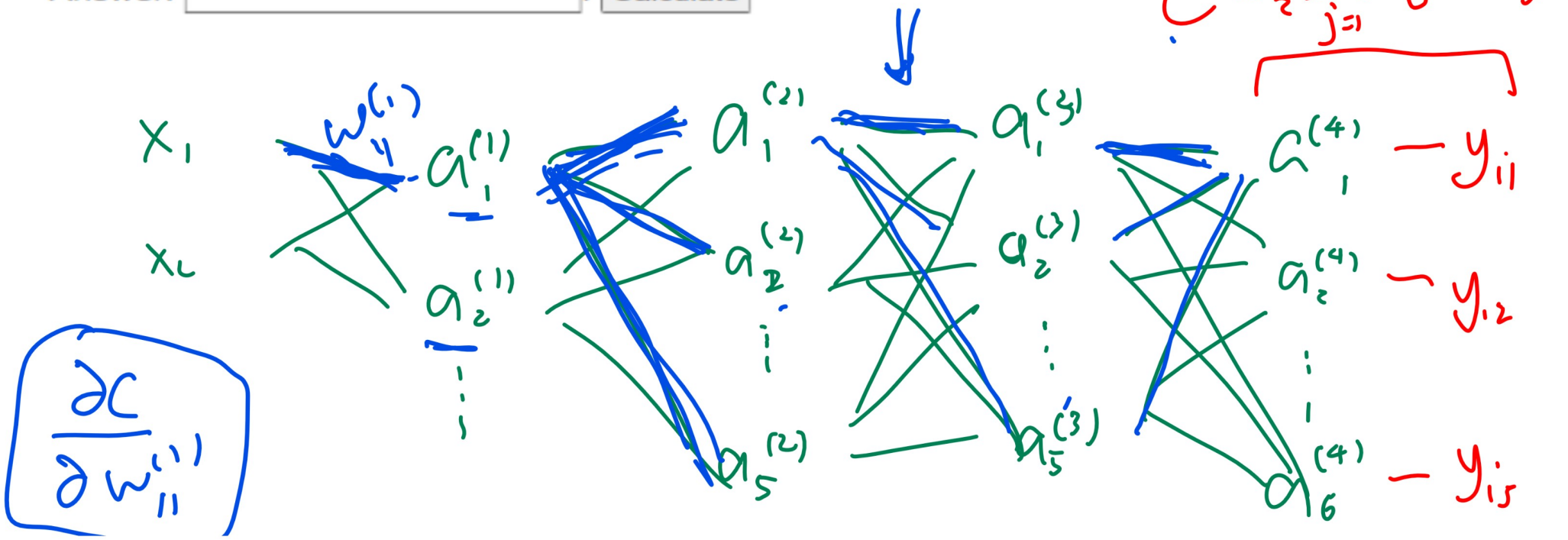
Question 1

• [3 points] Suppose you are given a neural network with 3 hidden layers, 9 input units, 6 output units, and $[10 \quad 5 \quad 5]$ hidden units. In one backpropogation step when computing the gradient of the cost (for example, squared loss) with respect to $w_{11}^{(1)}$, the weight in layer 1 connecting input 1 and hidden unit 1, how many weights (including $w_{11}^{(1)}$ itself, and including biases) are used in the backpropogation step of $\frac{\partial C}{\partial w_{11}^{(1)}}$?

• Note: the backpropogation step assumes the activations in all layers are already known so do not count the weights and biases in the forward step computing the activations.

• Answer: . Calculate



$+ \quad 5 + 5 \cdot 5 + 5 \cdot 6$

$\frac{\partial C}{\partial w_{11}^{(1)}} = a_1^{(1)}$ $\qquad$ $w_{11}^{(1)} = w_{11}^{(1)} - \alpha \frac{\partial C}{\partial w_{11}^{(1)}}$

Question 2

• [3 points] A tweet is ratioed if a reply gets more likes than the tweet. Suppose a tweet has 3 replies, and each one of these replies gets more likes than the tweet with probability 0.96 if the tweet is bad, and probability 0.11 if the tweet is good. Given a tweet is ratioed, what is the probability that it is a bad tweet? The prior probability of a bad tweet is 0.73.

• Answer: [ ] . Calculate

one of

$\Pr\{B\} = 0.73$ ✓ $\Pr\{\neg B\} = 0.27$ ✓

$\Pr\{R \mid B\} = 1 - (1-0.96)^3 = 1 - 0.04^3$ ✓

$\Pr\{\neg R \mid B\} = 0.04^3$

$Pr\{R \mid \neg B\} = 1 - (1-0.1)^3 = 1 - 0.9^3$ ✓

$Pr\{\neg R \mid \neg B\} = 0.9^3$

$$Pr\{B \mid R\} = \left[\frac{Pr\{B, R\}}{Pr\{R\}} \right] \rightarrow Pr\{R, B\} + Pr\{R, \neg B\}$$

$$= \left[\frac{Pr\{R \mid B\} \cdot Pr\{B\}}{Pr\{R \mid B\} \cdot Pr\{B\} + Pr\{R \mid \neg B\} \cdot Pr\{\neg B\}} \right]$$

$= \ldots\ldots$

Question 3

- [3 points] Given an infinite state sequence where the pattern "3 2 3 1 2 1 2 1" is repeated infinite number of time. What is the (maximum likelihood) estimated transition probability from state 1 to 3 (without smoothing)?
- Answer: [ ] Calculate

$\hat{P}_r\{3|13\} = \frac{C_{1\to 3}}{C_1} = \frac{1}{3}$

$\hat{P}_r\{2|13\} = \frac{2}{3}$ $\quad \hat{P}_r\{1|13\} = 0$

Question 4

Not perceptron.

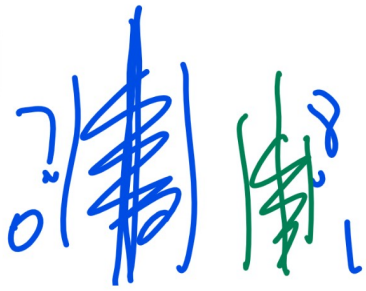
• [3 points] A hard margin support vector machine (SVM) is trained on the following dataset. Suppose we restrict $b = -1$, what is the value of w? Enter a single number, i.e. do not include b. Assume the SVM classifier is $1_{\{wx+b\geq 0\}}$.

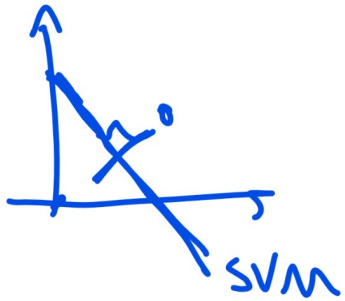
$w^T x + b = 0 \Rightarrow \hat{y} = 1$

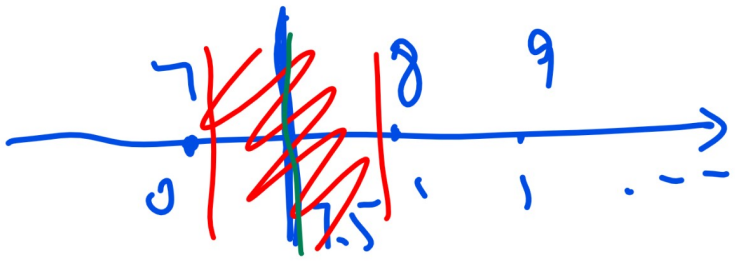
x_i	7	8	9	18	20
y_i	0	1	1	1	1

1 1 1 0 0 0

• Answer: [ ] Calculate







margin = 1

$x < 7.5 \Rightarrow 0$

$x \geq 7.5 \Rightarrow 1$

SVM

$$\left[\begin{array}{l} \mathbb{1}_{\{wx-1 \geq 0\}} \\ \Rightarrow \mathbb{1}_{\{wx \geq 1\}} \\ \Rightarrow \mathbb{1}_{\{x \geq \frac{1}{w}\}} \end{array}\right.$$

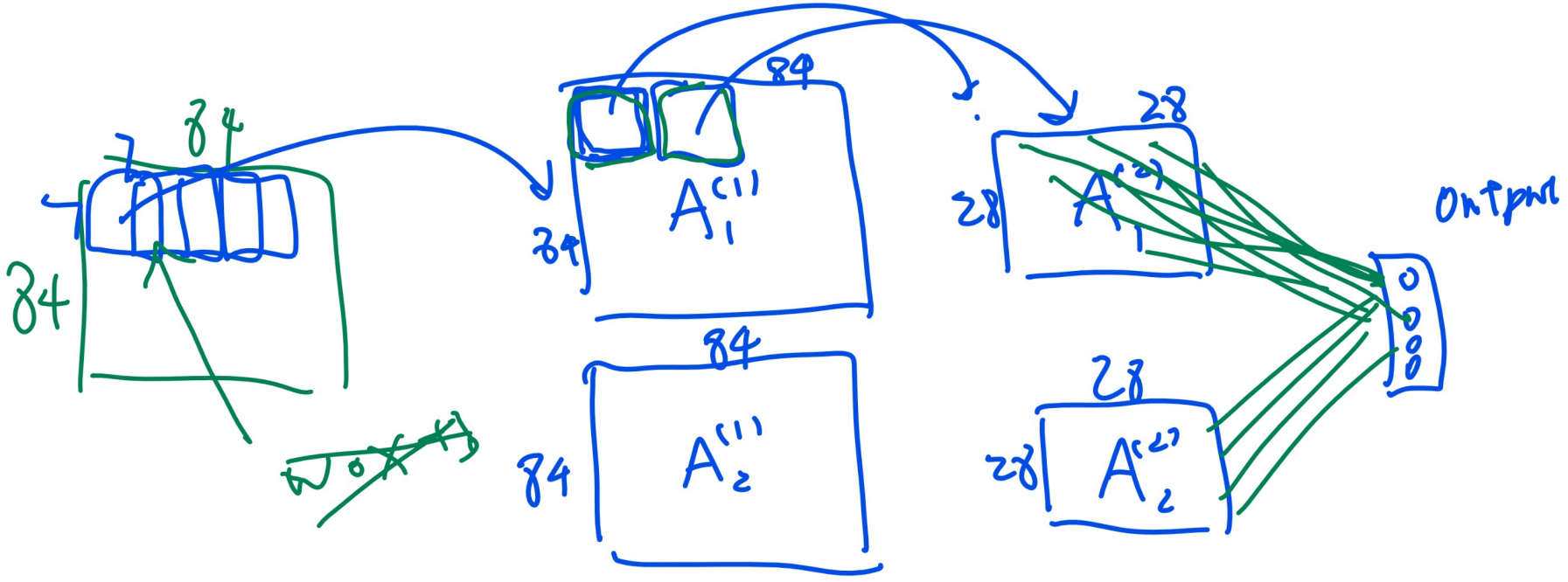
$$\frac{1}{w} = 7.5$$

$$w = \boxed{\frac{1}{7.5}}$$

Question 5

• [4 points] A convolutional neural network has input image of size 84 x 84 that is connected to a convolutional layer that uses a 7 x 7 filter, zero padding of the image, and a stride of 1. There are 2 activation maps. (Here, zero-padding implies that these activation maps have the same size as the input images.) The convolutional layer is then connected to a pooling layer that uses 3 x 3 max pooling, a stride of 3 (non-overlapping, no padding) of the convolutional layer. The pooling layer is then fully connected to an output layer that contains 4 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.

• Answer: . Calculate



2 filters → 7·7 = 49 weights

0 weights

28·28·4 two of these

$2 \cdot 7 \cdot 7 + 2 \cdot 28 \cdot 28 \cdot 4$

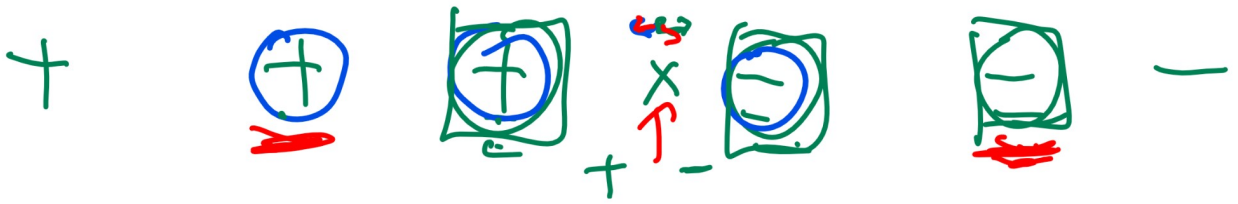
2021 MIA-C Q4

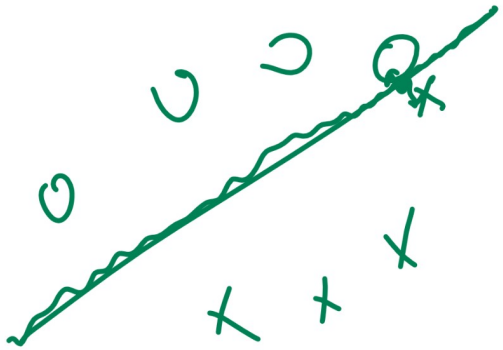
Question 4

• [4 points] You are given a training set of six points and their 2-class classifications (+ or -): $(-3.56, +)$, $(-2.33, +)$, $(-1.72, +)$, $(1.49, -)$, $(3.24, -)$, $(4.4, -)$. What is the decision boundary associated with this training set using 3NN (3 Nearest Neighbor)? Note: there is one more point compared to the question from the homework.

1NN

• Answer: [ ] Calculate





$$\frac{(-2.33 + 3.24)}{2}$$

→ midpt

decision boundary

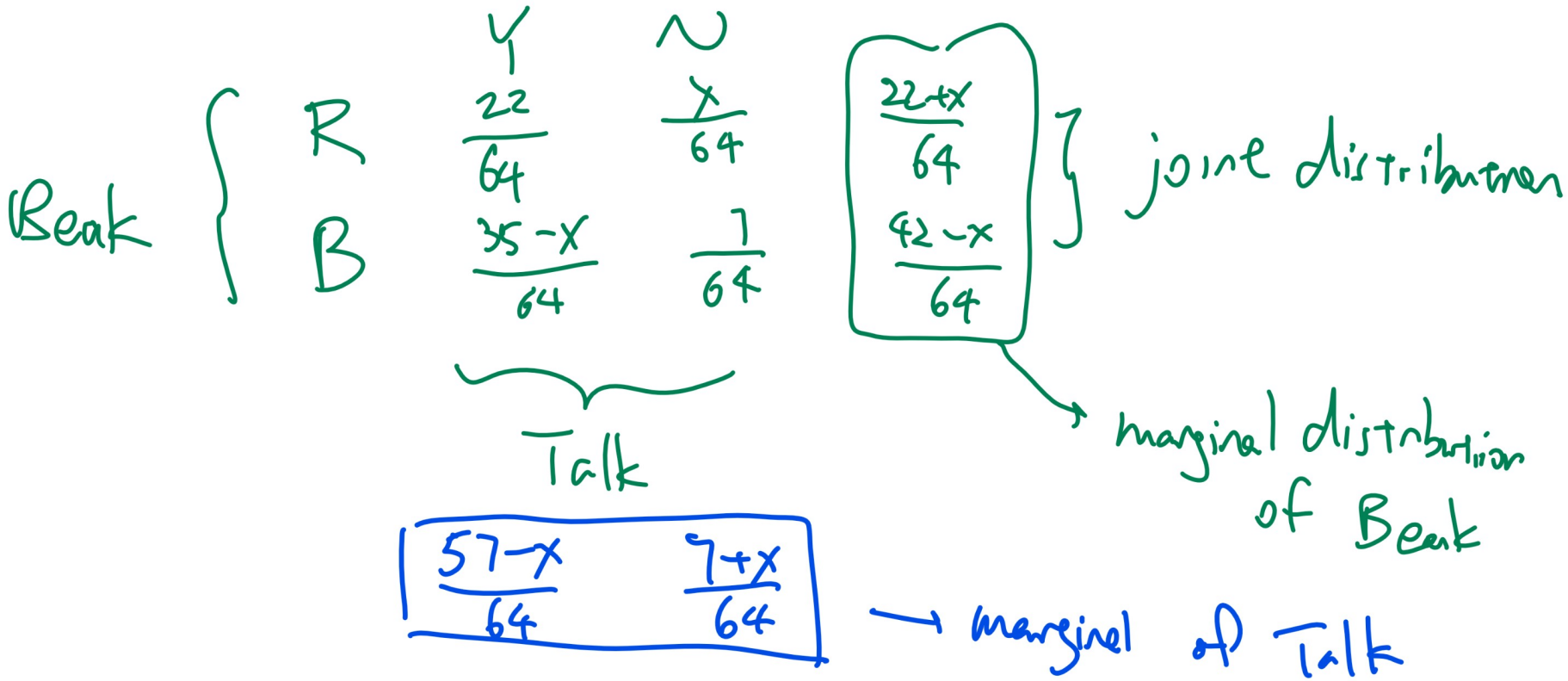
Question 2

• [4 points] There are 64 parrots. They have either a red beak or a black beak. They can either talk or not. Complete the two cells in the following table so that the mutual information (i.e. information gain) between "Beak" and "Talk" is ~~0.01~~ 0:

independent

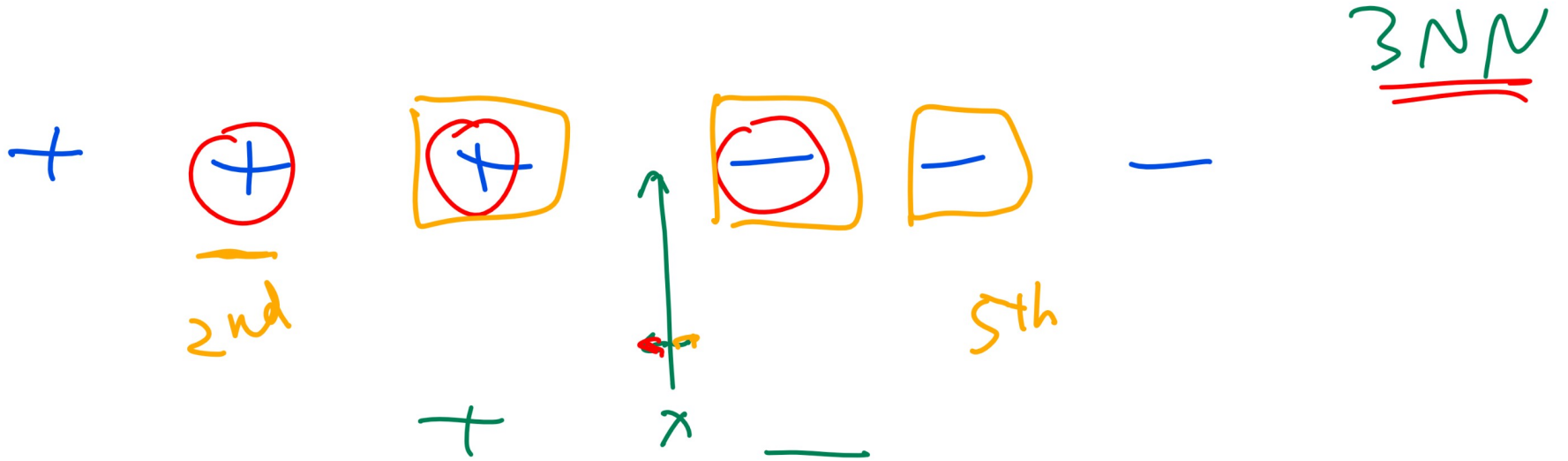
Number of parrots	Beak	Talk
22	Red	Yes
? x	Red	No
?? $64-22-7-x=35-x$	Black	Yes
7	Black	No

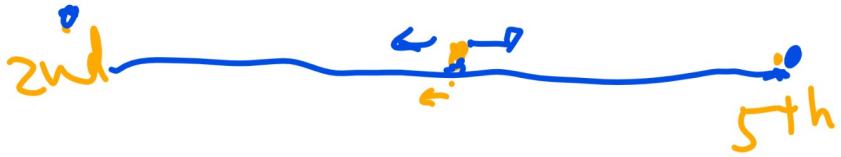
• Answer (comma separated vector): . Calculate



independence: $P_r\{R, Y\} = P_r\{R\} \cdot P_r\{Y\}$

$$\frac{22}{64} = \frac{22+x}{64} \cdot \frac{57-x}{64} \Big] \Rightarrow \text{solve for } x.$$





Question 9

• [4 points] In a convolutional neural network, suppose the activation map of a convolution layer is $\begin{bmatrix} -2 & -2 & -4 & 2 \\ 5 & 3 & 10 & -2 \\ 7 & 6 & 10 & 7 \\ -9 & -2 & 6 & 2 \end{bmatrix}$. What is the activation map after a non-overlapping (stride 2) 2 by 2 max-pooling layer?

• Answer (matrix with multiple lines, each line is a comma separated vector):

Calculate

$\begin{bmatrix} 5 & 10 \\ 7 & 10 \end{bmatrix}$

5,10 换行 7,10