CSE 4308 001 – Fall 2018
Exam 3, Wednesday 12/12/2018

Name: SOLUTION

Student ID:

(Not providing this information: -10 Points)
(Name missing in Individual pages: -5 Points)

Name: _____

Total exam points: 100.

Score table:

Question	Points	Out of
1		10
2		15
3		15
4		15
5		15
6		10
7		5
8		10
9		5
Total		100

Name: _____

Question 1 – 10 Points

Color	Price $10 to $39	Price $40 to $69	Price $70 to $100
red	0.15	0.25	0.05
green	0.05	0.1	0.1
blue	0.1	0.05	0.15

a.

The above joint distribution table shows the probability of prices and colors for hats. For example, the probability that a hat is red and costs $40 to $69 is 0.25. Using that table calculate the following values (you do not need to calculate exact numerical values. Just give the variable free forumlas)

Part a (5 points): Determine the probability that a hat is not red and costs less than $70.

$$P(\neg red \cap < \$70) = 0.05 + 0.1 + 0.1 + 0.05$$

$$= 0.3$$

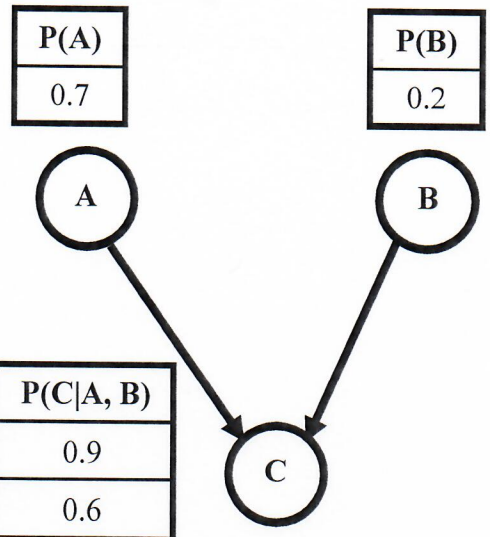
Part b (5 points): Determine the probability that a hat is red given that it costs less than $70

$$P(red / \$70) = \frac{P(red \cap \$70)}{P(\$70)}$$

$$= \frac{0.15 + 0.25}{0.15 + 0.25 + 0.05 + 0.1 + 0.1 + 0.05}$$

$$= \frac{0.4}{0.7} = \frac{4}{7}$$

Question 2 – 15 points



A	B	P(C|A, B)
T	T	0.9
T	F	0.6
F	T	0.5
F	F	0.2

In your answers for this question, you do not have to carry out numerical calculations, but you do have to provide completely specified numerical expressions, that only include numbers and mathematical operations on numbers. **Your final answers should have no variables in them**.

Part a (5 points):
For this Bayesian network, compute P((A=true) AND (B=false) AND (C=false)).

$P(A \wedge \neg B \wedge \neg C) =$

$= P(\neg C / A, \neg B)\, P(A)\, P(\neg B)$

$= (1 - 0.6) \cdot 0.7 \; (1 - 0.2)$

$= 0.4 \times 0.7 \times 0.8$

Name: _____

Part b (5 points): For this Bayesian network, compute P((B=true) | (A=false)) **using enumeration**

$$P(B/\neg A) = \frac{P(B \cap \neg A)}{P(\neg A)}$$

$$= \frac{P(B \cap \neg A \cap C) + P(B \cap \neg A \cap \neg C)}{P(\neg A)}$$

$$= \frac{P(B) \cdot P(\neg A) \cdot P(C/B, \neg A) + P(B) \cdot P(\neg A) \cdot P(\neg C/B, \neg A)}{P(\neg A)}$$

$$= \frac{P(B) \cdot P(\neg A) \left[P(C/B, \neg A) + P(\neg C/B, \neg A) \right]}{P(\neg A)}$$

$$= P(B) \cdot P(\neg A) / P(\neg A) = P(B) = 0.7.$$

Part c (5 points): For this Bayesian network, is B independent of A? Justify your answer.

As shown above $P(B/\neg A) = P(B)$

Similarly we can show $P(B/A) = P(B)$

So B is independent of A

ALT Ans.

From the N/w we can see that B, A have no parents in common, so they are independent of each other.

Question 3 – 15 points

Consider the Bayesian networks on the left and on the right:

P(A)
0.4

A	P(B|A)
T	0.2
F	0.7

P(B)
X1

B	P(A|B)
T	X2
F	X3

Are there any values of X1, X2, X3, that make the network on the right express the exact same joint probability distribution table values [Ex: P(A=t, B=f) etc.] as the network on the left? If yes, provide variable-free formulas for the values of X1, X2 and X3. If not, justify why not.

$P(A_1, B_1) = P(A_1)\, P(B_1/A_1)$

$P(A_2, B_2) = P(B_2) \cdot P(A_2/B_2)$

So $0.4 \times 0.2 = X_1 \cdot X_2$.

Similarly

$\cancel{(1-0.4)\times(1-0.7)} =$

$P(A_1, \neg B_1) = P(A_2, \neg B_2)$

So $0.4 \times (1-0.2) = (1-X_1) \cdot X_3$.

$P(\neg A_1, B_1) = P(\neg A_2, B_2)$

$(1-0.4) \cdot 0.7 = X_1 \cdot (1-X_2)$.

So

$X_1 \cdot X_2 = 0.08$

$X_3 - X_1 \cdot X_3 = 0.32$

$X_1 - X_1 \cdot X_2 = 0.42$

Soln.

$$X_1 = 0.42 + 0.08$$

$$= 0.5$$

$$X_2 = \frac{0.08}{0.5} = 0.16$$

$$X_3 = \frac{0.32}{1-0.5} = \frac{0.32}{0.5} = 0.64.$$

Name: _____

Question 4 – 15 points

We want to build a decision tree that determines whether a certain pattern is of type X or type Y. The decision tree can use only tests that are based on attributes A1, A2, A3 and A4. We have the following training examples (each row corresponds to a training example).

Class	A1	A2	A3	A4
X	T	F	F	T
X	T	F	F	T
X	F	F	F	T
X	F	F	F	T
X	T	F	F	F
X	F	F	F	T
Y	F	T	F	F
Y	F	T	F	F
Y	T	T	F	F
Y	F	T	F	T
Y	T	T	F	F
Y	T	T	F	F

Part a. (10 points) What is the information gain if we use A1, A2, A3 or A4 as the decision test for the root node? **You do not have to compute logarithms, but you should write the correct formula and replace variables with numbers wherever you can**

ENTROPY BEFORE SPLIT

$$H = -\frac{6}{12}\log_2\frac{6}{12} - \frac{6}{12}\log_2\frac{6}{12} = 1$$

Using A1

$A_1 = T$: $X: 3$ $Y = 3$

$A = F$: $X = 3$ $Y = 3$

$$H_{A1=T} = -\frac{3}{6}\log_2\frac{3}{6} - \frac{3}{6}\log_2\frac{3}{6} = 1$$

$$H_{A_1=F} = -\frac{3}{6}\log_2\frac{3}{6} - \frac{3}{6}\log_2\frac{3}{6} = 1$$

$$I_{A_1} = H - \frac{6}{12}H_{A_1=T} - \frac{6}{12}H_{A_1=F} = 1 - 0.5 - 0.5$$

$$= 0$$

Using A_2

$A_2 = T \quad x: 0 \quad y = 6$

$A_2 = F \quad x: 6 \quad y = 0$

$$H_{A_2=T} = -\frac{0}{6}\log_2\frac{0}{6} - \frac{6}{6}\log_2\frac{6}{6} = 0.$$

$$H_{A_2=F} = -\frac{6}{6}\log_2\frac{6}{6} - \frac{0}{6}\log_2\frac{0}{6} = 0.$$

$$I_{A_2} = H - \frac{6}{12}H_{A_2=T} - \frac{6}{12}H_{A_2=F}$$

$$= 1 - 0 - 0 = 1$$

Using A_3

$A_3 = T \quad x: 0 \quad y = 0$

$A_3 = F \quad x: 6 \quad y = 6.$

$$H_{A_3=T} = -\frac{0}{0}\log_2\frac{0}{0} - \frac{0}{0}\log_2\frac{0}{0} = 0$$

$$H_{A_3=F} = -\frac{6}{12}\log_2\frac{6}{12} - \frac{6}{12}\log_2\frac{6}{12} = 1.$$

$$I_{A_3} = H - \frac{0}{12}\times H_{A_3=T} - \frac{12}{12}\times H_{A_3=F} = 1 - 0 - 1$$

$$= 0$$

Using A_4

$A_4 = T \quad x = 5 \quad y = 1$

$A_4 = F \quad x = 1 \quad y = 5$

$$H_{A_4=T} = -\frac{5}{6}\log_2\frac{5}{6} - \frac{1}{6}\log_2\frac{1}{6}$$

$$H_{A_4=F} = -\frac{1}{6}\log_2\frac{1}{6} - \frac{5}{6}\log_2\frac{5}{6}$$

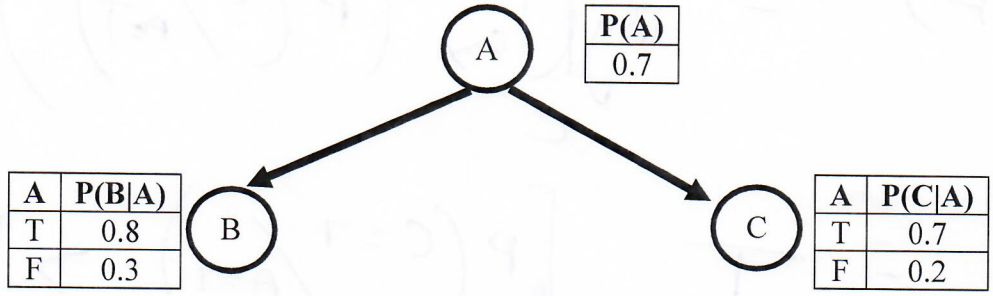
$$I_{A_4} = H - \frac{6}{12}H_{A_4=T} - \frac{6}{12}H_{A_4=F}$$

Part b. (5 points) Among attributes A1, A2, A3 and A4, which one should be used as a test for the root node according to the information (entropy) gain criterion? You should be able to answer this question without explicitly computing the entropy gain of each attribute. Justify your answer.

A_2 will have highest Information gain as it splits the set into completely lopsided subsets. So it will have highest Information Gain and so will be used.

Name: _____

Question 5 – 15 points



P(A)
0.7

A	P(B|A)
T	0.8
F	0.3

A	P(C|A)
T	0.7
F	0.2

The Bayesian network shown above illustrates the relationships between A, B, and C. We want to define a **Bayesian** classifier predicting the value of C based on the values of A and B. A classifier for this problem is a function F(x, y), with two boolean inputs: x specifies whether A is true or false, and y specifies whether B is true or false. The output of F is a Boolean, estimating whether C is true or false and is given by:

$$F(x,y) = \underset{c}{\operatorname{argmax}}\, P\left({}^{C}/_{A=x,B=y}\right)$$

Where x and y can be true or false. Define the **Bayesian** classifier for this problem, by specifying the output of the classifier for all possible inputs.

$$P\left({}^{C}/_{A,B}\right) = \frac{P(C\ A\ B)}{P(A,B)}$$

$$= \frac{P(C, A, B)}{P(A, B, C) + P(A, B, \neg C)}$$

$$= \frac{P(C/A)\,P(B/A)\,P(A)}{P(C/A)\,P(B/A)\,P(A) + P(\neg C/A)\,P(B/A)\,P(A)}$$

$$= \frac{P(C/A)\,P(B/A)\,P(A)}{P(B/A)\,P(A)\left[P(C/A) + P(\neg C/A)\right]} = P(C/A)$$

$$F(T, F) = \underset{C}{\operatorname{Arg\,Max}} \left(P\left(C / A=T\right) \right)$$

$$= T \quad \left[P\left(C=T / A=T\right) > P\left(C=F / A=T\right) \right]$$

Similarly

$$F(T, T) = T$$

$$F(F, T) = F$$

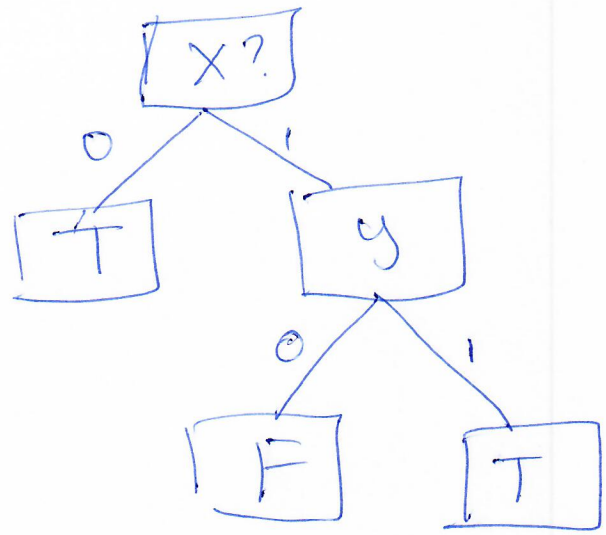
$$F(F, F) = F \ldots$$

Name: _____

Question 6 - 10 points

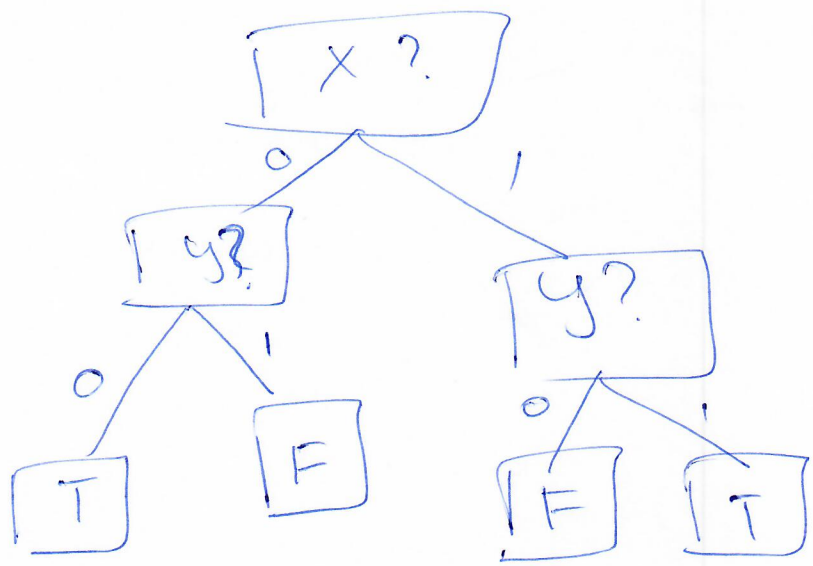
Part a. (5 points) Consider the function: **X ⇒ Y**. That function would take two Boolean inputs X and Y which can have values 0 or 1 and would output 0 or 1 based whether it satisfies the equation. Can you represent this function as a Decision Tree? If so give me a Tree that implements this function.

X	y	X ⇒ y
0	0	1
1	0	0
0	1	1
1	1	1



Part b. (5 points) Consider the function: **X ⇔ Y**. That function would take two Boolean inputs X and Y which can have values 0 or 1 and would output 0 or 1 based whether it satisfies the equation. Can you represent this function as a Decision Tree? If so give me a Tree that implements this function.

X	y	X ⇔ y
0	0	1
0	1	0
1	0	0
1	1	1



Question 7 - 5 points

We have a classification problem with N classes. Let us say we know that a true Bayes classifier for this problem has an error rate of 0.15.

Let us say build a decision tree using a training set of size 1000 to handle this problem, what is the lowest error rate you could possibly achieve. Can you improve this error rate by using more training samples? Justify your answer.

True Bayes Classifier is the best classifier possible for the problem so the lowest error rate the decision tree can reach is.

0.15

It is not possible to build a better tree by adding more examples

Name: _____

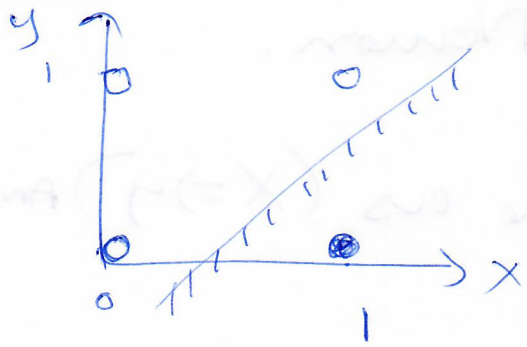
Question 8 – 10 points

Can you represent the functions in Question 6 as a single neuron? If so, give the weights of the neuron. If not, can you represent it as a network or neurons. If so, give the network.

For all perceptrons:

- Assume that the activation function outputs **0** if the weighted sum of inputs is LESS THAN 0, and outputs **1** otherwise.
- Assume that the bias input is always set to +1.

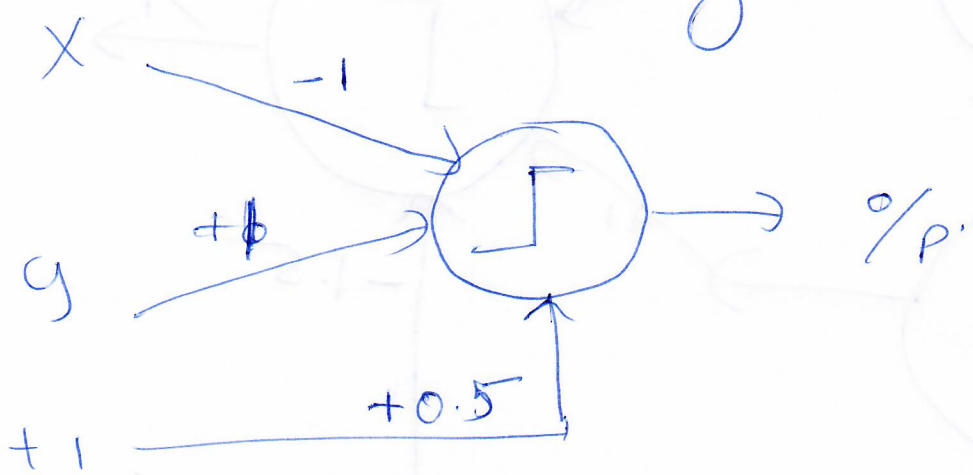
X => Y



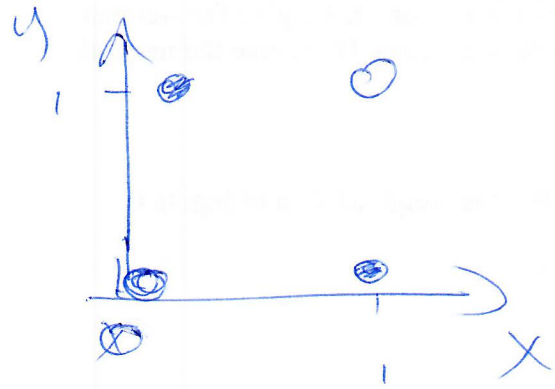
• : Neuron gives 0

o : Neuron gives 1.

Outputs are linearly separable, so we can use sigh neuron.



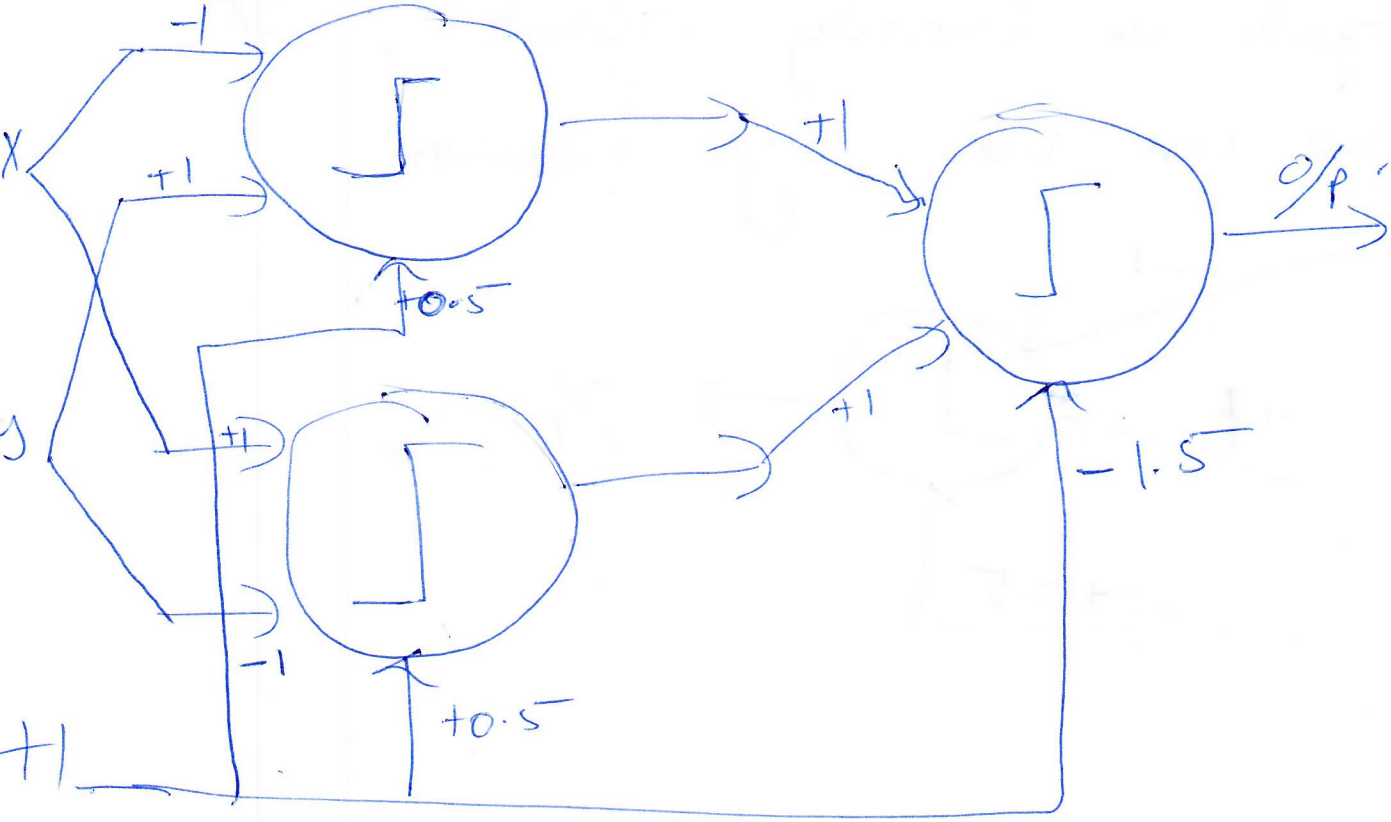
$X \Leftrightarrow y$



This is not linearly separable

Cannot use single Neuron.

$X \Leftrightarrow y$ is the same as $(X \Rightarrow y)$ AND $(y \Rightarrow X)$



Question 9 – 5 points

We want to check if for a given X and Y value, whether **6X – 5Y – 14 ≤ 10**

Can we do this using a single Perceptron or does it need a network of multiple perceptrons? They should take X and Y as input and return 1 if the equation holds true and 0 otherwise.

- If you can use a single perceptron, define its parameters.
- If you cannot design a single perceptron, design a network of multiple perceptrons that will achieve the same results.

For all perceptrons:

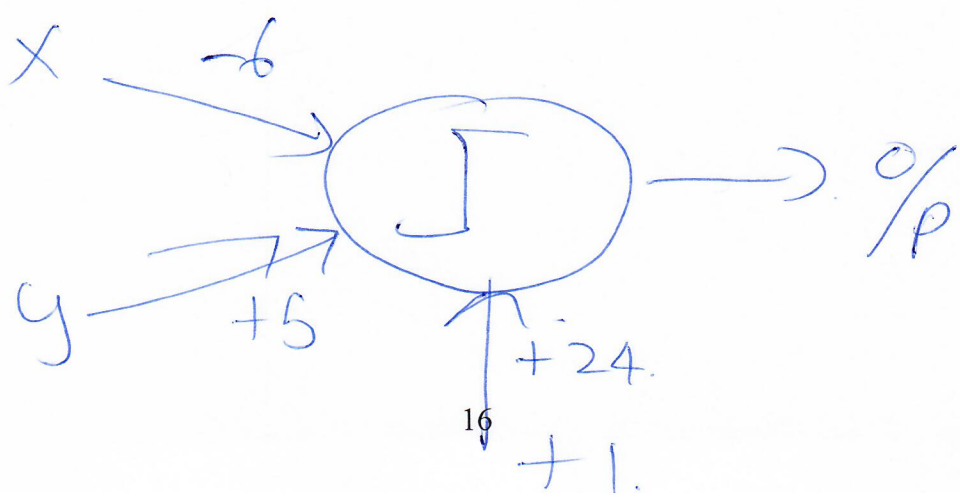
- Assume that the activation function outputs **0** if the weighted sum of inputs is LESS THAN 0, and outputs **1** otherwise.
- Assume that the bias input is always set to +1.

This is linearly separable so we can use a single ~~line~~ neuron.

$$6x - 5y - 14 \leq 10$$

$$-6x + 5y + 24 \geq 0$$

So Neuron is.



Name: _____

SCRATCH