CS 540: Introduction to Artificial Intelligence Summer 2020

Midterm Review

Instructor: Young Wu TA: Dandi Chen

1. **(M2Q4) (Spring 2017 Final Q3, Spring 2018 Final Q7, video for M2Q5)** Consider a Linear Threshold Unit (LTU) perceptron with initial weights $w = [-0.8, -0.5, -0.8]^T$ and bias $b = 0.5$. Given a new input $x = [0, 0, 1]^T$ and $y = 1$, learning rate $\alpha = 1$, compute the updated weights and bias $[w_1, w_2, w_3, b]$.

   Solution: According to Perceptron Algorithm,

$$w = w - \alpha(a_i - y_i)x_i$$
$$b = b - \alpha(a_i - y_i)$$

Since

$$\begin{aligned} a_i &= \mathbb{1}_{\{w^T x + b \geq 0\}} \\ &= \mathbb{1}_{\{-0.8\times 0 - 0.5 \times 0 - 0.8 \times 1 + 0.5 \geq 0\}} \\ &= \mathbb{1}_{\{-0.3 \geq 0\}} \\ &= 0 \end{aligned}$$

w1x1 + w2x2 + w3x3 + b

Then the updated weights and bias are

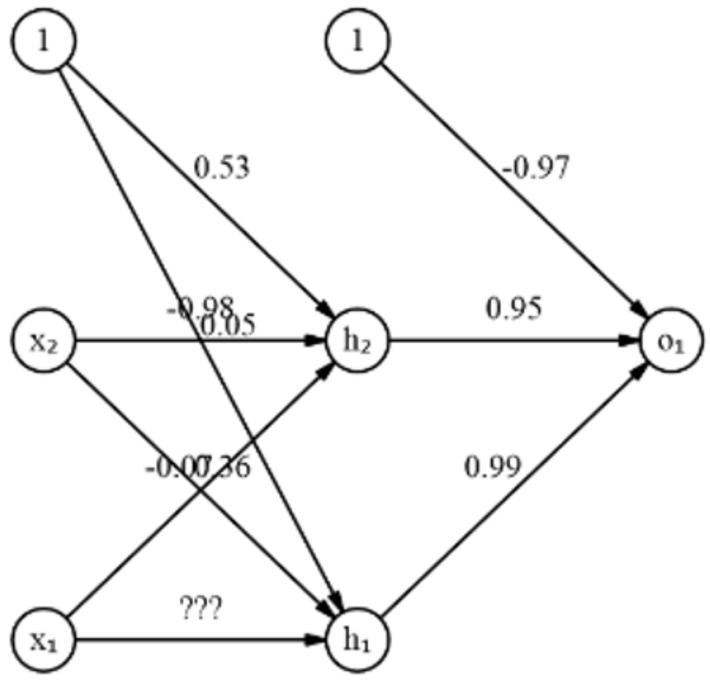
$$\begin{aligned} w &= w - \alpha(a_i - y_i)x_i \\ &= \begin{bmatrix} -0.8 \\ -0.5 \\ -0.8 \end{bmatrix} - 1 \cdot (0 - 1) \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -0.8 \\ -0.5 \\ 0.2 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} b &= b - \alpha(a_i - y_i) \\ &= 0.5 - 1 \cdot (0 - 1) \\ &= 1.5 \end{aligned}$$

2. (M3Q7, video for M3Q6) (Fall 2010 Final Q17)

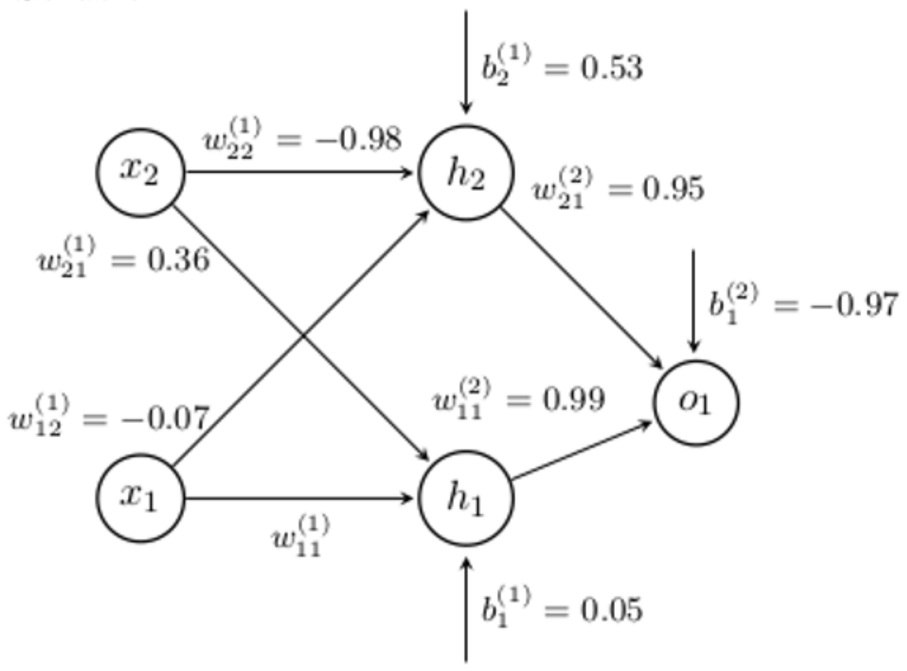
Fill in the missing weight below so that it computes the following function. All inputs takes value 0 or 1, and the perceptrons are linear threshold units.

x_1	x_2	y or o_1
0	0	1
0	1	1
1	0	0
1	1	0



Hint: if the weights are not shown clearly, you could move the nodes around with mouse or touch.

Solution:



h1, h2, o1 can be either 0 or 1

$$h_1 = \mathbb{1}_{\{w_{11}^{(1)} x_1 + w_{21}^{(1)} x_2 + b_1^{(1)} \geq 0\}}$$
$$h_2 = \mathbb{1}_{\{w_{12}^{(1)} x_1 + w_{22}^{(1)} x_2 + b_2^{(1)} \geq 0\}}$$
$$o_1 = \mathbb{1}_{\{w_{11}^{(2)} h_1 + w_{21}^{(2)} h_2 + b_1^{(2)} \geq 0\}}$$

x_1	x_2	$w_{11}^{(1)} x_1 + w_{21}^{(1)} x_2 + b_1^{(1)}$	h_1	$w_{12}^{(1)} x_1 + w_{22}^{(1)} x_2 + b_2^{(1)}$	h_2	$w_{11}^{(2)} h_1 + w_{21}^{(2)} h_2 + b_1^{(2)}$	o_1
0	0	$w_{11}^{(1)} \cdot 0 + 0.36 \cdot 0 + 0.05$	1	$-0.07 \cdot 0 - 0.98 \cdot 0 + 0.53$	1	$0.99 \cdot 1 + 0.95 \cdot 1 - 0.97$	1
0	1	$w_{11}^{(1)} \cdot 0 + 0.36 \cdot 1 + 0.05$	1	$-0.07 \cdot 0 - 0.98 \cdot 1 + 0.53$	0	$0.99 \cdot 1 + 0.95 \cdot 0 - 0.97$	1
1	0	$w_{11}^{(1)} \cdot 1 + 0.36 \cdot 0 + 0.05$	$\mathbb{1}_{\{w_{11}^{(1)} + 0.05 \geq 0\}} = 0$	$-0.07 \cdot 1 - 0.98 \cdot 0 + 0.53$	1	$0.99 \cdot 0 + 0.95 \cdot 1 - 0.97$	0
1	1	$w_{11}^{(1)} \cdot 1 + 0.36 \cdot 1 + 0.05$	$\mathbb{1}_{\{w_{11}^{(1)} + 0.41 \geq 0\}} = 0$	$-0.07 \cdot 1 - 0.98 \cdot 1 + 0.53$	0	$0.99 \cdot 0 + 0.95 \cdot 0 - 0.97$	0

Therefore,

If h1 = 1, o1 cannot be 0 when x1 = 1.

Therefore, h1 has to be 0.

$$w_{11}^{(1)} + 0.05 < 0$$
$$w_{11}^{(1)} + 0.41 < 0$$

i.e.

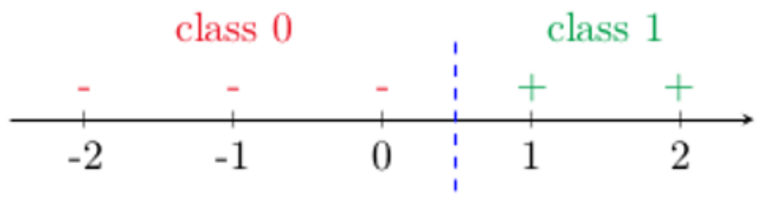
$$w_{11}^{(1)} < -0.41$$

3. **(Summer 2019 Midterm A Q28)** Given the following training set S, suppose n instances are removed, and 1 nearest neighbor with Manhattan distance is trained on the remaining set S', and tested on the original training set S. If an instance is equally distant from two points with different labels, the 1 nearest neighbor labels the instance as 0 with probability $\frac{1}{2}$ and 1 with probability $\frac{1}{2}$. If the accuracy is 100 percent (for certain) on S, what is the maximum possible value for n?

tie-breaking rule

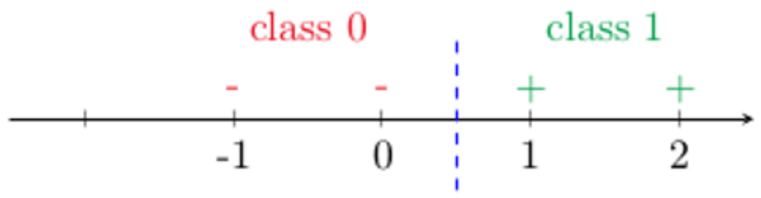
$$S = (x_i, y_i)_{i=1,2,...,5} = \{(-2,0), (-1,0), (0,0), (1,1), (2,1)\}$$

Solution:

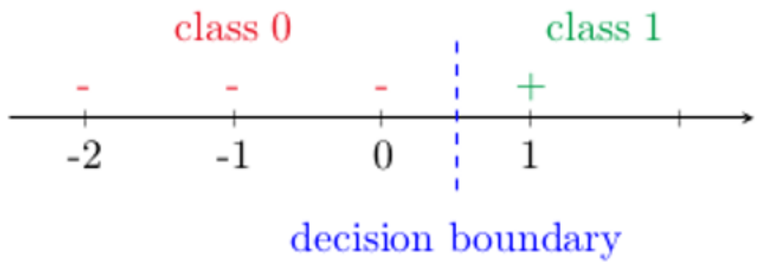


This is a linearly separable dataset. It trains on $5 - n$ instances (i.e. set S'), and tests on 5 instances (i.e. set S). Removing $(-2,0)$ or removing $(-1,0)$ or removing $(2,1)$ or removing $(-2,0), (-1,0), (2,1)$ all maintain the same decision boundary.

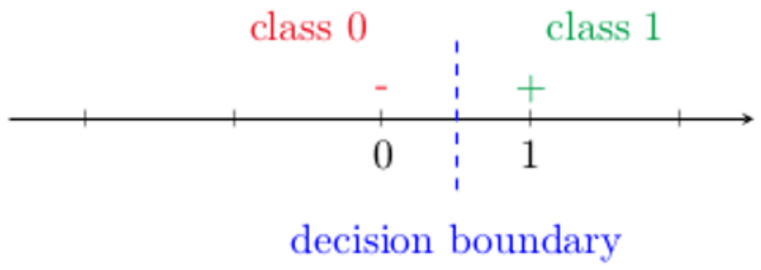
- If $(-2,0)$ is removed: training set (set S') is $\{(-1,0), (0,0), (1,1), (2,1)\}$; testing set (set S): $\{(-2,0), (-1,0), (0,0), (1,1), (2,1)\}$. Then $x = -2$ will be labeled to be 0 as its nearest neighbor is $(-1,0)$. Therefore, its predicted label and its actual label are both 0, i.e. 100% testing accuracy is guaranteed.



- If $(2,1)$ is removed:



- If $(-2,0), (-1,0), (2,1)$ are removed:

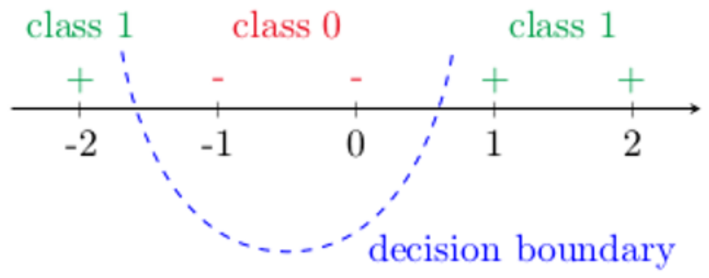


Therefore, the maximum possible value for n is 3, i.e. removing $(-2,0), (-1,0), (2,1)$.

4. **(Summer 2019 Midterm A Q29)** Continue from the previous question. Same assumptions as the previous question: what is the maximum possible value for n if the training set S is changed to the following?

$$S = (x_i, y_i)_{i=1,2,...,5} = \{(-2,1), (-1,0), (0,0), (1,1), (2,1)\}$$

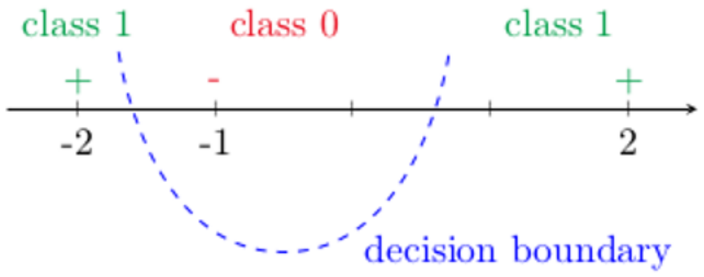
Solution:



This dataset is not linearly separable. It still trains on $5-n$ instances (i.e. set S'), and tests on 5 instances (i.e. set S). Removing $(2,1)$ or removing $(0,0)$ and $(1,1)$ can maintain the same decision boundary.

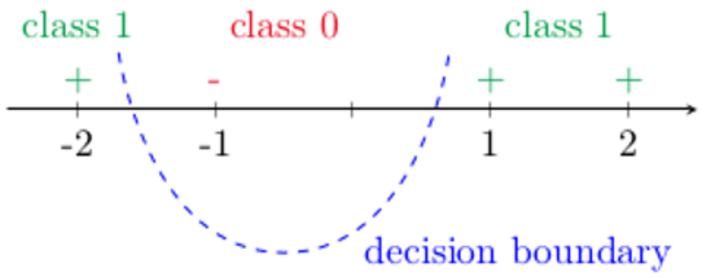
In general, training set and testing set are not overlapped.

- If $(0,0)$ and $(1,1)$ are removed: training set (set S') is $\{(-2,1), (-1,0), (2,1)\}$; testing set (set S) is $\{(-2,1), (-1,0), (0,0), (1,1), (2,1)\}$. Then $x=0$ will be labeled 0 as its nearest neighbor is $(-1,0)$. $x=1$ will be labeled 1 as its nearest neighbor is $(2,1)$. Therefore, their predicted labels and actual labels are the same, i.e. 100% testing accuracy is guaranteed.



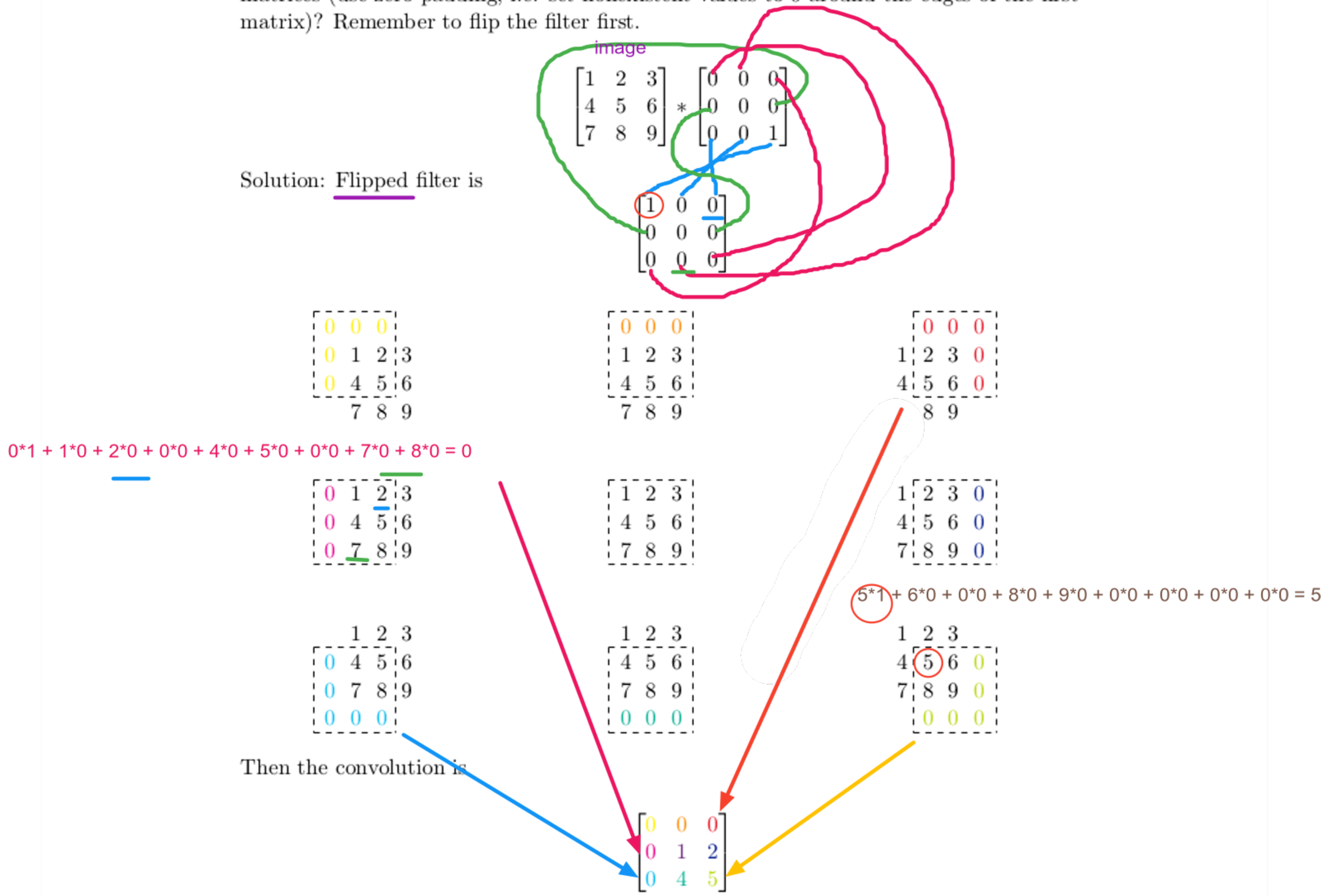
a different tie-breaking rule

- If $(0,0)$ is removed: training set (set S') is $\{(-2,1), (-1,0), (1,1), (2,1)\}$; testing set (set S) is $\{(-2,1), (-1,0), (0,0), (1,1), (2,1)\}$. Then $x=0$ will be labeled 0 as its nearest neighbor is $(-1,0)$ if the tie-breaking rule is to use the label of the lower indexed instance when having equal distance (this is the default tie-breaking rule this year while it was not for year 2019). However, if we follow the tie-breaking rule in this problem ("the 1 nearest neighbor labels the instance as 0 with probability $\frac{1}{2}$ and 1 with probability $\frac{1}{2}$"), we cannot guarantee whether its label will be either 1 or 0, since the Manhattan distance between $(0,0)$ to its nearest neighbors $(-1,0)$ and $(1,1)$ are the same while they are from two different classes.



Therefore, the maximum possible value for n is 2.

5. **(Summer 2019 Midterm A Q19)** What is the convolution between the following two matrices (use zero padding, i.e. set nonexistent values to 0 around the edges of the first matrix)? Remember to flip the filter first.



Solution: Flipped filter is

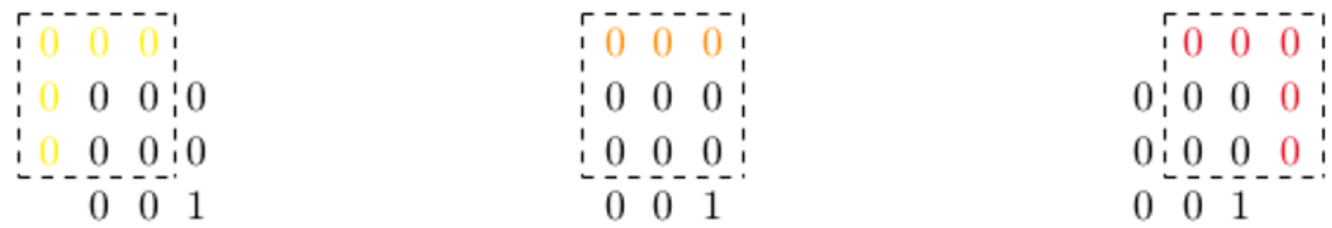
Then the convolution is

6. **(Summer 2019 Midterm A Q20)** Continue from the previous question, what is the convolution between the following two matrices? Remember to flip the filter first.

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Solution: Flipped filter is

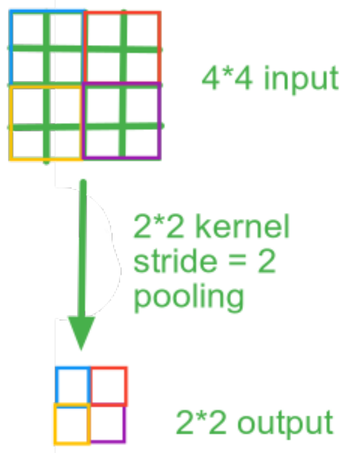
$$\begin{bmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix}$$



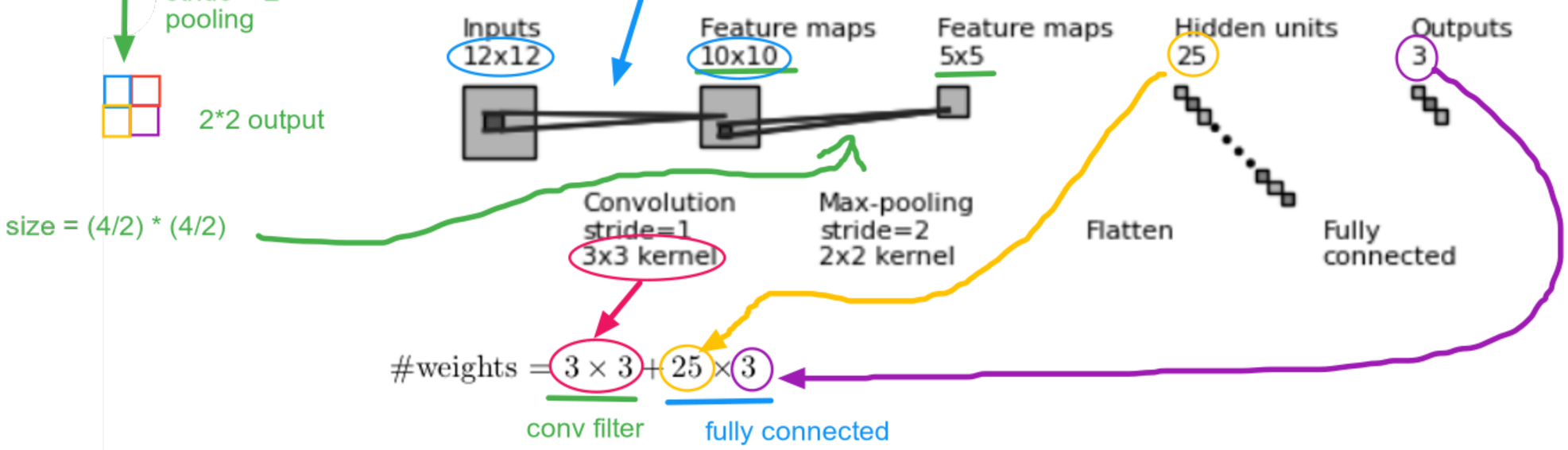
Then the convolution is

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 4 & 5 \end{bmatrix}$$

7. **(Fall 2019 Final Q15)** A convolutional neural network (CNN) has an input image of size 12×12 that is connected to a convolutional layer that uses a 3×3 filter, no padding of the image, and a stride of 1. The convolutional layer is then connected to a pooling layer that uses 2×2 max pooling, a stride of 2, and no padding of the convolutional layer. The pooling layer is then fully connected to an output layer that contains 3 output units. There are no hidden layers between the pooling layer and the output layer. How many different weights must be learned in this whole network, not including any bias weights, if any?



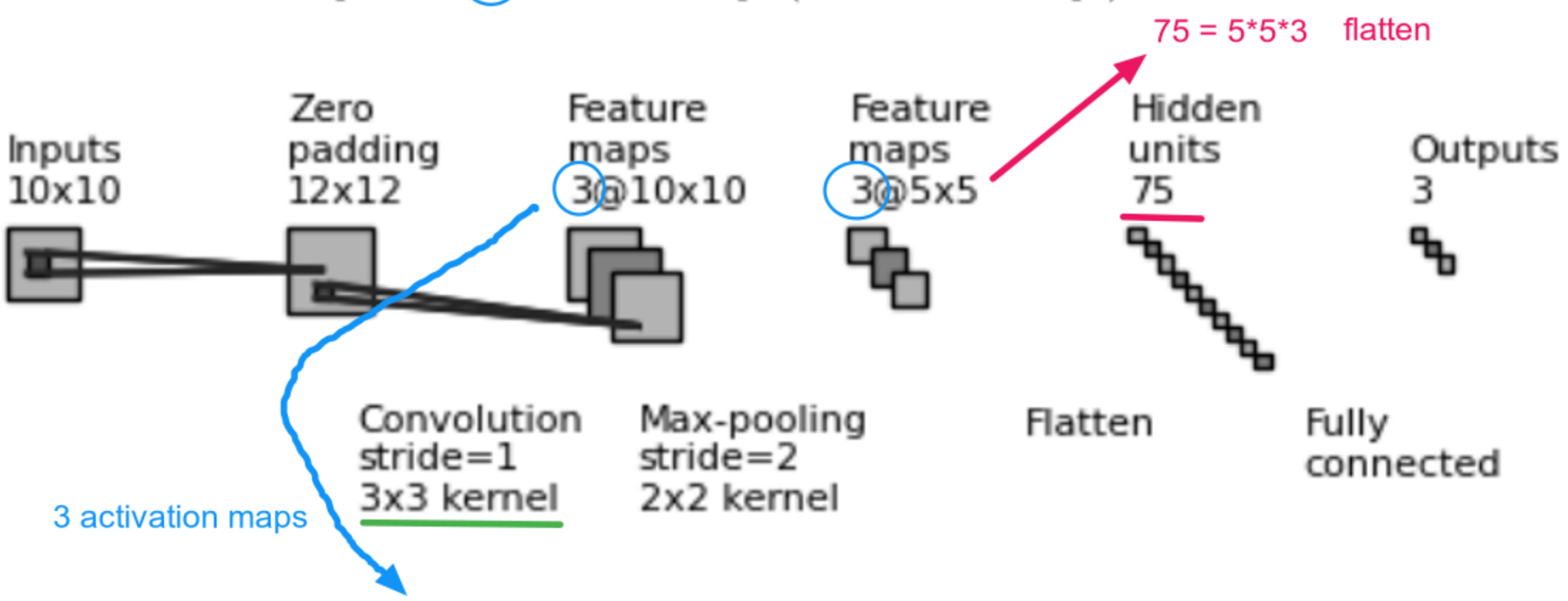
Solution:



$$\#\text{weights} = 3 \times 3 + 25 \times 3$$

conv filter — fully connected

One more example with 3 activation maps (i.e. 3 feature maps):

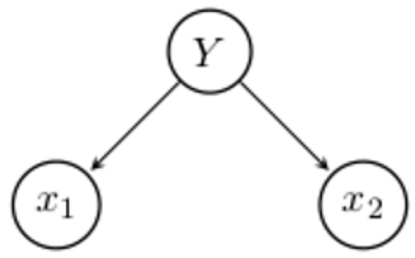


$$\#\text{weights} = 3 \times 3 \times 3 + 75 \times 3$$

convolution — fully connected

8. **(M7Q1, video) (Fall 2016 Final Q18, Fall 2011 Midterm Q20)** Consider a classification problem with $n = 28$ classes, $y \in \{1, 2, ..., n\}$, and two binary features $x_1, x_2 \in \{0, 1\}$. Suppose $\mathbb{P}(Y = y) = \frac{1}{28}$, $\mathbb{P}(x_1 = 1|Y = y) = \frac{y}{46}$, $\mathbb{P}(x_2 = 1|Y = y) = \frac{y}{52}$. Which class will naive Bayes classifier produce on a test item with $x_1 = 1$ and $x_2 = 1$? Hint: make sure you read and understand Fall 2016 Final Q18.

Solution:



P(A|B) = P(A, B)/P(B)

$$\begin{aligned}
\underset{y}{\text{argmax}}\, \mathbb{P}(Y = y|x_1 = 1, x_2 = 1) &= \underset{y}{\text{argmax}}\, \frac{\mathbb{P}(x_1 = 1, x_2 = 1, Y = y)}{\mathbb{P}(x_1 = 1, x_2 = 1)} \\
&= \underset{y}{\text{argmax}}\, \frac{\mathbb{P}(x_1 = 1, x_2 = 1|Y = y)\mathbb{P}(Y = y)}{\mathbb{P}(x_1 = 1, x_2 = 1)} \\
&= \underset{y}{\text{argmax}}\, \frac{\mathbb{P}(x_1 = 1|Y = y)\mathbb{P}(x_2 = 1|Y = y)\mathbb{P}(Y = y)}{\mathbb{P}(x_1 = 1, x_2 = 1)} \\
&= \underset{y}{\text{argmax}}\, \mathbb{P}(x_1 = 1|Y = y)\mathbb{P}(x_2 = 1|Y = y) \\
&= \underset{y}{\text{argmax}}\, \frac{y}{46} \cdot \frac{y}{52} \\
&= \underset{y}{\text{argmax}}\, y^2
\end{aligned}$$

y^2 is increasing when y is positive. Since $y \in 1, 2, ..., n$ and $n = 28$,

$$\underset{y}{\text{argmax}}\, y^2 = 28$$

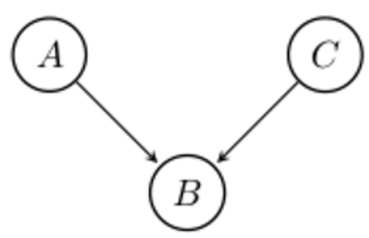
9. **(M7Q2) (Fall 2014 Final Q9)** Consider the following directed graphical model over binary variables: $A \rightarrow B \leftarrow C$. Given the CPTs:

Variable	Probability	Variable	Probability
$\mathbb{P}(A=1)$	0.68		
$\mathbb{P}(C=1)$	0.64		
$\mathbb{P}(B=1\|A=1,C=1)$	0.85	$\mathbb{P}(B=1\|A=0,C=1)$	0.95
$\mathbb{P}(B=1\|A=1,C=0)$	0.99	$\mathbb{P}(B=1\|A=0,C=0)$	0.44

What is the probability that $\mathbb{P}(A=0,B=0,C=1)$?

Solution:

$$\begin{aligned}\mathbb{P}(\neg A,\neg B,C) &= \mathbb{P}(\neg B|\neg A,C)\cdot\mathbb{P}(\neg A,C)\\ &= \mathbb{P}(\neg B|\neg A,C)\cdot\mathbb{P}(\neg A)\cdot\mathbb{P}(C)\\ &= (1-\mathbb{P}(B|\neg A,C))\cdot(1-\mathbb{P}(A))\cdot\mathbb{P}(C)\\ &= (1-0.95)\times(1-0.68)\times 0.64\end{aligned}$$



common effect

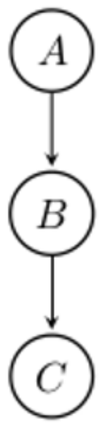
10. **(M7Q3) (Fall 2014 Final Q9)** Consider the following directed graphical model over binary variables: $A \to B \to C$. Given the CPTs:

Variable	Probability	Variable	Probability
$\mathbb{P}(A=1)$	0.17		
$\mathbb{P}(B=1\|A=1)$	0.2	$\mathbb{P}(B=1\|A=0)$	0.34
$\mathbb{P}(C=1\|B=1)$	0.01	$\mathbb{P}(C=1\|B=0)$	0.33

What is the probability that $\mathbb{P}(A=1|C=1)$?

Solution:

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



causal chain

While

$$\begin{aligned}\mathbb{P}(A,B,C) &= \mathbb{P}(C|A,B)\cdot\mathbb{P}(A,B)\\ &= \mathbb{P}(C|A,B)\cdot\mathbb{P}(B|A)\cdot\mathbb{P}(A)\\ &= \mathbb{P}(C|B)\cdot\mathbb{P}(B|A)\cdot\mathbb{P}(A)\end{aligned}$$

P(A|B) = P(A, B)/P(B)

P(A, B) = P(A|B) * P(B)

In other words,

$$\begin{aligned}\mathbb{P}(A|C) &= \frac{\mathbb{P}(A,C)}{\mathbb{P}(C)}\\ &= \frac{\mathbb{P}(A,B,C)+\mathbb{P}(A,\neg B,C)}{\mathbb{P}(A,B,C)+\mathbb{P}(A,\neg B,C)+\mathbb{P}(\neg A,B,C)+\mathbb{P}(\neg A,\neg B,C)}\\ &= \frac{\mathbb{P}(C|B)\cdot\mathbb{P}(B|A)\cdot\mathbb{P}(A)+\mathbb{P}(C|\neg B)\cdot\mathbb{P}(\neg B|A)\cdot\mathbb{P}(A)}{\mathbb{P}(C|B)\cdot\mathbb{P}(B|A)\cdot\mathbb{P}(A)+\mathbb{P}(C|\neg B)\cdot\mathbb{P}(\neg B|A)\cdot\mathbb{P}(A)+\mathbb{P}(C|B)\cdot\mathbb{P}(B|\neg A)\cdot\mathbb{P}(\neg A)+\mathbb{P}(C|\neg B)\cdot\mathbb{P}(\neg B|\neg A)\cdot\mathbb{P}(\neg A)}\\ &= \frac{0.01\times0.2\times0.17+0.33\times(1-0.2)\times0.17}{0.01\times0.2\times0.17+0.33\times(1-0.2)\times0.17+0.01\times0.34\times(1-0.17)+0.33\times(1-0.34)\times(1-0.17)}\end{aligned}$$

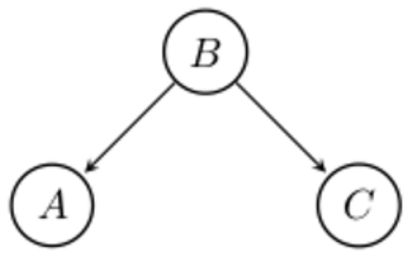
11. **(M7Q4, video) (Fall 2014 Final Q9)** Consider the following directed graphical model over binary variables: $A \leftarrow B \rightarrow C$. Given the CPTs:

Variable	Probability	Variable	Probability
$\mathbb{P}(B=1)$	0.66		
$\mathbb{P}(C=1\|B=1)$	0.12	$\mathbb{P}(C=1\|B=0)$	0.26
$\mathbb{P}(A=1\|B=1)$	0.28	$\mathbb{P}(A=1\|B=0)$	0.53

What is the probability that $\mathbb{P}(A=0|C=1)$?

Solution:

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



common cause

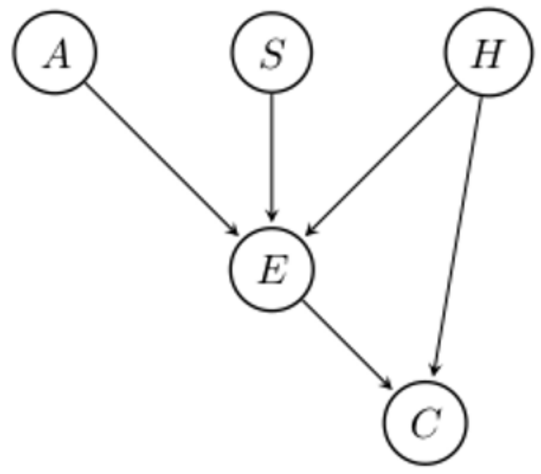
While

$$\begin{aligned}\mathbb{P}(\neg A, B, C) &= \mathbb{P}(\neg A, C|B) \cdot \mathbb{P}(B) \\ &= \mathbb{P}(\neg A|B) \cdot \mathbb{P}(C|B) \cdot \mathbb{P}(B)\end{aligned}$$

In other words,

$$\begin{aligned}\mathbb{P}(\neg A|C) &= \frac{\mathbb{P}(\neg A, C)}{\mathbb{P}(C)} \\ &= \frac{\mathbb{P}(\neg A, B, C) + \mathbb{P}(\neg A, \neg B, C)}{\mathbb{P}(\neg A, B, C) + \mathbb{P}(\neg A, \neg B, C) + \mathbb{P}(A, B, C) + \mathbb{P}(A, \neg B, C)} \\ &= \frac{\mathbb{P}(\neg A|B) \cdot \mathbb{P}(C|B) \cdot \mathbb{P}(B) + \mathbb{P}(\neg A|\neg B) \cdot \mathbb{P}(C|\neg B) \cdot \mathbb{P}(\neg B)}{\mathbb{P}(\neg A|B) \cdot \mathbb{P}(C|B) \cdot \mathbb{P}(B) + \mathbb{P}(\neg A|\neg B) \cdot \mathbb{P}(C|\neg B) \cdot \mathbb{P}(\neg B) + \mathbb{P}(A|B) \cdot \mathbb{P}(C|B) \cdot \mathbb{P}(B) + \mathbb{P}(A|\neg B) \cdot \mathbb{P}(C|\neg B) \cdot \mathbb{P}(\neg B)} \\ &= \frac{(1-0.28) \times 0.12 \times 0.66 + (1-0.53) \times 0.26 \times (1-0.66)}{(1-0.28) \times 0.12 \times 0.66 + (1-0.53) \times 0.26 \times (1-0.66) + 0.28 \times 0.12 \times 0.66 + 0.53 \times 0.26 \times (1-0.66)}\end{aligned}$$

12. **(Spring 2017 Final Q8)**



Consider the following Bayesian Network containing 5 Boolean random variables:

(a) Write an expression for computing $\mathbb{P}(A, \neg S, H, E, \neg C)$ given only information that is in the associated CPTs for this network.

(c) How many numbers must be stored in total in all CPTs associated with this network (excluding numbers that can be calculated from other numbers)?

Solution: (a)

$$\begin{aligned}\mathbb{P}(A, \neg S, H, E, \neg C) &= \mathbb{P}(\neg C|A, \neg S, H, E) \cdot \mathbb{P}(A, \neg S, H, E) \\ &= \mathbb{P}(\neg C|H, E) \cdot \mathbb{P}(E|A, \neg S, H) \cdot \mathbb{P}(A, \neg S, H) \\ &= \mathbb{P}(\neg C|H, E) \cdot \mathbb{P}(E|A, \neg S, H) \cdot \mathbb{P}(A) \cdot \mathbb{P}(\neg S) \cdot \mathbb{P}(H)\end{aligned}$$

(c) $1 + 1 + 1 + 2^3 + 2^2$

1	$\mathbb{P}(A = 1)$
1	$\mathbb{P}(S = 1)$
1	$\mathbb{P}(H = 1)$
2^3	$\mathbb{P}(E = 1\|A, S, H), \mathbb{P}(E = 1\|\neg A, S, H), ..., \mathbb{P}(E = 1\|\neg A, \neg S, \neg H)$
2^2	$\mathbb{P}(C = 1\|E, H), \mathbb{P}(C = 1\|\neg E, H), \mathbb{P}(C = 1\|E, \neg H), \mathbb{P}(C = 1\|\neg E, \neg H)$

A = 0, A = 1; S = 0, S= 1; H = 0, H=1

P(C = 0|E = 1, H = 1) = 1 - P(C=1 | E = 1, H=1)

A S H

0, 0, 0
0, 0, 1

0, 1, 0
0, 1, 1

1, 0, 0
1, 0, 1
1, 1, 0
1, 1, 1