Lecture 40 – final exam review

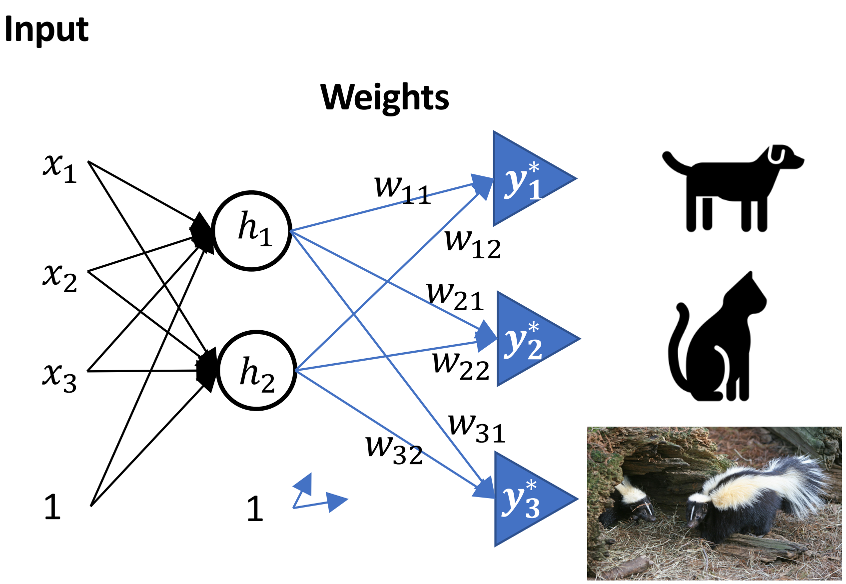
Mark Hasegawa-Johnson

5/6/2020

Some sample problems

- DNNs: Practice Final, question 23
- Reinforcement learning: Practice Final, question 24
- Games: Practice Final, question 25
- Game theory: Practice Final, question 26

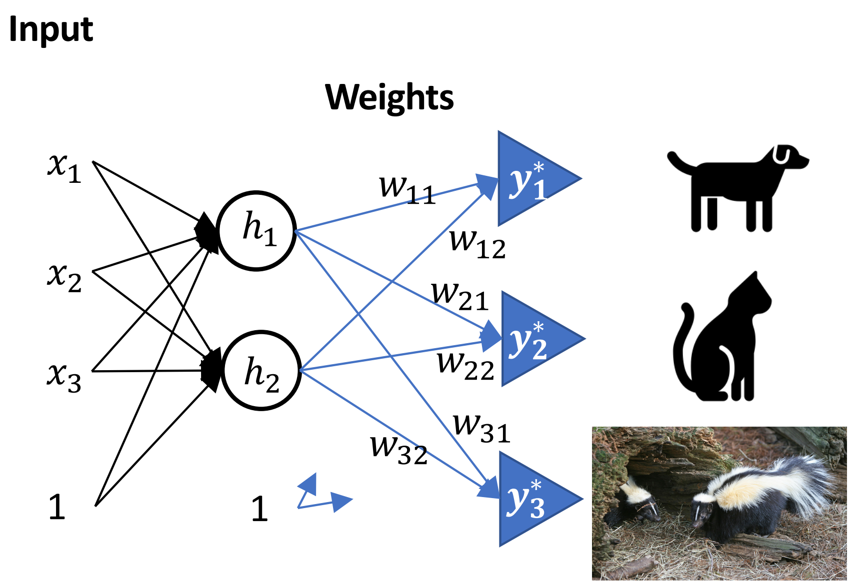
Practice Exam, question 23



By http://www.birdphotos.com - Own work, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=4409510

You have a two-layer neural network trained as an animal classier. The input feature vector is $\vec{x} = [x_1, x_2, x_3, 1]$, where x_1, x_2, and x_3 are some features, and 1 is multiplied by the bias. There are two hidden nodes, and three output nodes, $\vec{y}^* = [y_1^*, y_2^*, y_3^*,]$, corresponding to the three output classes $y_1^* = \mathrm{Pr}(\mathrm{dog}|\,\vec{x})$, $y_2^* = \mathrm{Pr}(\mathrm{cat}|\,\vec{x})$, $y_3^* = \mathrm{Pr}(\mathrm{skunk}|\,\vec{x})$. Hidden node activations are sigmoid; output node activations are softmax.

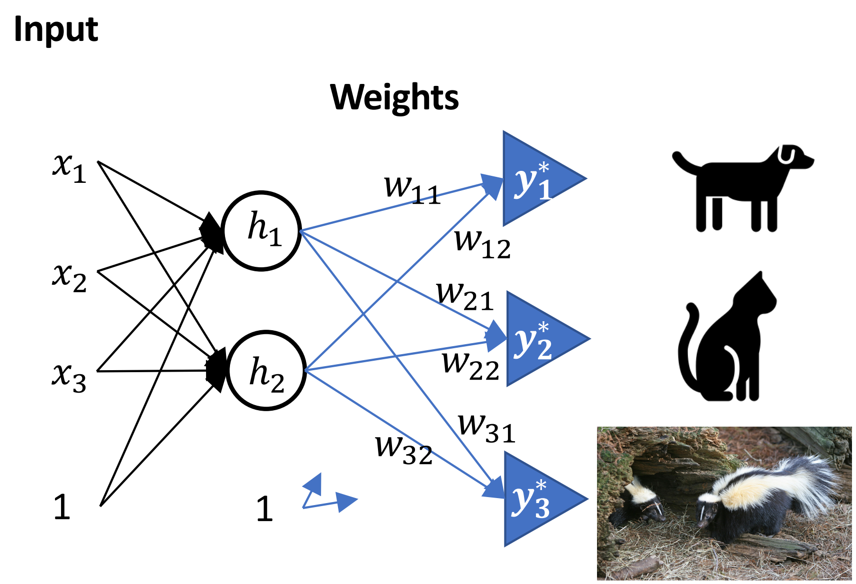
Practice Exam, question 23



By http://www.birdphotos.com - Own work, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=4409510

(a) A Maltese puppy has feature vector $\vec{x} = [2,20,-1,1]$. All weights and biases are initialized to zero. What is $\vec{y}^*$?

Practice Exam, question 23



By http://www.birdphotos.com - Own work, CC BY 3.0, https://commons.wikimedia.org /w/index.php?curid=4409510

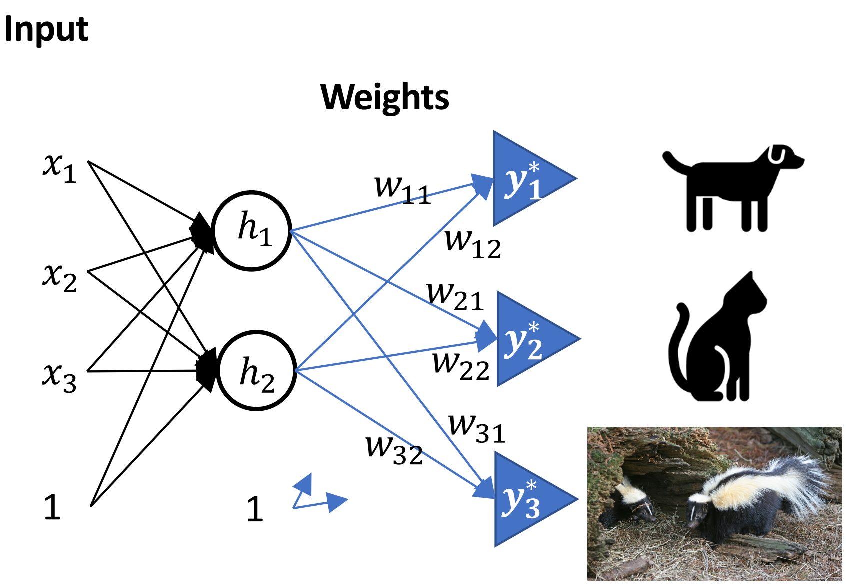
(a) A Maltese puppy has feature vector $\vec{x} = [2,20,-1,1]$. All weights and biases are initialized to zero. What is $\vec{y}^*$?

Hidden node excitations are both: $0\times\vec{x} = 0$

Therefore, hidden node activations are both:

$$\frac{1}{1+e^{-0}} = \frac{1}{1+1} = \frac{1}{2}$$

Practice Exam, question 23



By http://www.birdphotos.com - Own work, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=4409510

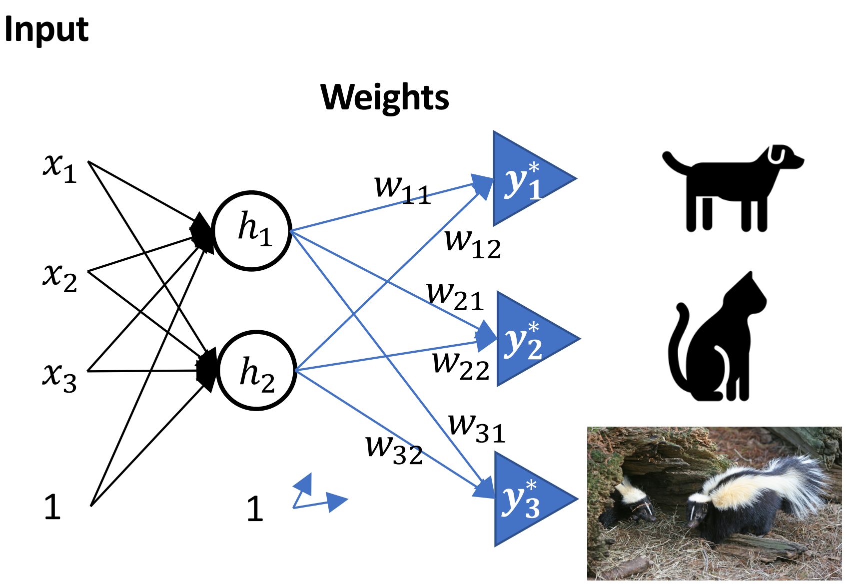
(a) A Maltese puppy has feature vector $\vec{x} = [2,20,-1,1]$. All weights and biases are initialized to zero. What is $\vec{y}^*$?

Output node excitations are all:
$0 \times \vec{h} = 0$

Therefore, output node activations are all:

$$\frac{e^0}{\sum_{i=1}^{3} e^0} = \frac{1}{3}$$

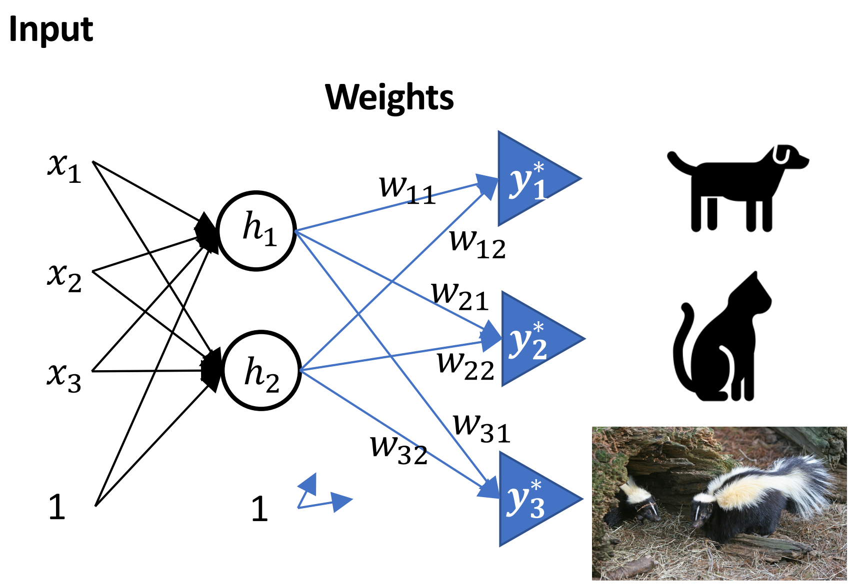
Practice Exam, question 23



By http://www.birdphotos.com - Own work, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=4409510

(b) Let w_{ij} be the weight connecting the ith output node to the jth hidden node. What is $\frac{dy_2^*}{dw_{21}}$? Write your answer in terms of y_i^*, w_{ij}, and/or h_j for appropriate values of i and/or j.

Practice Exam, question 23



By http://www.birdphotos.com - Own work, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=4409510

(b) What is $\frac{dy_2^*}{dw_{21}}$?

Answer: OK, first we need the definition of softmax. Let's write it in lots of parts, so it will be easier to differentiate.

$$y_2^* = \frac{\text{num}}{\text{den}}$$

Where "num" is the numerator of the softmax function:

$$\text{num} = \exp(f_2)$$

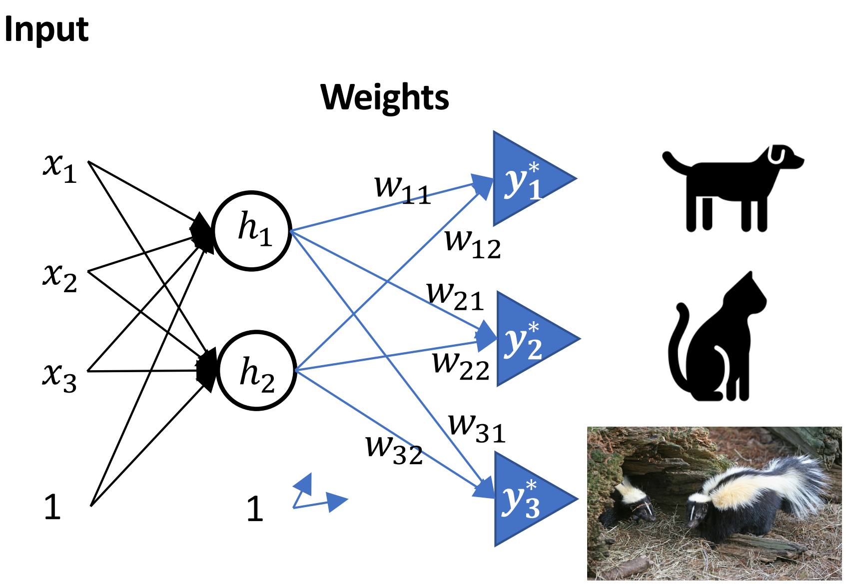
"den" is the denominator of the softmax function:

$$\text{den} = \sum_{i=1}^{3} \exp(f_i)$$

And both of those are written in terms of the softmax excitations, let's call them f_i:

$$f_i = \sum_j w_{ij} h_i$$

Practice Exam, question 23



By http://www.birdphotos.com - Own work, CC BY 3.0, https://commons.wikimedia.org /w/index.php?curid=4409510

(b) What is $\frac{dy_2^*}{dw_{21}}$?

Now we differentiate each part:

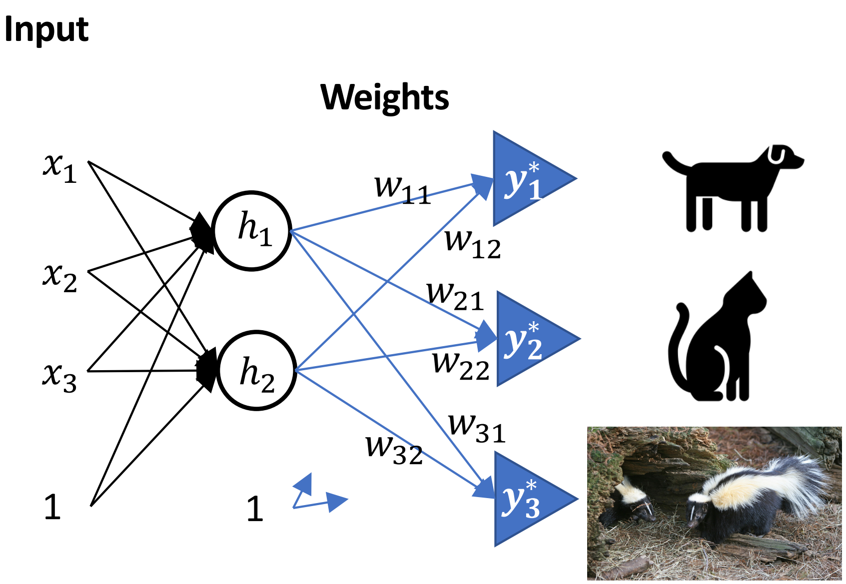
$$\frac{dy_2^*}{dw_{21}} = \left(\frac{1}{\text{den}}\right)\frac{d\text{num}}{dw_{21}} - \left(\frac{\text{num}}{\text{den}^2}\right)\frac{d\text{den}}{dw_{21}}$$

$$\frac{d\text{num}}{dw_{21}} = \exp(f_2)\frac{df_2}{dw_{21}}$$

$$\frac{d\text{den}}{dw_{21}} = \sum_{i=1}^{3}\exp(f_i)\frac{df_i}{dw_{21}} = \exp\left(f_2\right)\frac{df_2}{dw_{21}}$$

$$\frac{df_2}{dw_{21}} = h_1$$

Practice Exam, question 23



By http://www.birdphotos.com - Own work, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=4409510

(b) What is $\frac{dy_2^*}{dw_{21}}$?

Putting it all back together again:

$$\frac{dy_2^*}{dw_{21}}$$

$$= \left(\frac{1}{\sum_{i=1}^{3}\exp(f_i)}\right)\exp(f_2)h_1$$

$$-\left(\frac{\exp(f_2)}{\left(\sum_{i=1}^{3}\exp(f_i)\right)^2}\right)\exp(f_2)h_1$$

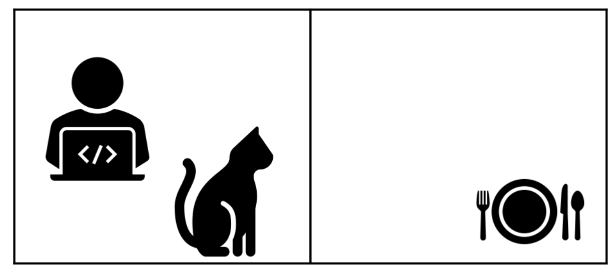
$$\frac{dy_2^*}{dw_{21}} = y_2^* h_1 - (y_2^*)^2 h_1$$

Some sample problems

- DNNs: Practice Final, question 23
- Reinforcement learning: Practice Final, question 24
- Games: Practice Final, question 25
- Game theory: Practice Final, question 26

Practice Exam, question 24

A cat lives in a two-room apartment. It has two possible actions: purr, or walk. It starts in room s0 = 1, where it receives the reward r0 = 2 (petting). It then implements the following sequence of actions: a0 =walk, a1 =purr. In response, it observes the following sequence of states and rewards: s1 = 2, r1 = 5 (food), s2 = 2.

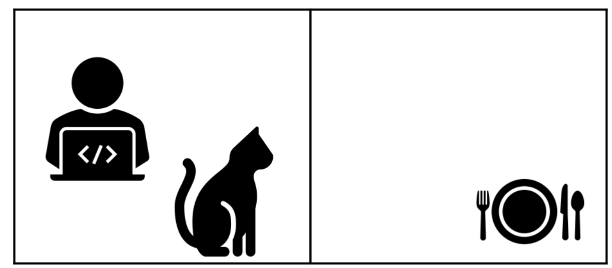


Practice Exam, question 24

(a) The cat starts out with a Q-table whose entries are all Q(s,a) = 0.

- …then performs one iteration of TD-learning using each of the two SARS sequences described above.
- …it uses a relatively high learning rate (alpha = 0.05) and a relatively low discount factor (gamma = 3/4).

Which entries in the Q-table have changed, after this learning, and what are their new values?



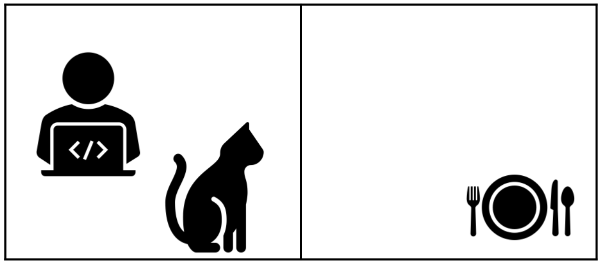
Practice Exam, question 24

Time step 0:

$$SARS = (1, walk, 2,2)$$
$$Qlocal \ = \ R(1) + \gamma \max_a Q(2,a)$$
$$= \ 2 \ + \left(\frac{3}{4}\right)\max(0,0) = 2$$
$$Q(1,w) \ = \ Q(1,w) \ + \alpha(Qlocal - Q(1,w))$$
$$= \ 0 \ + \ 0.05 * (2-0) \ = \ 0.1$$

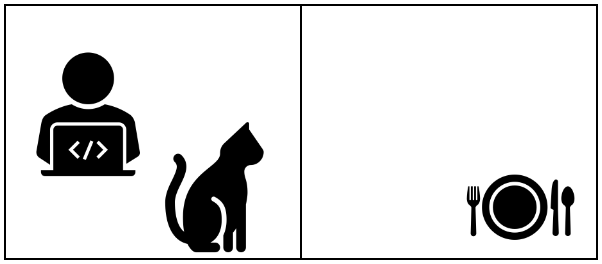
Time step 1:

$$SARS = (2, purr, 5,2)$$
$$Qlocal \ = \ R(2) + \gamma \max_a Q(2,a)$$
$$= 5 \ + \left(\frac{3}{4}\right)\max(0,0) = 5$$
$$Q(2,purr) \ = \ Q(2,p) \ + \alpha(Qlocal - Q(2,p))$$
$$= \ 0 \ + \ 0.05 * (5-0) \ = \ 0.25$$



Practice Exam, question 24

(b) The cat decides, instead, to use model-based learning. Based on these two observations, it estimates P(s'|s,a) with Laplace smoothing, where the smoothing constant is k=1. Find P(s'|2,purr).



Time step 0:

$$SARS = (1, walk, 2,2)$$

Time step 1:

$$SARS = (2, purr, 5,2)$$

Practice Exam, question 24

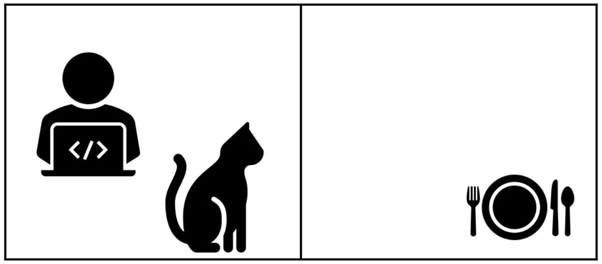
(b) Find P(s'|2,purr).

$$\mathrm{P}(s' = 1 | s = 2, a = purr)$$

$$= \frac{1 + Count(s = 2, a = purr, s' = 1)}{2 + \sum Count(s = 2, a = purr, s')} = \frac{1}{2+1}$$

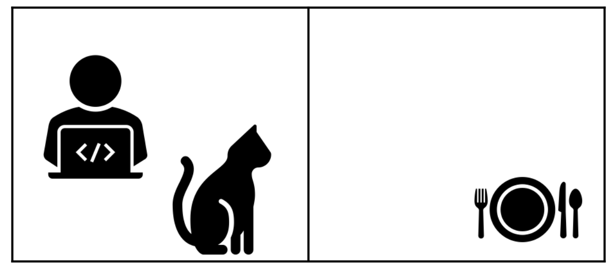
$$\mathrm{P}(s' = 2 | s = 2, a = purr)$$

$$= \frac{1 + Count(s = 2, a = purr, s' = 2)}{2 + \sum Count(s = 2, a = purr, s')} = \frac{1+1}{2+1}$$



Practice Exam, question 24

(c) The cat estimates R(1)=2, R(2)=5, and the following P(s'|s,a) table. It chooses the policy pi(1)=purr, pi(2)=walk. What is the policy-dependent utility of each room? Write two equations in the two unknowns U(1) and U(2); don't solve.



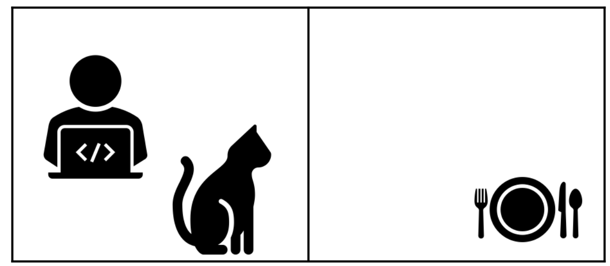
	a=purr		a=walk	
	s=1	s=2	s=1	s=2
s'=1	2/3	1/3	1/3	2/3
s'=2	1/3	2/3	2/3	1/3

Practice Exam, question 24

(c) Answer: policy-dependent utility is just like Bellman's equation, but without the max operation. The equations are

$$U(1) = R(1) + \gamma \sum_{s'} P(s'|s = 1, \pi(1)) U(s')$$

$$U(2) = R(2) + \gamma \sum_{s'} P(s'|s = 2, \pi(2)) U(s')$$



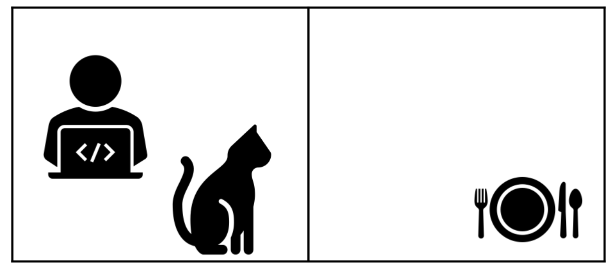
	a=purr		a=walk	
	s=1	s=2	s=1	s=2
s'=1	2/3	1/3	1/3	2/3
s'=2	1/3	2/3	2/3	1/3

Practice Exam, question 24

(c) Answer: So to solve, we just plug in the values for all variables except U(1) and U(2):

$$U(1) = 2 + \left(\frac{3}{4}\right)\left(\frac{2}{3}U(1) + \frac{1}{3}U(2)\right)$$

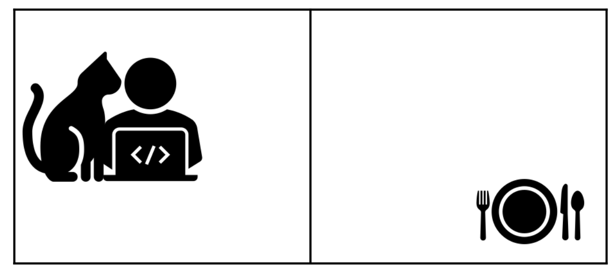
$$U(2) = 5 + \left(\frac{3}{4}\right)\left(\frac{2}{3}U(1) + \frac{1}{3}U(2)\right)$$



	a=purr		**a=walk**	
	s=1	s=2	s=1	s=2
s’=1	2/3	1/3	1/3	2/3
s’=2	1/3	2/3	2/3	1/3

Practice Exam, question 24

(d) Since it has some extra time, and excellent python programming skills, the cat decides to implement deep reinforcement learning, using an actor-critic algorithm. Inputs are one-hot encodings of state and action. What are the input and output dimensions of the actor network, and of the critic network?

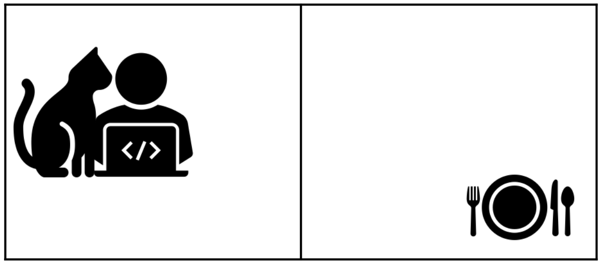


Practice Exam, question 24

(d)

Actor network is $\pi_a\ (s) =$ probability that action a is the best action, where a=1 or a=2. So output has two dimensions.

Input is the state, s. If there are two states, encoded using a one-hot vector, then state 1 is encoded as $s = [1,0]$, state 2 is encoded as $s = [0,1]$. So, two dimensions.

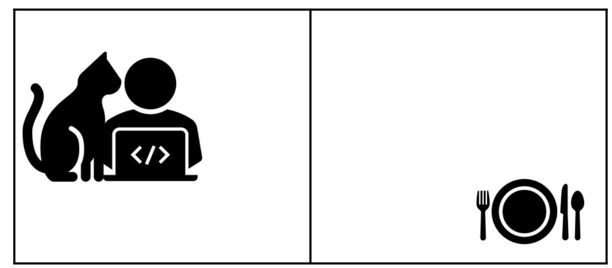


Practice Exam, question 24

(d)

Critic network is $Q(s, a) =$ quality of action a in state s. Quality is a scalar (for any given action and state), so output has one dimension (scalar).

Input is the state, s, and the action, a. Problem statement says that each is a one-hot vector, so $s = [1,0]$ or $s = [0,1]$, concatenated with $a = [1,0]$ or $a = [0,1]$, for a total of 4 dimensions.



Some sample problems

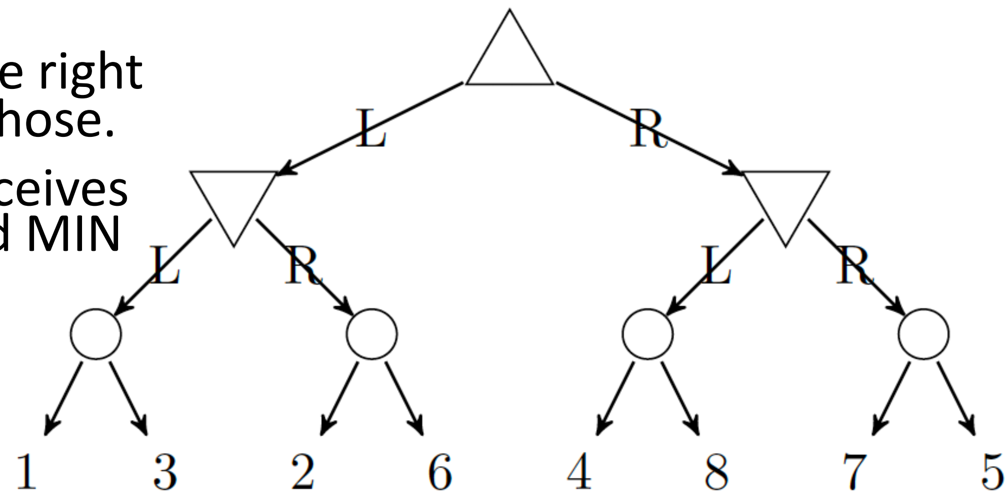
- DNNs: Practice Final, question 23
- Reinforcement learning: Practice Final, question 24
- Games: Practice Final, question 25
- Game theory: Practice Final, question 26

Practice Exam, question 25

Consider a game with eight cards, sorted onto the table in four stacks of two cards each. MAX and MIN each know the contents of each stack, but they don't know which card is on top. The game proceeds as follows.

1. MAX chooses either the left or the right pair of stacks.
2. MIN chooses either the left or the right stack, within the pair that MAX chose.
3. The top card is revealed. MAX receives the face value of the card (c), and MIN receives 9-c.

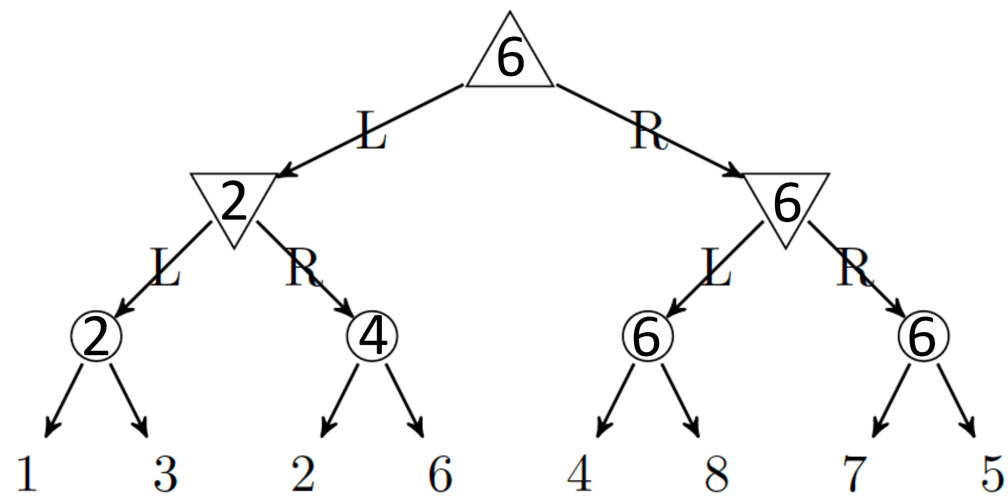
Girl with Cards by Lucius Kutchin, 1933, Smithsonian American Art Museum



Practice Exam, question 25

(a) What is the value of the MAX node?

Girl with Cards by Lucius Kutchin, 1933, Smithsonian American Art Museum



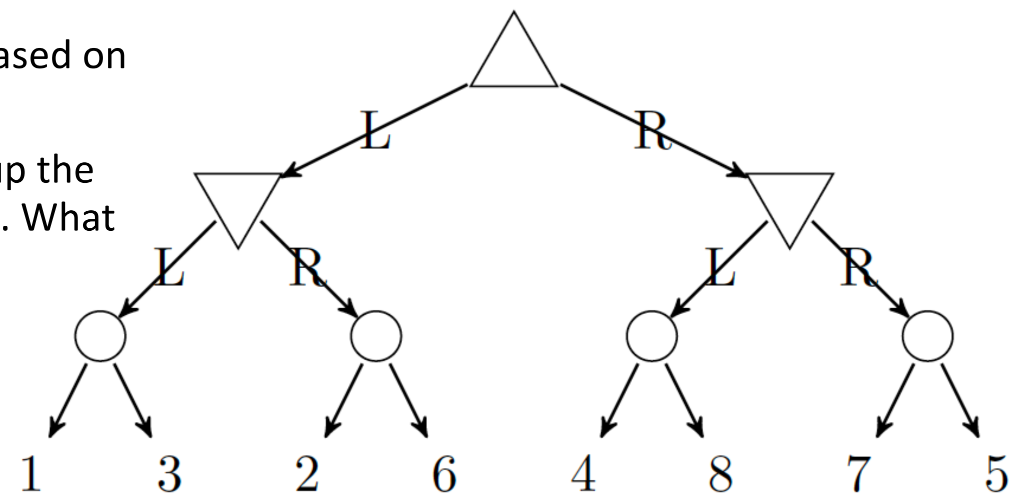
Practice Exam, question 25

Girl with Cards by Lucius Kutchin, 1933, Smithsonian American Art Museum

Rule change: after MAX chooses a pair of stacks, he is permitted to look at the top card in any one stack. He must show the card to MIN, then replace it, so that it remains the top card in that stack. Define the belief state, b, to be the set of all possible outcomes of the game, i.e., the starting belief state is the set b = {1,2,3,4,5,6,7,8}.

1. PREDICT operation modifies the belief state based on the action of a player.
2. OBSERVE operation modifies the belief state based on MAX's observation.

Suppose MAX chooses the action R. He then turns up the top card in the rightmost deck, revealing it to be a 7. What is the resulting belief state?



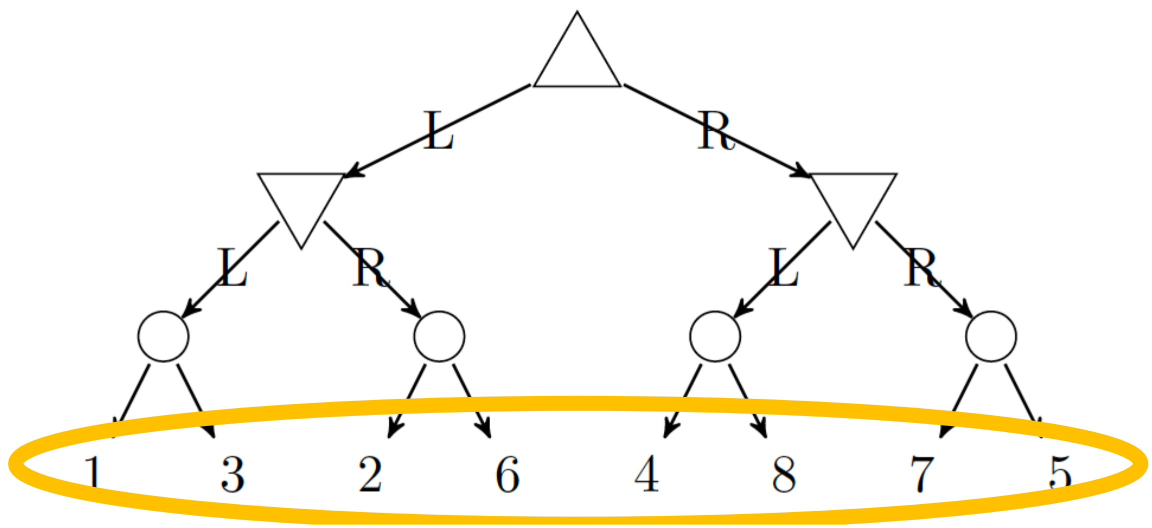
Practice Exam, question 25

Starting belief state is the set b = {1,2,3,4,5,6,7,8}.

1. PREDICT operation modifies the belief state based on the action of a player. (MAX chooses the action R).
2. OBSERVE operation modifies the belief state based on MAX's observation. (MAX observes that 7 is on top).

Girl with Cards by Lucius Kutchin, 1933, Smithsonian American Art Museum

Starting belief state is the set b = {1,2,3,4,5,6,7,8}.



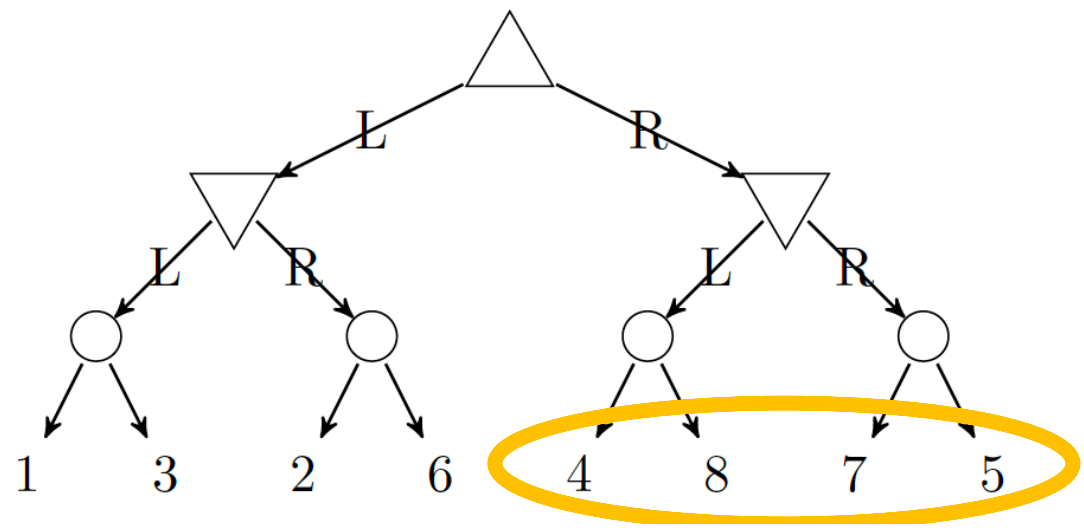
Practice Exam, question 25

Starting belief state is the set b = {1,2,3,4,5,6,7,8}.

1. PREDICT operation modifies the belief state based on the action of a player. (MAX chooses the action R).
2. OBSERVE operation modifies the belief state based on MAX's observation. (MAX observes that 7 is on top).

Girl with Cards by Lucius Kutchin, 1933, Smithsonian American Art Museum

MAX chooses the action R.



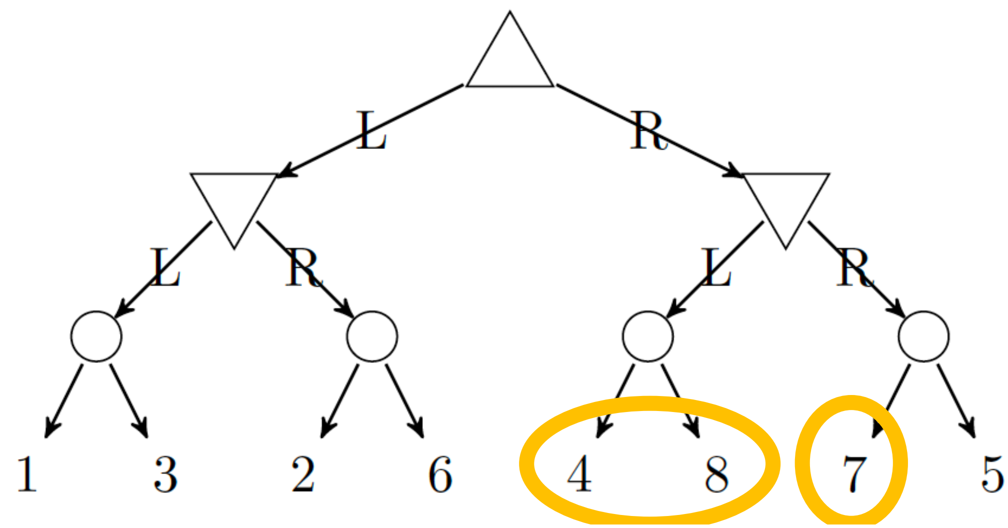
Practice Exam, question 25

Girl with Cards by Lucius Kutchin, 1933, Smithsonian American Art Museum

Starting belief state is the set b = {1,2,3,4,5,6,7,8}.

1. PREDICT operation modifies the belief state based on the action of a player. (MAX chooses the action R).
2. OBSERVE operation modifies the belief state based on MAX's observation. (MAX observes that 7 is on top).

Final belief state is therefore b={4,8,7}.



MAX observes that 7 is on top of 5.

Some sample problems

- DNNs: Practice Final, question 23
- Reinforcement learning: Practice Final, question 24
- Games: Practice Final, question 25
- Game theory: Practice Final, question 26

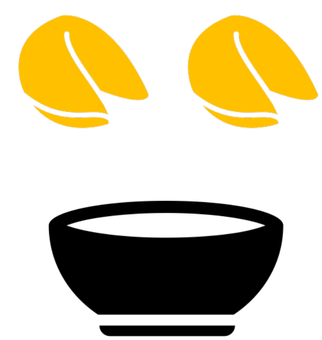
Practice exam, question 26

(a). Two cookies, three roommates.

We decide to use a VCG auction, with proceeds going into a cookie fund.

…and the bids are:

$5 $3 $6



Calculate the net value (value received minus price paid) of each roommate, and of the cookie fund.

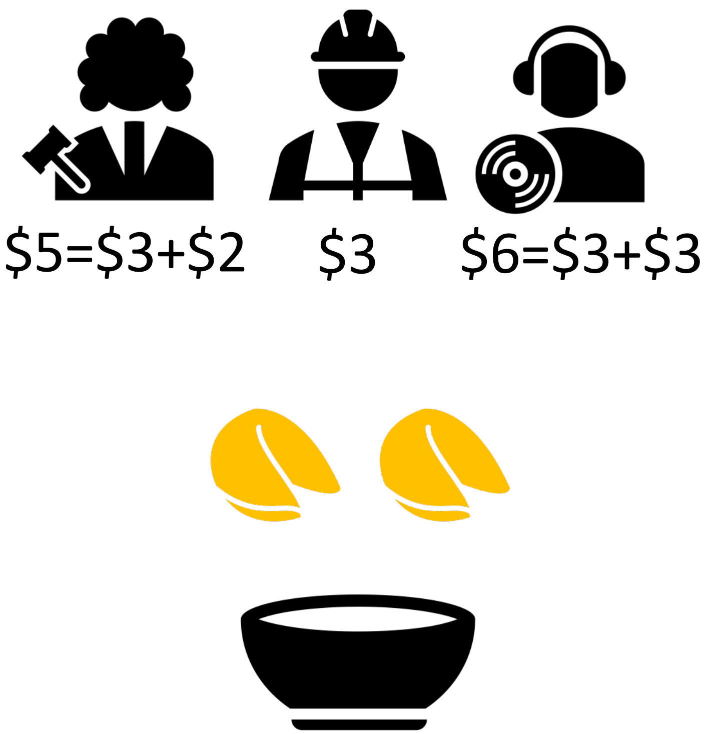
Practice exam, question 26

VCG auction:

Cookies go to the N highest bidders, i.e., the judge and the DJ.

They each pay b(N+1), i.e., \$3.

Because they each pay b(N+1), it's a dominant strategy to bid what the cookie is really worth to each of them, so we can assume that's what they've done.



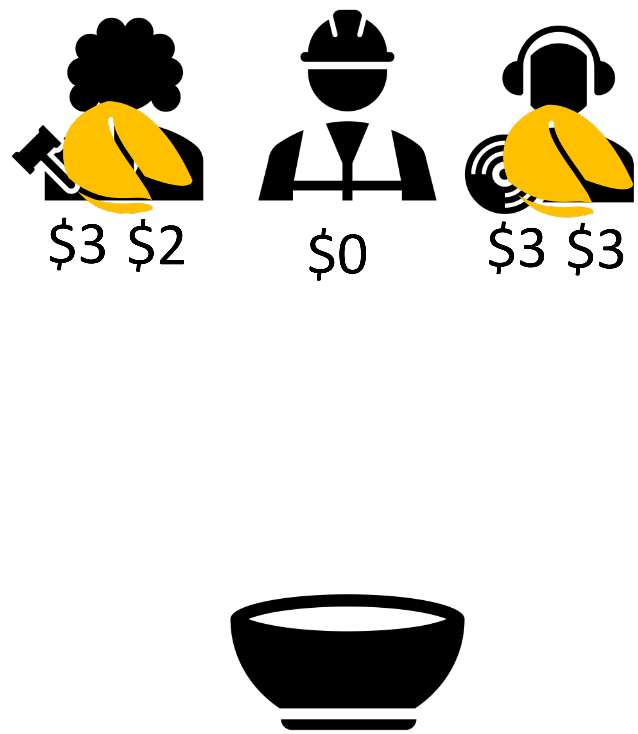
Practice exam, question 26

Value to the construction worker: $0, because they didn't get a cookie, or spend any money.

Value to the judge: $5 (value of the cookie) - $3 (price paid) = $2

Value to the DJ: $6 (value of the cookie) - $3 (price paid) = $3

Value to the cookie fund: 2 * $3 = $6



Practice exam, question 26

(b). Three cookies, two roommates.

One cookie is deluxe, worth \$10.

The other two are regular, worth \$1 each.

Possible outcomes:

1. A chooses deluxe (\$10), B chooses regular, then B gets the third (\$2), or vice versa.
2. A and B each choose a regular, then they split the deluxe (\$6 each).
3. A and B each choose deluxe, then they fight, and the dog eats all of the cookies (\$0).

Practice exam, question 26

Find the mixed-strategy Nash equilibrium.

	Regular	Deluxe
Regular	6, 6	2, 10
Deluxe	10, 2	0, 0

Practice exam, question 26

	Regular	Deluxe
Regular	6, 6	2, 10
Deluxe	10, 2	0, 0

Find the mixed-strategy Nash equilibrium.

If  chooses deluxe with probability p, then it is rational for  to choose randomly only if
$$2p + 6(1-p) = 0p + 10(1-p)$$

…in other words, random choice is rational for  only if $p = 2/3$.

Practice exam, question 26

	Regular	Deluxe
Regular	6, 6	2, 10
Deluxe	10, 2	0, 0

 : Random choice is rational only if  chooses deluxe with probability $p = 2/3$.

 : Random choice is rational only if  chooses deluxe with probability $q = 2/3$.

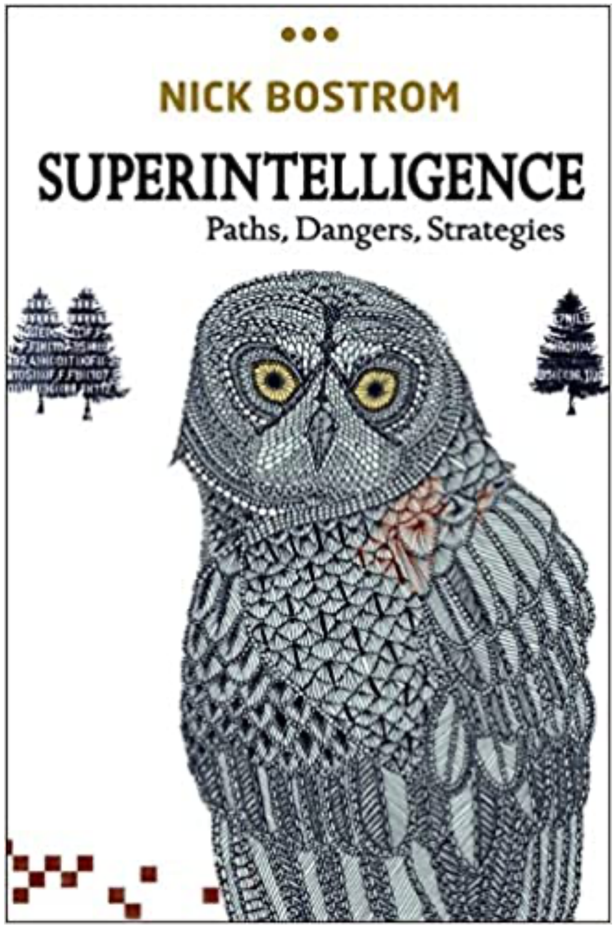
So $p = 2/3$, $q = 2/3$ is a Nash equilibrium.

Final thoughts…

- Some books worth reading

Superintelligence (2014)

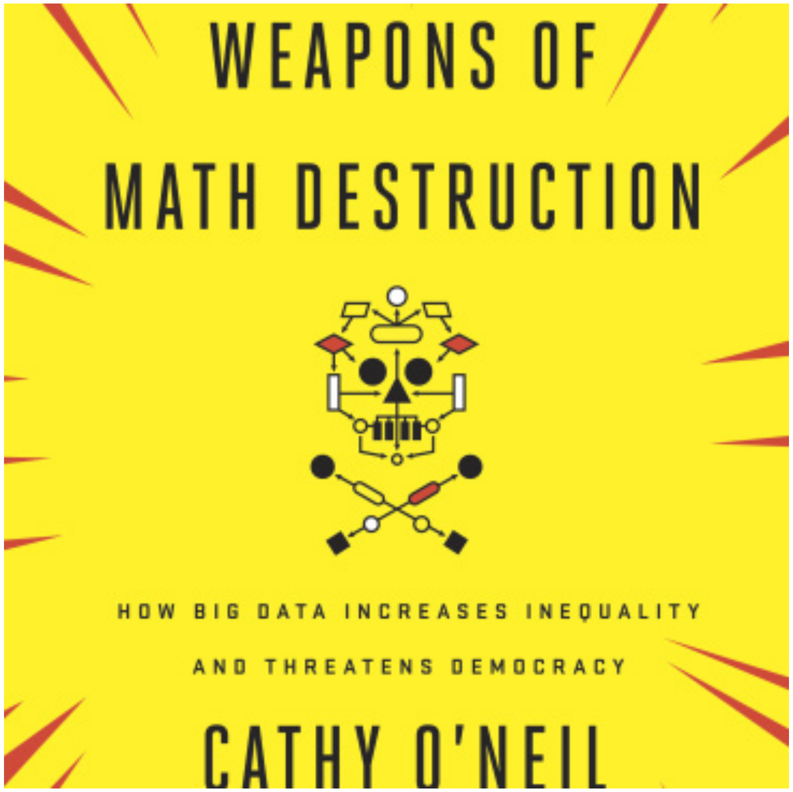
What would happen if we produced an AI with the goal of making as many paper clips as possible… and it succeeded?



Weapons of math destruction (2016)

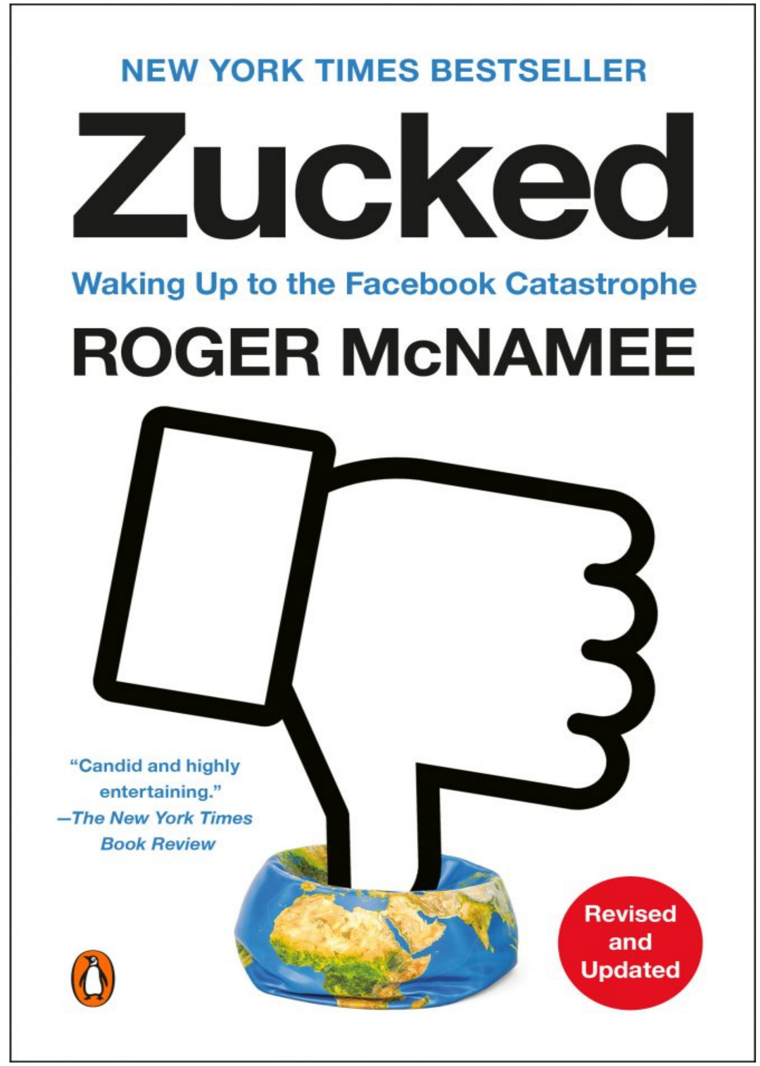
A “weapon of math destruction” is a statistical model used in a way that is

- Scaled beyond the level for which it was designed
- Measures a proxy-measure, rather than the thing it’s actually trying to optimize
- Blind to the actual outcomes it produces



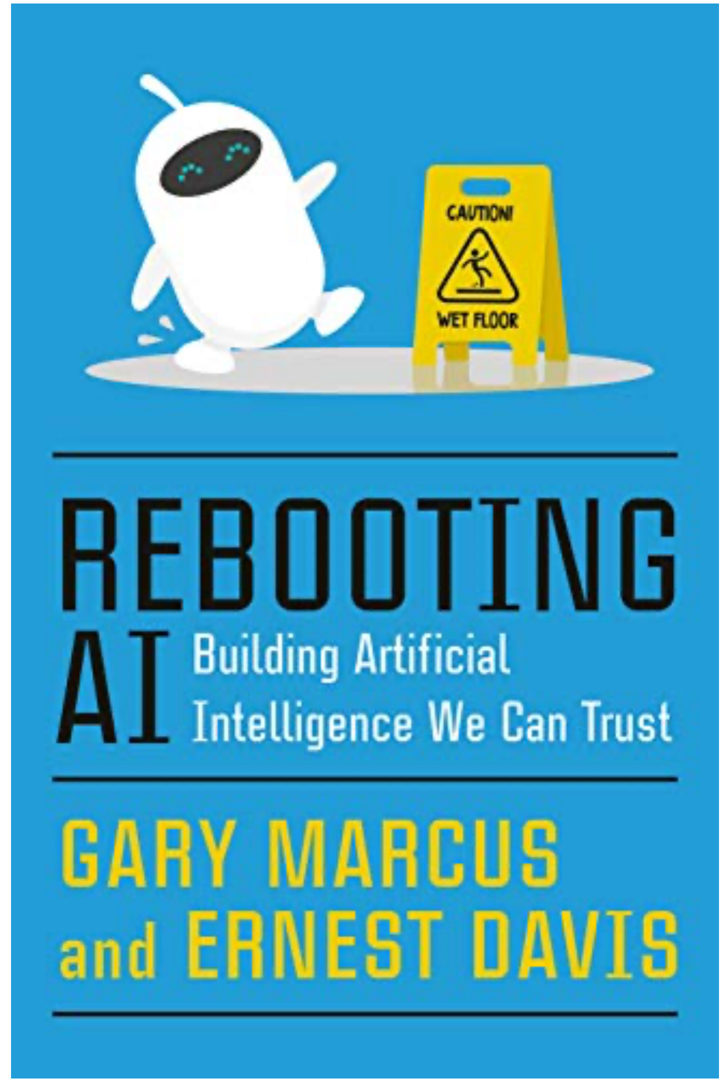
Zucked (2019)

Uses Facebook as an illustrative model of the way in which the drive to provide customers what they want is often, but not always, in the best interest of society.

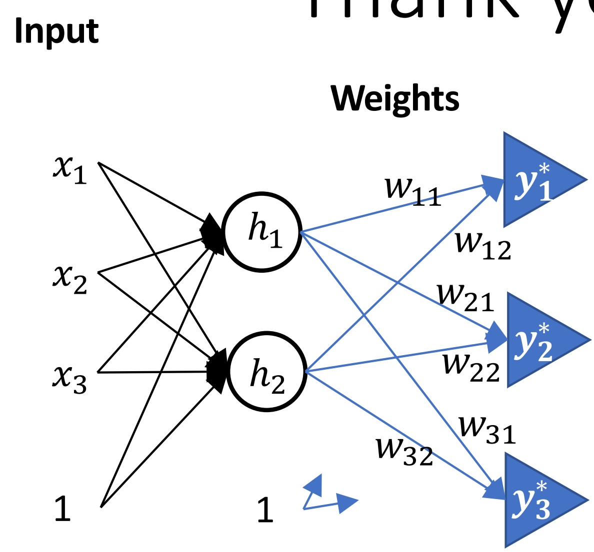


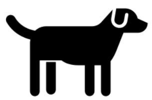
Rebooting AI (2019)

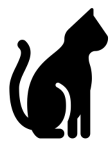
Argues that the greatest threat of AI is not that it will replace human beings, but that it will fail outrageously, in ways human beings are unable to predict, because no human would ever fail in that way.



Thank you! Have a happy summer!







By http://www.birdphotos.com - Own work, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=4409510