Mouse Select Text Draw Stamp Spotlight Eraser Format Undo Redo Clear Save

Who can see what you share here? Recording On

Learning

- Adapt through interaction with the world: rote memory to developing a complex strategy
- Types of learning:
  1. Supervised learning (dense feedback)
  2. Unsupervised learning (no feedback)
  3. Reinforcement learning (sparse feedback, environment altering), etc.
- Advantages (two, among many):
  1. Fault tolerance ✓
  2. No need for a complete specification to begin with ✓
- Becoming a central focus of AI.

1990 ~ 2000

ML.

2005~2010

Deep Learning

Not specifically related to deep learning, but anyway, so it's machine learning is really one of the most popular and powerful technologies

You are screen sharing Stop Share

What Is Machine Learning?

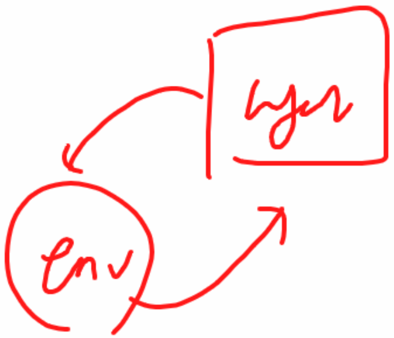
dominating the field.

- A subfield of AI that is rapidly growing in importance.
- Performance of a system is improved based on learning experience.
- Learning from data.

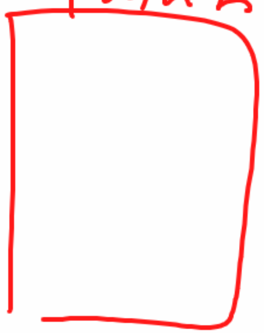
What Is Machine Learning?

dominating the field.

- A subfield of AI that is rapidly growing in importance.
- Performance of a system is improved based on learning experience.
- Learning from data.



Data file



Mouse Select Text Draw Stamp Spotlight Eraser Format Undo Redo Clear Save

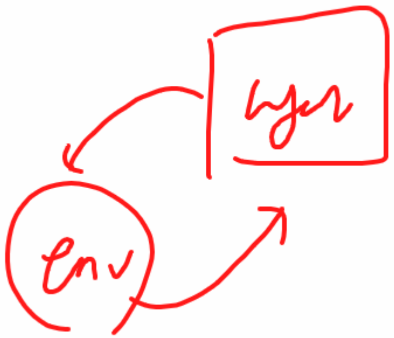
Who can see what you share here? Recording On

What Is Machine Learning?

dominating the field.

- A subfield of AI that is rapidly growing in importance.
- Performance of a system is improved based on learning experience.
- Learning from data.

Data file



So all of these are different forms of experience. So learning from data is basically one form of experience

You are screen sharing Stop Share

Huge amount of data

"Big Data"

Why Machine Learning?

- Abundance of data: the data deluge.
  - Scientific instruments.
  - Data acquisition devices.
  - Internet and the web.
  - All sectors of human society producing and digitizing data (e.g., your cell phone).
- Not enough human expertise or human power to make sense of such huge amounts of data.

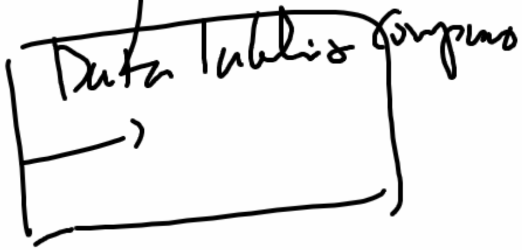
ML

Amazon's Mechanical Turk

¢

label

requires human effort



Machine Learning in the News

IBM's Watson   Google DeepMind's AlphaGo

- IBM's Watson beats human champions: Jeopardy (game show)
- Google detects cats from YouTube videos.
- Google Glass app recognizes people it sees.
- Legal, medical, financial applications.
- Google DeepMind: Atari 2600 game playing, AlphaGo, AlphaStar

Machine Learning in the News

IBM's Watson   Google DeepMind's AlphaGo

- IBM's Watson beats human champions: Jeopardy (game show)
- Google detects cats from YouTube videos.
- Google Glass app recognizes people it sees.
- Legal, medical, financial applications.
- Google DeepMind: Atari 2600 game playing, AlphaGo, AlphaStar

Mouse Select Text Draw Stamp Spotlight Eraser Format Undo Redo Clear Save

Who can see what you share here? Recording On

2015

Deep R.L

I

Atari 2600

Machine Learning in the News

2016

IBM's Watson   Google DeepMind's AlphaGo

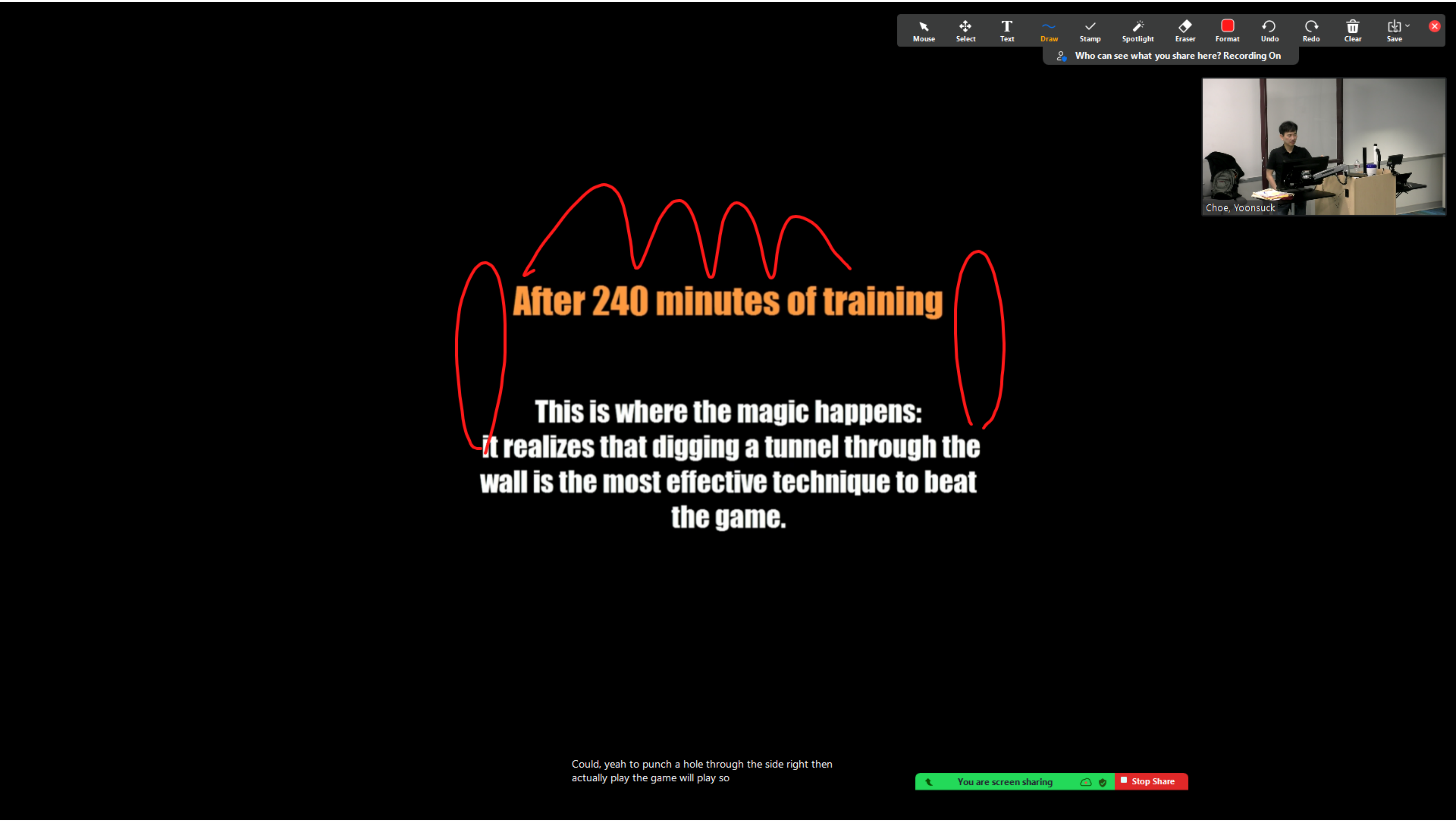
Google Deepmind

UK

- IBM's Watson beats human champions: Jeopardy (game show)
- Google detects cats from YouTube videos.
- Google Glass app recognizes people it sees.
- Legal, medical, financial applications.
- Google DeepMind: Atari 2600 game playing, AlphaGo, AlphaStar

alpha goal, and then applied that to atari, 26, 3,600 games

You are screen sharing Stop Share


Mouse
Select
Text
Draw
Stamp
Spotlight
Eraser
Format
Undo
Redo
Clear
Save
Who can see what you share here? Recording On
Choe, Yoonsuck
After 240 minutes of training
This is where the magic happens:
it realizes that digging a tunnel through the
wall is the most effective technique to beat
the game.
Could, yeah to punch a hole through the side right then
actually play the game will play so
You are screen sharing
Stop Share

Machine Learning in the News

IBM's Watson   Google DeepMind's AlphaGo

- IBM's Watson beats human champions: Jeopardy (game show)
- Google detects cats from YouTube videos.
- Google Glass app recognizes people it sees.
- Legal, medical, financial applications.
- Google DeepMind: Atari 2600 game playing, AlphaGo, AlphaStar

$P(\text{Earthquake} \mid \text{HouseShaking}) = 0.9$

observation

Probabilistic Inference

- **Diagnostic inferences**: $P(Cause|Effect)$
- **Causal inferences**: $P(Effect|Cause)$
- **Intercausal inferences**: causes of a common effect (**explaining away**: cause has already been found)

  $P(Cause|Effect) >> P(Cause|Effect \wedge OtherCause)$

- **Mixed inferences**: combining two or more of the above

  $P(A|CauseOfA \wedge EffectOfA)$

Construction

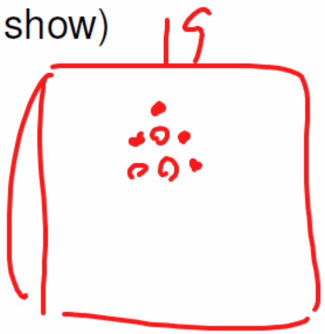
$P(\text{Earthquake} \mid \text{House shaking, Construction}) = 0.2$

Machine Learning in the News

IBM's Watson   Google DeepMind's AlphaGo

Deep Q Network

- IBM's Watson beats human champions: Jeopardy (game show)
- Google detects cats from YouTube videos.
- Google Glass app recognizes people it sees.
- Legal, medical, financial applications.
- Google DeepMind: Atari 2600 game playing, AlphaGo, AlphaStar



Machine Learning in the News

IBM's Watson   Google DeepMind's AlphaGo

Deep Network

15

15

- IBM's Watson beats human champions: Jeopardy (game show)
- Google detects cats from YouTube videos.
- Google Glass app recognizes people it sees.
- Legal, medical, financial applications.
- Google DeepMind: Atari 2600 game playing, AlphaGo, AlphaStar

Network, combined with something called Monte Carlo. 3
search has been able to actually beat this former champion

Mouse Select Text Draw Stamp Spotlight Eraser Format Undo Redo Clear Save

Who can see what you share here? Recording On

Machine Learning in the News

IBM's Watson   Google DeepMind's AlphaGo

- IBM's Watson beats human champions: Jeopardy (game show)
- Google detects cats from YouTube videos.
- Google Glass app recognizes people it sees.
- Legal, medical, financial applications.
- Google DeepMind: Atari 2600 game playing, AlphaGo, AlphaStar

And he's just thing the game dictated by Alpha go

You are screen sharing   Stop Share

Machine Learning in the News

IBM's Watson Google DeepMind's AlphaGo

- IBM's Watson beats human champions: Jeopardy (game show)
- Google detects cats from YouTube videos.
- Google Glass app recognizes people it sees.
- Legal, medical, financial applications.
- Google DeepMind: Atari 2600 game playing, AlphaGo, AlphaStar

ACM Turing Award 2018: Deep Learning

Association for Computing Machinery

ACM

ACM recognizes excellence

ACM Home ACM A.M. Turing Award Turing 50

ACM AWARDS | ADVANCED GRADES OF MEMBERSHIP | SIG AWARDS | REGIONAL AWARDS | NOMINATING PROCESS | AWARDS

Awards Home | ACM A.M. Turing Award | Spotlight on Turing Laureates | ACM Prize in Computing

Home > Latest Awards News > 2018 Turing Award

Fathers of the Deep Learning Revolution Receive ACM A.M. Turing Award

Bengio, Hinton and LeCun Ushered in Major Breakthroughs in Artificial Intelligence

Communications of the ACM

- ACM Turing Award 2018 goes to Deep Learning pioneers!

So this is a good way to find your research interested to int

What Does It Take to do ML?

A lot of math:

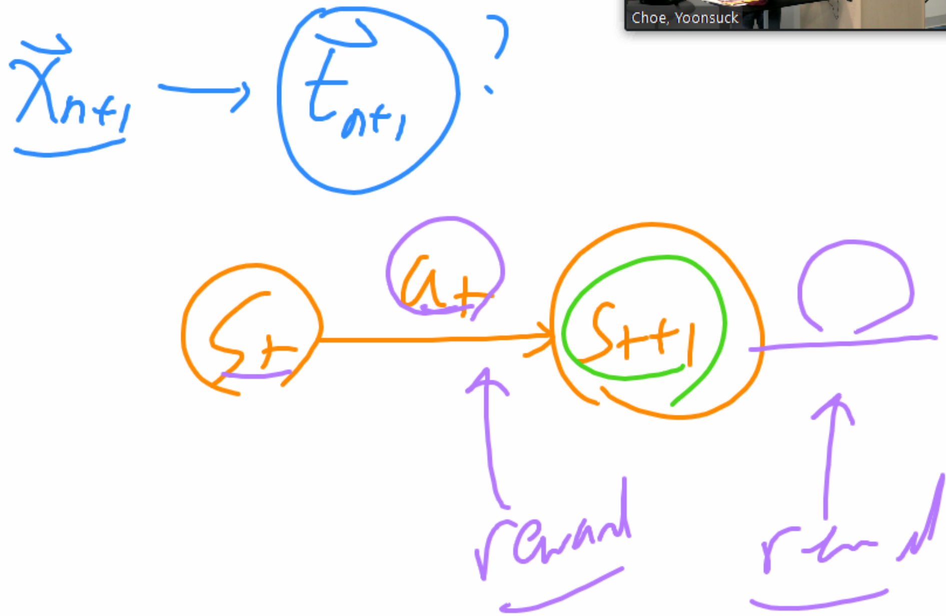
- Linear algebra
- Calculus
- Probability and statistics
- Differential geometry
- Numerical methods

$A \cdot B \quad (A)^{-1}$

$$\begin{cases} (\bigcirc * \bigcirc) + \bigcirc \\ \bigcirc + (\bigcirc * \bigcirc) \end{cases}$$

Types of Machine Learning

1. Supervised learning
   - Input-Target pairs
   - $\{\langle \vec{x}_i, \vec{t}_i \rangle | i = 1, 2, .., n\}$
2. Unsupervised learning
   - A bunch of inputs (unlabeled)
   - $\{\vec{x}_i | i = 1, 2, ..., n\}$
3. Reinforcement learning
   - state$_1$ $\xrightarrow{\text{action}_1}$ state$_2$ $\xrightarrow{\text{action}_2}$ state$_3$, ... , reward
   - $s_{t+1} = \underbrace{\delta(s_t, a_t)}_{\text{env}}$, $r_{t+1} = \underbrace{\rho(s_t, a_t)}_{\text{env}}$



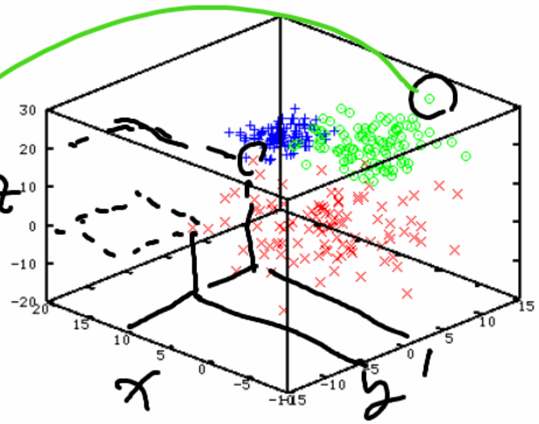
Types of Machine Learning

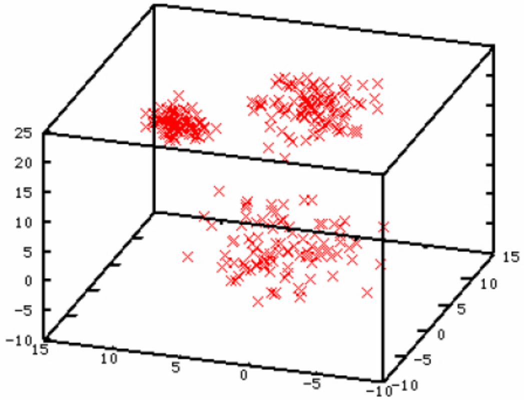
1. Supervised learning
   - Input-Target pairs
   - $\{\langle \vec{x}_i, \vec{t}_i \rangle | i = 1, 2, .., n\}$
2. Unsupervised learning
   - A bunch of inputs (unlabeled)
   - $\{\vec{x}_i | i = 1, 2, ..., n\}$
3. Reinforcement learning
   - $\text{state}_1 \xrightarrow[\text{action}_1]{} \text{state}_2 \xrightarrow[\text{action}_2]{} \text{state}_3$, ... , reward
   - $s_{t+1} = \delta(s_t, a_t), r_{t+1} = \rho(s_t, a_t)$

So you're the local reward would actually go down. But then you get a good grade. Then that that actually give you a much more

Example Data

x	y	z	Color
10	5	0	×
10	1	20	+
			O





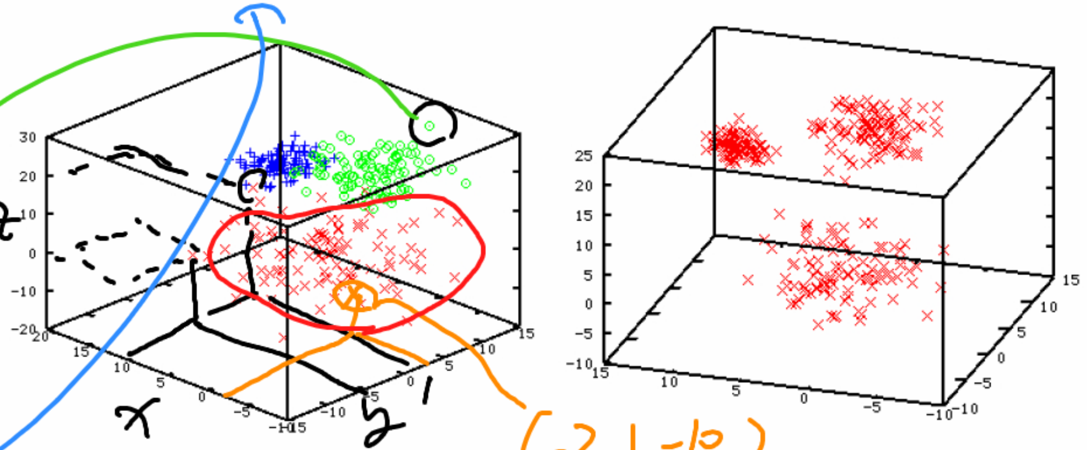
- Left: supervised
- Right: unsupervised
- Typically very high dimensional (10,000, 1 million [or more]).

Example Data

"X"

x	y	z	color
10	5	0	X
10	1	20	+
			O
			⋮
-2, 1, -10			X ?

1ᴺ4



- Left: supervised
- Right: unsupervised
- Typically very high dimensional (10,000, 1 million [or more]).

cond. indep

$P(\text{Ache} \mid \text{Cav}, \text{Catch})$

cause

cause

Probabilistic Inference

- **Diagnostic inferences**: $P(Cause|Effect)$
- **Causal inferences**: $P(Effect|Cause)$
- **Intercausal inferences**: causes of a common effect (**explaining away**: cause has already been found)

  $P(Cause|Effect) >> P(Cause|Effect \land OtherCause)$

Intercausal Inference

- **Mixed inferences**: combining two or more of the above

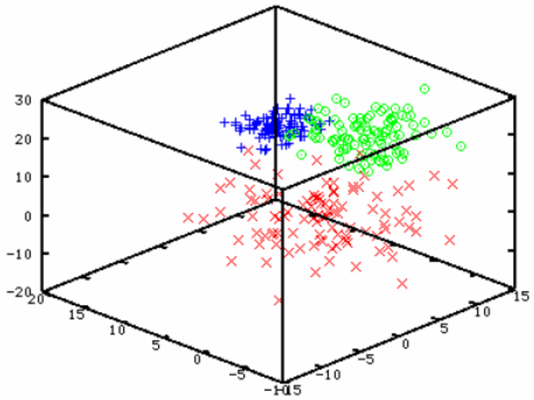
  $P(A|CauseOfA \land EffectOfA)$

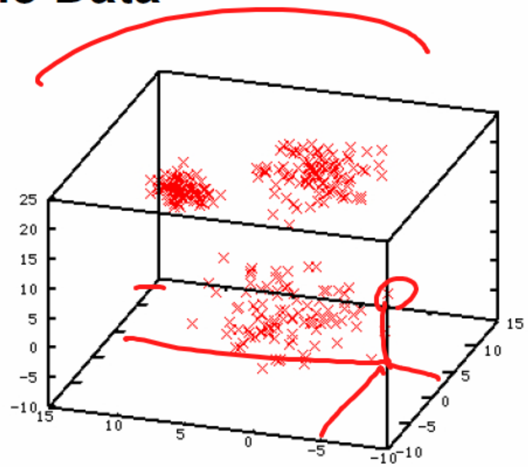
$P(\text{Earthquake} \mid \text{Houseshaking}, \text{Construction})$

cause

cause

Example Data



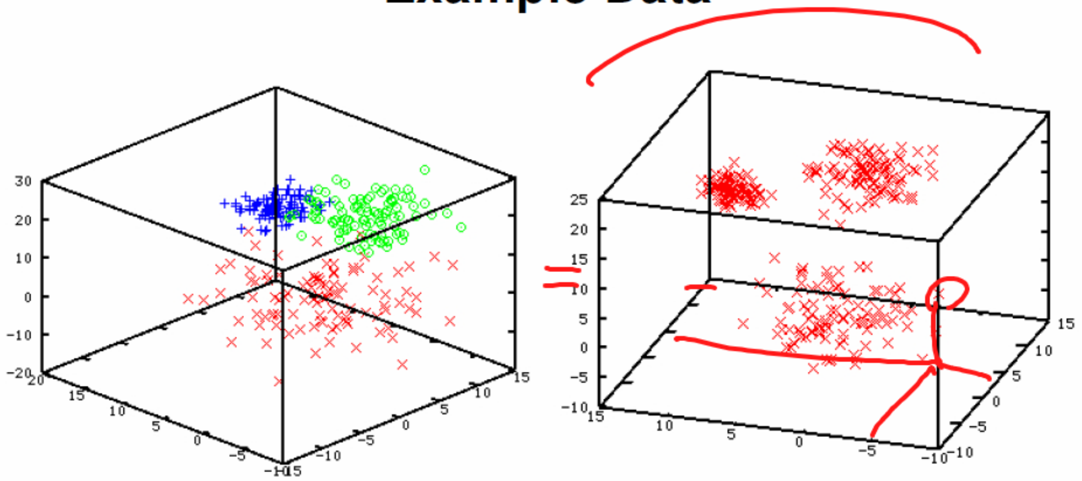


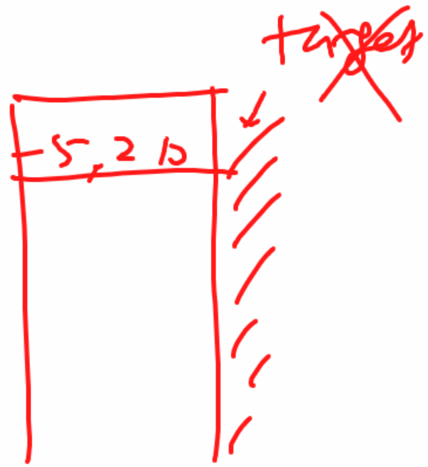
~~target~~

-5, 2 10

- Left: supervised
- Right: unsupervised
- Typically very high dimensional (10,000, 1 million [or more]).

Example Data





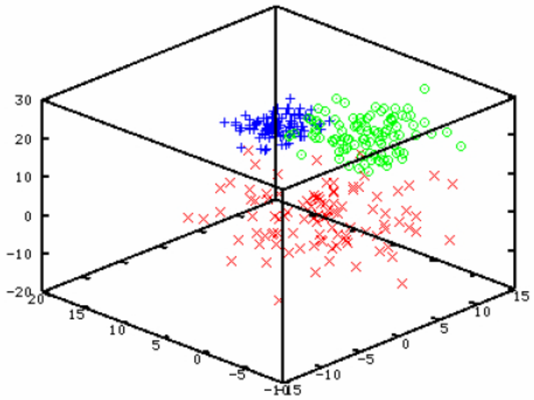
- Left: supervised
- Right: unsupervised
- Typically very high dimensional (10,000, 1 million [or more]).

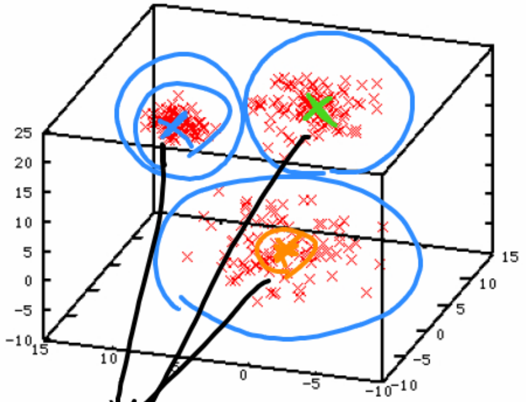
So, in fact, these 2 are the same data set. but here I just removed the label on the target value

Mouse Select Text Draw Stamp Spotlight Eraser Format Undo Redo Clear Save

Who can see what you share here? Recording On

Example Data





features

clusters. - - structure.

- Left: supervised
- Right: unsupervised
- Typically very high dimensional (10,000, 1 million [or more]).

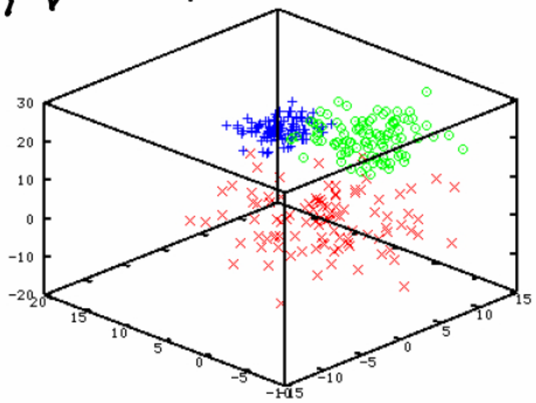
So there's something that you can say about this kind of data set

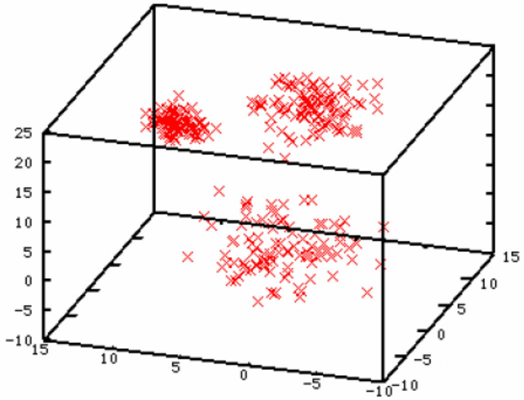
You are screen sharing Stop Share

Example Data

3 D input.

x	y	z





x	y	z

3 D

- Left: supervised
- Right: unsupervised
- Typically very high dimensional (10,000, 1 million [or more]).

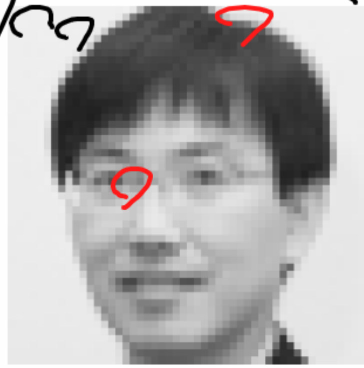
Can you imagine that this kind of data exist

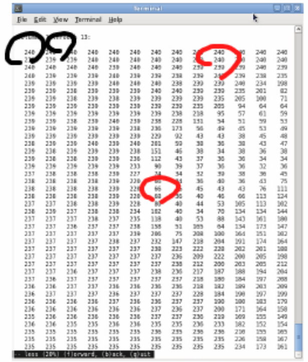
$\rightarrow$ 2500 Values

High-dimensional Data

50

50





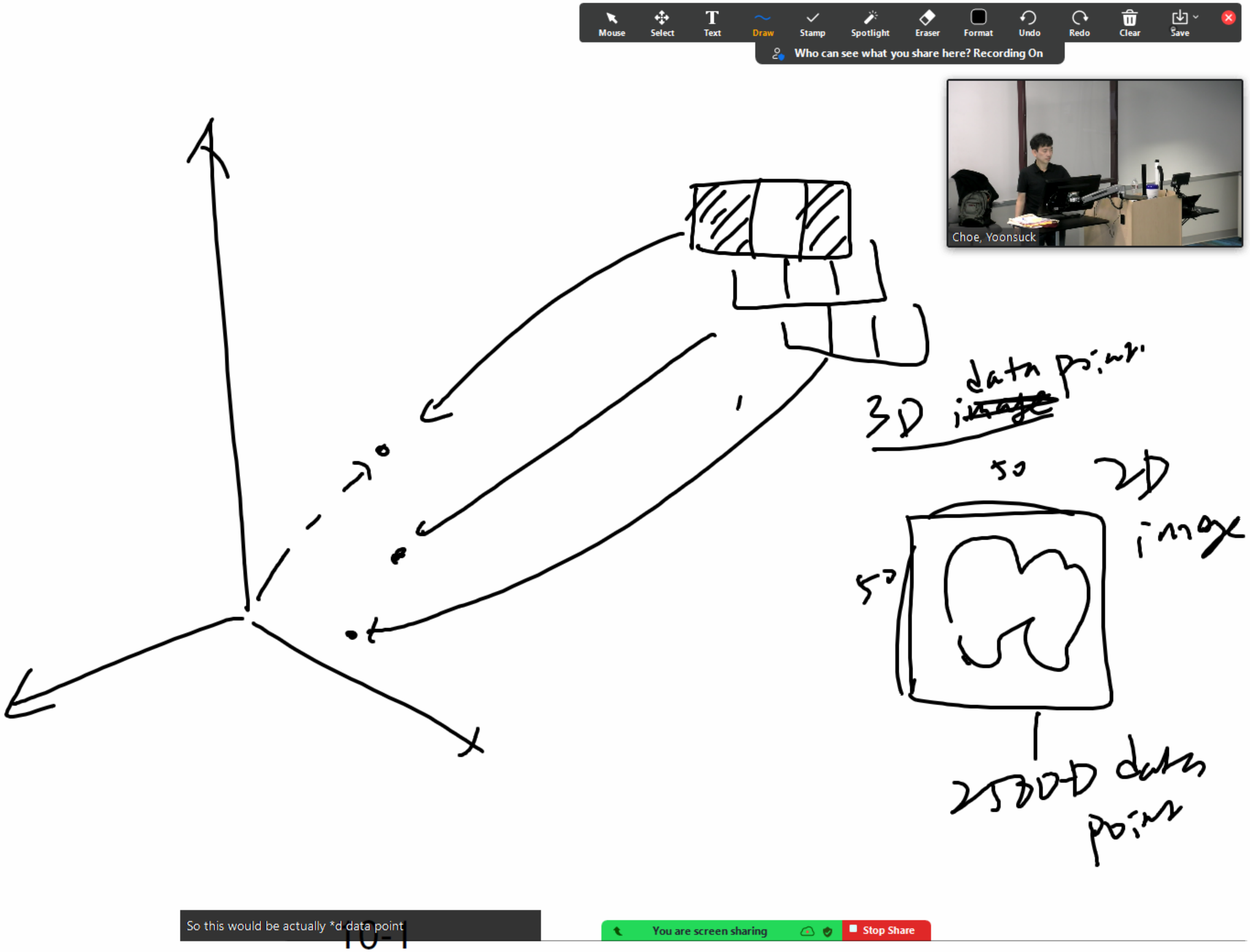
50

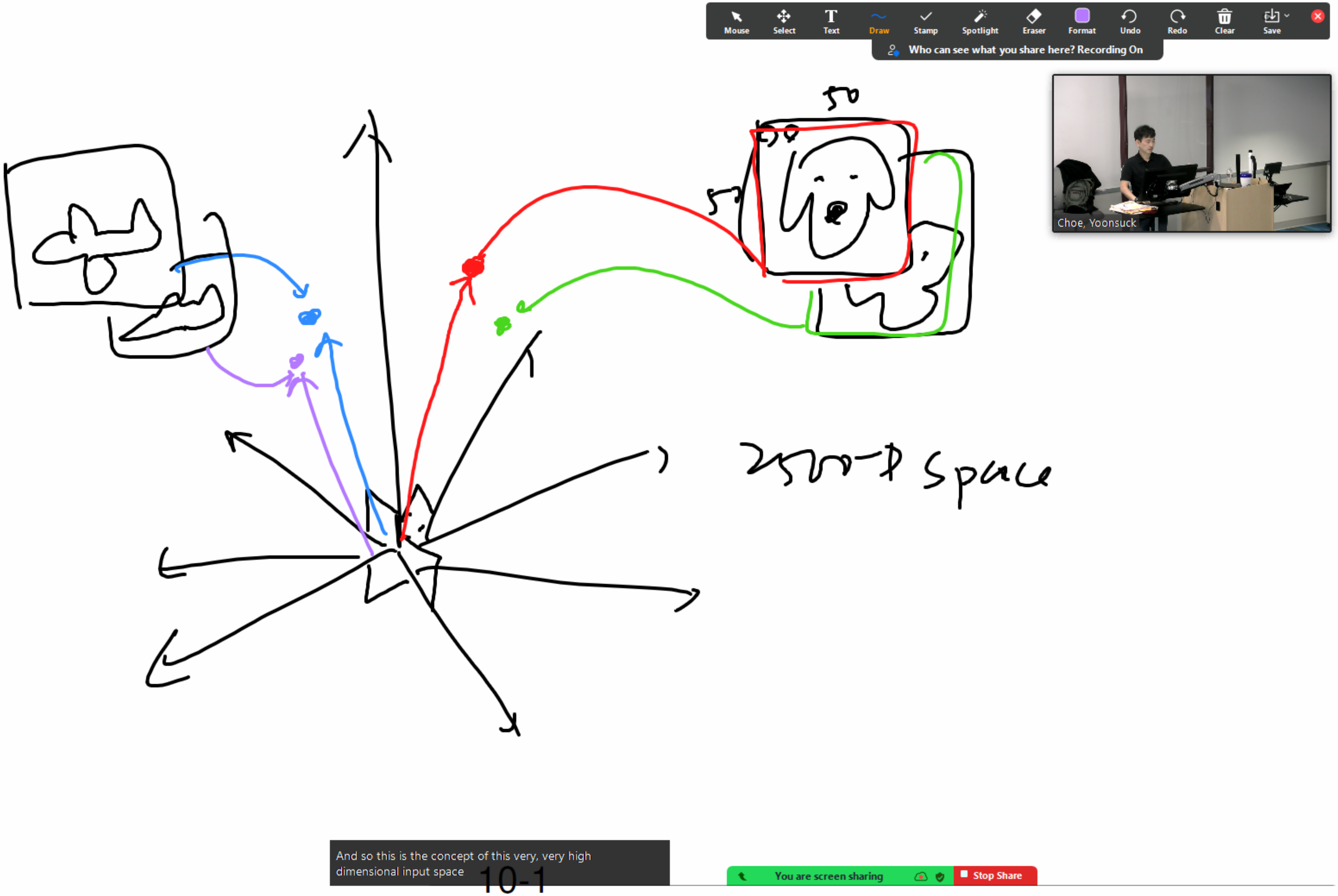
50

ML → "Choe"

$$\longrightarrow \underbrace{[240, 240, \ldots, 232, \ldots.]}_{\text{2,500-D vector}}$$

- Images: these are 2D images, but ...
- These are 50×50 = 2,500-dimensional vectors.
  - Each such image is a single point in 2,500-dimensional space.


Mouse
Select
Text
Draw
Stamp
Spotlight
Eraser
Format
Undo
Redo
Clear
Save
Who can see what you share here? Recording On
Choe, Yoonsuck
3D ~~image~~ data point
50
2D image
50
2500-D data point
So this would be actually *d data point
You are screen sharing
Stop Share


Mouse
Select
Text
Draw
Stamp
Spotlight
Eraser
Format
Undo
Redo
Clear
Save
Who can see what you share here? Recording On
Choe, Yoonsuck
50
50
2500-D space
And so this is the concept of this very, very high dimensional input space
You are screen sharing
Stop Share

Supervised Learning

① Supervised learning

- Regression: approximating $y = f(x)$
- Classification: face recognition, hand-written character recognition, credit risk assessment, etc.
- Techniques:
  - Neural networks
  - Decision tree learning
  - Support vector machines
  - Radial basis functions
  - Naive Bayes learning
  - k-nearest neighbor

② Unsupervised learning

③ Reinforcement learning

input	$\mathbb{R}$
x_1	y_1
x_2	y_2
	⋮

input target

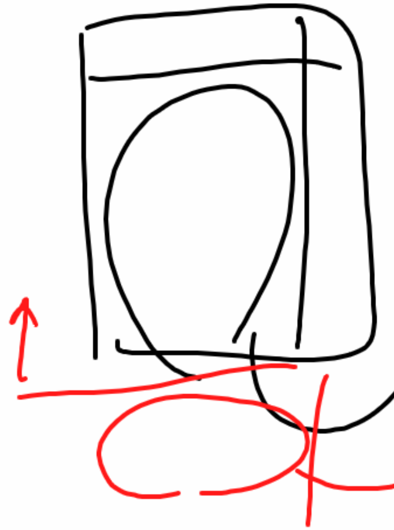
input	target
	"cat"
	"dog"

input target

ML → error

Supervised Learning Issues

error

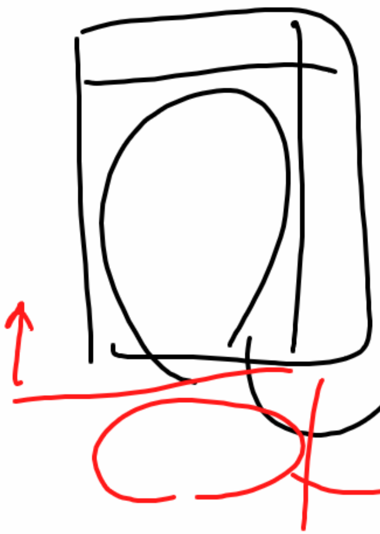


- How well will it do on training inputs?
- How well will it do on novel inputs?
  - Generalization.
- How many samples needed for sufficient performance and generalization?
  - Sample complexity
  - Curse of dimensionality
  - Computational learning theory
- Catastrophic forgetting (online learning hard).

ML → error

Supervised Learning Issues

error



- How well will it do on training inputs?
- How well will it do on novel inputs?
  - Generalization.
- How many samples needed for sufficient performance and generalization?
  - Sample complexity
  - Curse of dimensionality
  - Computational learning theory
- Catastrophic forgetting (online learning hard).

cond. indep

P(Ache | Cav, Catch)

cause

cause

Probabilistic Inference

- **Diagnostic inferences**: $P(Cause|Effect)$
- **Causal inferences**: $P(Effect|Cause)$
- **Intercausal inferences**: causes of a common effect (**explaining away**: cause has already been found)

  $P(Cause|Effect) >> P(Cause|Effect \wedge OtherCause)$

Intercausal Inference

- **Mixed inferences**: combining two or more of the above

  $P(A|CauseOfA \wedge EffectOfA)$

P(Earthquake | Houseshaking, Construction)

cause

cause

Supervised Learning Issues

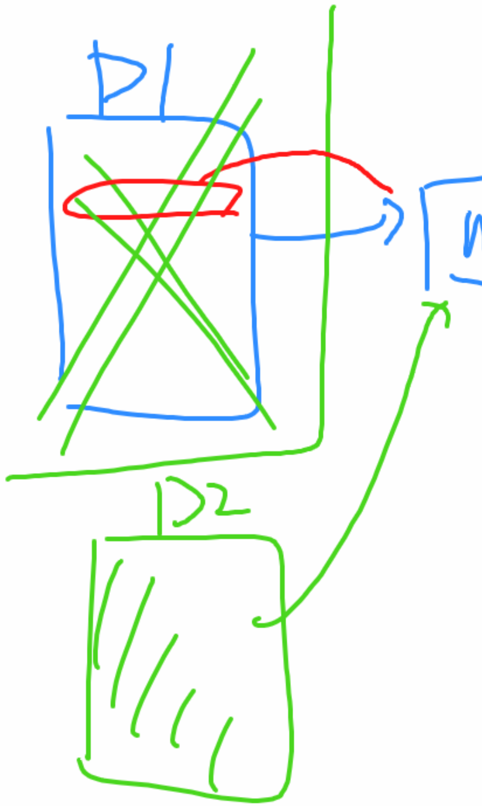
D1 D2 ML 98.5% 90.2% 98%

D1 D2 ML 98% 98.5% 97.5%

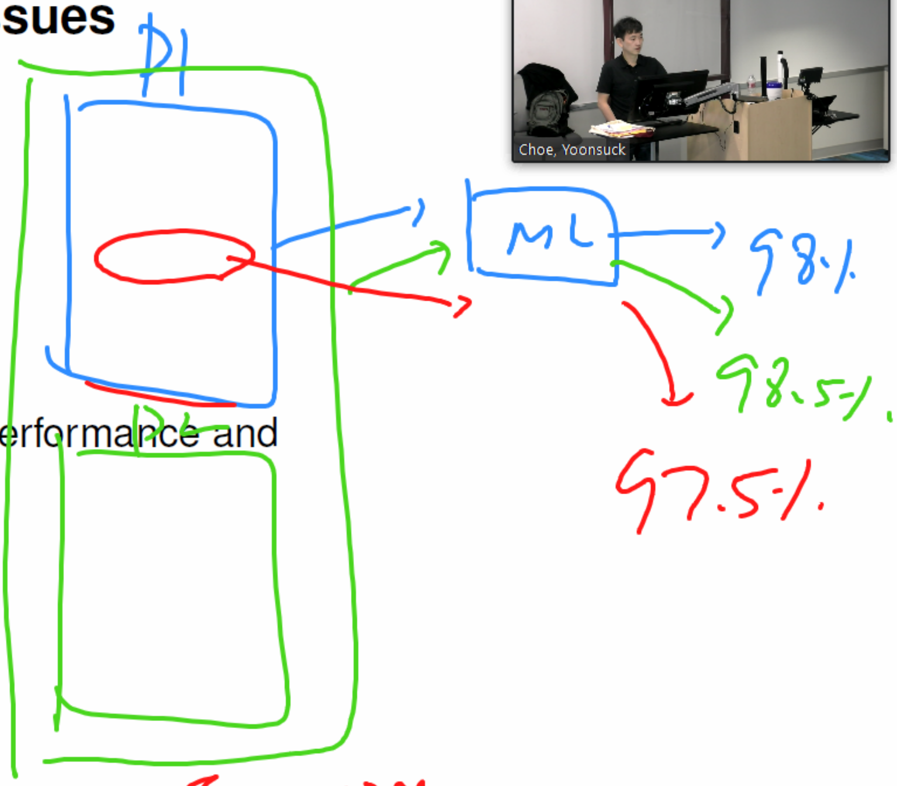
Expensive

- How well will it do on training inputs?
- How well will it do on novel inputs?
  - Generalization.
- How many samples needed for sufficient performance and generalization?
  - Sample complexity
  - Curse of dimensionality
  - Computational learning theory
- Catastrophic forgetting (online learning hard).

Supervised Learning Issues



- How well will it do on training inputs?
- How well will it do on novel inputs?
  - Generalization.
- How many samples needed for sufficient performance and generalization?
  - Sample complexity
  - Curse of dimensionality
  - Computational learning theory
- Catastrophic forgetting (online learning hard).



Expensive

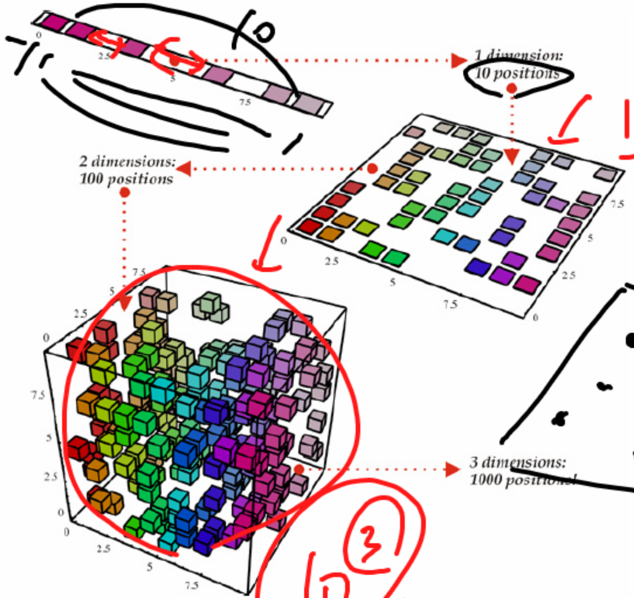
Continual learning

That came out to overcome this kind of issue, and this is called continual learning.

Addendum: Curse of Dimensionality

X

1D 10 10^2 2D



10^3

10^{2500}

X b

n=10

sparse

From: Yoshua Bengio's page

- Exponentially many points needed to achieve same density of training samples.

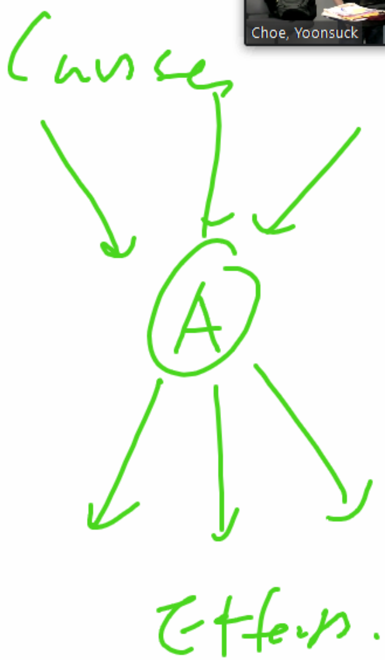
Probabilistic Inference

- **Diagnostic inferences**: $P(Cause|Effect)$
- **Causal inferences**: $P(Effect|Cause)$
- **Intercausal inferences**: causes of a common effect (**explaining away**: cause has already been found)

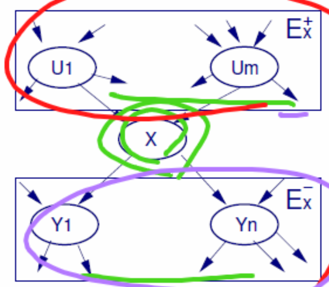
$$P(Cause|Effect) >> P(Cause|Effect \wedge OtherCause)$$

- **Mixed inferences**: combining two or more of the above

$$P(A|CauseOfA \wedge EffectOfA)$$



Answering Queries: A Brief Outline



Given a set of evidence E, find the conditional probability $\mathbf{P}(X|E)$ where X is the query.

- Recursively deterimine the **causal support** E_X^+ for X.
- Recursively deterimine the **evidential support** E_X^- for X.

Note: This is only when the graph is **singly connected**.

So. And there are certain conditions. For example, the graph has to be singular

Using Belief Networks and Probabilistic Inference

- Making decisions based on the derived probabilities and an agent's utility function.
- Deciding which additional variables to include in the model.
- Performing **sensitivity analysis** to find out which node is most important and thus should be more accurate.
- Explaining the results of reasoning.

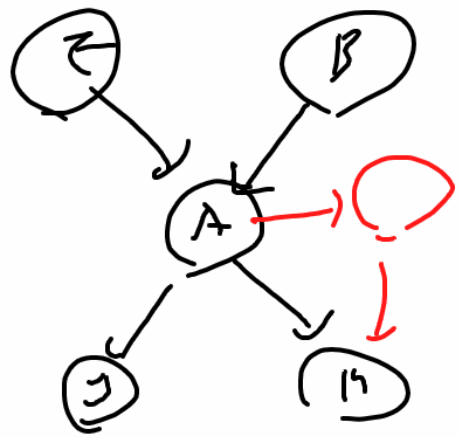
$p(X \mid E)$

$\text{Utility}(X = x_i)$

$\rightarrow$ decision

Using Belief Networks and Probabilistic Inference

- Making decisions based on the derived probabilities and an agent's utility function.
- Deciding which additional variables to include in the model.
- Performing **sensitivity analysis** to find out which node is most important and thus should be more accurate.
- Explaining the results of reasoning.



Knowledge Engineering for Uncertain Reasoning

- Decide what to talk about (i.e. what to be included in the model). Gradually add more factors that can influence the current collection of events.
- Determine the variables to use and the range of the values.
- Encode general knowledge about the dependence between variables:
  1. qualitative: which variable depends on some other variable
  2. quantitative: probability value of the dependence (from experience, or from data gathered from a sample space)

graph

cond. prob.

- Encode a description of the specific problem instance: assign values to the variables.

$p(X=?\mid E=?)$

- Pose queries to the inference procedure and get answers: what is the probability of X? how sensitive are the values in the conditional probability tables to perturbations?

Learning

- Adapt through interaction with the world: rote memory to developing a complex strategy
- Types of learning:
  1. Supervised learning (dense feedback)
  2. Unsupervised learning (no feedback)
  3. Reinforcement learning (sparse feedback, environment altering), etc.
- Advantages (two, among many):
  1. Fault tolerance
  2. No need for a complete specification to begin with
- Becoming a central focus of AI.