CS 446: Machine Learning

A. G. Schwing & M. Telgarsky

University of Illinois at Urbana-Champaign, 2018

L1: Introduction

Goals of this lecture

Goals of this lecture

- Introduction to machine learning

Goals of this lecture

- Introduction to machine learning
- Organizational details

Goals of this lecture

- Introduction to machine learning
- Organizational details
- k-Nearest Neighbor

Goals of this lecture

- Introduction to machine learning
- Organizational details
- k-Nearest Neighbor

Reading Material

Goals of this lecture

- Introduction to machine learning
- Organizational details
- k-Nearest Neighbor

Reading Material

- K. Murphy; Machine Learning: A Probabilistic Perspective; Chapter 1

Goals of this lecture

- Introduction to machine learning
- Organizational details
- k-Nearest Neighbor

Reading Material

- K. Murphy; Machine Learning: A Probabilistic Perspective; Chapter 1

Style of this lecture and all others

Goals of this lecture

- Introduction to machine learning
- Organizational details
- k-Nearest Neighbor

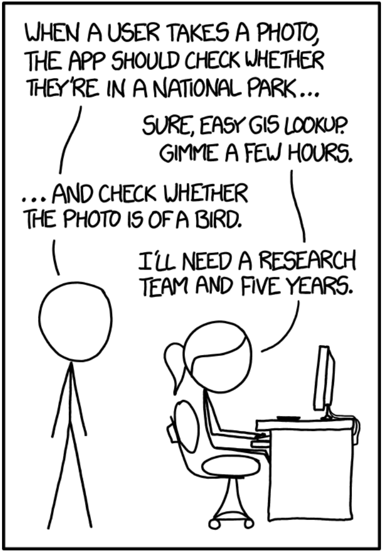
Reading Material

- K. Murphy; Machine Learning: A Probabilistic Perspective; Chapter 1

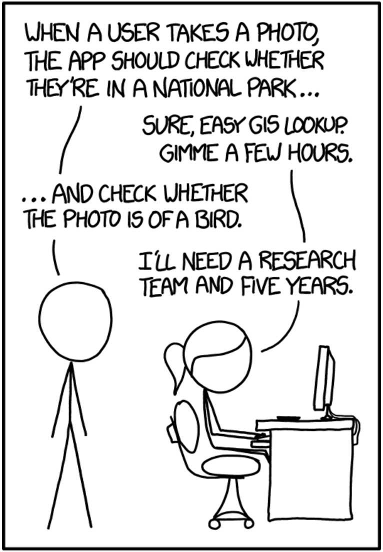
Style of this lecture and all others

- Interactive

Machine Learning



IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.



IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

somewhat dated joke

What is machine learning? What applications?

Discuss with your neighbor

Where machine learning shows up:

Where machine learning shows up:

- Chess play (Deep Blue, IBM, 1997)

Where machine learning shows up:

- Chess play (Deep Blue, IBM, 1997)
- Jeopardy (Watson, IBM, 2011)

Where machine learning shows up:

- Chess play (Deep Blue, IBM, 1997)
- Jeopardy (Watson, IBM, 2011)
- Playing Atari games (DQN, Deepmind, 2015)

Where machine learning shows up:

- Chess play (Deep Blue, IBM, 1997)
- Jeopardy (Watson, IBM, 2011)
- Playing Atari games (DQN, Deepmind, 2015)
- Game of Go (AlphaGo, Deepmind, 2016)

Where machine learning shows up:

- Chess play (Deep Blue, IBM, 1997)
- Jeopardy (Watson, IBM, 2011)
- Playing Atari games (DQN, Deepmind, 2015)
- Game of Go (AlphaGo, Deepmind, 2016)
- Autonomous Driving (Audi, BMW, Mercedes, Uber, Waymo, Lyft, NVIDIA, Intel, ...)

Example:

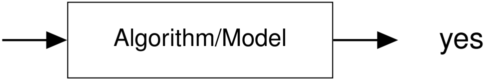
Let's get a computer to recognize whether there is a cat in the image.

Example:

Let’s get a computer to recognize whether there is a cat in the image.

Example:

Let's get a computer to recognize whether there is a cat in the image.

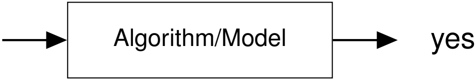


Example:

Let's get a computer to recognize whether there is a cat in the image.

Example:

Let's get a computer to recognize whether there is a cat in the image.

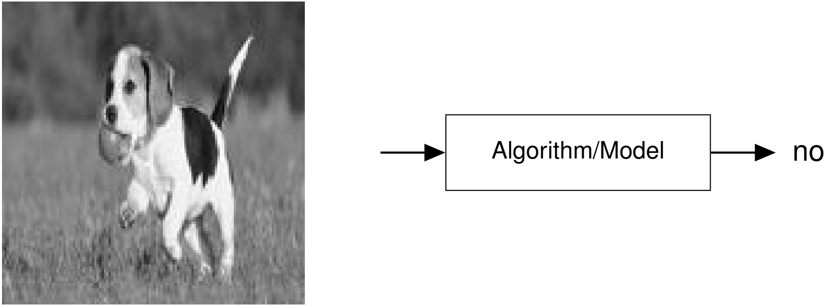


Example:

Let's get a computer to recognize whether there is a cat in the image.

Example:

Let's get a computer to recognize whether there is a cat in the image.



Formulation of machine learning for now:

Formulation of machine learning for now:

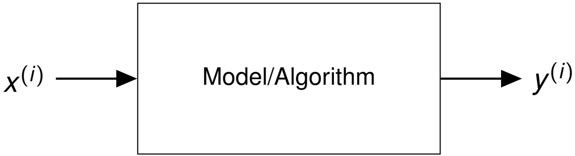
- input data/value/vector: $x^{(i)}$

Formulation of machine learning for now:

- input data/value/vector: $x^{(i)}$
- label/output: $y^{(i)}$

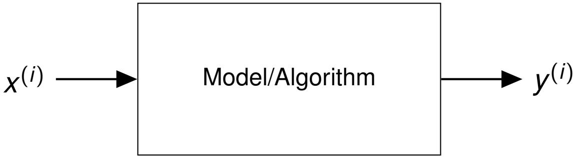
Formulation of machine learning for now:

- input data/value/vector: $x^{(i)}$
- label/output: $y^{(i)}$



Formulation of machine learning for now:

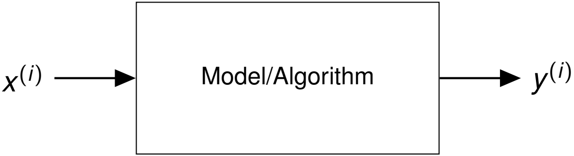
- input data/value/vector: $x^{(i)}$
- label/output: $y^{(i)}$



How do we call this process?

Formulation of machine learning for now:

- input data/value/vector: $x^{(i)}$
- label/output: $y^{(i)}$



How do we call this process?

- Inference

Formulation of machine learning for now:

- input data/value/vector: $x^{(i)}$
- label/output: $y^{(i)}$

$$x^{(i)} \longrightarrow \boxed{\text{Model/Algorithm}} \longrightarrow y^{(i)}$$

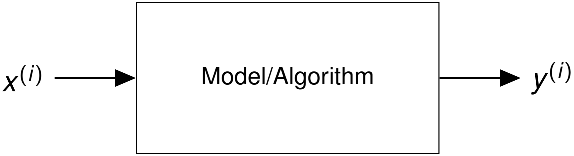
How do we call this process?

- Inference
- Prediction

Where is the machine learning here?

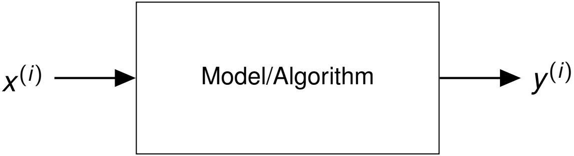
$$x^{(i)} \longrightarrow \boxed{\text{Model/Algorithm}} \longrightarrow y^{(i)}$$

Where is the machine learning here?



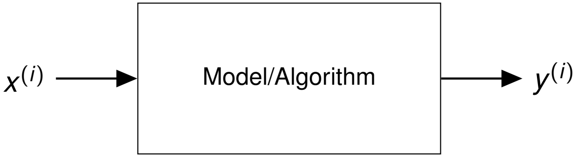
- Model/Algorithm depends on parameters $\boldsymbol{w}$

Where is the machine learning here?



- Model/Algorithm depends on parameters $\boldsymbol{w}$
- Learning/Fitting of parameters $\boldsymbol{w}$

Where is the machine learning here?



- Model/Algorithm depends on parameters $\boldsymbol{w}$
- Learning/Fitting of parameters $\boldsymbol{w}$
- Based on dataset $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$

How could we recognize a cat? What should the algorithm do?

How does an image look like for a computer?

How does an image look like for a computer?

Demo

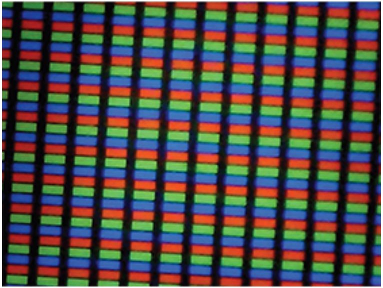
How does an image look like for a computer?

Demo

69	69	70	70	71	72	72
79	80	80	81	82	83	83
87	88	88	89	90	91	91
95	96	96	97	98	99	99
108	108	109	110	110	111	112
119	119	120	120	121	122	122
130	131	131	132	133	134	134
144	144	145	145	146	147	148
158	158	159	160	162	163	164
172	172	173	174	175	176	177
187	188	189	190	191	192	193
198	198	199	200	202	203	204
207	207	208	209	210	212	212

How does an image look like for a computer?

Cell phone screen:



Given that we know how a computer ‘sees’ an image, how can it recognize cats?

Given that we know how a computer 'sees' an image, how can it recognize cats?

Given that we know how a computer ‘sees’ an image, how can it recognize cats?



Still very hard to describe what a cat looks like.

Given that we know how a computer ‘sees’ an image, how can it recognize cats?



Still very hard to describe what a cat looks like.

Let’s look at tons of examples (Dataset).

Dataset (Thousands of examples):





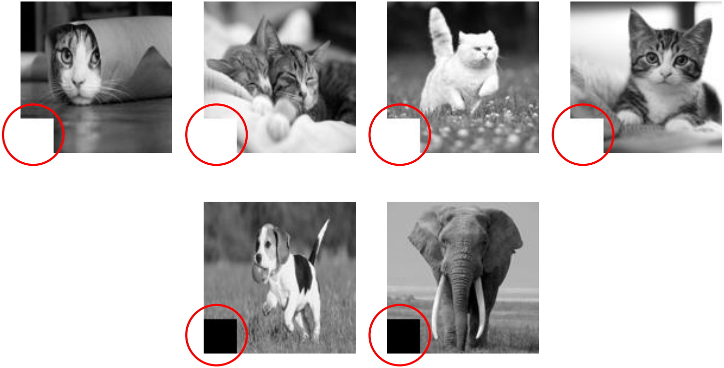
Dataset (Thousands of examples):





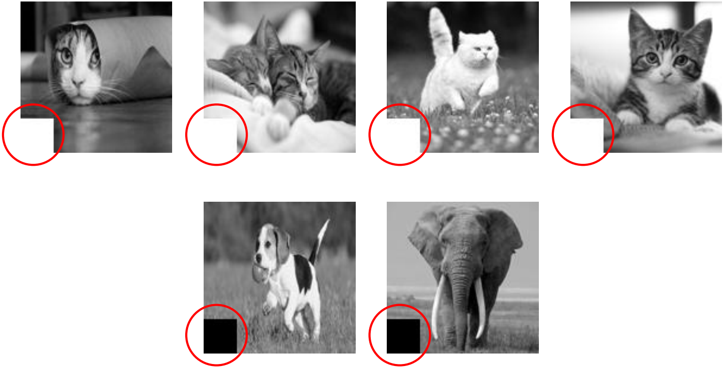
Instead of asking to recognize cats in general, let's recognize cats in this dataset

Dataset (Thousands of examples):



Instead of asking to recognize cats in general, let's recognize cats in this dataset

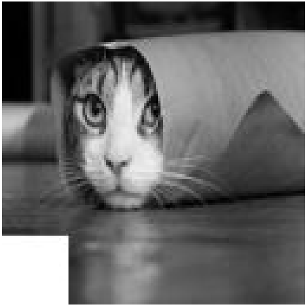
Dataset (Thousands of examples):



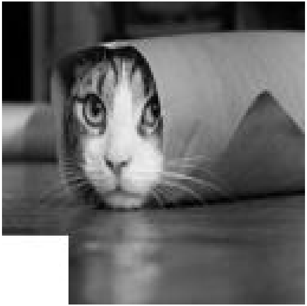
Instead of asking to recognize cats in general, let's recognize cats in this dataset

Algorithm: if bottom left corner is black (0) say no, otherwise say yes

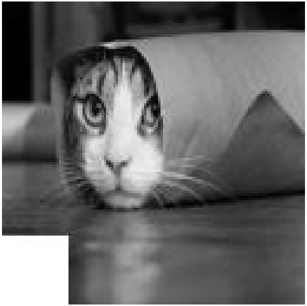
Algorithm: if bottom left corner is black (0) say no, otherwise say yes



Algorithm: if bottom left corner is black (0) say no, otherwise say yes

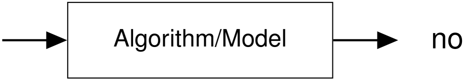


Algorithm: if bottom left corner is black (0) say no, otherwise say yes

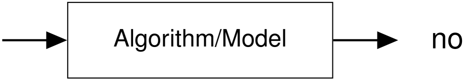


Algorithm: if bottom left corner is black (0) say no, otherwise say yes

Algorithm: if bottom left corner is black (0) say no, otherwise say yes



Algorithm: if bottom left corner is black (0) say no, otherwise say yes

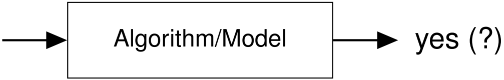


Works perfect on our dataset. :)

Algorithm: if bottom left corner is black (0) say no, otherwise say yes

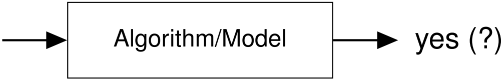
Works perfect on our dataset. :)

Algorithm: if bottom left corner is black (0) say no, otherwise say yes



Works perfect on our dataset. :)

Algorithm: if bottom left corner is black (0) say no, otherwise say yes



Works perfect on our dataset. :)
But does not generalize to other data. :(

Conclusion:

Conclusion:

We designed a simple "classifier" that works on this dataset but doesn't work on real data.

Conclusion:

We designed a simple "classifier" that works on this dataset but doesn't work on real data.

To our rescue:

Conclusion:

We designed a simple "classifier" that works on this dataset but doesn't work on real data.

To our rescue:

Machine learning found mechanisms to search for mappings which generalize.

Conclusion:

We designed a simple "classifier" that works on this dataset but doesn't work on real data.

To our rescue:

Machine learning found mechanisms to search for mappings which generalize.

Scope of this class:

In this lecture we talk about mechanisms. A detailed treatment about generalization is left to lectures on learning theory.

Categorization of pattern recognition algorithms according to

Categorization of pattern recognition algorithms according to

- Available annotated data (supervised vs. unsupervised)

Categorization of pattern recognition algorithms according to

- Available annotated data (supervised vs. unsupervised)
- Complexity of model (linear vs. non-linear)

Categorization of pattern recognition algorithms according to

- Available annotated data (supervised vs. unsupervised)
- Complexity of model (linear vs. non-linear)
- Structure of output (independent vs. structured)

Categorization of pattern recognition algorithms according to

- Available annotated data (supervised vs. unsupervised)
- Complexity of model (linear vs. non-linear)
- Structure of output (independent vs. structured)
- Modeling of data ($x^{(i)}$) or label ($y^{(i)}$) (generative vs. discriminative)

- Syllabus

- Syllabus
- Assignments
- Piazza
- Office hours
- Midterm & Final exam

- Syllabus
- Assignments
- Piazza
- Office hours
- Midterm & Final exam
- Grading policy
  - 3-point: 50% Final + 50% Homework or 25% Final + 25% Midterm + 50% Homework
  - 4-point: 25% Final + 25% Midterm + 50% Homework

Classification tasks:

Classification tasks:

- Spam filter

Classification tasks:

- Spam filter
- Cancer classification

Classification tasks:

- Spam filter
- Cancer classification
- ...

Classification tasks:

- Spam filter
- Cancer classification
- ...

How to address those problems?

Nearest Neighbor

Nearest Neighbor

- Dataset:

Nearest Neighbor

- Dataset: $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$

Nearest Neighbor

- Dataset: $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$
- New datapoint: x

Nearest Neighbor

- Dataset: $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$
- New datapoint: x
- Label of new datapoint: y

Nearest Neighbor

- Dataset: $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$
- New datapoint: x
- Label of new datapoint: y

How to choose y?

Nearest Neighbor

- Dataset: $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$
- New datapoint: x
- Label of new datapoint: y

How to choose y?

$y = y^{(k)} \quad \text{where} \quad k =$

Nearest Neighbor

- Dataset: $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$
- New datapoint: x
- Label of new datapoint: y

How to choose y?

$$y = y^{(k)} \quad \text{where} \quad k = \arg\min_{i \in \{1,\ldots,N\}} \|x^{(i)} - x\|_2^2$$

Nearest Neighbor

- Dataset: $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$
- New datapoint: x
- Label of new datapoint: y

How to choose y?

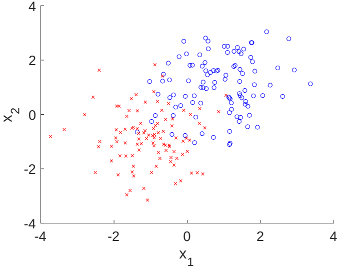
$$y = y^{(k)} \quad \text{where} \quad k = \arg\min_{i \in \{1,\ldots,N\}} \|x^{(i)} - x\|_2^2 = \arg\min_{i \in \{1,\ldots,N\}} d(x^{(i)}, x)$$

Nearest Neighbor

- Dataset: $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$
- New datapoint: x
- Label of new datapoint: y

How to choose y?

$$y = y^{(k)} \quad \text{where} \quad k = \arg\min_{i \in \{1,\ldots,N\}} \|x^{(i)} - x\|_2^2 = \arg\min_{i \in \{1,\ldots,N\}} d(x^{(i)}, x)$$

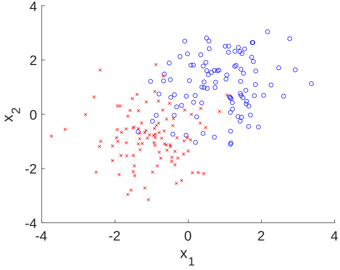


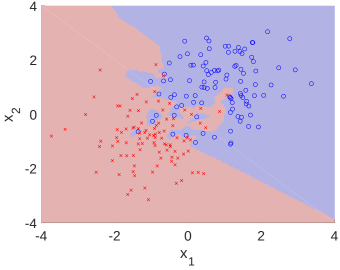
Nearest Neighbor

- Dataset: $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^N$
- New datapoint: x
- Label of new datapoint: y

How to choose y?

$$y = y^{(k)} \quad \text{where} \quad k = \arg\min_{i \in \{1,\ldots,N\}} \|x^{(i)} - x\|_2^2 = \arg\min_{i \in \{1,\ldots,N\}} d(x^{(i)}, x)$$



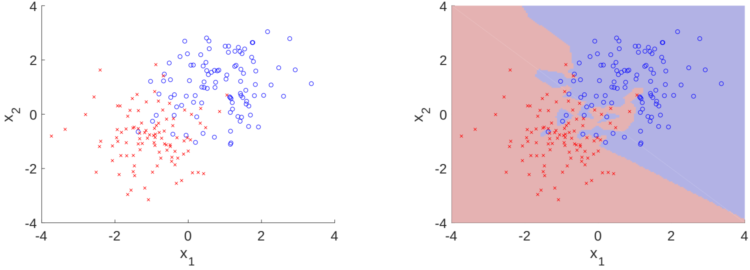


Nearest Neighbor

- Dataset: $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$
- New datapoint: x
- Label of new datapoint: y

How to choose y?

$$y = y^{(k)} \quad \text{where} \quad k = \arg\min_{i\in\{1,\ldots,N\}} \|x^{(i)} - x\|_2^2 = \arg\min_{i\in\{1,\ldots,N\}} d(x^{(i)}, x)$$



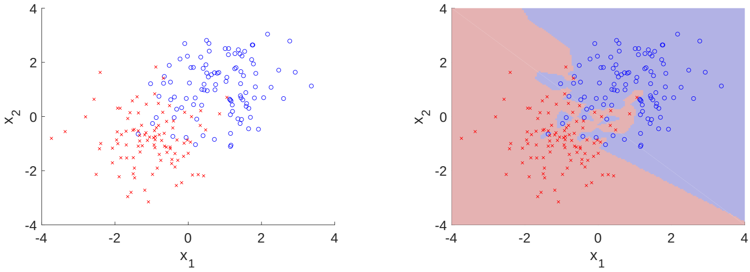
Shortcomings?

Nearest Neighbor

- Dataset: $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$
- New datapoint: x
- Label of new datapoint: y

How to choose y?

$$y = y^{(k)} \quad \text{where} \quad k = \arg\min_{i \in \{1,\ldots,N\}} \|x^{(i)} - x\|_2^2 = \arg\min_{i \in \{1,\ldots,N\}} d(x^{(i)}, x)$$



Shortcomings?

k-Nearest Neighbors

More applications far beyond Nearest Neighbor. How would you solve them with Nearest Neighbor?

Example: Differentiating between 1000 image categories

Example: Differentiating between 1000 image categories

Which object is illustrated?

Example: Differentiating between 1000 image categories

Which object is illustrated?

- Car

Example: Differentiating between 1000 image categories

Which object is illustrated?

- Car
- Truck

Example: Differentiating between 1000 image categories

Which object is illustrated?

- Car
- Truck
- Recreational Vehicle

Example: Differentiating between 1000 image categories

Which object is illustrated?

- Car
- Truck
- Recreational Vehicle
- Ambulance truck

Example: Differentiating between 1000 image categories

Which object is illustrated?

- Car
- Truck
- Recreational Vehicle
- Ambulance truck
- Fire truck

Example: Instance level video segmentation

Example: Describing an image

Works

Example: Describing an image

Works

a group of people standing around a room with remotes
`logprob: -9.17`

a young boy is holding a baseball bat
`logprob: -7.61`

a cow is standing in the middle of a street
`logprob: -8.84`

Works, ...

Works, ...

a cat is sitting on a toilet seat
logprob: -7.79

a display case filled with lots of different types of donuts
logprob: -7.78

a group of people sitting at a table with wine glasses
logprob: -6.71

Kind of works, ...

Kind of works, ...

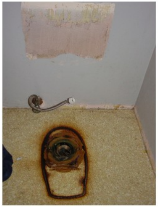
a man standing next to a clock on a wall
logprob: -10.08

a young boy is holding a baseball bat
logprob: -7.65

a cat is sitting on a couch with a remote control
logprob: -12.45

And fails occasionally too

And fails occasionally too



a toilet with a seat up in a bathroom
logprob: -13.44

a woman holding a teddy bear in front of a mirror
logprob: -9.65

a horse is standing in the middle of a road
logprob: -10.34

Creativity

Example: Describing an image

Example: Describing an image

Object Labels: 'person'
AG-CVAE sentences:
a man and a woman standing in a room
a man and a woman are playing a game
a man standing next to a woman in a room
a man standing next to a woman in a field
a man standing next to a woman in a suit

Object Labels: 'person', 'remote'
AG-CVAE sentences:
a man and a woman playing a video game
a man and a woman are playing a video game
a man and woman are playing a video game
a man and a woman playing a game with a remote
a woman holding a nintendo wii game controller

Object Labels: 'person','bus'
AG-CVAE sentences:
a man and a woman sitting on a bus
a man and a woman sitting on a train
a man and woman sitting on a bus
a man and a woman sitting on a bench
a man and a woman are sitting on a bus

Object Labels: 'person', 'train'
AG-CVAE sentences:
a man and a woman sitting on a train
a woman and a woman sitting on a train
a woman sitting on a train next to a train
a woman sitting on a bench in a train
a man and a woman sitting on a bench

Example: Asking questions about an image

Example: Asking questions about an image

So far:

What can a computer do for you?

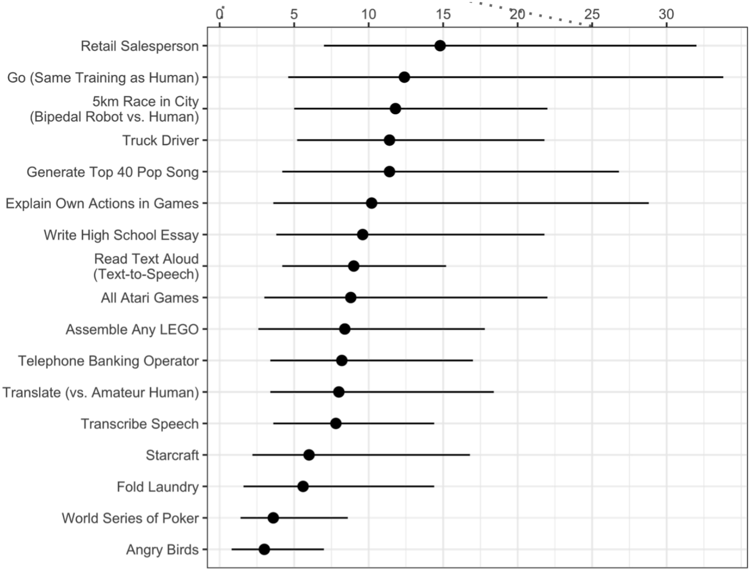
So far:

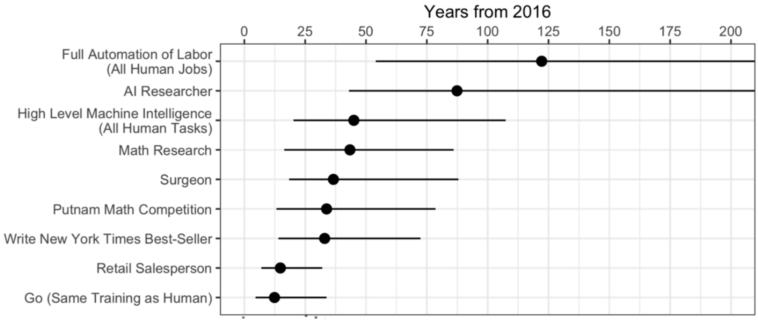
What can a computer do for you?

Now:

What will a computer be able to do for you?

Future: (years from 2016; from https://arxiv.org/pdf/1705.08807.pdf)




Years from 2016
0
25
50
75
100
125
150
175
200
Full Automation of Labor
(All Human Jobs)
AI Researcher
High Level Machine Intelligence
(All Human Tasks)
Math Research
Surgeon
Putnam Math Competition
Write New York Times Best-Seller
Retail Salesperson
Go (Same Training as Human)

Videos

Amazing future ahead. Let's make it happen.

Quiz:

Quiz:

- What is nearest neighbor?

Quiz:

- What is nearest neighbor?
- How do we formulate nearest neighbor?

Important topics of this lecture:

Important topics of this lecture:

- Getting to know k-Nearest Neighbor

Important topics of this lecture:

- Getting to know k-Nearest Neighbor
- Basic mechanism to address many tasks

Important topics of this lecture:

- Getting to know k-Nearest Neighbor
- Basic mechanism to address many tasks
- Think about how to solve a problem with Nearest Neighbor