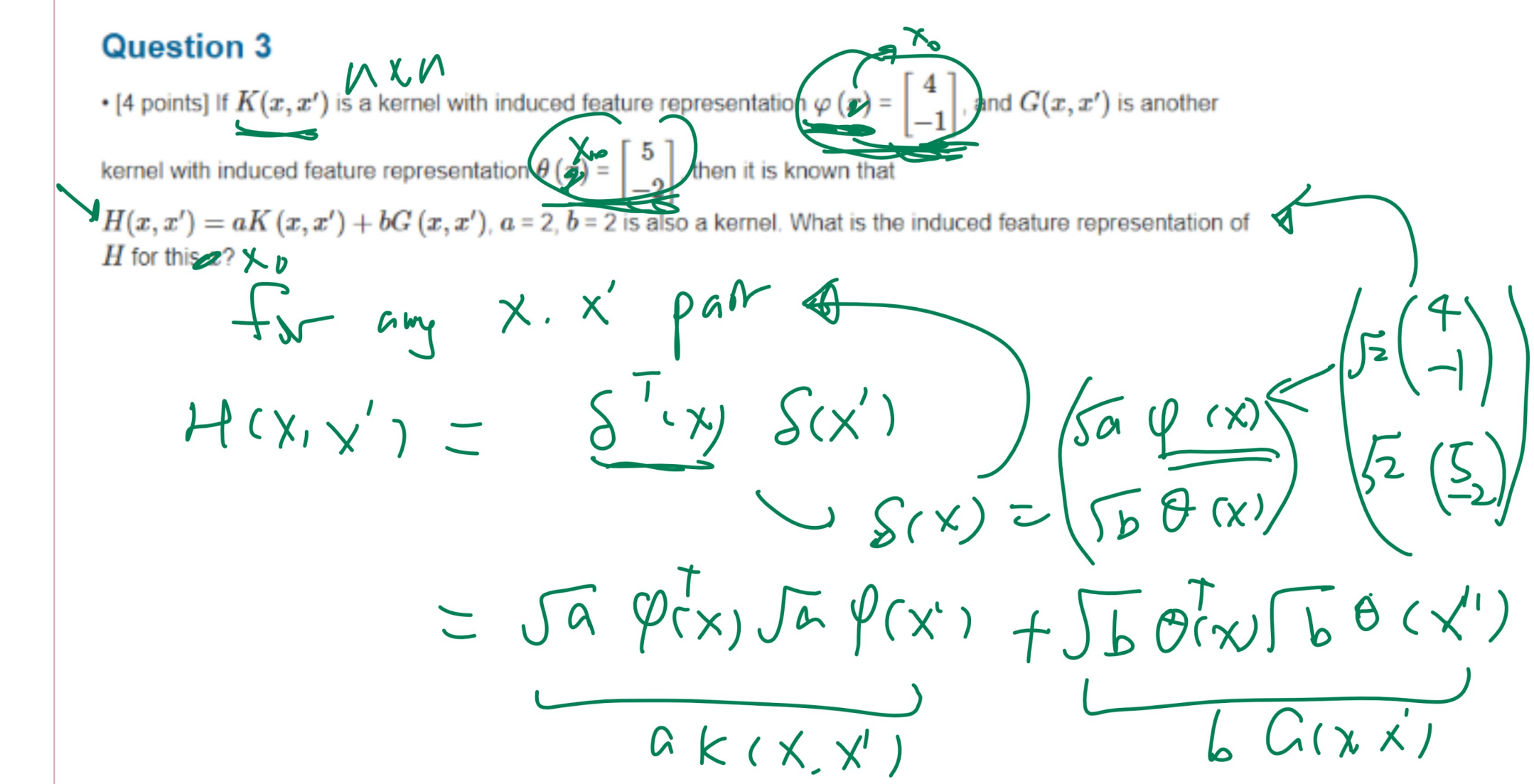
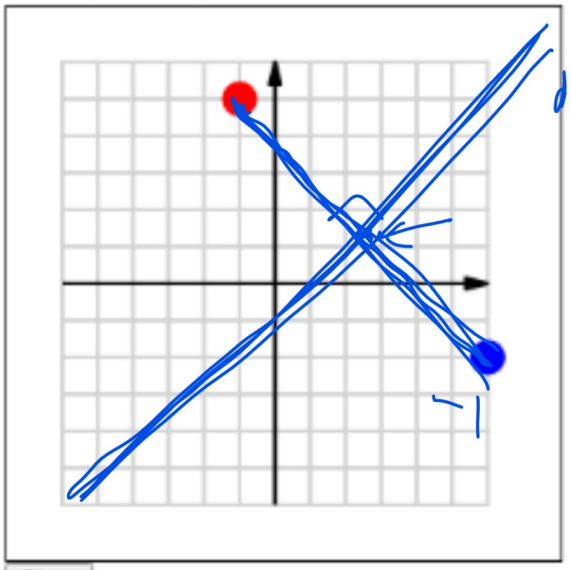
Question 3

• [4 points] If $K(x, x')$ is a kernel with induced feature representation $\varphi(x) = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$, and $G(x, x')$ is another kernel with induced feature representation $\theta(x) = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$, then it is known that $H(x, x') = aK(x, x') + bG(x, x')$, $a = 2$, $b = 2$ is also a kernel. What is the induced feature representation of H for this a?



Question 8

• [6 points] A linear SVM (Support Vector Machine) with with weights w_1, w_2, b is trained on the following data set: $x_1 = \begin{bmatrix} -1 \\ 5 \end{bmatrix}$, $y_1 = 0$ and $x_2 = \begin{bmatrix} 6 \\ -2 \end{bmatrix}$, $y_2 = 1$. The attributes (i.e. features) are two dimensional (x_1, x_2) and the label y is binary. The classification rule is $\hat{y} = 1_{\{w_1 x_1 + w_2 x_2 + b \geq 0\}}$. Assuming $b = -28$, what is (w_1, w_2) ? The drawing is not graded.



Clear

x_{i1}, x_{i2}

$w_1 x_{i1} + w_2 x_{i2} + b = 0$

midpoint : $\begin{pmatrix} 2.5 \\ 1.5 \end{pmatrix}$

slope : $\frac{4}{4} = 1$

$y = mx + b$

$x_{i2} = 1 \cdot x_{i1} - 1$

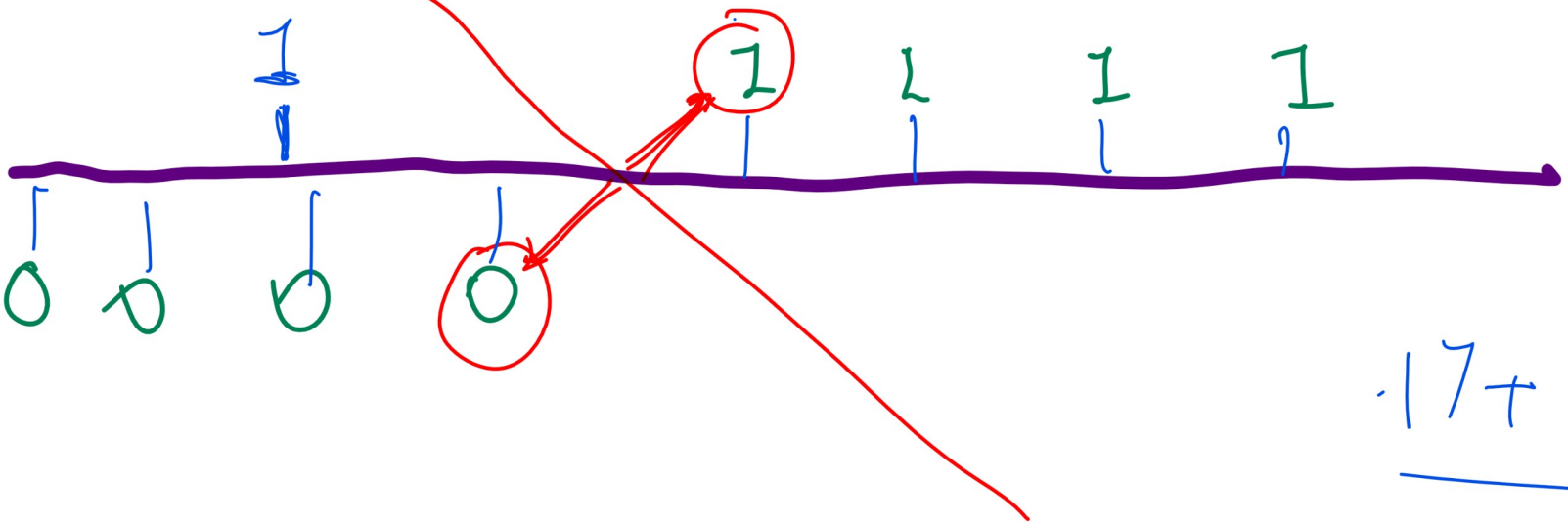
$x_{i1} - x_{i2} - 1 = 0$

$28\, x_{i1} - 28\, x_{i2} - 28 = 0$

$w = \begin{pmatrix} 28 \\ 28 \end{pmatrix}$

Question 9

• [4 points] Given a linear SVM (Support Vector Machine) that perfectly classifies a set of training data containing 10 positive examples and 7 negative examples with 2 support vectors. After adding one more positively labeled training example and retraining the SVM, what is the maximum possible number of support vectors possible in the new SVM.



17 + 1 = 18

Question 8

2-4 OR 5-7

• [3 points] The RDA Corporation has a prison with many cells. Without justification, you're about to be randomly thrown into a cell with equal probability. Cells 1 to 4 have Toruks that eat prisoners. Cells 5 to 7 are safe. With sufficient bribe, the warden will answer your question "Will I be in cell 1?" What's the mutual information (we call it information gain) between the warden's answer and your encounter with the Toruks? (I didn't write the stories in these questions, so I don't know the reference too.)

$$IG(Y|X) = H(Y) - H(Y|X)$$

$$Y = \begin{cases} 1 & \text{encounter} = \frac{4}{7} \\ 0 & \text{no} \quad \frac{3}{7} \end{cases}$$

$$X = \begin{cases} 1 & \text{yes in cell 1} \quad \frac{1}{7} \\ 0 & \text{no} \quad \frac{6}{7} \end{cases}$$

$$H(Y) = -\frac{4}{7}\log_2\left(\frac{4}{7}\right) - \frac{3}{7}\log_2\left(\frac{3}{7}\right)$$

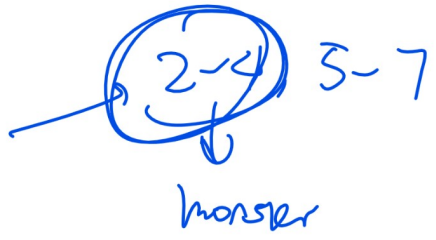
$$H(Y|X) = \frac{1}{7}H(Y|X=1) + \frac{6}{7}H(Y|X=0)$$

$H(Y|X=1) = 0$

$H(Y|X=0) = 1$

$$Y|X=1 = \begin{cases} 1 & \text{prob} = 1 \\ 0 & \text{prob} = 0 \end{cases}$$

$$Y|X=0 = \begin{cases} 1 & \text{prob} = \frac{3}{6} \\ 0 & \text{prob} = \frac{1}{2} \end{cases}$$



$$H(Y|X=1) = -1\log_2 1 - 0\log_2 0$$

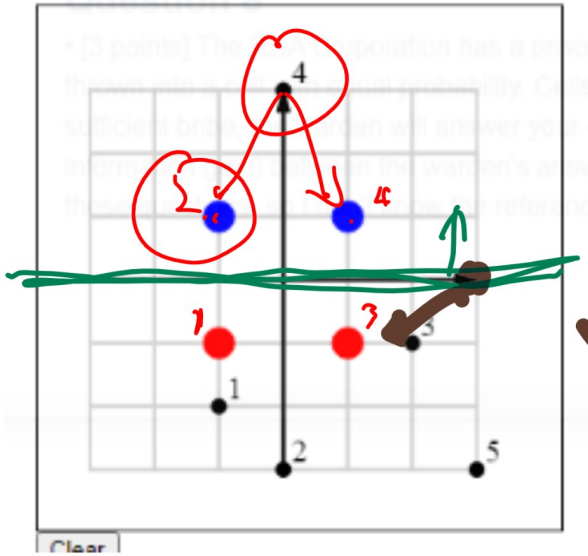
$$H(Y|X=0) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

Question 3

• [3 points] Consider points in 2D and binary labels. Given the training data in the table, and use Manhattan distance with 1NN (Nearest Neighbor), which of the following points in 2D are classified as 1? Answer the question by first drawing the decision boundaries. The drawing is not graded.

index	x_1	x_2	label
1	-1	-1	1
2	-1	1	0
3	1	-1	1
4	1	1	0

-1, 1 0
-1, -1 1



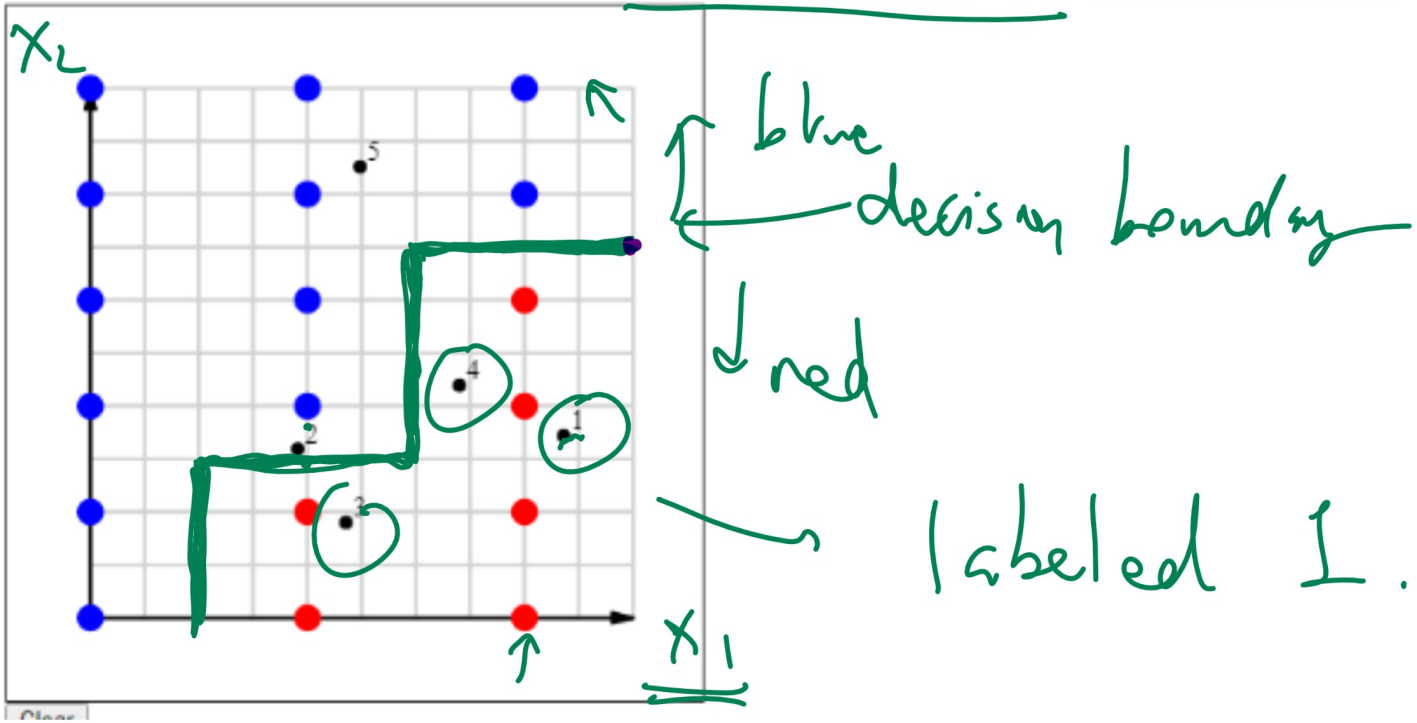
index 2 < index 4

copy y_2.

on the line → class 1

Question 2

• [3 points] Consider binary classification in 2D where the intended label of a point $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is positive (1) if $x_1 > x_2$ and negative (0) otherwise. Let the training set be all points of the form $x = \begin{bmatrix} 4 & a \\ 2 & b \end{bmatrix}$ where a, b are integers. Each training item has the correct label that follows the rule above. With a 1NN (Nearest Neighbor) classifier (Euclidean distance), which ones of the following points are labeled positive? The drawing is not graded.



Clear

$H(Y) - H(Y|X)$

P2

$X = 1$
$X = 2$
$X = n$

$1, 2, \ldots n$

$X > t \longrightarrow 1$

$X \leq t \longrightarrow 0$

for each X

① try $X > 1$
$X \leq 1$

$H(Y|X>1) \cdot Pr\{X>1\}$
$+ H(Y|X\leq 1) \cdot Pr\{X\leq 1\}$

Compute IG.

② try $X > 2$
$X \leq 2$

IG

... $t = 1, 2 \ldots 9$

find largest IG.