Q1 Local Optima and Gradient Descent

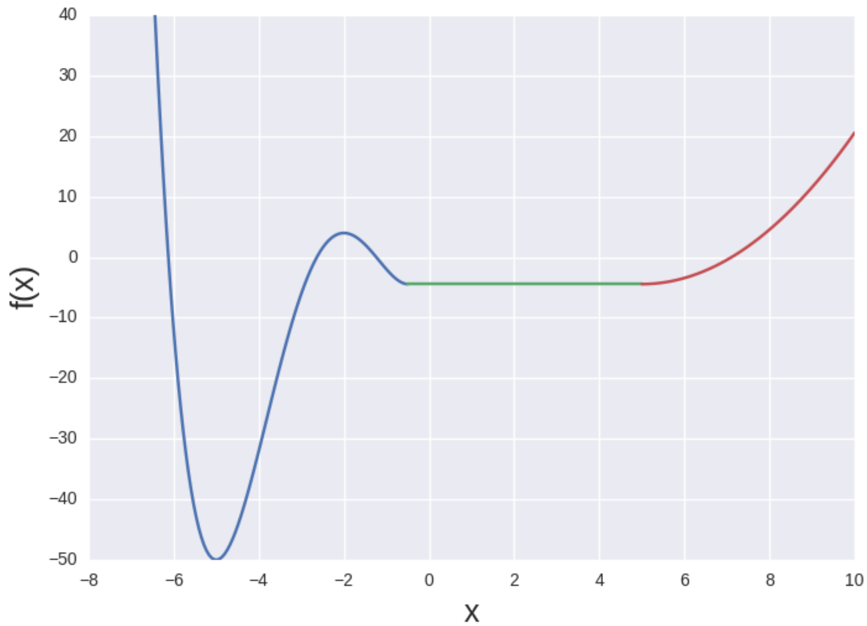
43 Points

After a busy year of chasing ghosts, Pacman and Paclady are planning to visit the Kakslauttanen Arctic Resort for their winter vacation. Paclady who is particularly fond of skiing, excitedly begins planning ahead. Pacman, who is apprehensive of skiing (when asked why, he rambles on about the Aspen Red Ghost Chase of 2012, but we won't get into that), reluctantly agreed to go skiing, but under one condition: Paclady must tell Pacman how steep the slopes are at several points of interest.

Paclady asks the resort for terrain details, and receives the following graph. The resort says at any given location x, $f(x)$ models the terrain height. Specifically:

- when $x \leq -\frac{1}{2}$, $f(x) = \frac{1}{2}x^4 + 5x^3 + \frac{27}{2}x^2 + 10x$,
- when $-\frac{1}{2} \leq x \leq 5$, $f(x) = -\frac{71}{16}$,
- and when $x \geq 5$, $f(x) = x^2 - 10x + \frac{329}{16}$.

See below for a plot:



The local optima for f lie at $x = -5$ and $x = -2$, with a plateau in the region $-1/2 \leq x \leq 5$.

Q1.1
10 Points

Paclady decides to compute derivatives to measure how steep slopes are.
Evaluate $f'(-6)$.

-44

Q1.2
10 Points

Evaluate $f'(0)$.

0

Q1.3
10 Points

Evaluate $f'(8)$.

6

Q1.4
6 Points

Pacman and Paclady get to the resort, and have a fantastic time skiing,
but get lost. Unfortunately, a blizzard kicks in right then, reducing
visibility. As Pacman panics and brings up the Aspen Red Ghost Chase of
2012, Paclady remembers that their glass igloo cabin is located at the
global minimum elevation point of the resort ($x = -5$). The blizzard complicates
things, since they can't ski due to the reduced visibility for safety.
After thinking for a minute, Pacman says, "Aha! We can get home in
that case by following gradient descent, as long as we employ a small
step size -- once we hit a gradient of 0, we know we're home!"
Paclady pauses and says, "Your algorithm almost works, but it depends on
where in the resort we currently are."

Check all regions where Pacman and Paclady can be, and still find their
igloo, assuming that they employ gradient descent with a small step
size and stop walking when they encounter a gradient of 0.

- [x] $x < -5$
- [x] $-5 < x < -2$
- [ ] $-2 < x < -1/2$
- [ ] $-1/2 < x < 3$
- [ ] $3 < x < 5$
- [ ] $x > 5$

Note: Make sure you select all of the correct options--there may be more than one!

Q1.5
7 Points

While slowly trudging to their igloo via gradient descent, Pacman and Paclady get into an argument. Pacman complains that trudging down a hill is tiresome, and that they instead should have gotten an igloo closer to $x = 3$. Paclady says that Pacman's previous gradient descent algorithm wouldn't lead them to the igloo in this case, unless they were already at the igloo. Why is this the case?

Gradient descent would cause Pacman and Paclady to reach $x = -2$ rather than $x = 3$, since it is at a local maximum.

Gradient descent terminates when it reaches a gradient of 0, and neighboring regions around $x = 3$ all have a gradient of 0, so Pacman and Paclady would stop searching outside of $x = 3$, within the plateau.

When gradient descent is stuck in a plateau, it searches for regions with negative rather than zero gradient.

When gradient descent is stuck in a plateau, it searches for regions with positive rather than zero gradient.

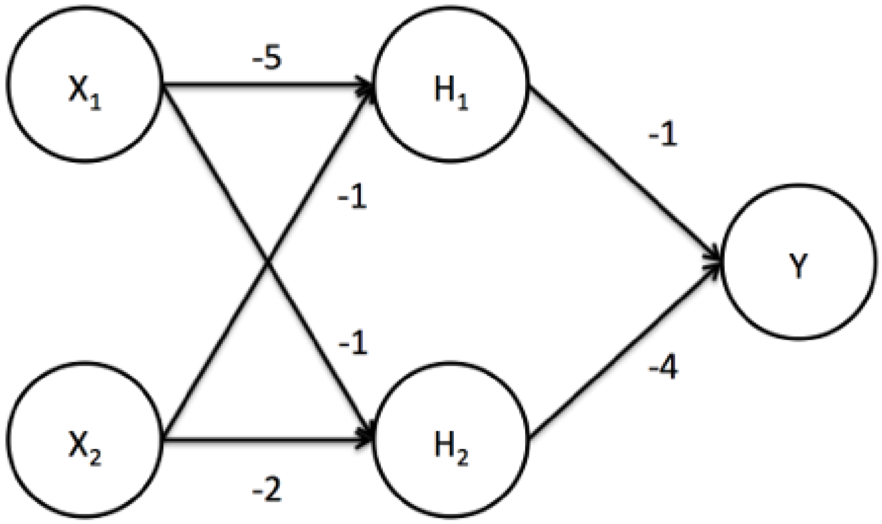
Gradient descent seeks to maximize a function, which would lead Pacman and Paclady either to $-\infty$ or to ∞.

Q2 Neural Networks and Logic Gates

36 Points

As you probably know, Pacumus Maximus Corporation (PMC) is the most well known company that manufactures the gears that let Pacman open and close its mouth. Recently, the Vice President of Science in PMC decided to replace all of its low-level NAND, AND, and NOR gates with mini neural networks. Unfortunately, there was a Pacman uprising, where the Pacmen took over the factory, eating all of the neural network documentation. The scientists were able to salvage the following neural networks weights, but don't remember which gates these neural networks corresponded to. They've hired you to help them recover this information.

Here is the first network that you're given:



Above, the nodes H_1, H_2, and Y have sigmoid activations and each have biases of 0.5. The inputs are placed in X_1 and X_2. To convert the output Y into a boolean value, we round Y. This will be the case for all problems in this section.

Concretely, we have the following, where w_{AB} denotes the weight between nodes A and B:

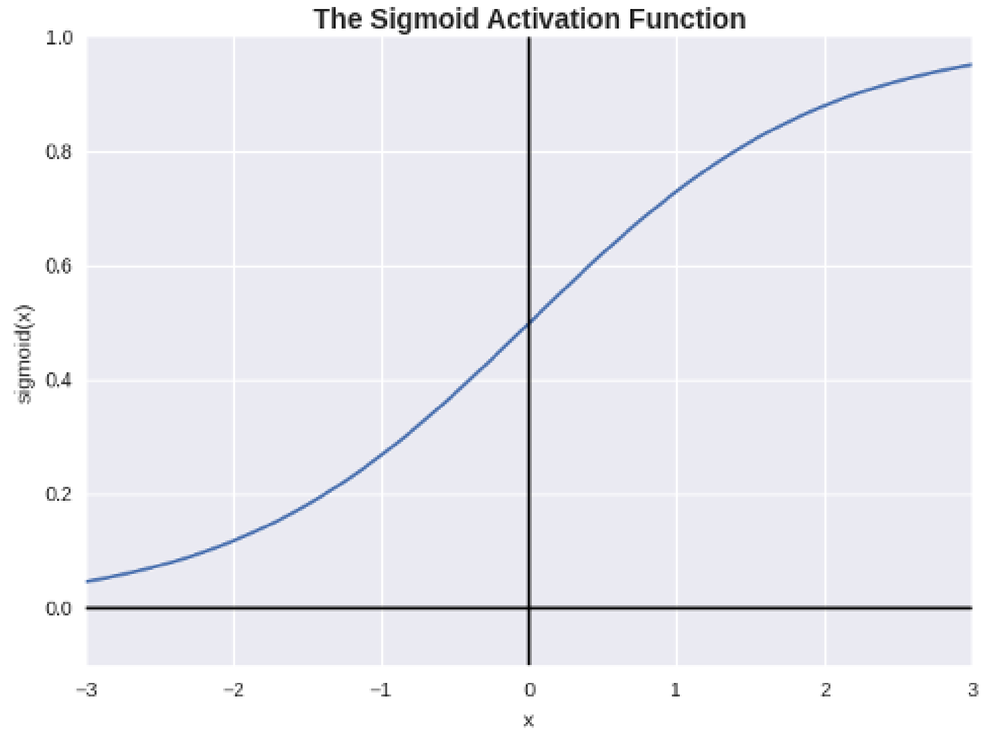
$$H_1(x) = \sigma(w_{H_1X_1} \cdot X_1 + w_{H_1X_2} \cdot X_2 + 0.5)$$

$$H_2(x) = \sigma(w_{H_2X_1} \cdot X_1 + w_{H_2X_2} \cdot X_2 + 0.5)$$

$$Y(x) = \text{round}\{\sigma(w_{YH_1} \cdot H_1 + w_{YH_2} \cdot H_2 + 0.5)\}$$

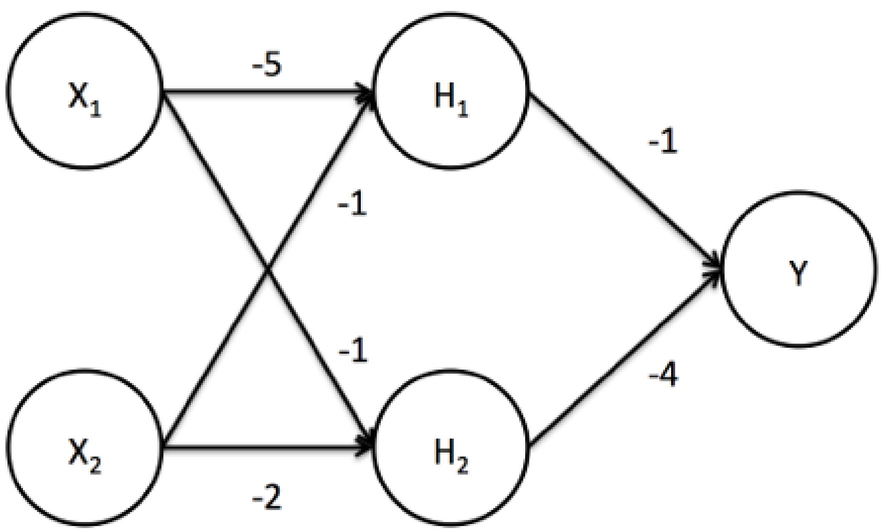
$$\sigma(x) = \frac{1}{1+e^{-x}}$$

Recall that the sigmoid function, $\sigma(x)$, looks like this:



Q2.1
9 Points

Here's the first network again, for convenience:



What does this first network correspond to?

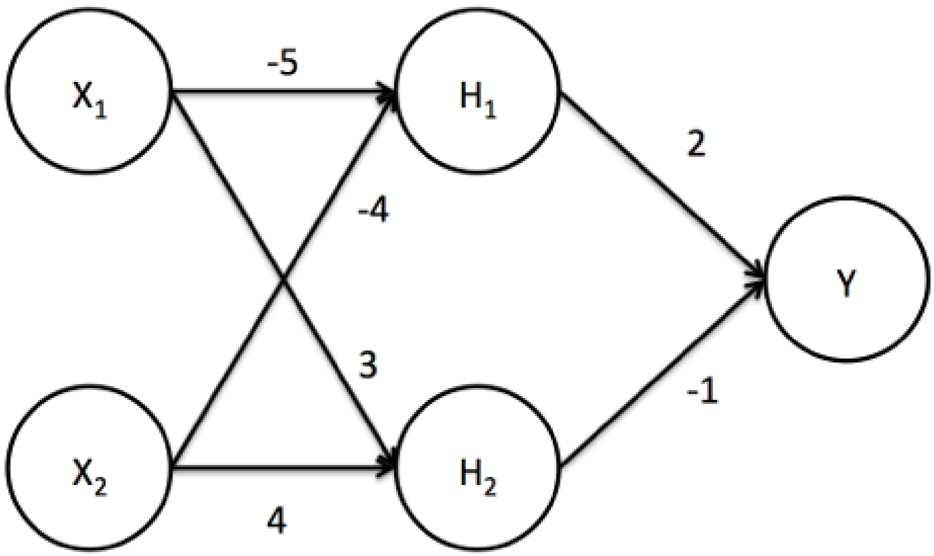
NAND

AND

NOR

Q2.2
9 Points

Here is the second network that you're given:



What does this network correspond to?

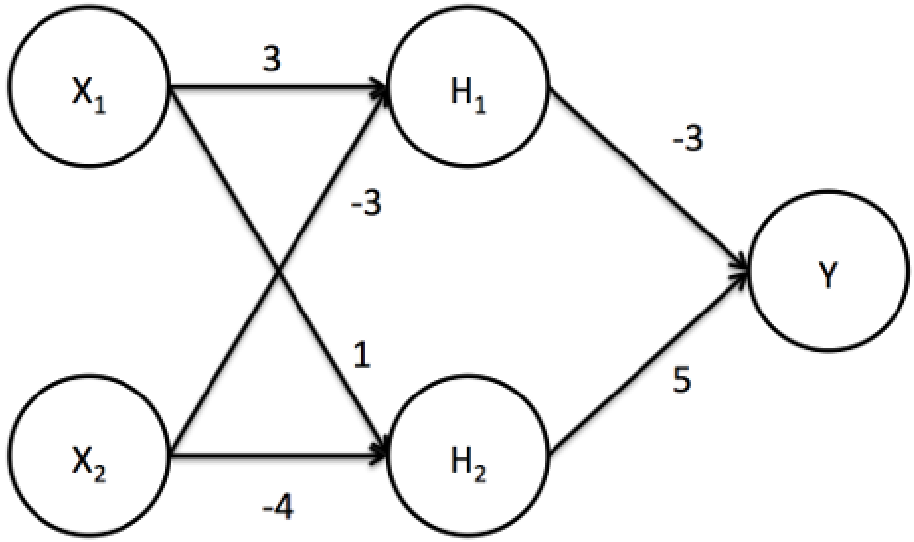
NAND

AND

NOR

Q2.3
9 Points

Here is the third network that you're given:



What does this network correspond to?

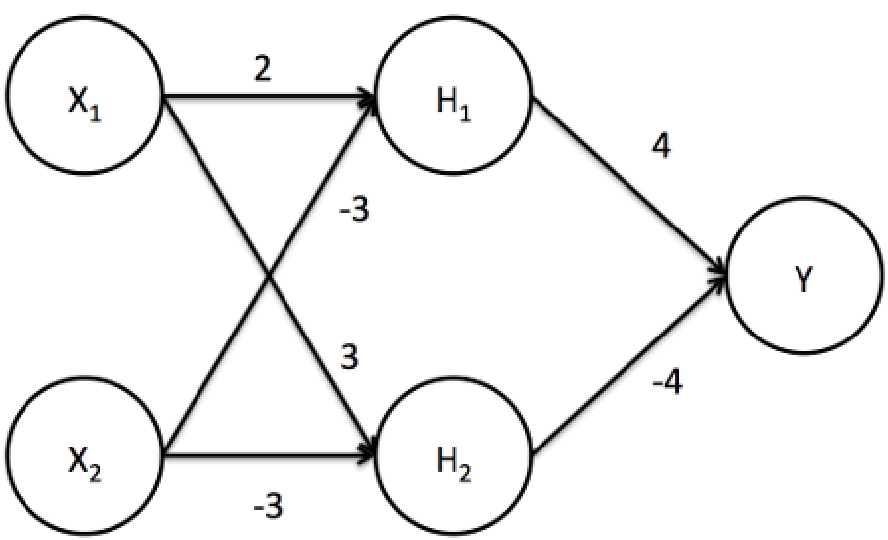
NAND

AND

NOR

Q2.4
9 Points

Here is the final network that you're given:



What does this network correspond to?

NAND

AND

NOR

Q3 Learning Rates

21 Points

[Video Link](http://www.youtube.com/watch?v=FaDgovU4_0o)

http://www.youtube.com/watch?v=FaDgovU4_0o

Watch the above YouTube video. There are three sub-panels. We will refer to the left one by (A), the middle one by (B), and the right one by (C). The same objective is being optimized by gradient descent, but has different learning rates in each subpanel.

Q3.1

7 Points

Which animation corresponds to the lowest learning rate?

Animation (A)

Animation (B)

Animation (C)

Q3.2
7 Points

Which animation corresponds to the medium learning rate?

- Animation (A)
- Animation (B)
- Animation (C)

Q3.3
7 Points

Which animation corresponds to the largest learning rate?

- Animation (A)
- Animation (B)
- Animation (C)

HW 11 (Electronic Component)

Graded

Student

ریحانه شاهرخیان

Total Points

100 / 100 pts

Question 1

Local Optima and Gradient Descent		**43** / 43 pts
1.1	(no title)	**10** / 10 pts
1.2	(no title)	**10** / 10 pts

1.3	(no title)	**10** / 10 pts
1.4	(no title)	**6** / 6 pts
1.5	(no title)	**7** / 7 pts

Question 2

Neural Networks and Logic Gates **36** / 36 pts

2.1	(no title)	**9** / 9 pts
2.2	(no title)	**9** / 9 pts
2.3	(no title)	**9** / 9 pts
2.4	(no title)	**9** / 9 pts

Question 3

Learning Rates **21** / 21 pts

3.1	(no title)	**7** / 7 pts
3.2	(no title)	**7** / 7 pts
3.3	(no title)	**7** / 7 pts