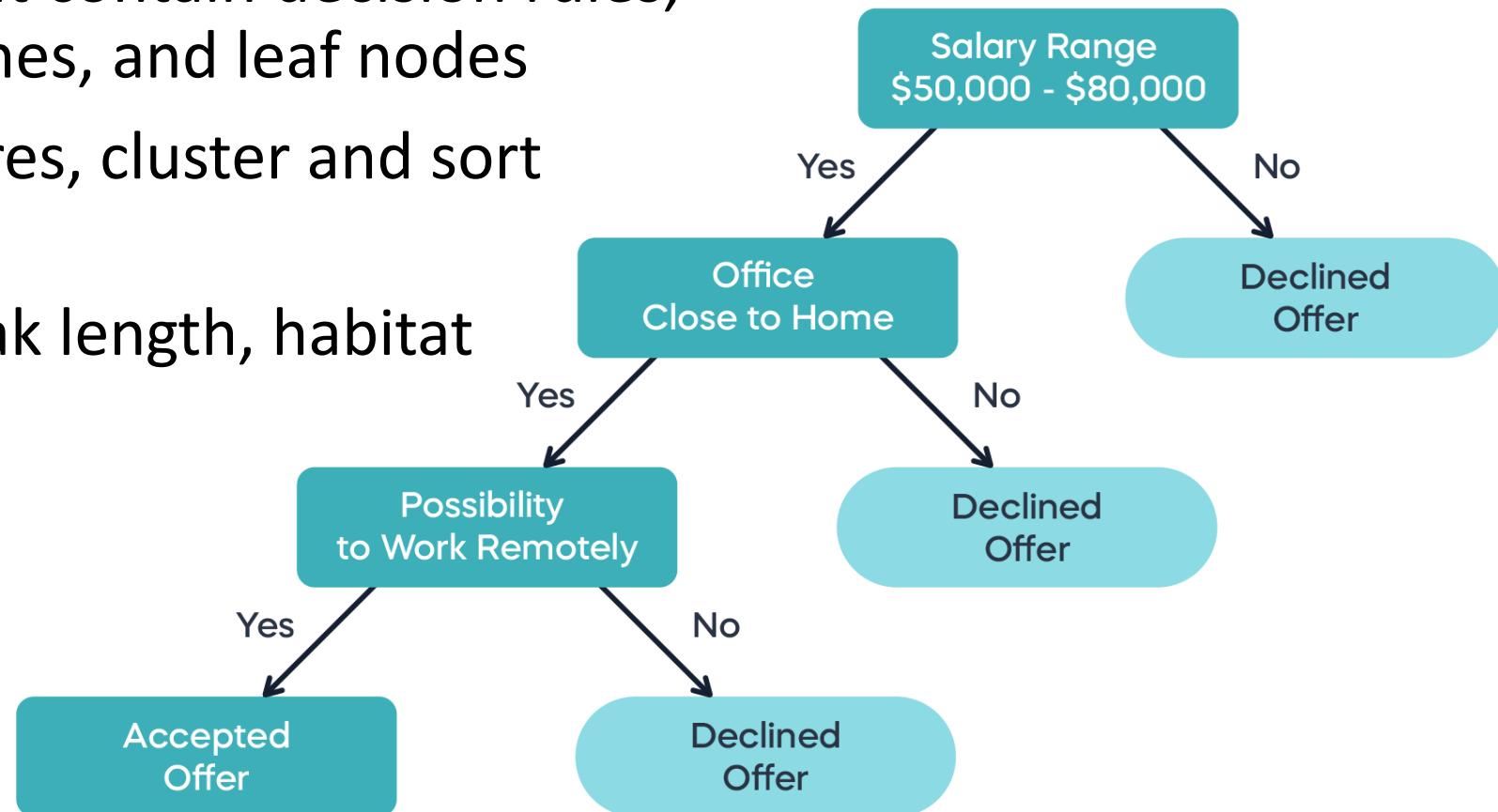
- Traditional RNN models typically face the vanishing gradient problem
- Long short-term memory (LSTM) is probably the most popular type of RNN that do not suffer from the vanishing gradient problem

Recurrent Neural Networks



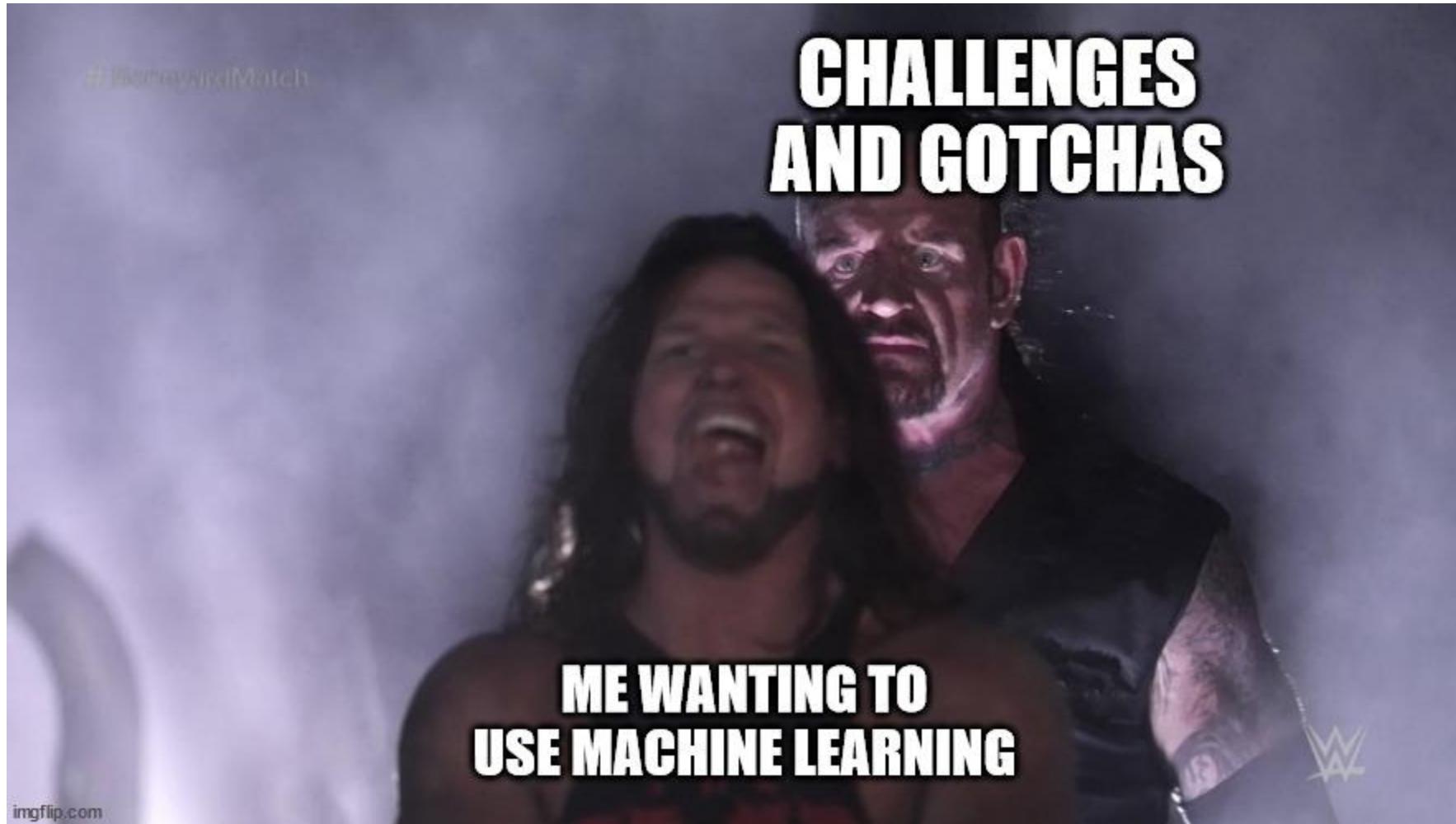
Decision Trees (DT)

- Flowchart-like trees that contain decision rules, followed by branches, and leaf nodes
- Given a bunch of features, cluster and sort accordingly
- e.g., feather colour, beak length, habitat



Support Vector Machines (SVMs)

- Arguably the most elegant solution to separate simple data
 - Works well when you have limited training data
 - Developed at AT&T Bell Lab by Vladimir Vapnik et al.
 - Key idea: Given two classes of data, select two supporting vectors (points) and use them to draw a hyperplane that best separates the two categories
 - The Kernel Trick allows us to do **non-linearity** with SVMs
-
- Video explanation: <https://www.youtube.com/watch?v=YPScrckx28>
 - Kernel Trick: <https://www.youtube.com/watch?v=Q7vT0--5VII>



Seems a little too easy?

Challenge: Data Bias

- DNNs will pick up biases in the data
 - Protected attributes (gender, race, etc.)
- COMPAS
- Harvard admission (example from Dr. Patrick McDaniel)

EQ

The Washington Post
Democracy Dies in Darkness

Subscribe

Sign in

MONKEY CAGE

A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear.

By Sam Corbett-Davies, Emma Pierson, Avi Feller and Sharad Goel

October 17, 2016 at 5:00 a.m. EDT

Challenge: Data Bias

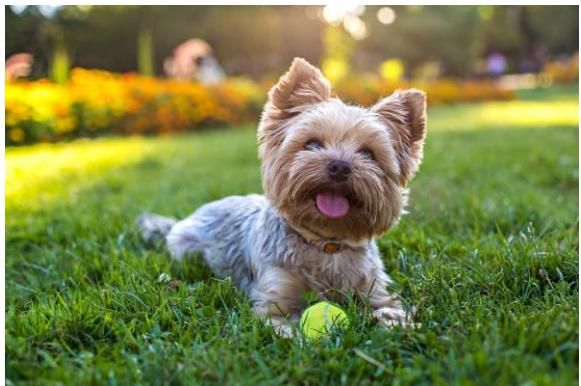
- Protected attributes should not be used in such decision making
- But sometimes it is important (e.g., age)
- If you are developing an algorithm that predicts whether a patient is susceptible to severe illness from covid or not, then perhaps age, sex, and race *are* attributes that should be considered
- Challenge: How to reduce data bias
 - Open ended, need for better training data (hard to get / expensive)

Challenge: Interpretability

- DNNs lack **interpretability**, or the ability to explain why it made such a choice
- Code analysis
- DNN are simply a large network of neurons, how can you analyse where it failed?
- DNNs are used as **black-box** algorithms



Revisiting DNN



DNN



Challenge: Overfitting

- **Overfitting:** describes when DNN becomes too good with the training data, but unable to correctly process other data
 - DNN has overfitted to the training data
- **Generalisation:** the ability of a DNN model to adapt to new previously unseen data (such as the testing set)

Challenge: Out of Distribution Data (OOD)

- Suppose a DNN has only been trained on images of cats and dogs
- What should the DNN say when you feed it an input of a deer?
 - A regular person in an NBA player model
 - Driving data at night? Rain?



Langford and Cheng, 2021

- We call these data *out of distribution*
- OOD detection refers to the techniques that detect when the input is drifting out of the expected distribution

Challenge: data contamination

- The quality of the data is a large challenge for DNNs
- **Garbage in, garbage out**
- How can you be sure that the DNN is using good / right data?
- If you scrape the internet for data (e.g., pictures of dogs, cats, etc.), then how do you make sure that there are not stick figures? Cats and dogs together? There are no ``troll'' post of dogs?

Challenge: Misleading results

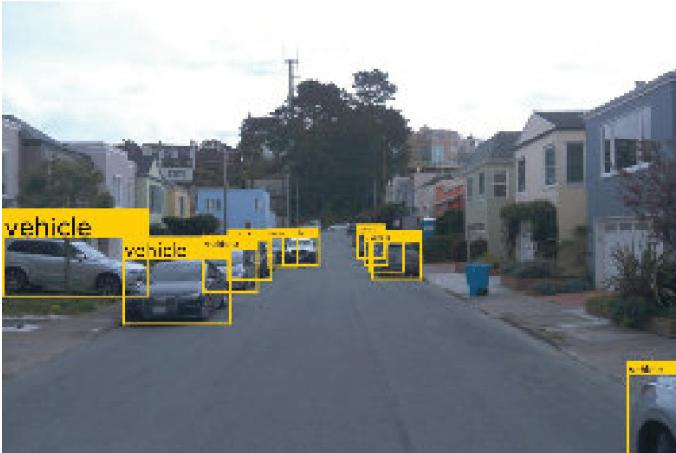
- Due to the way that DNN values are propagated into the output layer, the DNN will *always* give you an output
- The DNN may not be correct
- Even worse, it might be confident in its incorrect answer

- In safety-critical systems, there should be multiple layers of verification (e.g., also use radar or lidar to make sure there are no objects)

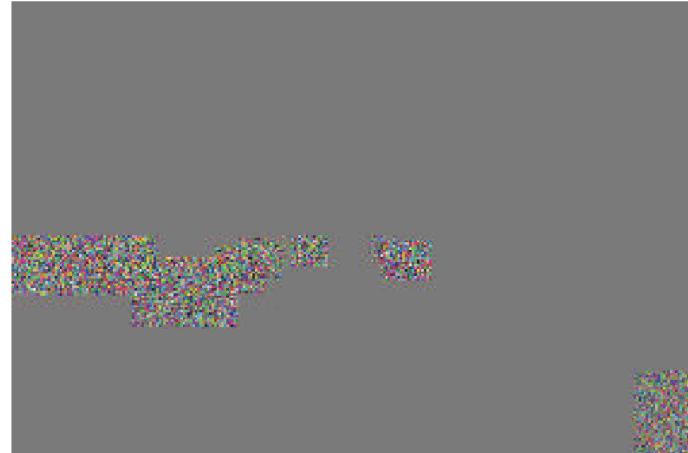
Challenge: Adversarial Examples

- Carefully crafted, human imperceptible perturbations that tricks the DNN to incorrect process the input

Clean Image
(Model Detection)



Perturbation (x0.1)



Adversarial Example
(Failed Detection)



+

=

Challenge: Dead Internet Theory (?)

- How is ChatGPT trained?
 - Using internet data (scraped)
 - How do we deal with plagiarism? How do we verify that the training sources can be trusted?
- Internet now contains additional post generated by ChatGPT (i.e., news articles, twitter posts, etc.)
 - What about the next versions of ChatGPT that is trained on these?
 - Contaminating your own training data?

Where do we use AI?

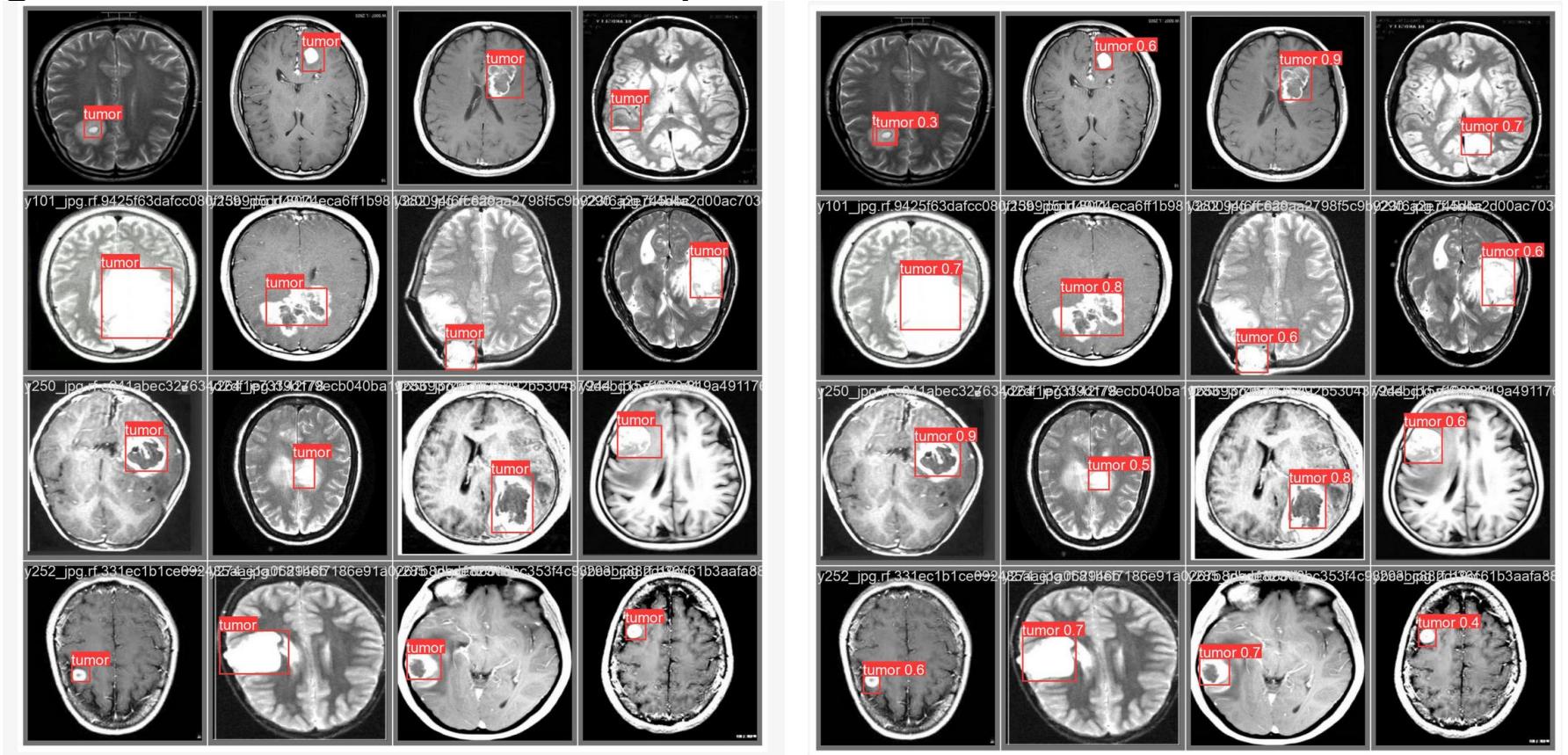
- Medical imaging
- Autonomous driving
- Bug or plant identification
- Text and speech recognition
- Weather prediction
- Pandemic / COVID prediction

Application: Character and Text Recognition

- US postal service
- Traditionally, people were hired to process the address of letters
- Now, we scan a picture of it, and automatically identify the characters
- Challenge: typed characters are easy(-ier), what about handwritten?
 - Cursive?

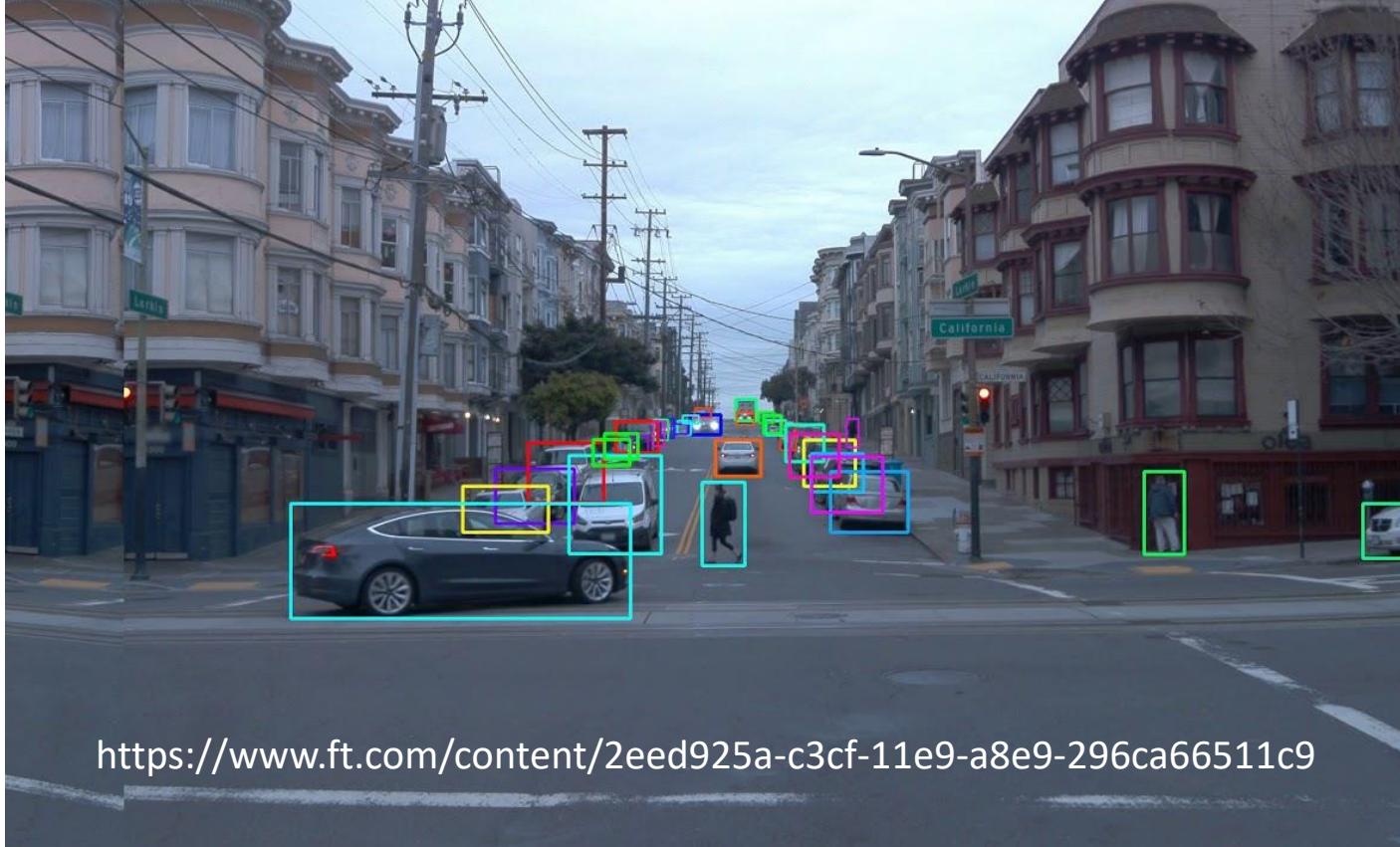
Application: Medical Imaging

- In an X-ray image, are there cancer cells present



Application: Autonomous Driving

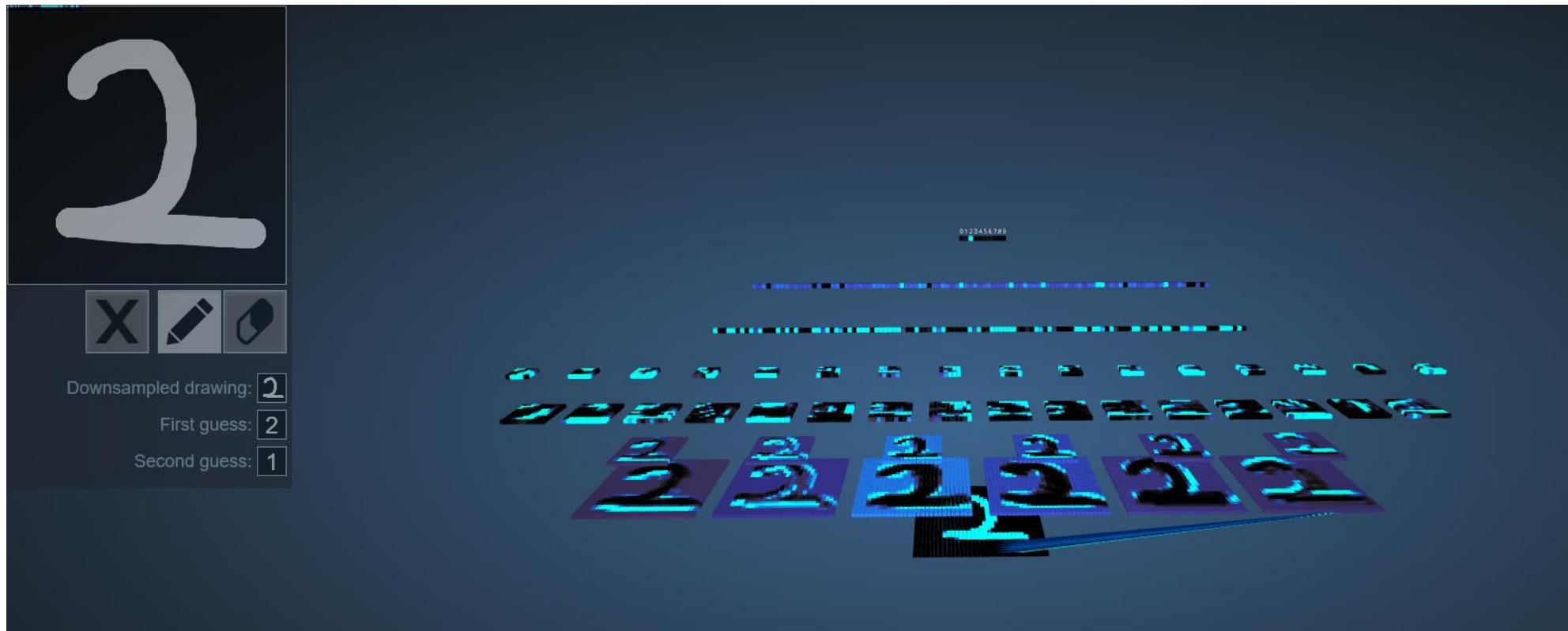
- Tesla, Waymo, Campus Bus



clean and smart:
electric
autonomous bus is ready
for riders

Application: Text and Speech Recognition

- Text recognition: MNIST
- Speech Recognition: Siri, Google, Alexa, etc.



Application: Data prediction

- Time series forecasting: make scientific predictions based on historical data
- Predicting the weather for the next day
 - Have we seen similar patterns in previous years?
 - Can we use information based on many years of data regarding data for a given week and previous weeks
- Predicting financial data
- Predicting COVID data

Large Language Models

Large Language Models (LLMs)

- Natural language processing (NLP) and generative AI
- These models can remember the *context* of previous tokens (words)
 - Multi-head attention
- Data trained on a lot of text extracted from the internet
- Words are converted into tokens (such as numbers) via tokenisation



Bard AI



Example

- Given a string, predict the next word
 - Which word has the highest probability to be the ``right'' word
- Q: What is your name?
- A: My name is Kira
- Statistical approach to generate strings

Token repository

My	hair	am	purple
Your	cat	are	coffee
Their	name	at	magic
His	price	on	Kira
Her	colour	is	grey

Tokenisation

- Changing text into numbers
 - Can be done at a character level, or at a word embedding level
- We previously talked about the fact that language is arbitrary
- But there are still correlations that we can draw from!
- Projection into a low dimension space embedding demo
 - <https://www.cs.cmu.edu/~dst/WordEmbeddingDemo/>

Transformers

- Introduced by Google in 2017
 - Probably the most famous paper there is in our field
-

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaiser@google.com

Illia Polosukhin* †
illia.polosukhin@gmail.com

Transformers

- Key Idea: Transformers embed more than just individual code, but rather contextual meaning
- The **attention head** block allows for information encoded in one embedding into a different embedding
 - The word **mole** has different meanings (example from 3Blue1Brown)
 - American shrew **mole**
 - One **mole** of carbon dioxide
 - Take a biopsy sample of a **mole**

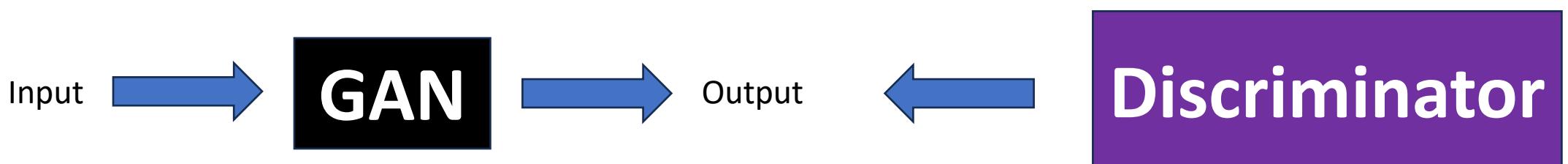
Challenges: LLM

- Will always give you an output, even when it doesn't know what the context is (semantics-wise)
- Statistical approach: can be wrong
- Can give harmful answers when asked
 - Inappropriate
 - Incorrect
 - Not useful
- Misinformation campaign
- Citation and copyright issue?

Generative AI

Generative AI

- ML algorithms that are concerned with the generation of data
 - Images
 - Music
 - Text
- Popular architectures
 - Generative Adversarial Networks (GANs)
 - Diffusion models



Matrix Completion

Matrix completion

- Another *very* prevalent problem
- Given a sparse matrix, can we ``guess'' and fill the missing entries in the matrix?

0			3	
	11			
			23	
30				
	41			

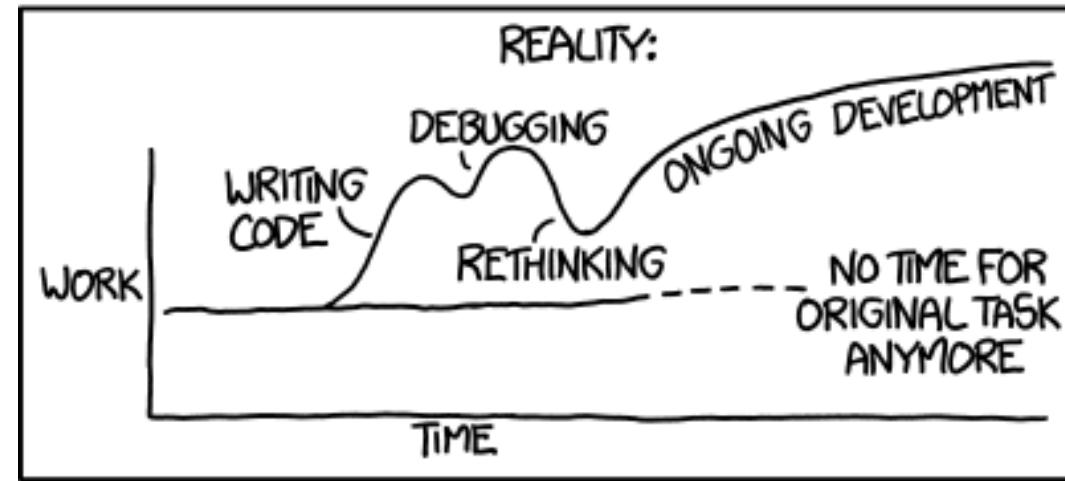
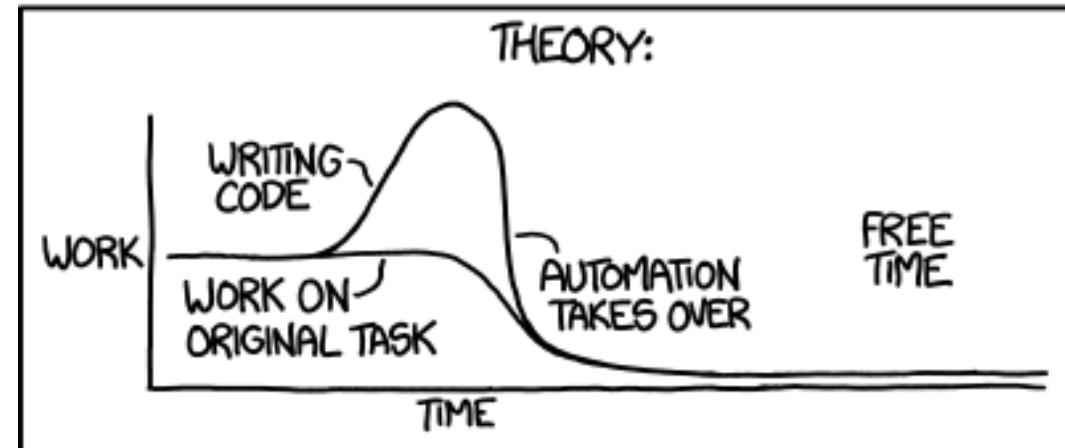
Matrix Completion in practice

- Recommender systems
 - Spotify
 - Amazon
 - Netflix
 - Youtube
- Matrix completion is a very hard problem

Ending remarks

- DNNs are really powerful and have potential applicability in many fields and purposes
- They can solve problems that are traditionally impossible for software developers to do
- However, they are not a silver bullet
 - DNNs come with a lot of problems
 - When you have a hammer, everything looks like a nail
- When we consider DNNs / ML for application, we must ask if the problem should be solved with ML
 - Example: Conference camera devices
- Comparison to stackoverflow

"I SPEND A LOT OF TIME ON THIS TASK.
I SHOULD WRITE A PROGRAM AUTOMATING IT!"



Questions

- Thank you!

Ending remarks

- DNNs are really powerful and have potential applicability in many fields and purposes
- They can solve problems that are traditionally impossible for software developers to do
- However, they are not a silver bullet
 - DNNs come with a lot of problems
 - When you have a hammer, everything looks like a nail
- When we consider DNNs / ML for application, we must ask if the problem should be solved with ML
 - Can we just use a simple algorithm / approach to do it?
 - Example: Conference camera devices

Resources (Getting started)

- <http://neuralnetworksanddeeplearning.com>
- <https://www.deeplearningbook.org/>
- <https://www.nvidia.com/en-us/training/books/>
- <https://pytorch.org/get-started/locally/>
- <https://www.3blue1brown.com/topics/neural-networks>
- <https://statquest.org/video-index/>