MLT: Week-7

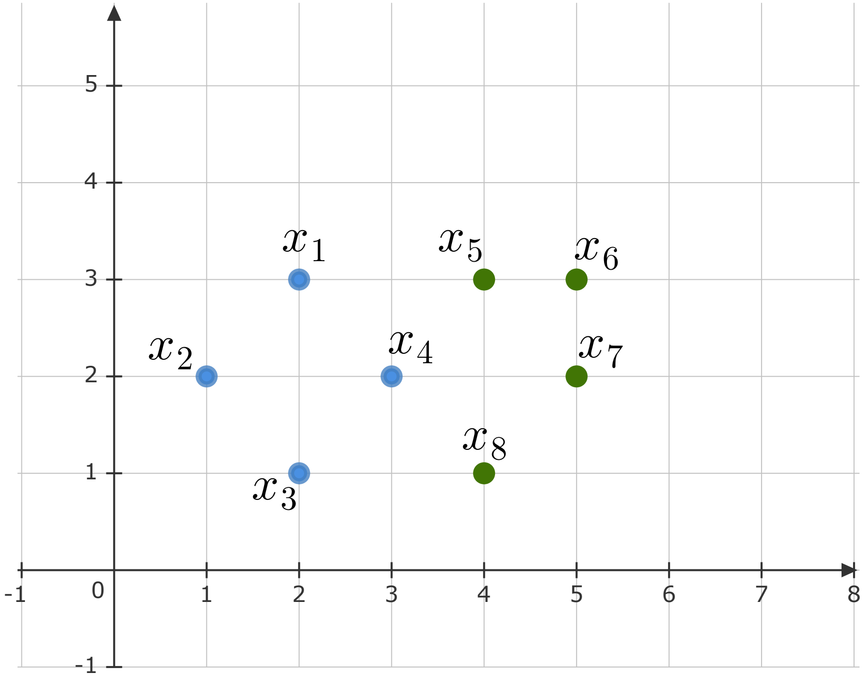
KNN and Decision Trees

A Aniruddha

Observe the dataset

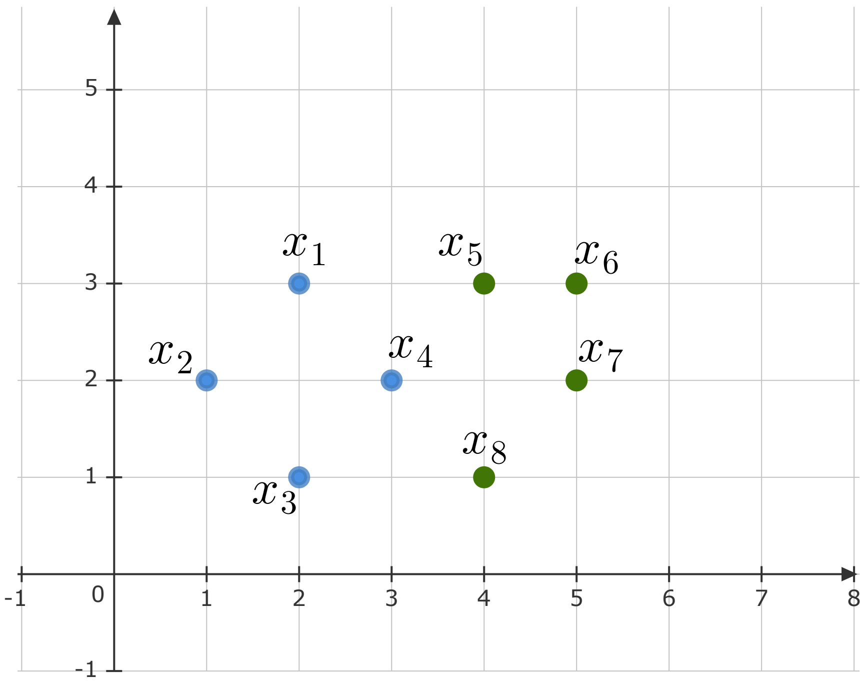
Observe the dataset

Consider the following dataset where blue points have class +1 and green points have class -1



Observe the dataset

Consider the following dataset where blue points have class +1 and green points have class -1



Point		Class
x_1	$(2,3)$	+1
x_2	$(1,2)$	+1
x_3	$(2,1)$	+1
x_4	$(3,2)$	+1
x_5	$(4,3)$	-1
x_6	$(5,3)$	-1
x_7	$(5,2)$	-1
x_8	$(4,1)$	-1

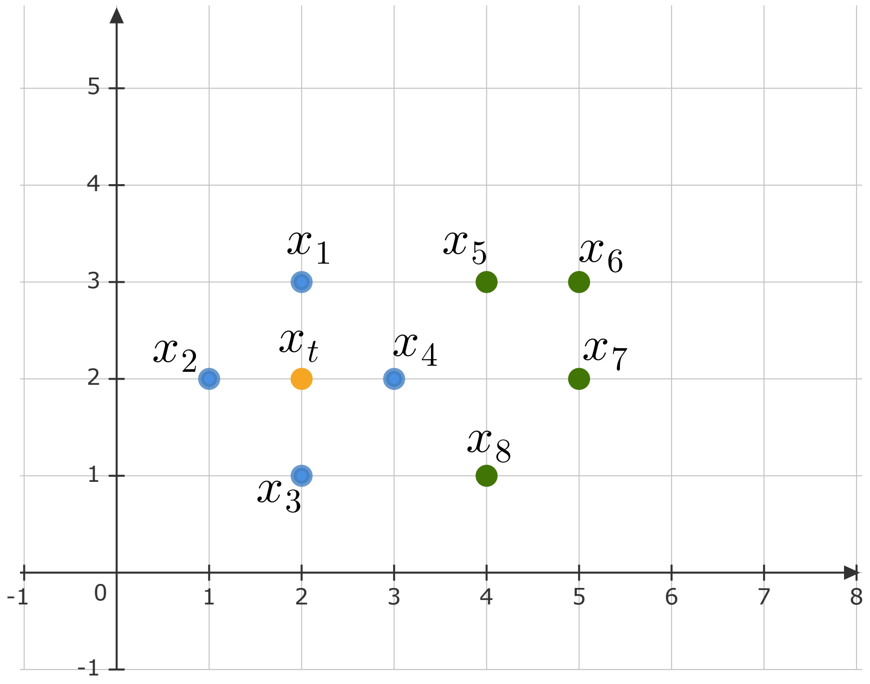
KNN - Procedure

KNN - Procedure

Consider a test point $x_t = (2, 2)$ and assign a label for different values of K

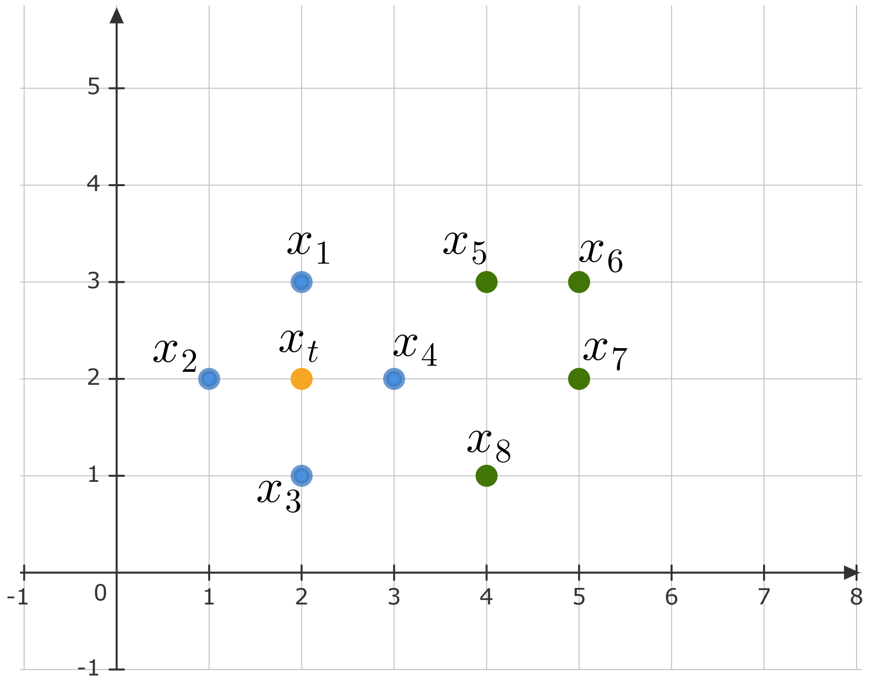
KNN - Procedure

Consider a test point $x_t = (2, 2)$ and assign a label for different values of K



KNN - Procedure

Consider a test point $x_t = (2,2)$ and assign a label for different values of K

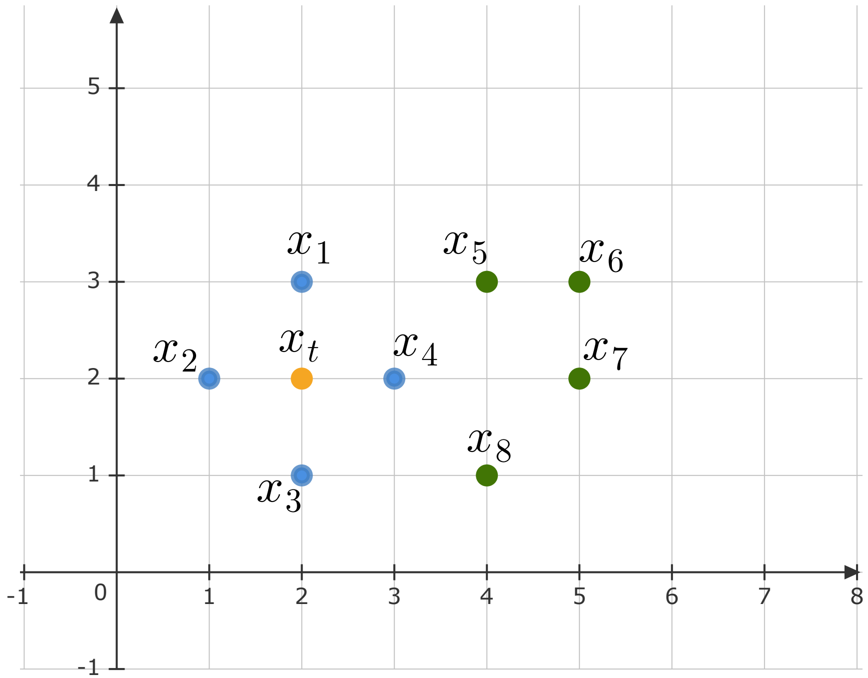


To assign a label to a test point,

1. Compute its distance from every other point in the dataset
2. Depending on the value of K, we choose the K closest points and assign the label corresponding to the majority of the points

KNN - Procedure

Consider a test point $x_t = (2, 2)$ and assign a label for different values of K



To assign a label to a test point,

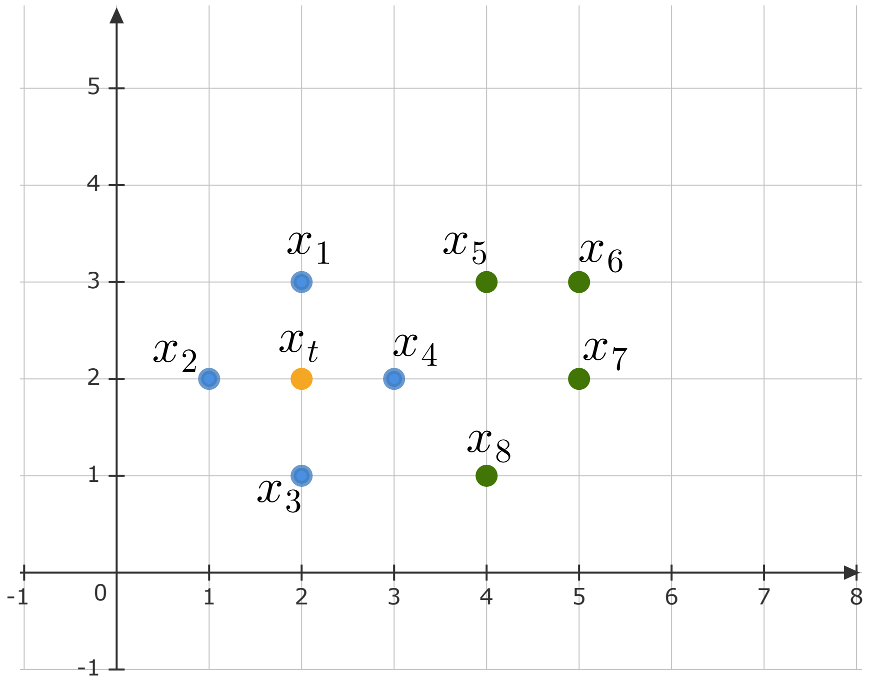
1. Compute its distance from every other point in the dataset
2. Depending on the value of K, we choose the K closest points and assign the label corresponding to the majority of the points

The distance of the point x_t from x_1 is given by,

$$\text{Distance} = \sqrt{(2-2)^2 + (2-3)^2}$$
$$= 1$$

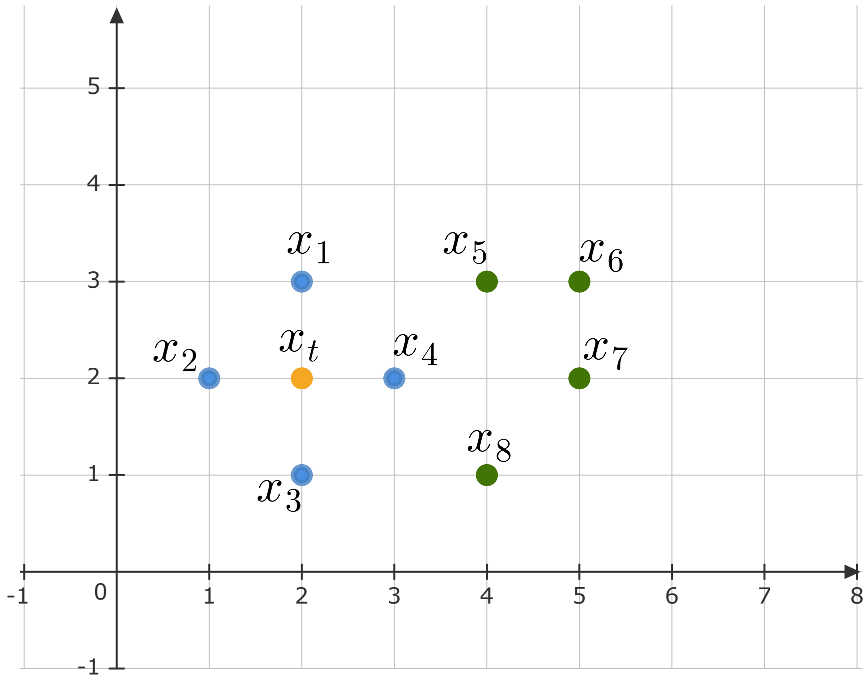
Distance of test point from training data

Distance of test point from training data



Distance of test point from training data

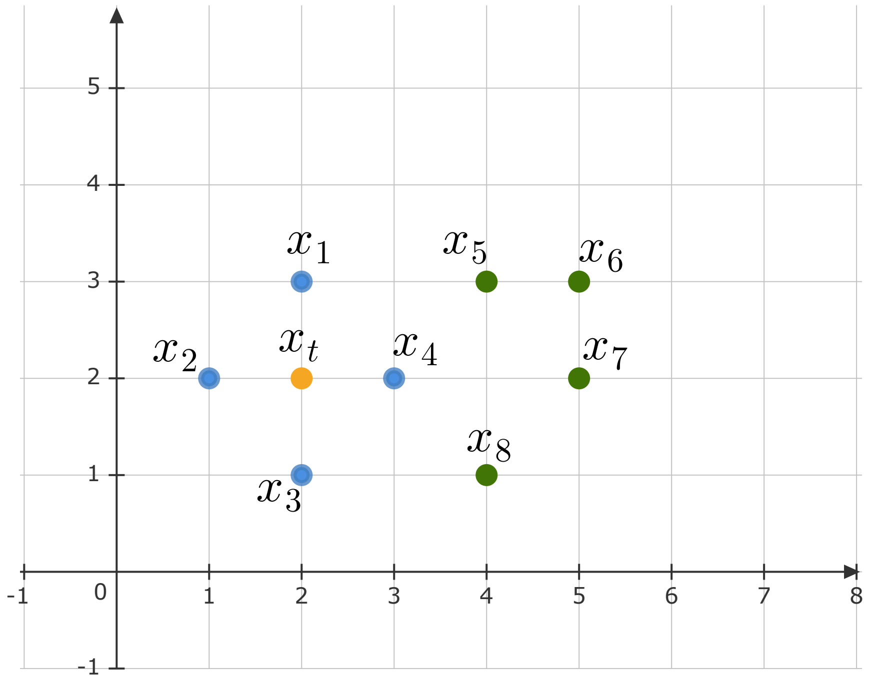
The distance of the test data point from each training data point is,



Point		Class	Distance
x_1	$(2, 3)$	+1	1
x_2	$(1, 2)$	+1	1
x_3	$(2, 1)$	+1	1
x_4	$(3, 2)$	+1	1
x_5	$(4, 3)$	-1	$\sqrt{5}$
x_6	$(5, 3)$	-1	$\sqrt{10}$
x_7	$(5, 2)$	-1	3
x_8	$(4, 1)$	-1	$\sqrt{5}$

Distance of test point from training data

The distance of the test data point from each training data point is,

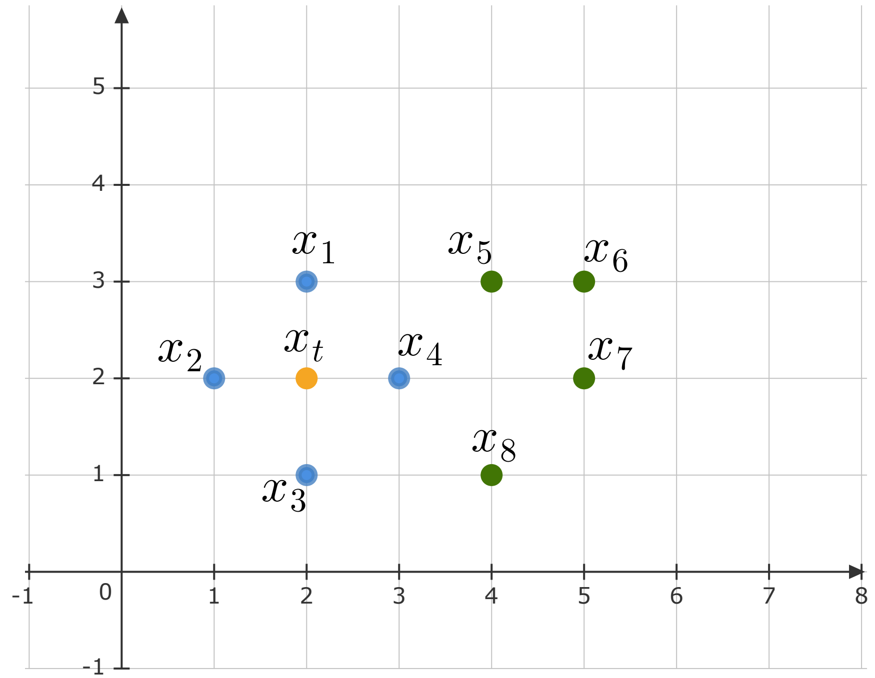


Point		Class	Distance
x_1	$(2,3)$	+1	1
x_2	$(1,2)$	+1	1
x_3	$(2,1)$	+1	1
x_4	$(3,2)$	+1	1
x_5	$(4,3)$	-1	$\sqrt{5}$
x_6	$(5,3)$	-1	$\sqrt{10}$
x_7	$(5,2)$	-1	3
x_8	$(4,1)$	-1	$\sqrt{5}$

We now consider different values of K and assign the label accordingly

Distance of test point from training data

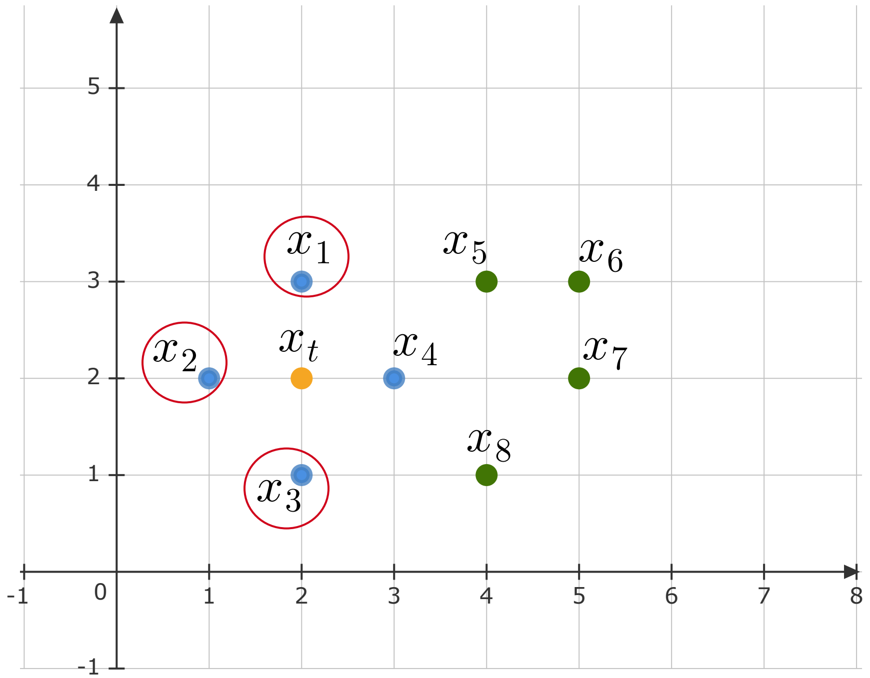
For K = 3, we take the three closest points and assign the class corresponding to the majority



Point		Class	Distance
x_1	$(2, 3)$	+1	1
x_2	$(1, 2)$	+1	1
x_3	$(2, 1)$	+1	1
x_4	$(3, 2)$	+1	1
x_5	$(4, 3)$	-1	$\sqrt{5}$
x_6	$(5, 3)$	-1	$\sqrt{10}$
x_7	$(5, 2)$	-1	3
x_8	$(4, 1)$	-1	$\sqrt{5}$

Label Assignment

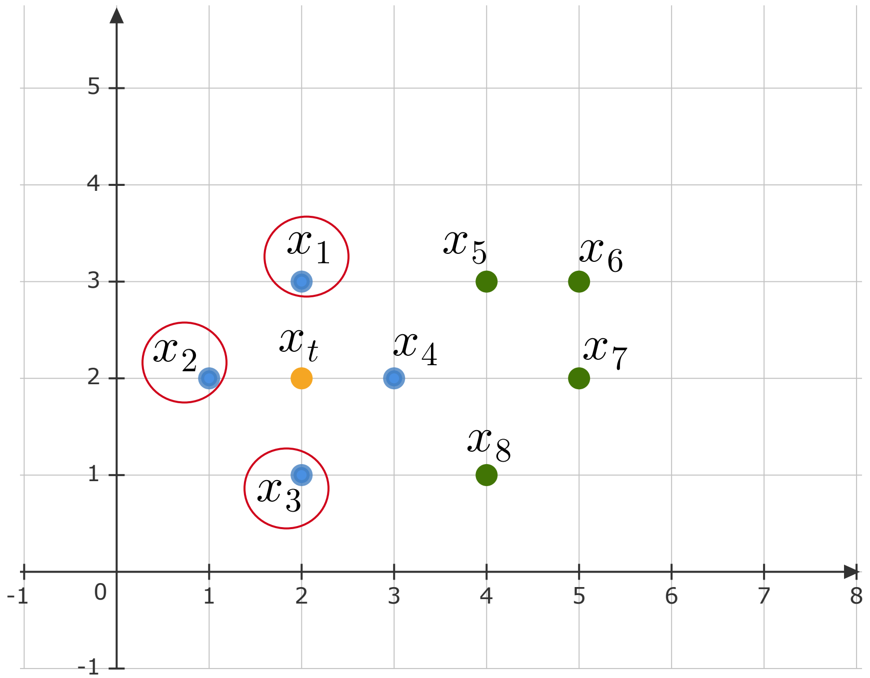
For K = 3, we take the three closest points and assign the class corresponding to the majority



Point		Class	Distance
x_1	$(2, 3)$	+1	1
x_2	$(1, 2)$	+1	1
x_3	$(2, 1)$	+1	1
x_4	$(3, 2)$	+1	1
x_5	$(4, 3)$	-1	$\sqrt{5}$
x_6	$(5, 3)$	-1	$\sqrt{10}$
x_7	$(5, 2)$	-1	3
x_8	$(4, 1)$	-1	$\sqrt{5}$

Label Assignment

For K = 3, we take the three closest points and assign the class corresponding to the majority

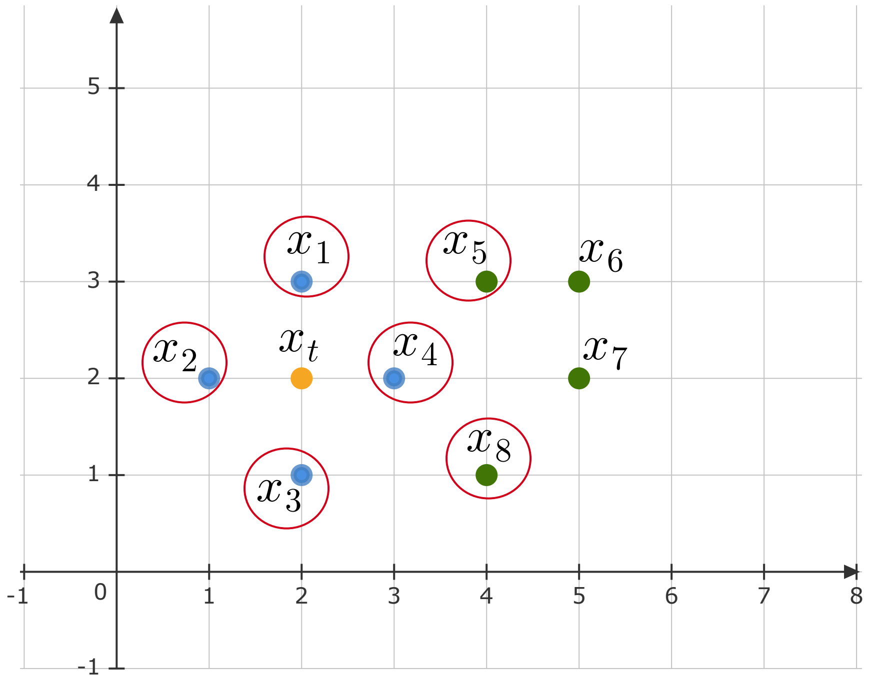


Point		Class	Distance
x_1	$(2,3)$	+1	1
x_2	$(1,2)$	+1	1
x_3	$(2,1)$	+1	1
x_4	$(3,2)$	+1	1
x_5	$(4,3)$	-1	$\sqrt{5}$
x_6	$(5,3)$	-1	$\sqrt{10}$
x_7	$(5,2)$	-1	3
x_8	$(4,1)$	-1	$\sqrt{5}$

Here, we see that the three closest points have a label of +1 and so we will assign the test point with the same label

Label Assignment

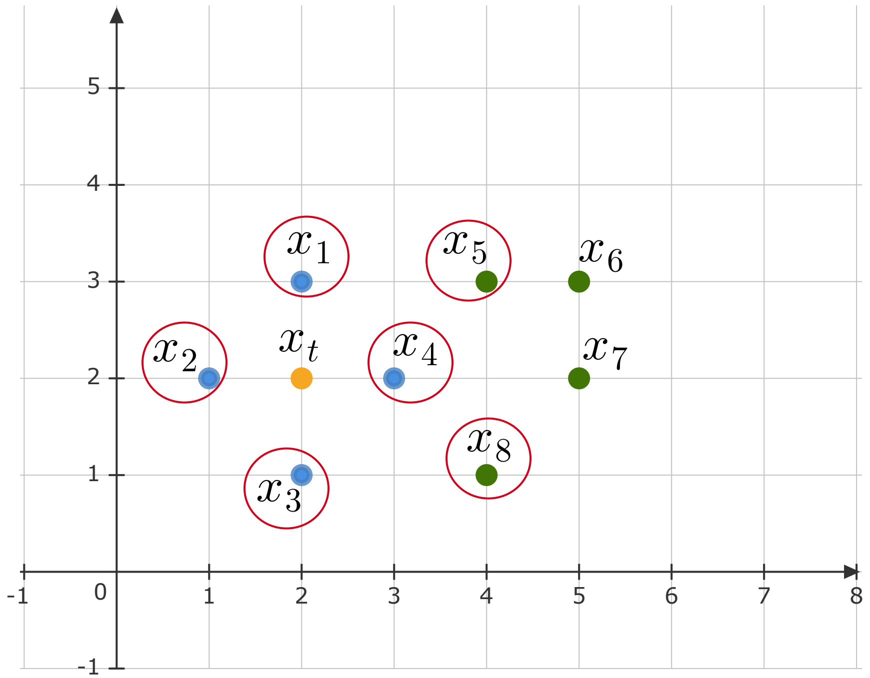
For K = 6, we take the six closest points and assign the class corresponding to the majority



Point		Class	Distance
x_1	$(2, 3)$	+1	1
x_2	$(1, 2)$	+1	1
x_3	$(2, 1)$	+1	1
x_4	$(3, 2)$	+1	1
x_5	$(4, 3)$	-1	$\sqrt{5}$
x_6	$(5, 3)$	-1	$\sqrt{10}$
x_7	$(5, 2)$	-1	3
x_8	$(4, 1)$	-1	$\sqrt{5}$

Label Assignment

For K = 6, we take the six closest points and assign the class corresponding to the majority

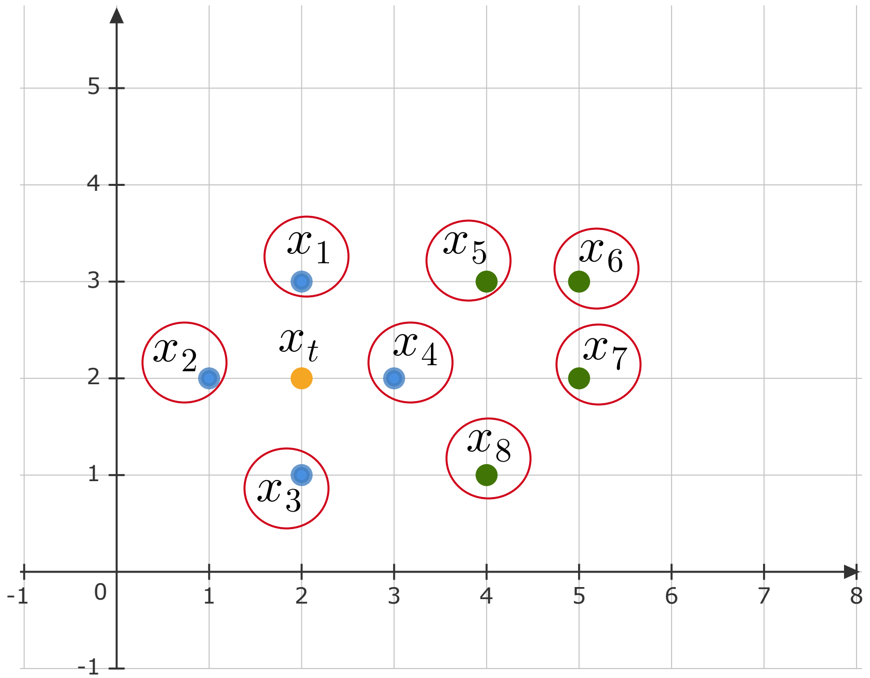


Here, we see that four points have a label of +1 and two points have a label of -1. Since the majority has label +1, we assign the same to x_t

Point		Class	Distance
x_1	$(2, 3)$	+1	1
x_2	$(1, 2)$	+1	1
x_3	$(2, 1)$	+1	1
x_4	$(3, 2)$	+1	1
x_5	$(4, 3)$	-1	$\sqrt{5}$
x_6	$(5, 3)$	-1	$\sqrt{10}$
x_7	$(5, 2)$	-1	3
x_8	$(4, 1)$	-1	$\sqrt{5}$

Label Assignment

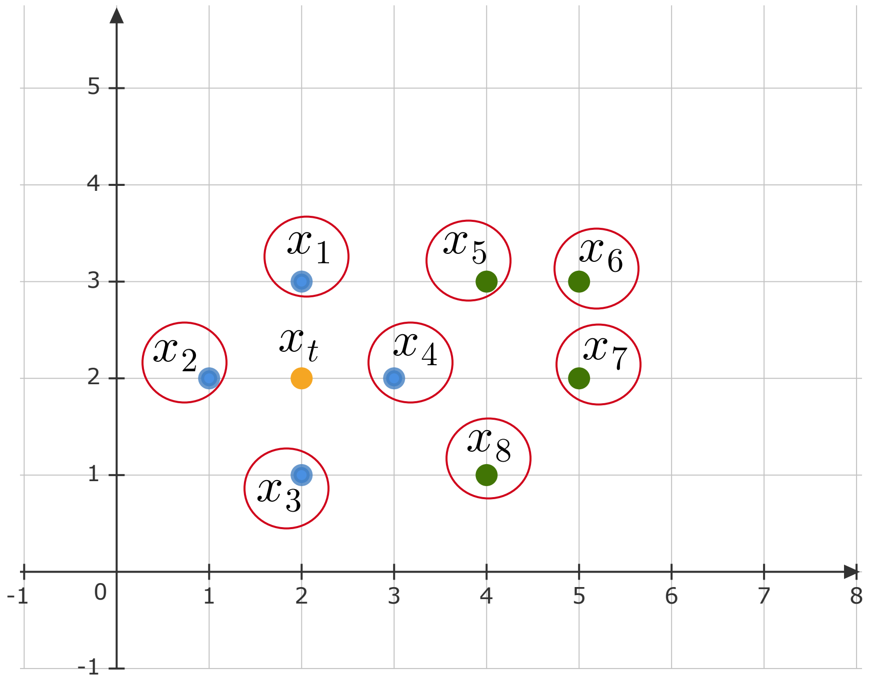
For K = 8, we take all the points and assign the class corresponding to the majority



Point		Class	Distance
x_1	$(2, 3)$	+1	1
x_2	$(1, 2)$	+1	1
x_3	$(2, 1)$	+1	1
x_4	$(3, 2)$	+1	1
x_5	$(4, 3)$	-1	$\sqrt{5}$
x_6	$(5, 3)$	-1	$\sqrt{10}$
x_7	$(5, 2)$	-1	3
x_8	$(4, 1)$	-1	$\sqrt{5}$

Label Assignment

For K = 8, we take all the points and assign the class corresponding to the majority

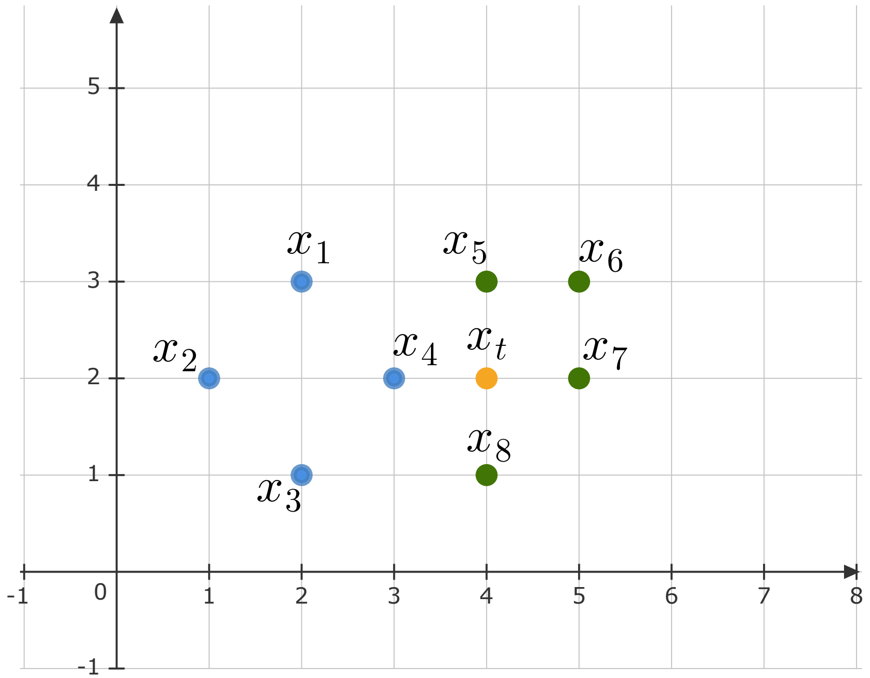


Here, we see that four points have a label of +1 and four points have a label of -1. Since there is no clear majority, we can assign any label

Point		Class	Distance
x_1	$(2,3)$	+1	1
x_2	$(1,2)$	+1	1
x_3	$(2,1)$	+1	1
x_4	$(3,2)$	+1	1
x_5	$(4,3)$	-1	$\sqrt{5}$
x_6	$(5,3)$	-1	$\sqrt{10}$
x_7	$(5,2)$	-1	3
x_8	$(4,1)$	-1	$\sqrt{5}$

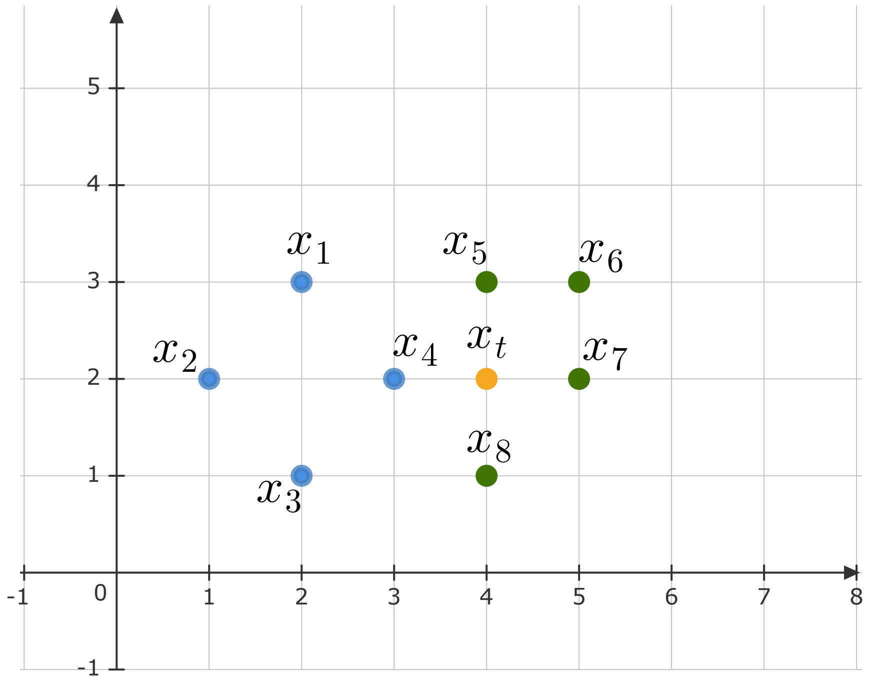
KNN - Procedure

Consider a test point $x_t = (4, 2)$ and assign a label for different values of K



Distance of test point from training data

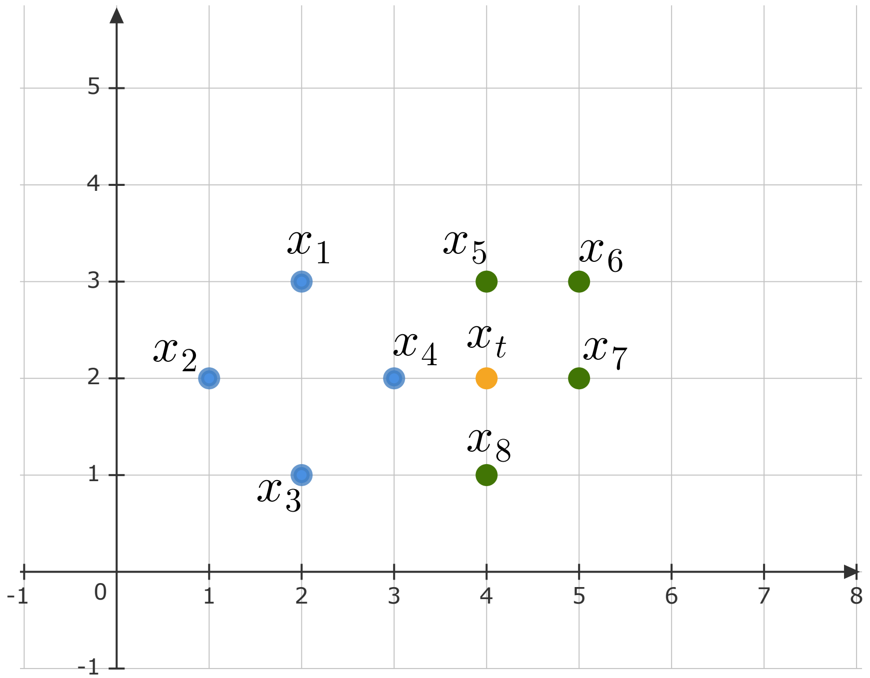
The distance of the test data point from each training data point is,



Point		Class	Distance
x_1	$(2, 3)$	+1	$\sqrt{5}$
x_2	$(1, 2)$	+1	3
x_3	$(2, 1)$	+1	$\sqrt{5}$
x_4	$(3, 2)$	+1	1
x_5	$(4, 3)$	-1	1
x_6	$(5, 3)$	-1	$\sqrt{2}$
x_7	$(5, 2)$	-1	1
x_8	$(4, 1)$	-1	1

Distance of test point from training data

The distance of the test data point from each training data point is,

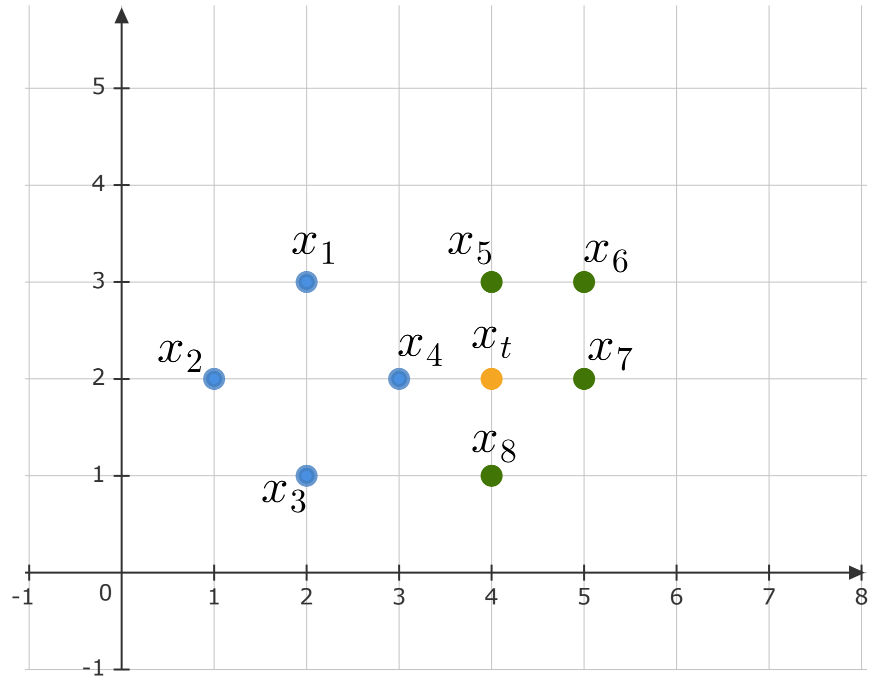


Point		Class	Distance
x_1	$(2, 3)$	+1	$\sqrt{5}$
x_2	$(1, 2)$	+1	3
x_3	$(2, 1)$	+1	$\sqrt{5}$
x_4	$(3, 2)$	+1	1
x_5	$(4, 3)$	-1	1
x_6	$(5, 3)$	-1	$\sqrt{2}$
x_7	$(5, 2)$	-1	1
x_8	$(4, 1)$	-1	1

We now consider different values of K and assign the label accordingly

Distance of test point from training data

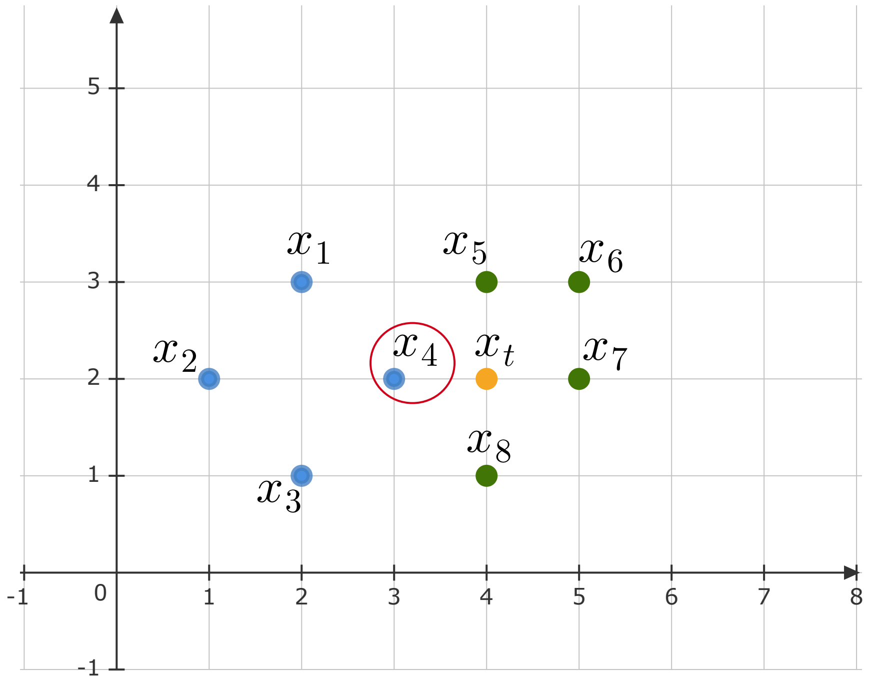
For K = 1, we take one of the closest points and assign its class to the test point



Point		Class	Distance
x_1	$(2, 3)$	+1	$\sqrt{5}$
x_2	$(1, 2)$	+1	3
x_3	$(2, 1)$	+1	$\sqrt{5}$
x_4	$(3, 2)$	+1	1
x_5	$(4, 3)$	-1	1
x_6	$(5, 3)$	-1	$\sqrt{2}$
x_7	$(5, 2)$	-1	1
x_8	$(4, 1)$	-1	1

Label Assignment

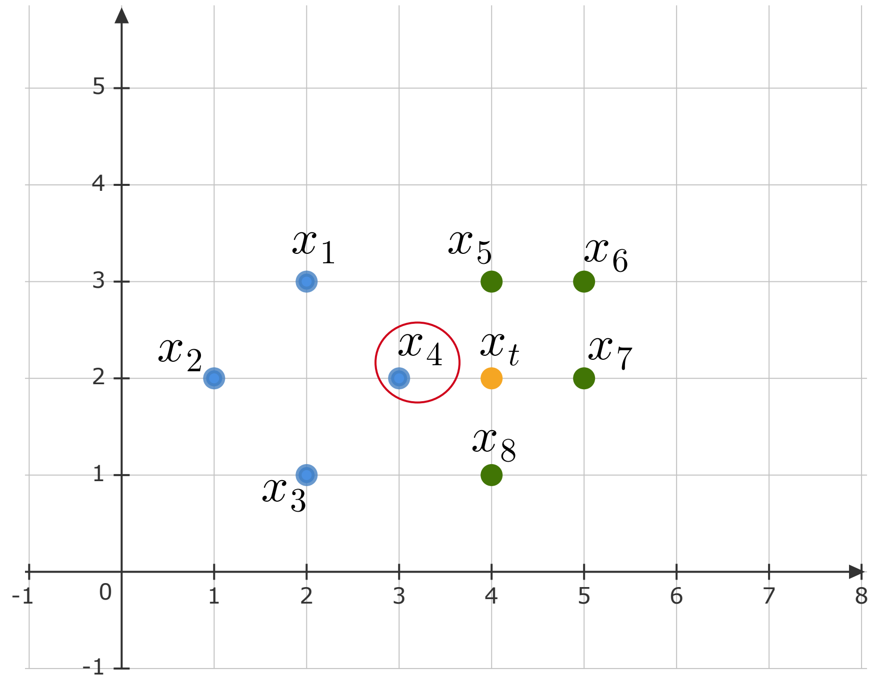
For K = 1, we take one of the closest points and assign its class to the test point



Point		Class	Distance
x_1	$(2, 3)$	+1	$\sqrt{5}$
x_2	$(1, 2)$	+1	3
x_3	$(2, 1)$	+1	$\sqrt{5}$
x_4	$(3, 2)$	+1	1
x_5	$(4, 3)$	-1	1
x_6	$(5, 3)$	-1	$\sqrt{2}$
x_7	$(5, 2)$	-1	1
x_8	$(4, 1)$	-1	1

Label Assignment

For K = 1, we take one of the closest points and assign its class to the test point

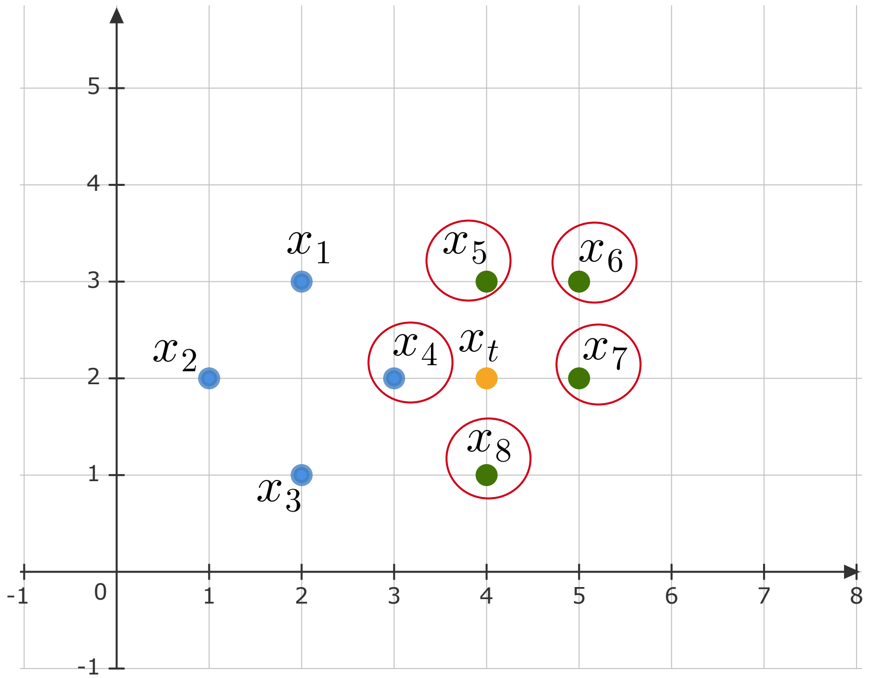


Here, we see that the point x_4 has a label of +1 and so we assign the same label to the test point

Point		Class	Distance
x_1	$(2, 3)$	+1	$\sqrt{5}$
x_2	$(1, 2)$	+1	3
x_3	$(2, 1)$	+1	$\sqrt{5}$
x_4	$(3, 2)$	+1	1
x_5	$(4, 3)$	-1	1
x_6	$(5, 3)$	-1	$\sqrt{2}$
x_7	$(5, 2)$	-1	1
x_8	$(4, 1)$	-1	1

Label Assignment

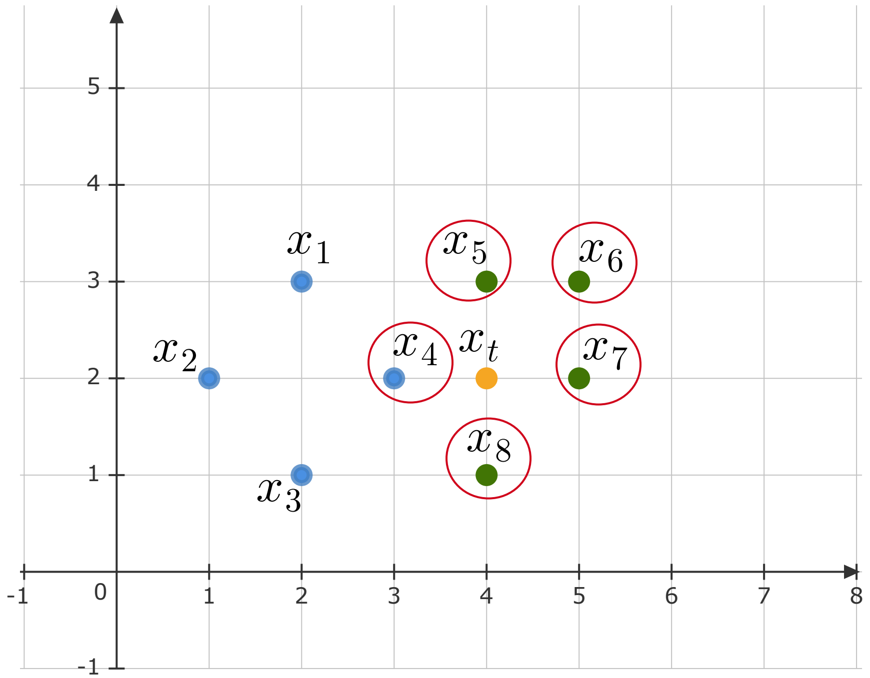
For K = 5, we take the five closest points and assign the class corresponding to the majority



Point		Class	Distance
x_1	$(2,3)$	+1	$\sqrt{5}$
x_2	$(1,2)$	+1	3
x_3	$(2,1)$	+1	$\sqrt{5}$
x_4	$(3,2)$	+1	1
x_5	$(4,3)$	-1	1
x_6	$(5,3)$	-1	$\sqrt{2}$
x_7	$(5,2)$	-1	1
x_8	$(4,1)$	-1	1

Label Assignment

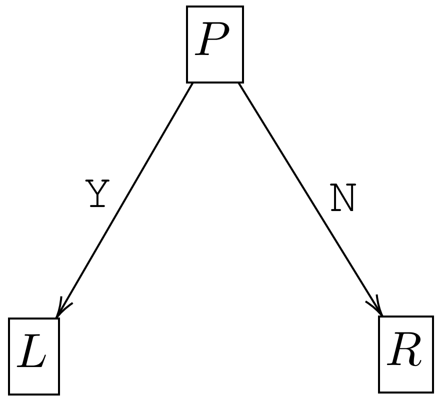
For K = 5, we take the five closest points and assign the class corresponding to the majority



Here, we see that four points have a label of -1 and one point has a label of +1. Since the majority has label -1, we assign the same to x_t

Point		Class	Distance
x_1	$(2, 3)$	+1	$\sqrt{5}$
x_2	$(1, 2)$	+1	3
x_3	$(2, 1)$	+1	$\sqrt{5}$
x_4	$(3, 2)$	+1	1
x_5	$(4, 3)$	-1	1
x_6	$(5, 3)$	-1	$\sqrt{2}$
x_7	$(5, 2)$	-1	1
x_8	$(4, 1)$	-1	1

Growing a Tree - Notations



- D: dataset at the parent
- $x_f < s$: question
- D_L and D_R: partitions
- p_P, p_L, p_R: proportions at P, L, R
- γ: proportions of points in L
- E_P: entropy of P
- E_L: entropy of L
- E_R: entropy of R
- IG: information gain

$$n_P = n_L + n_R \qquad \gamma = \frac{n_L}{n_P}$$

$$E = -p \log p - (1-p)\log(1-p)$$

$$IG = E_P - [\gamma E_L + (1-\gamma) E_R]$$

Observe the dataset

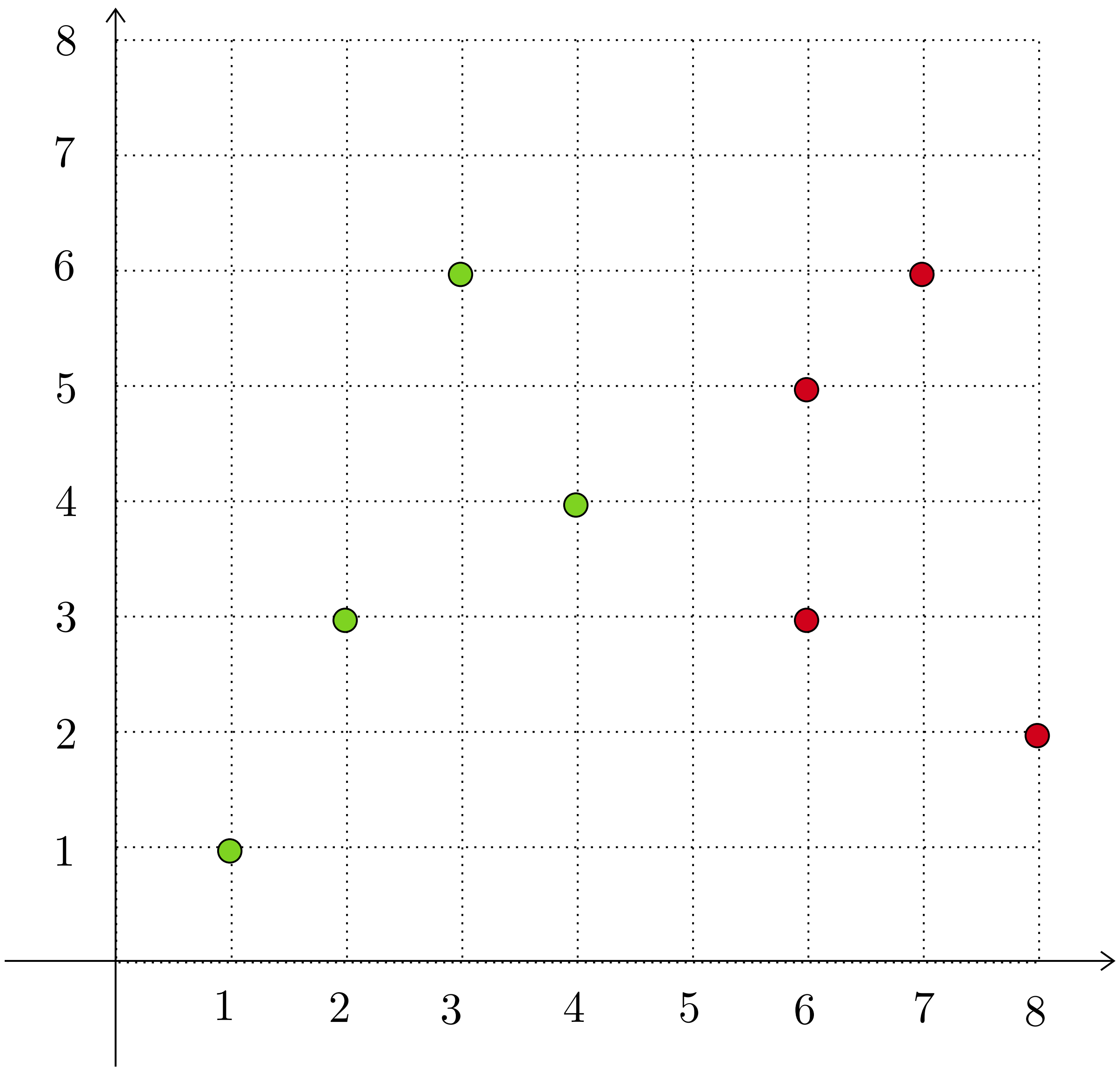
Consider a dataset $\{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)\}$ where $\mathbf{x}_i \in \mathbb{R}^d,\ y_i \in \{0, 1\}$

x_1	x_2	y
1	1	1
2	3	1
3	6	1
4	4	1
6	3	0
6	5	0
7	6	0
8	2	0

Observe the dataset

Consider a dataset $\{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)\}$ where $\mathbf{x}_i \in \mathbb{R}^d,\ y_i \in \{0, 1\}$

x_1	x_2	y
1	1	1
2	3	1
3	6	1
4	4	1
6	3	0
6	5	0
7	6	0
8	2	0

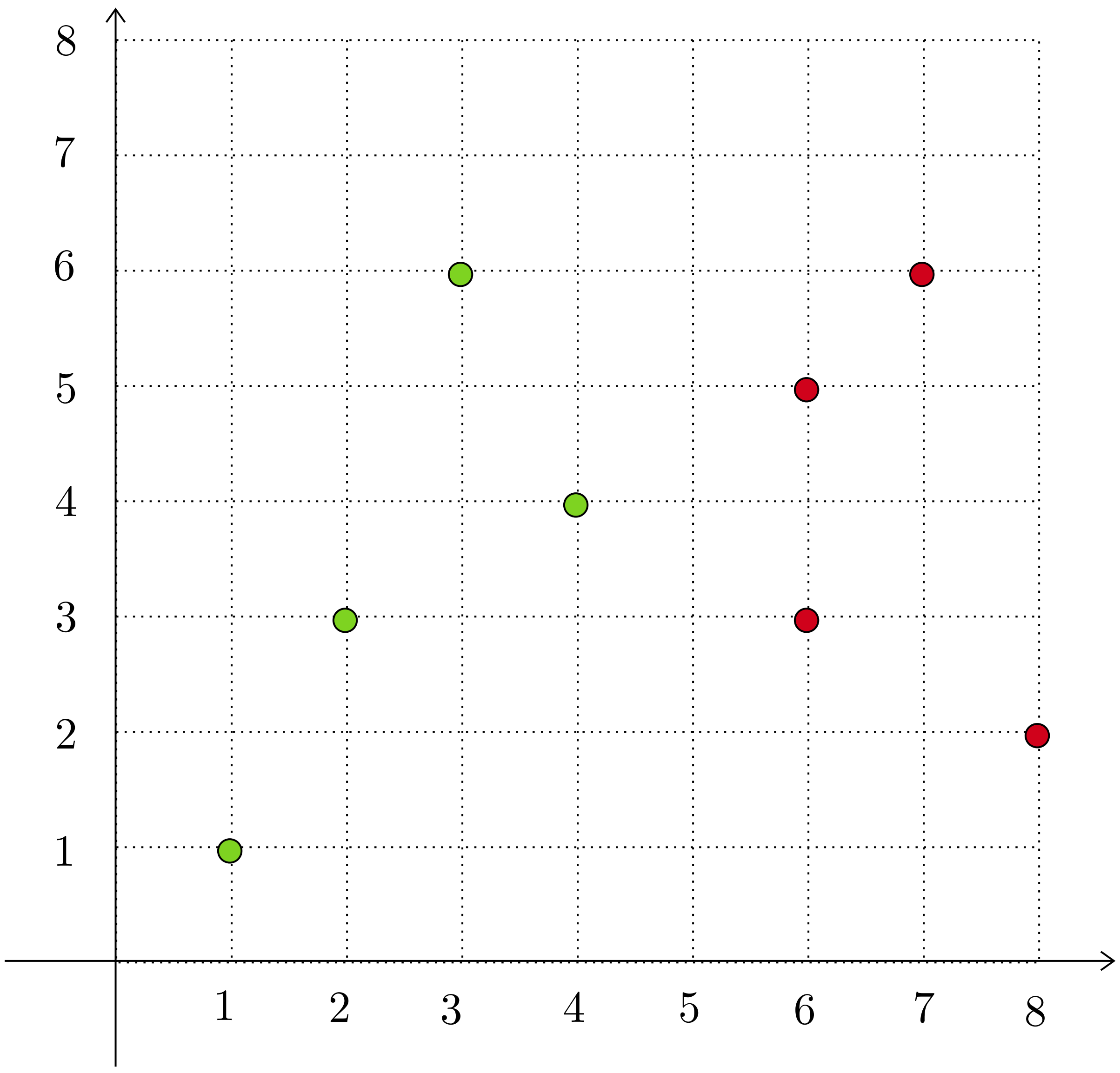


Observe the dataset

Consider a dataset $\{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)\}$ where $\mathbf{x}_i \in \mathbb{R}^d,\ y_i \in \{0, 1\}$

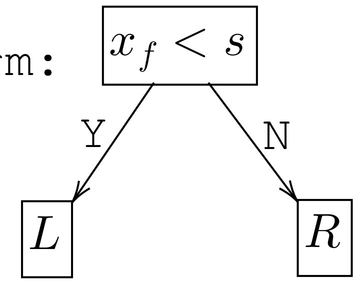
x_1	x_2	y
1	1	1
2	3	1
3	6	1
4	4	1
6	3	0
6	5	0
7	6	0
8	2	0

To grow the tree, we start with the node that leads to maximum information gain



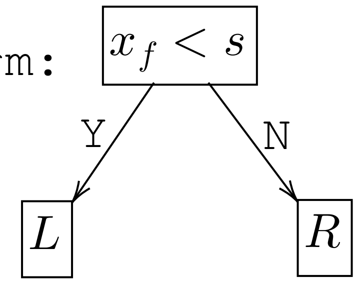
Best question

A question will be of the form:



Best question

A question will be of the form:



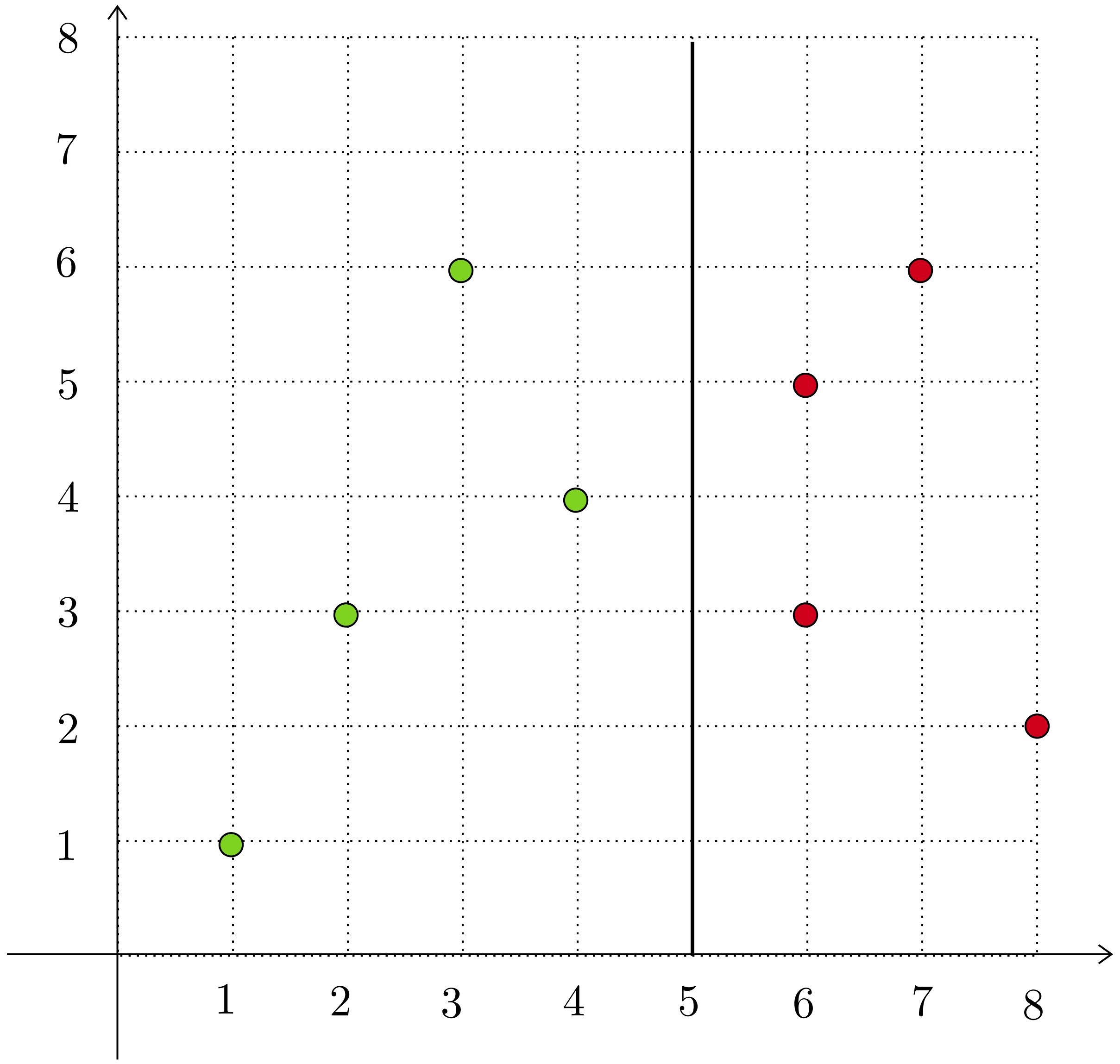
The IG for all questions is shown below and we find that $x_1 < 5$ has the most IG

$IG(f, s)$

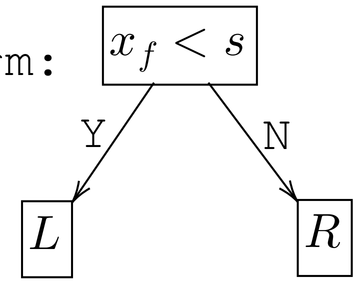
s	$IG(1, s)$
1.5	0.138
2.5	0.311
3.5	0.549
5	**1**
6.5	0.311
7.5	0.138

s	$IG(2, s)$
1.5	0.138
2.5	0
3.5	0
4.5	0.049
5.5	0

Best question



A question will be of the form:



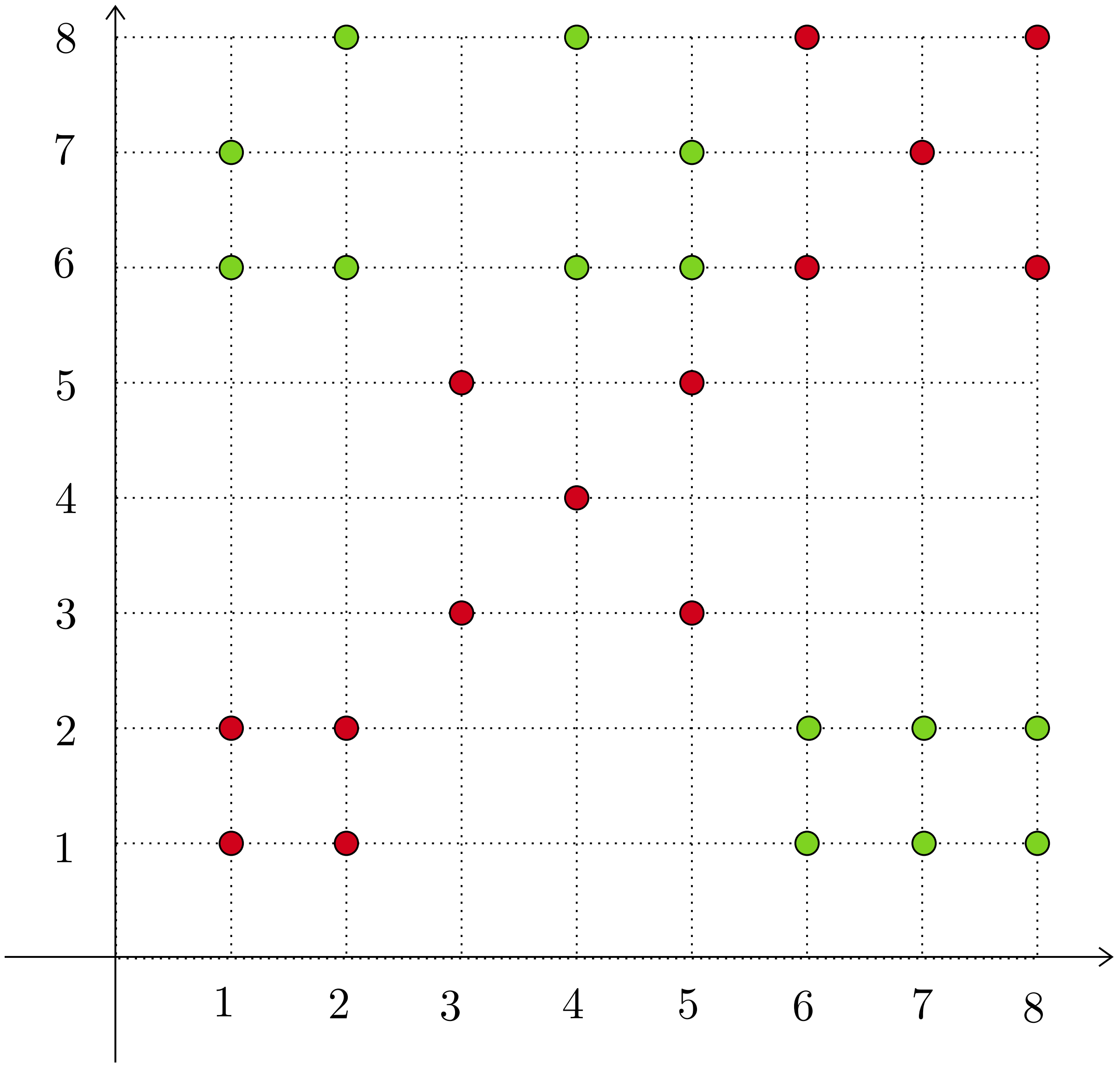
The IG for all questions is shown below and we find that $x_1 < 5$ has the most IG

$IG(f, s)$

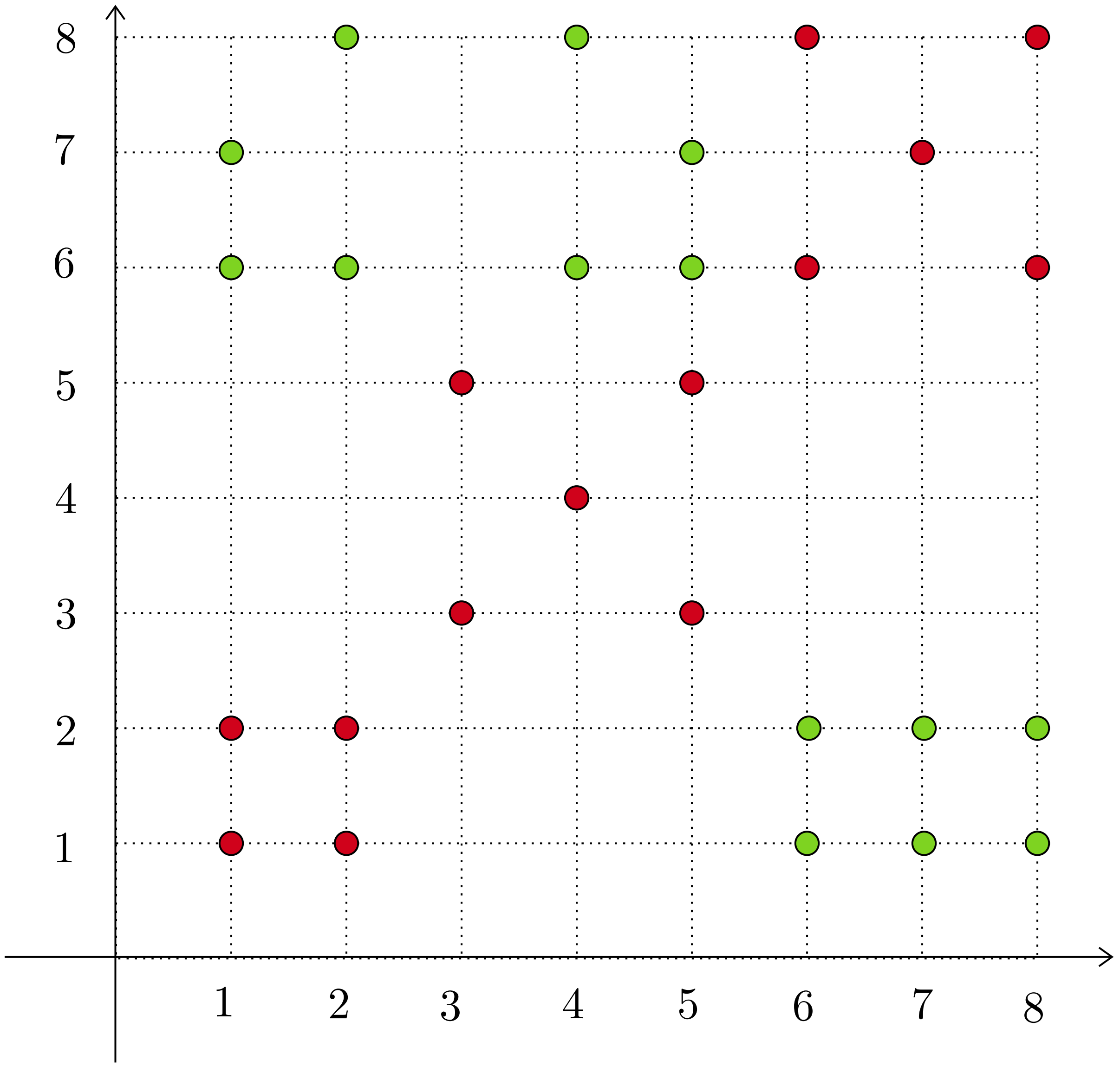
s	$IG(1, s)$
1.5	0.138
2.5	0.311
3.5	0.549
5	**1**
6.5	0.311
7.5	0.138

s	$IG(2, s)$
1.5	0.138
2.5	0
3.5	0
4.5	0.049
5.5	0

Growing a Tree - Example

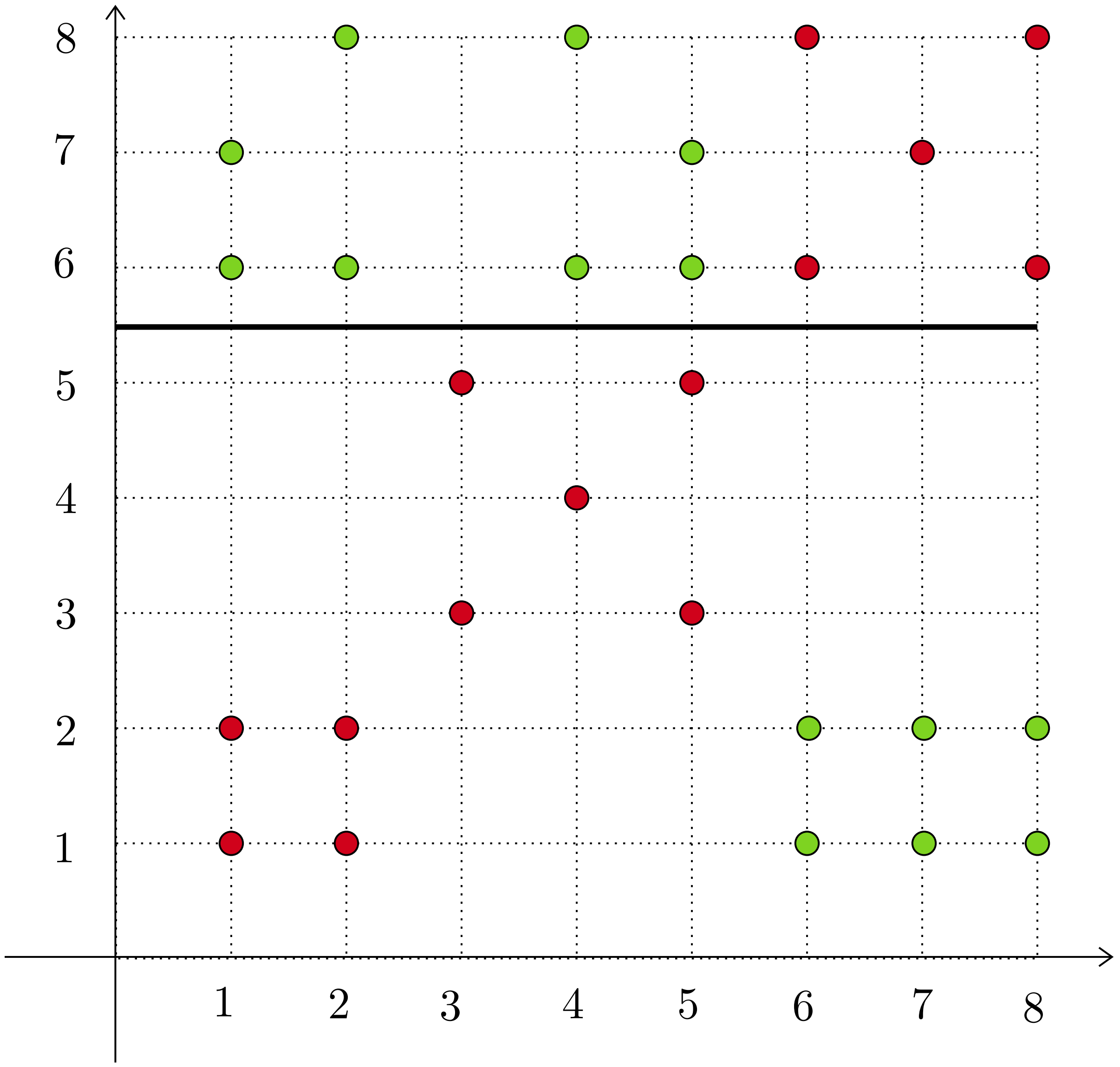


Growing a Tree

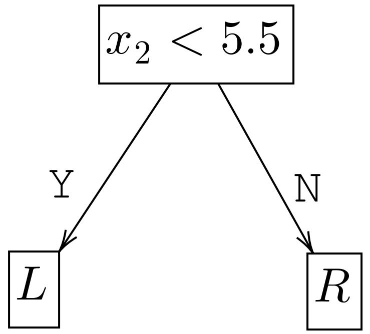


Consider the following dataset. We will now build a tree for this dataset

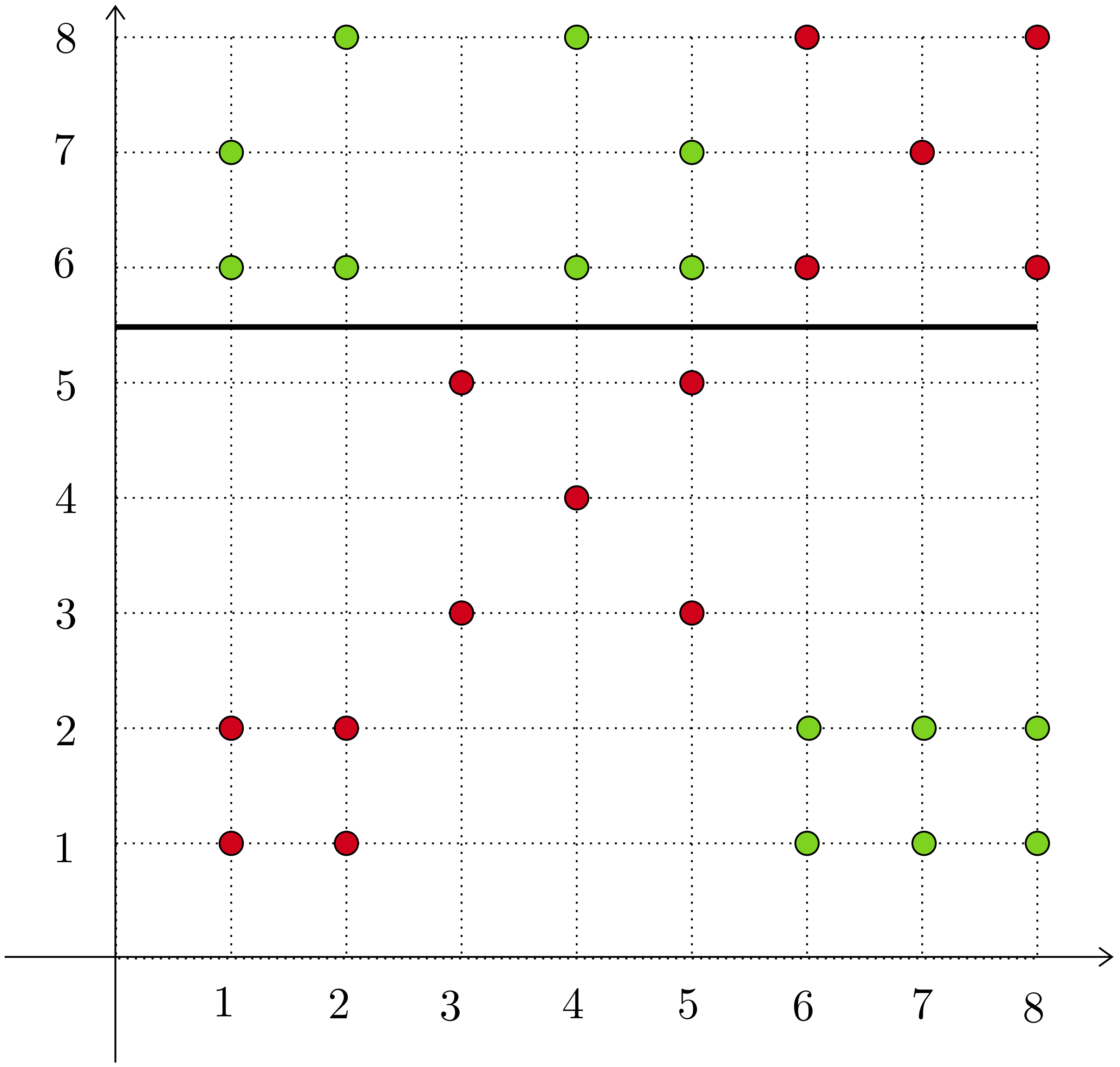
Growing a Tree



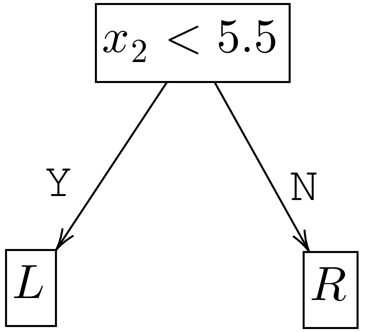
Consider the following dataset. We will now build a tree for this dataset



Growing a Tree



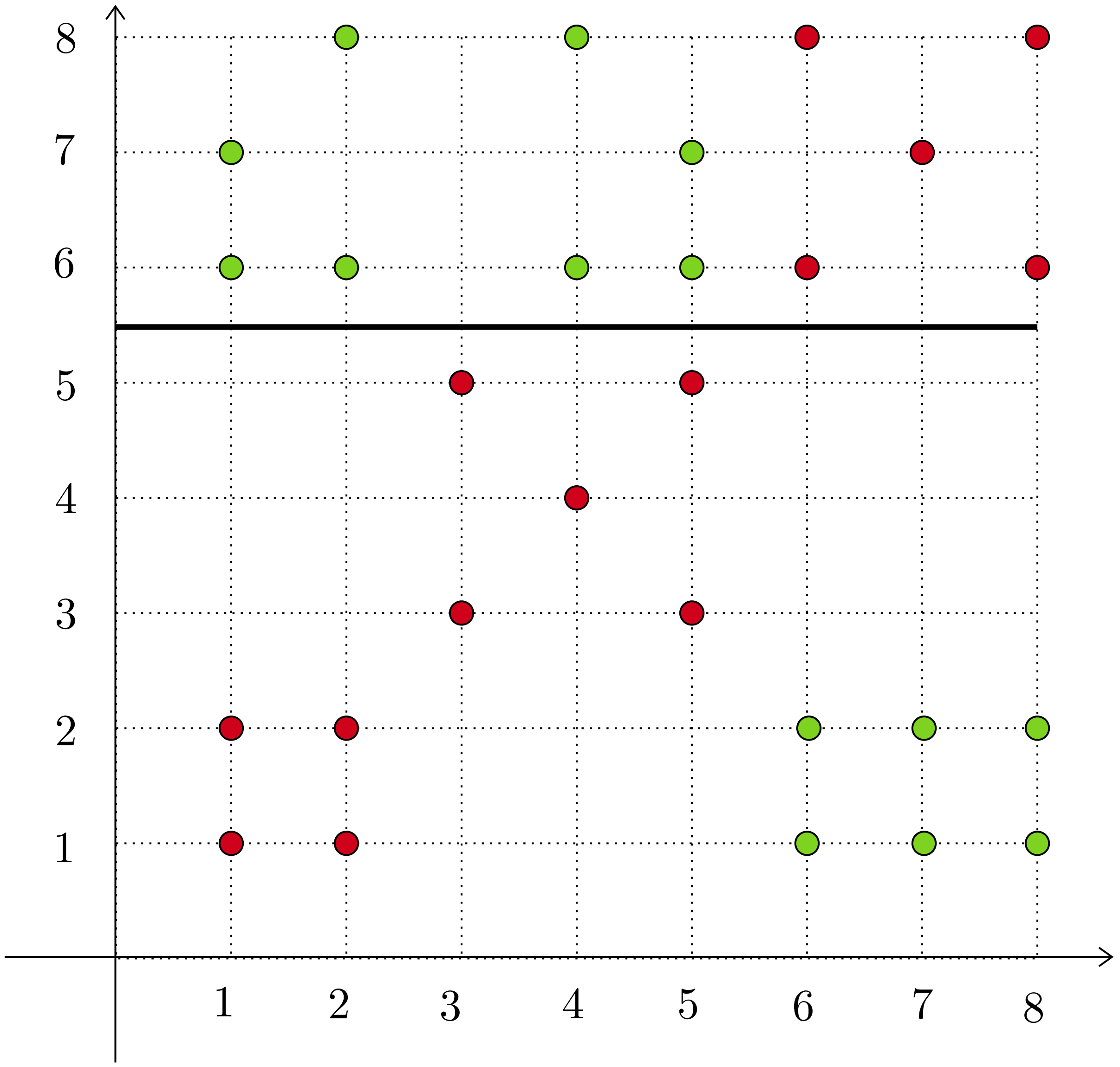
Consider the following dataset. We will now build a tree for this dataset



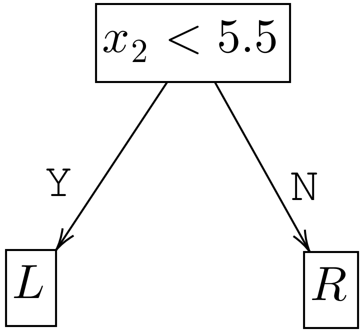
D_L= {G:6,R:9}

D_R= {G:8,R:5}

Growing a Tree



Consider the following dataset. We will now build a tree for this dataset



D_L = {G:6,R:9}

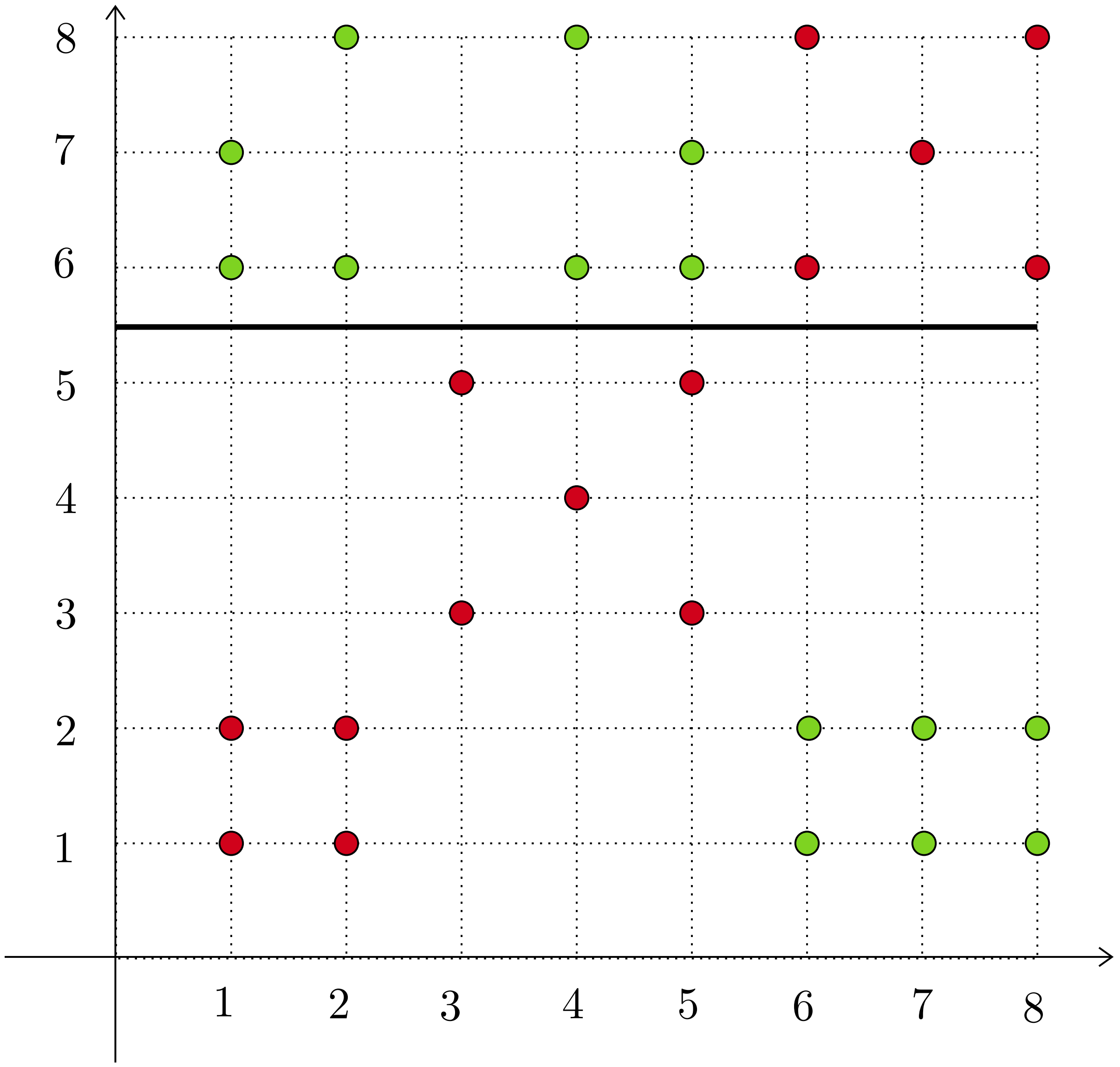
D_R = {G:8,R:5}

$$p_P = (14 / 28) = 0.5$$

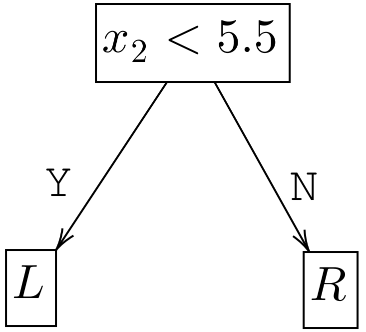
$$p_L = (6 / 15) = 0.4$$

$$p_R = (8 / 13) = 0.615$$

Growing a Tree



Consider the following dataset. We will now build a tree for this dataset



$D_L =$ {G:6,R:9}

$D_R =$ {G:8,R:5}

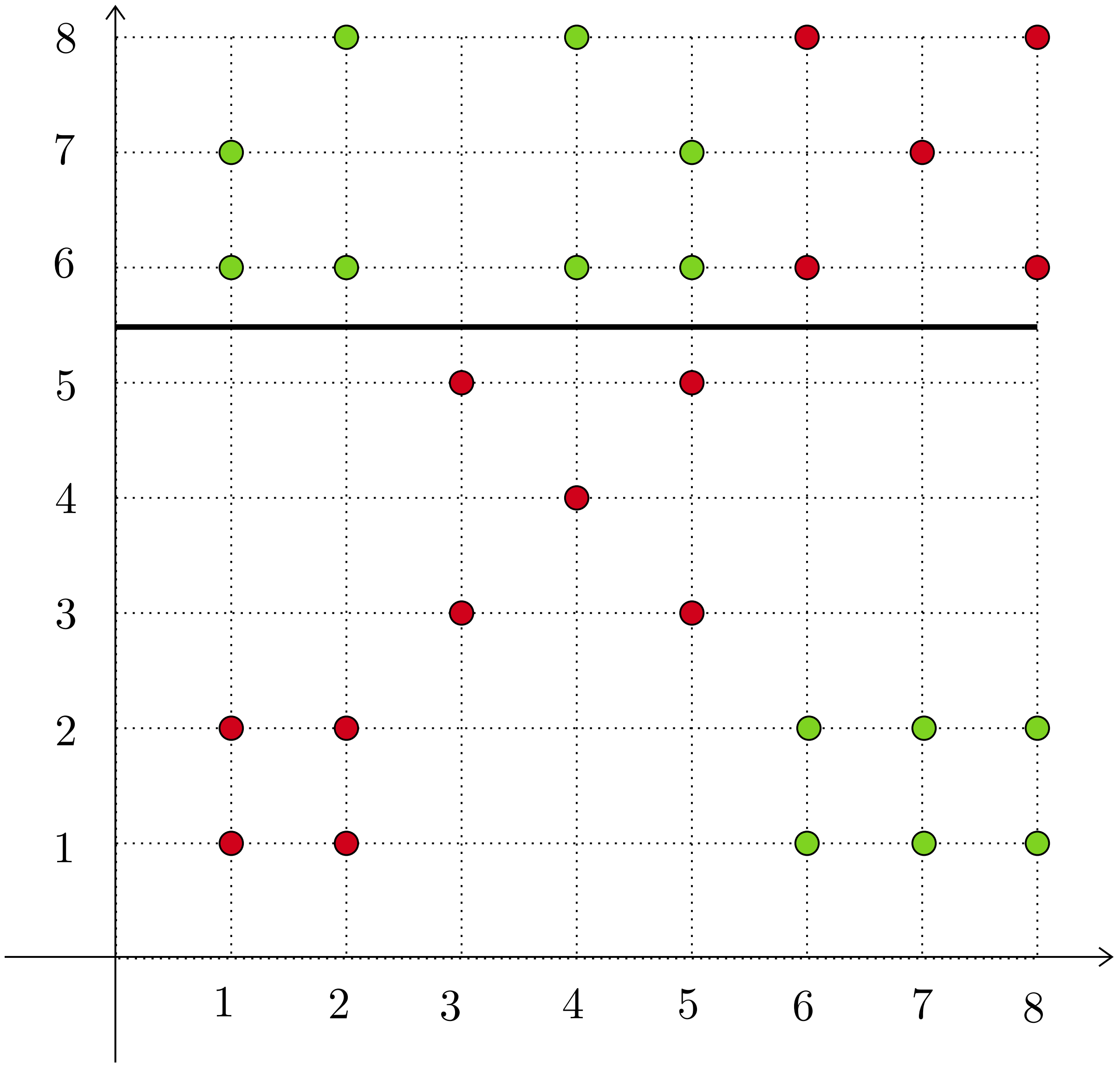
$$p_P = (14 / 28) = 0.5$$

$$p_L = (6 / 15) = 0.4$$

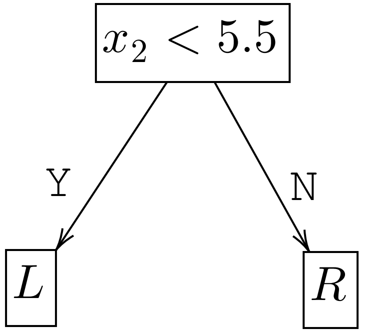
$$p_R = (8 / 13) = 0.615$$

$$E = -p \log p - (1-p)\log(1-p)$$

Growing a Tree



Consider the following dataset. We will now build a tree for this dataset



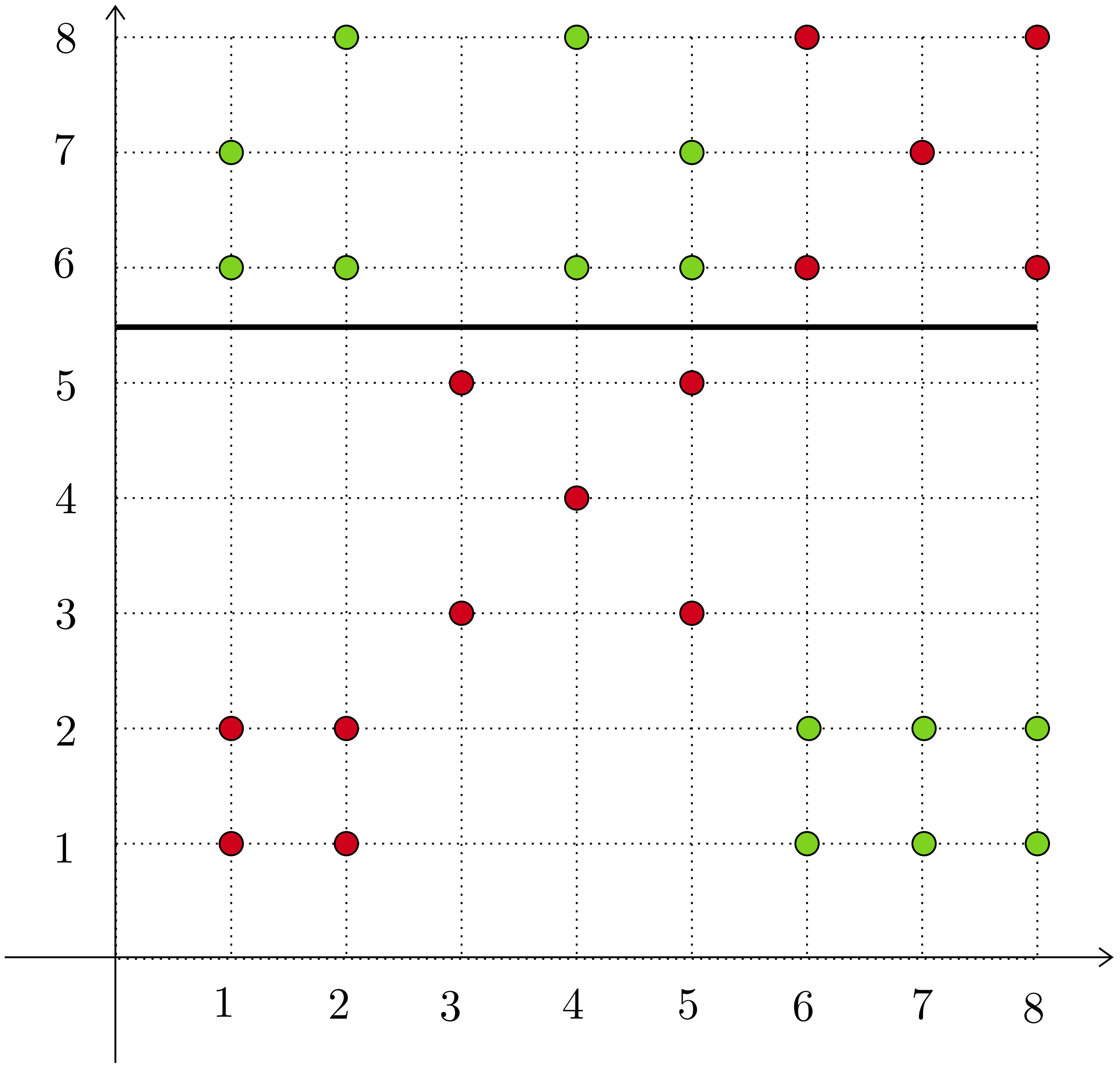
$D_L =$ {G:6,R:9} $\qquad$ $D_R =$ {G:8,R:5}

$$p_P = (14 \,/\, 28) = 0.5 \qquad p_L = (6 \,/\, 15) = 0.4 \qquad p_R = (8 \,/\, 13) = 0.615$$

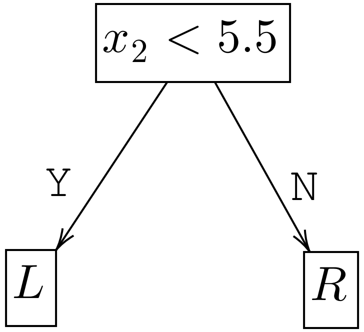
$$E = -p \log p - (1-p)\log(1-p)$$

$$\begin{aligned} E_P &= -0.5log(0.5) - (1-0.5)log(1-0.5) \\ &= -0.5log(0.5) - (0.5)log(0.5) \\ &= 1 \end{aligned}$$

Growing a Tree



Consider the following dataset. We will now build a tree for this dataset

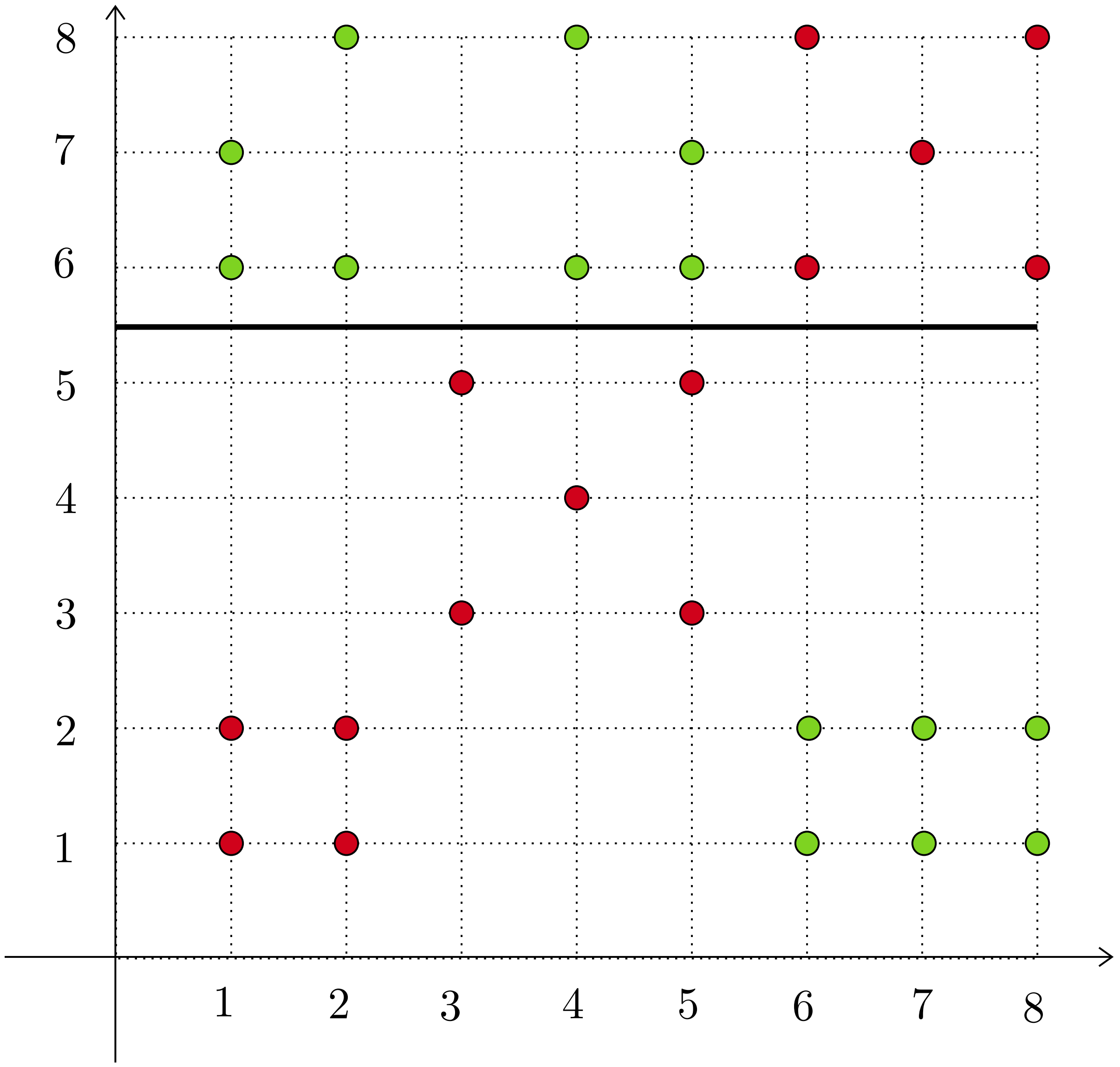


$D_L =$ {G:6,R:9} $D_R =$ {G:8,R:5}

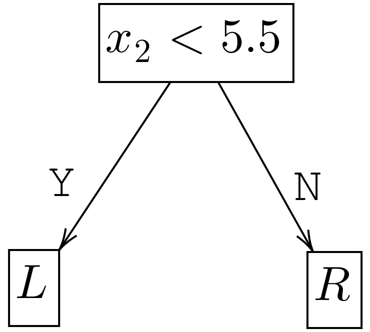
$p_P = (14 \,/\, 28) = 0.5$ $p_L = (6 \,/\, 15) = 0.4$ $p_R = (8 \,/\, 13) = 0.615$

$E_P = 1$ $E_L = 0.970$ $E_R = 0.961$

Growing a Tree



Consider the following dataset. We will now build a tree for this dataset



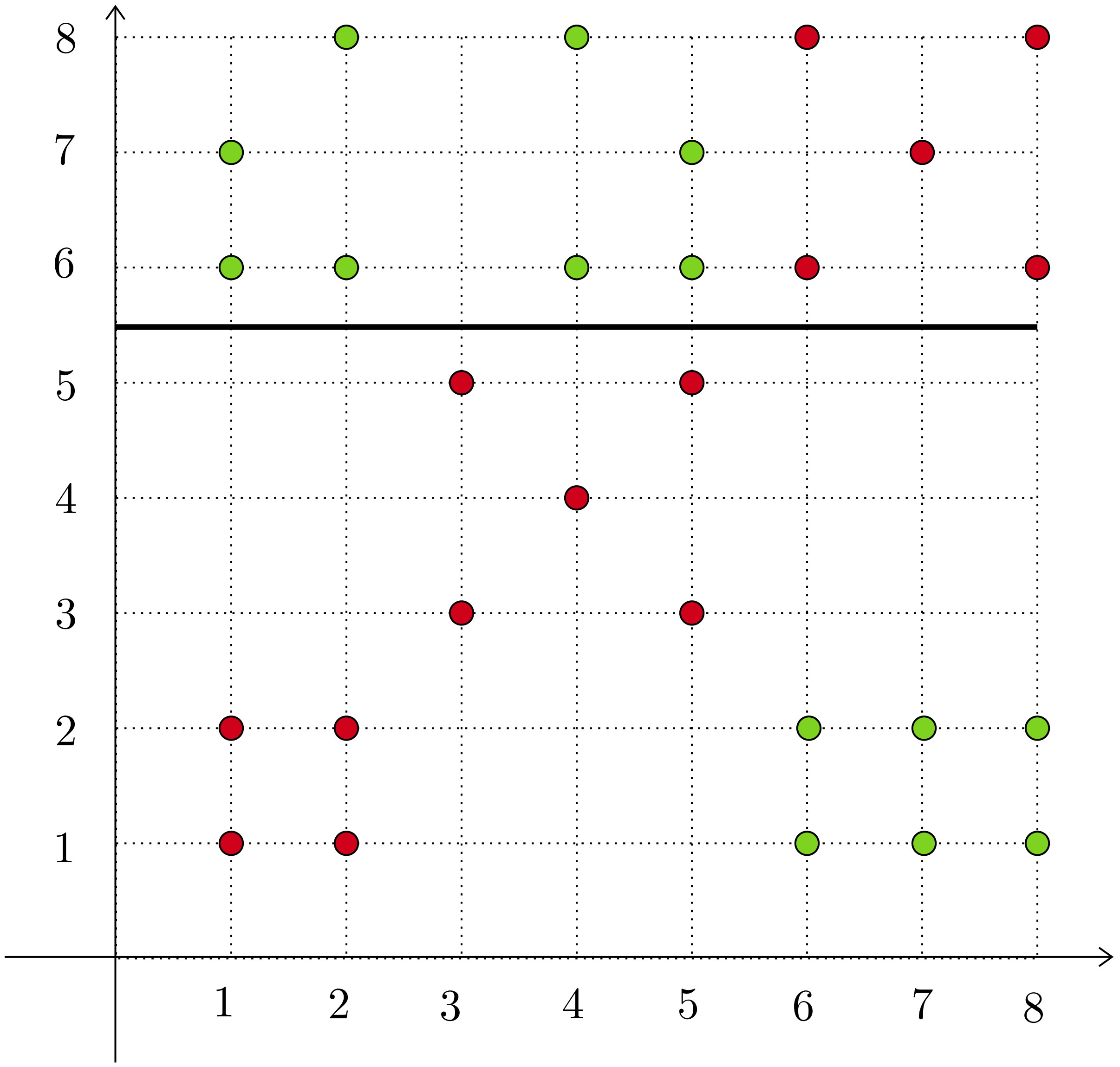
$D_L =$ {G:6,R:9} $D_R =$ {G:8,R:5}

$p_P = (14\,/\,28) = 0.5$ $p_L = (6\,/\,15) = 0.4$ $p_R = (8\,/\,13) = 0.615$

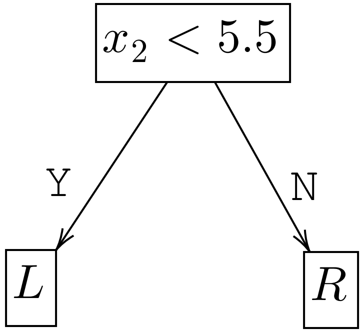
$E_P = 1$ $E_L = 0.970$ $E_R = 0.961$

$$IG = E_P - [\gamma E_L + (1-\gamma) E_R]$$

Growing a Tree



Consider the following dataset. We will now build a tree for this dataset



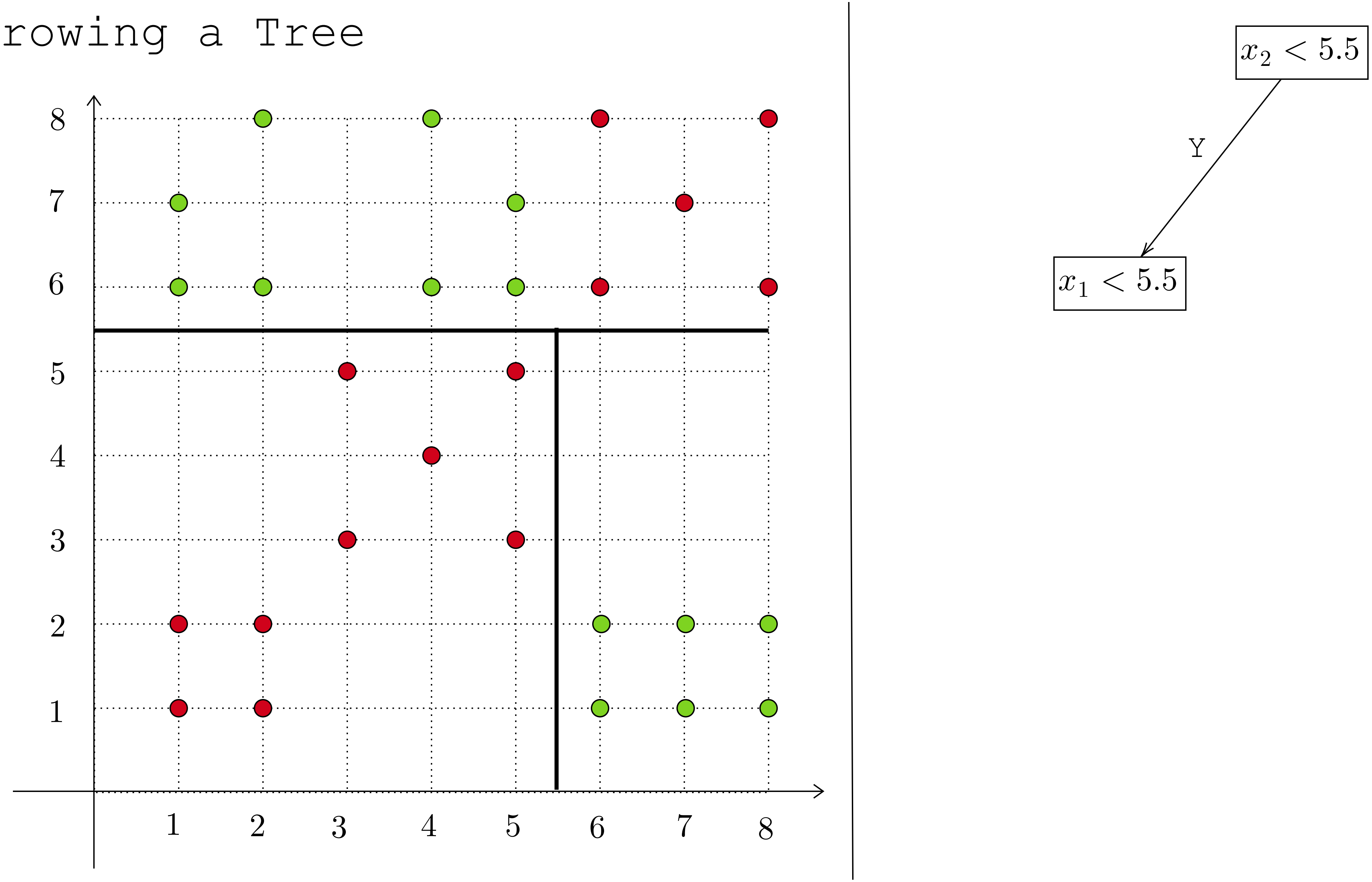
$D_L =$ {G:6,R:9} $D_R =$ {G:8,R:5}

$p_P = (14\,/\,28) = 0.5$ $p_L = (6\,/\,15) = 0.4$ $p_R = (8\,/\,13) = 0.615$

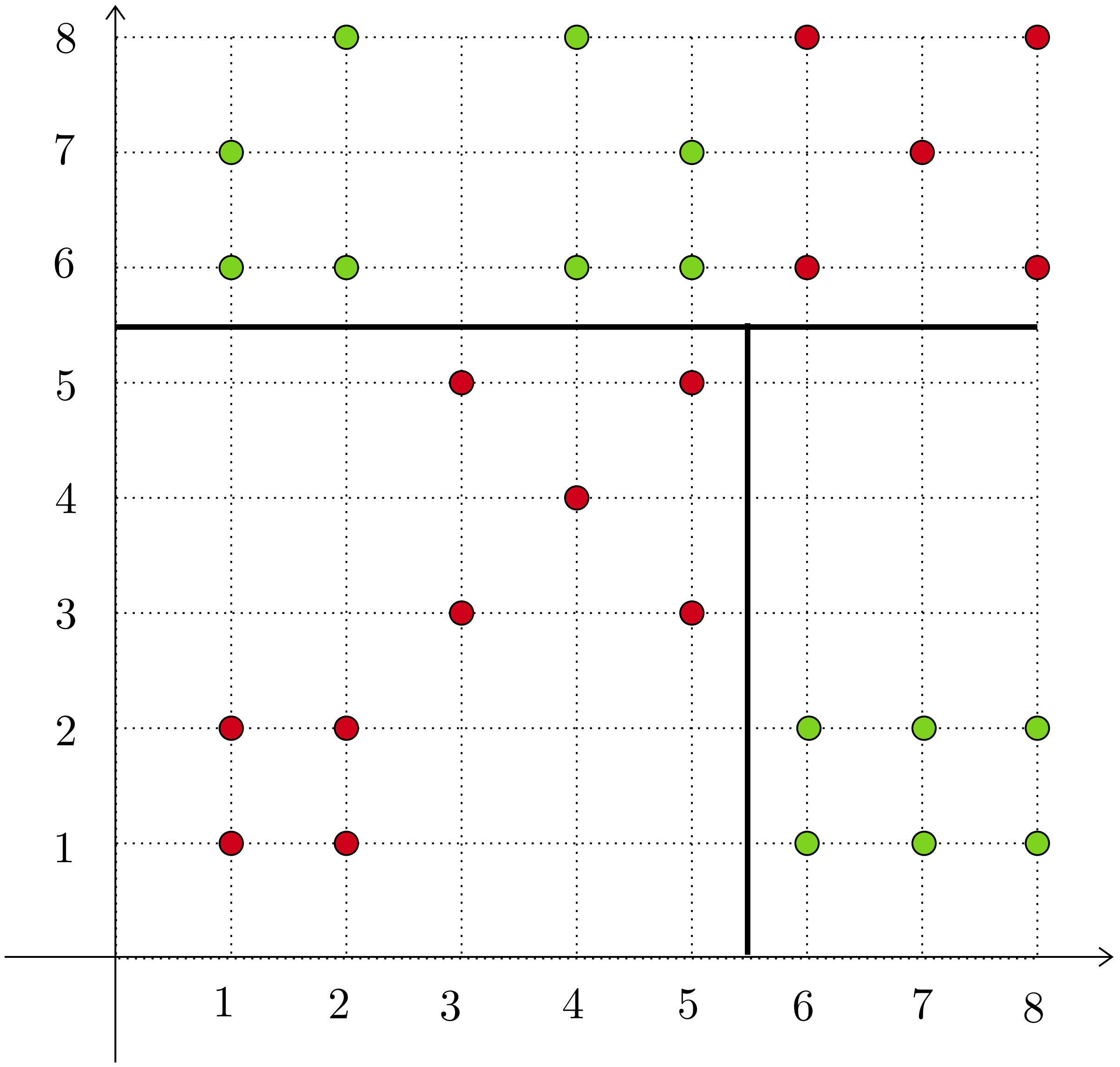
$E_P = 1$ $E_L = 0.970$ $E_R = 0.961$

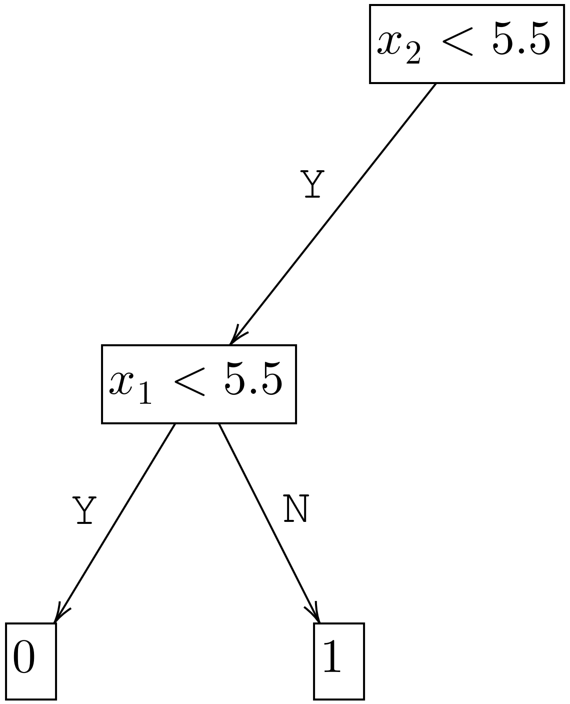
$$\begin{aligned} IG &= E_P - [\gamma E_L + (1-\gamma)E_R] \\ &= 1 - 0.535(0.970) + (0.465)(0.961) \\ &= 0.034 \end{aligned}$$

Growing a Tree

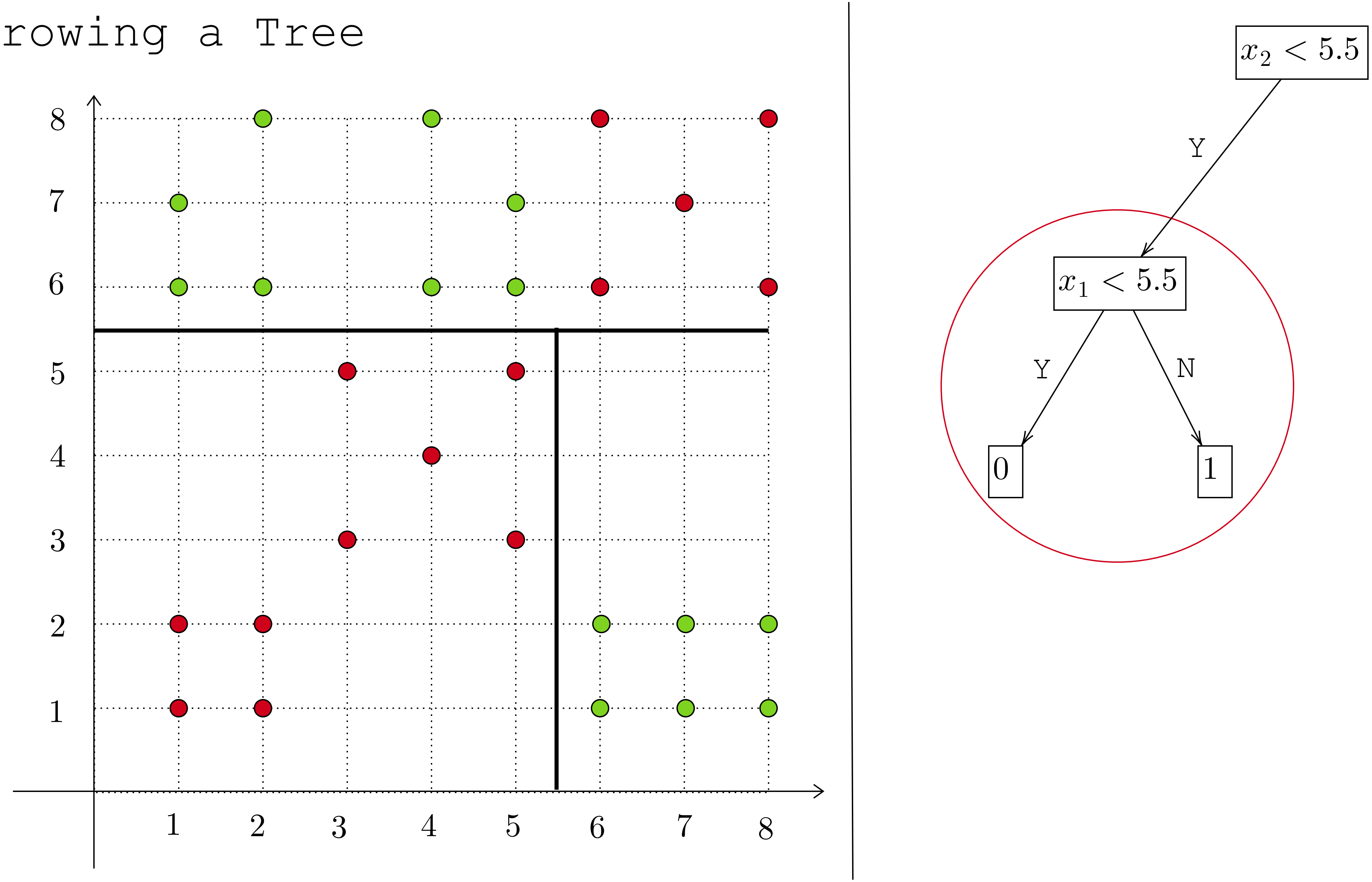


Growing a Tree

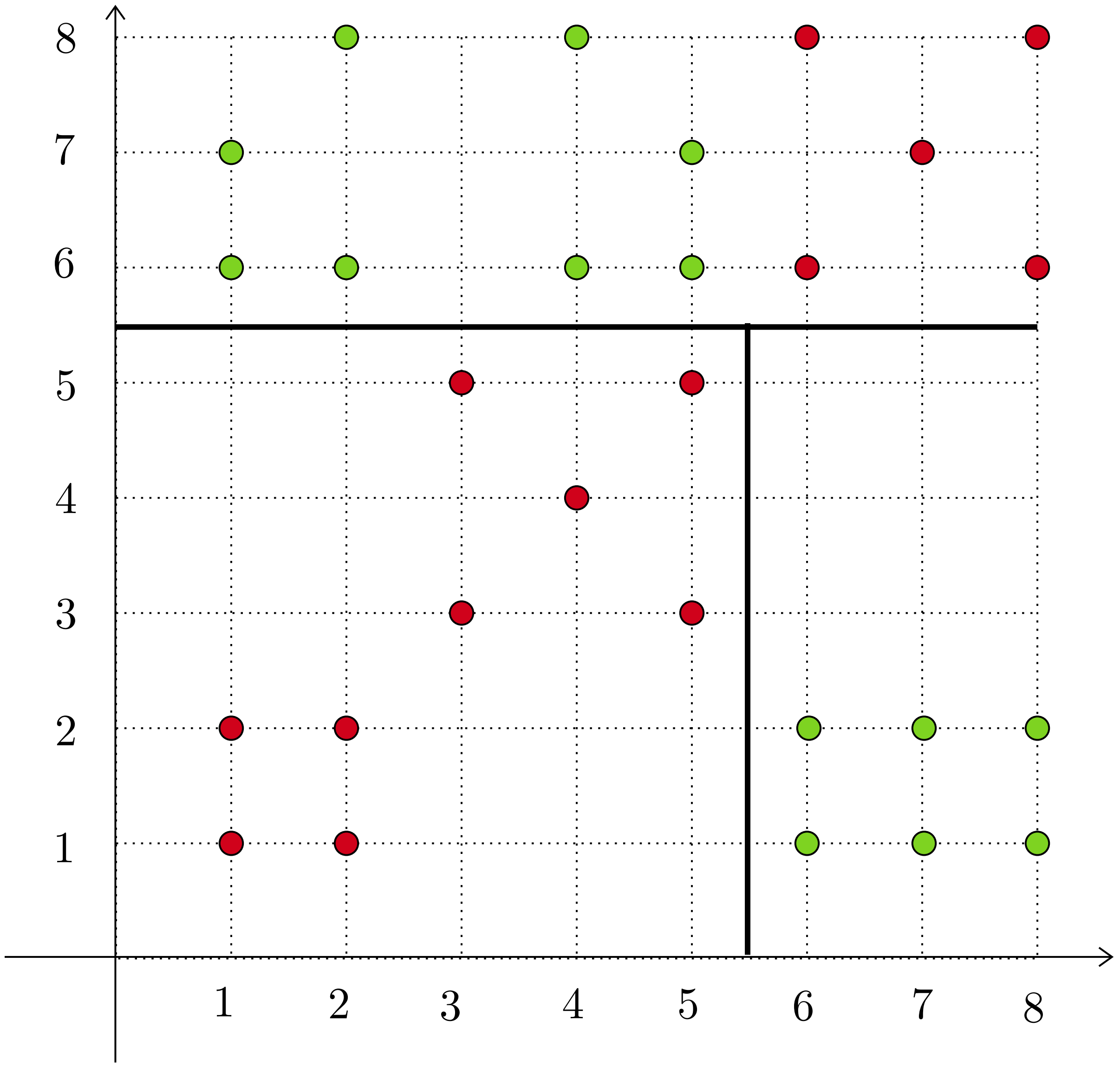


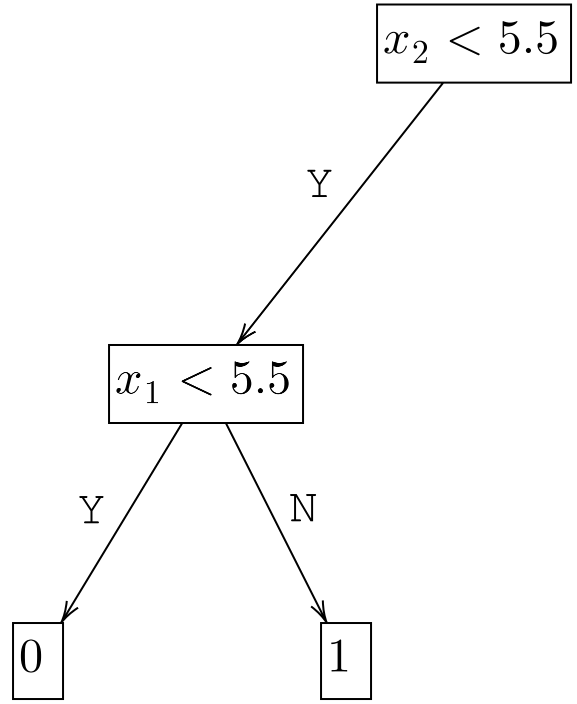


Growing a Tree



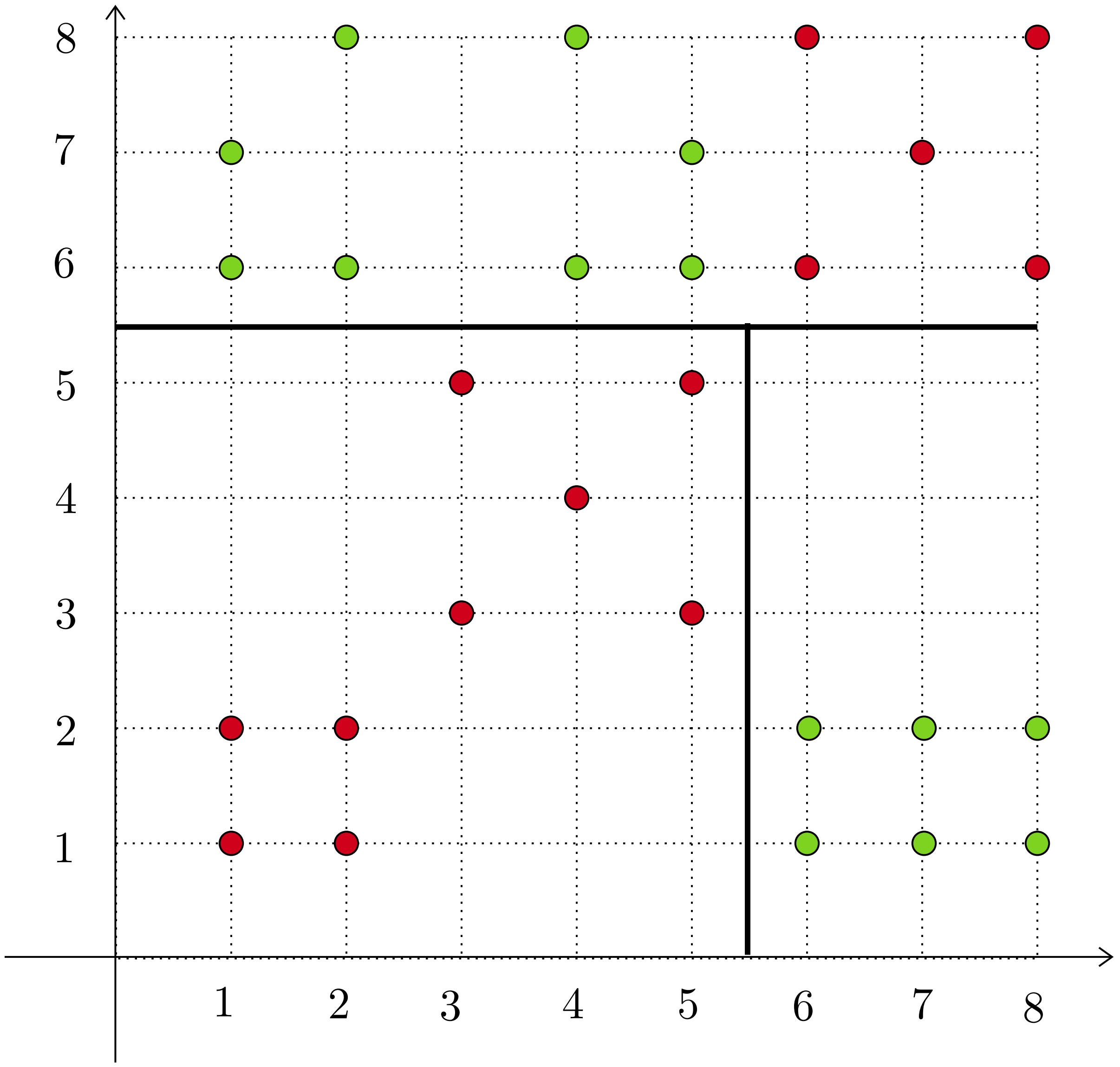
Growing a Tree

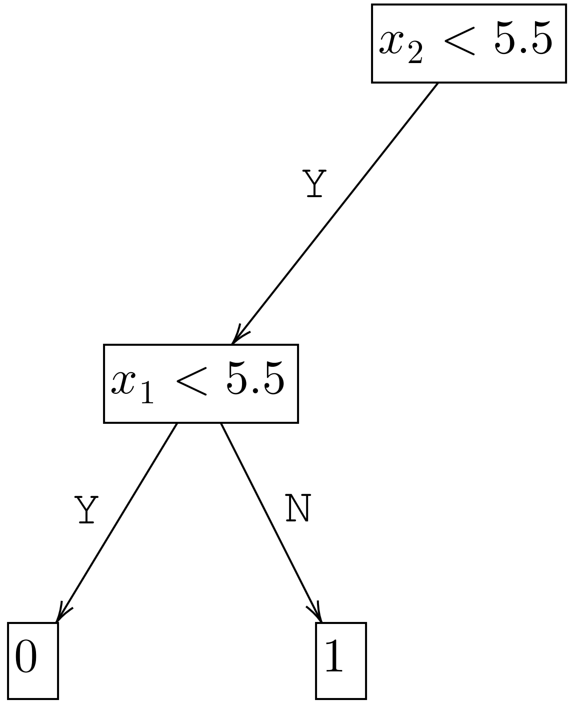




$D_L =$ {G:0,R:9}
$D_R =$ {G:6,R:0}

Growing a Tree





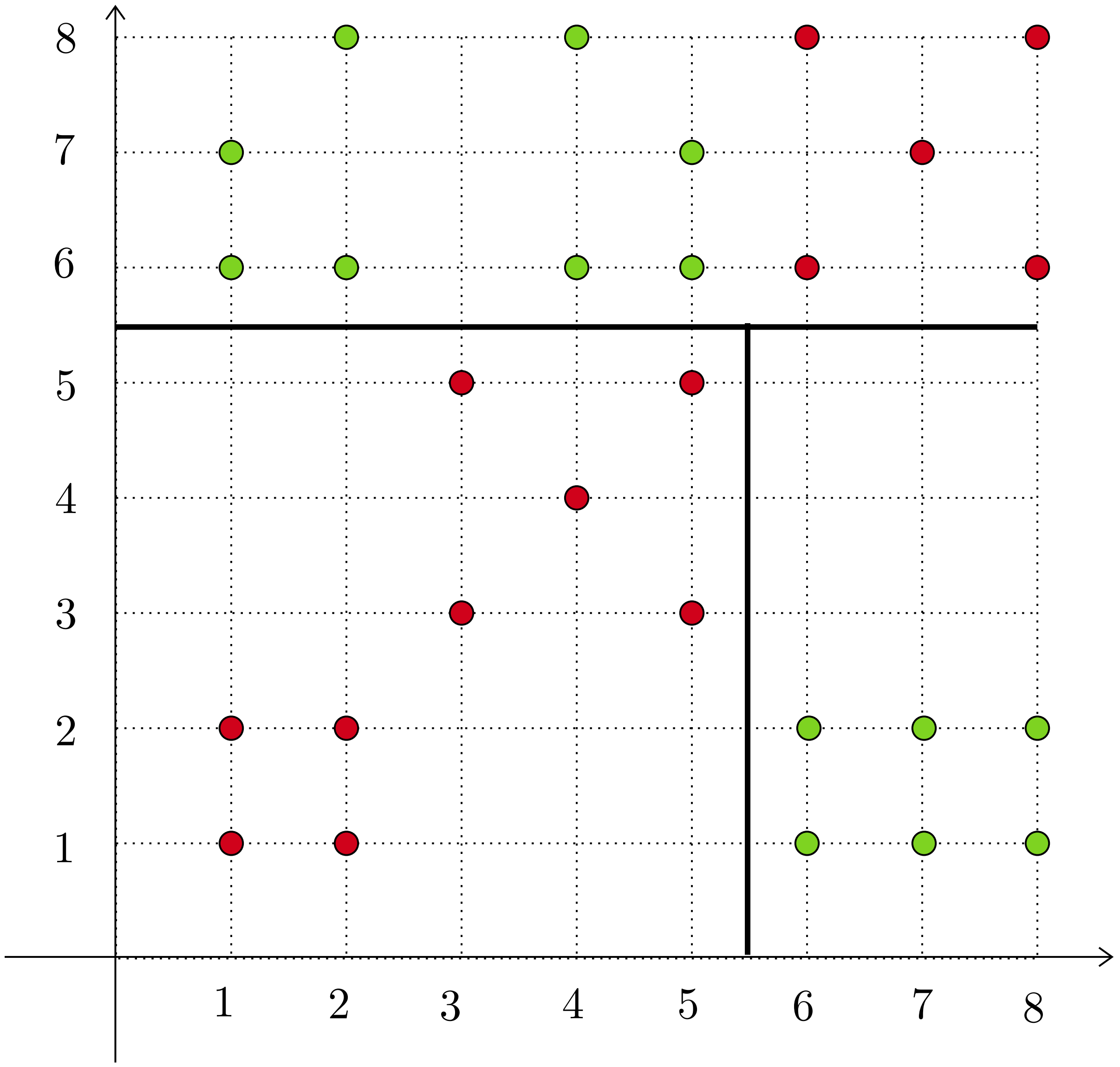
$D_L =$ {G:0,R:9}
$D_R =$ {G:6,R:0}

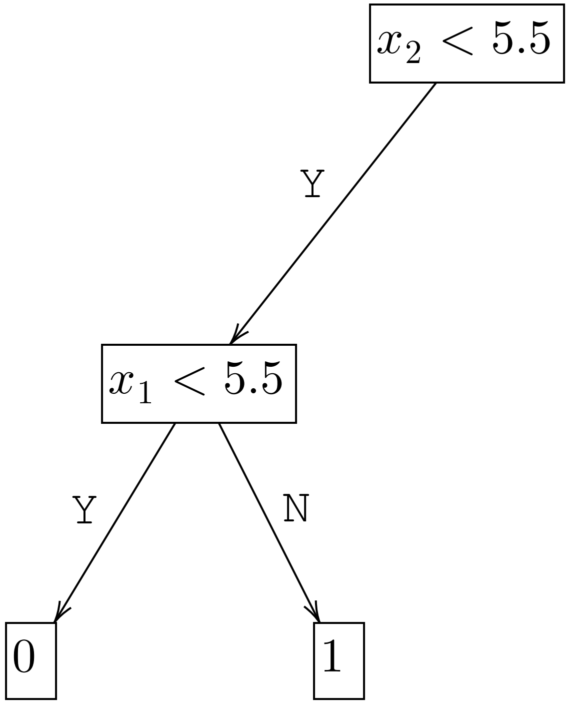
$p_P = (6 \,/\, 15) = 0.4$

$p_L = (0 \,/\, 9) = 0$

$p_R = (6 \,/\, 6) = 1$

Growing a Tree



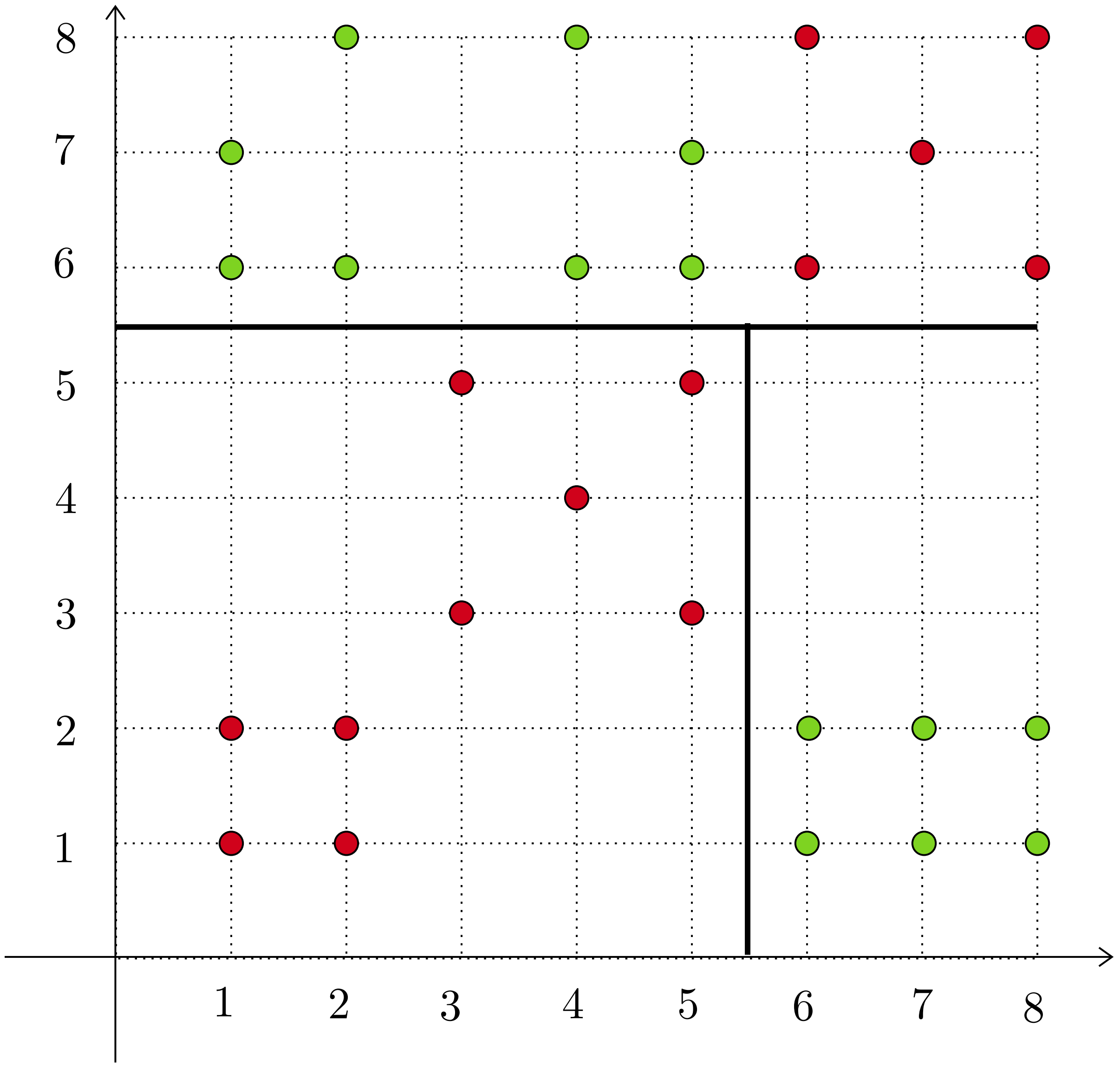


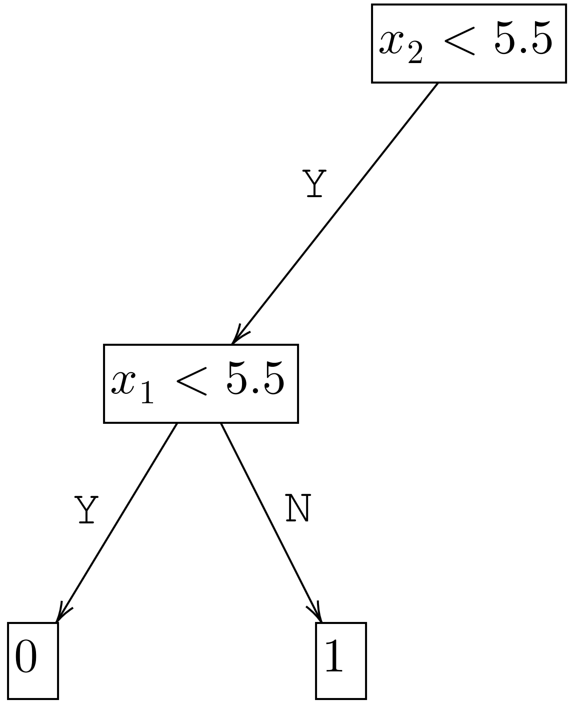
$D_L =$ {G:0,R:9}
$D_R =$ {G:6,R:0}

$$p_P = (6\,/\,15) = 0.4 \qquad p_L = (0\,/\,9) = 0 \qquad p_R = (6\,/\,6) = 1$$

$$E_P = 0.97 \qquad E_L = 0 \qquad E_R = 0$$

Growing a Tree





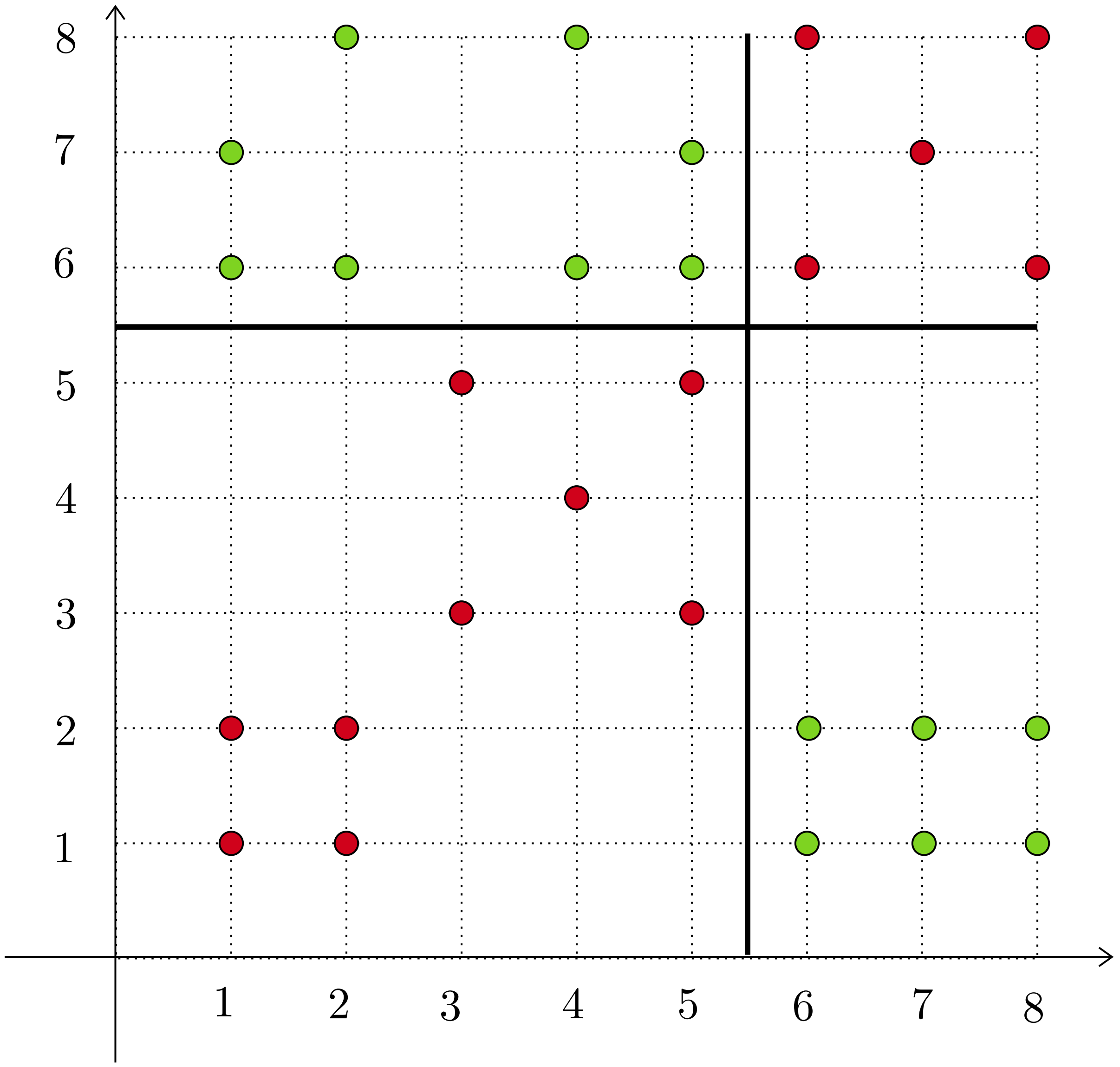
$D_L =$ {G:0,R:9}
$D_R =$ {G:6,R:0}

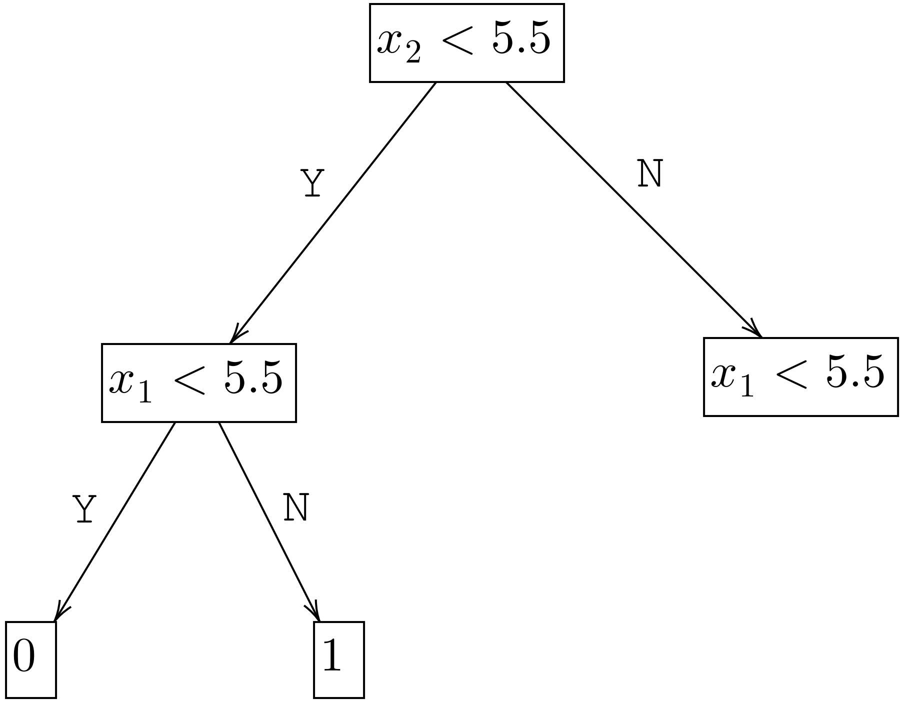
$$p_P = (6\,/\,15) = 0.4 \qquad p_L = (0\,/\,9) = 0 \qquad p_R = (6\,/\,6) = 1$$

$$E_P = 0.97 \qquad E_L = 0 \qquad E_R = 0$$

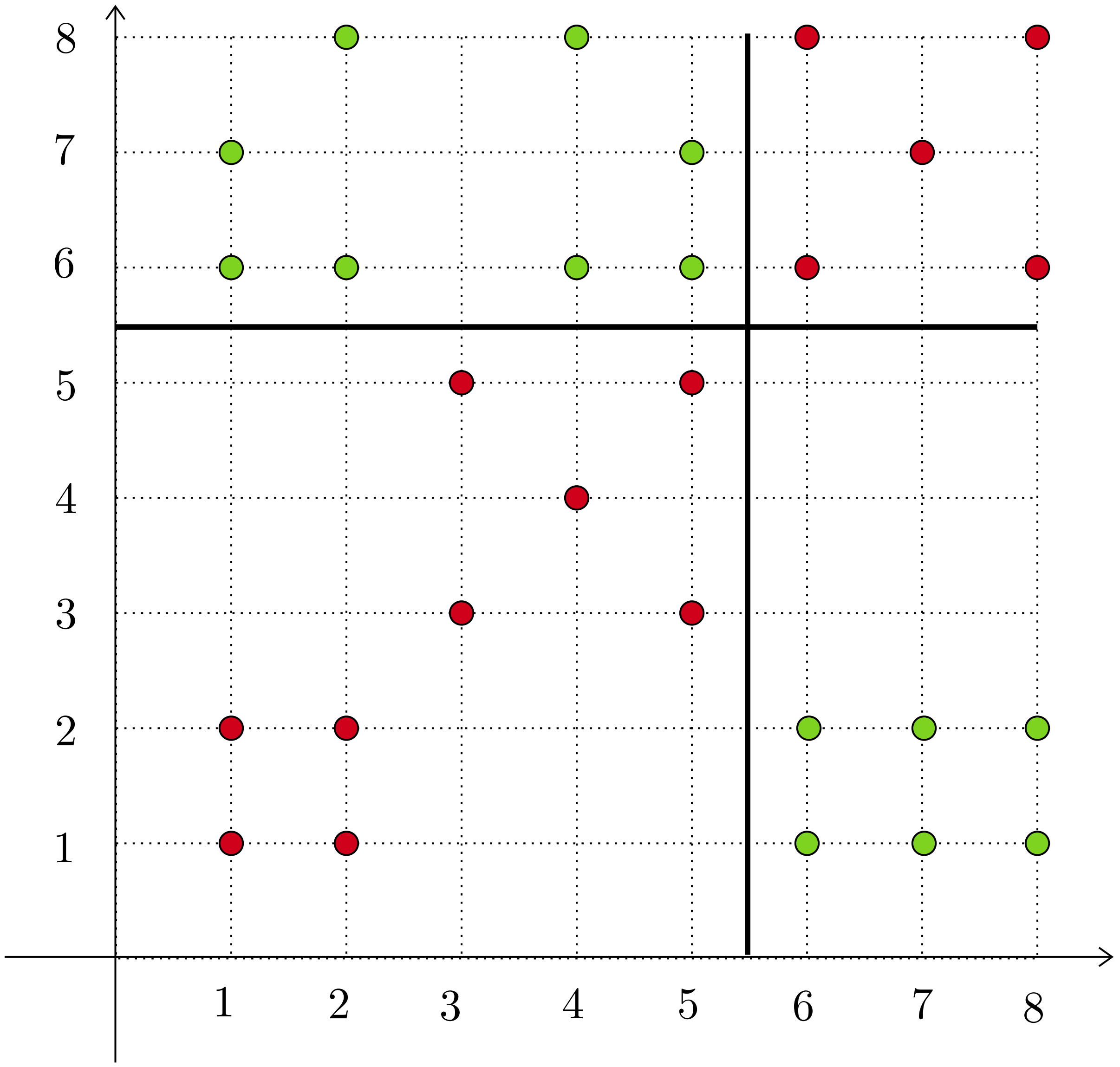
$$\begin{aligned} IG &= E_P - [\gamma E_L + (1-\gamma)E_R] \\ &= 0.97 - 0.6(0) + (0.4)(0) \\ &= 0.97 \end{aligned}$$

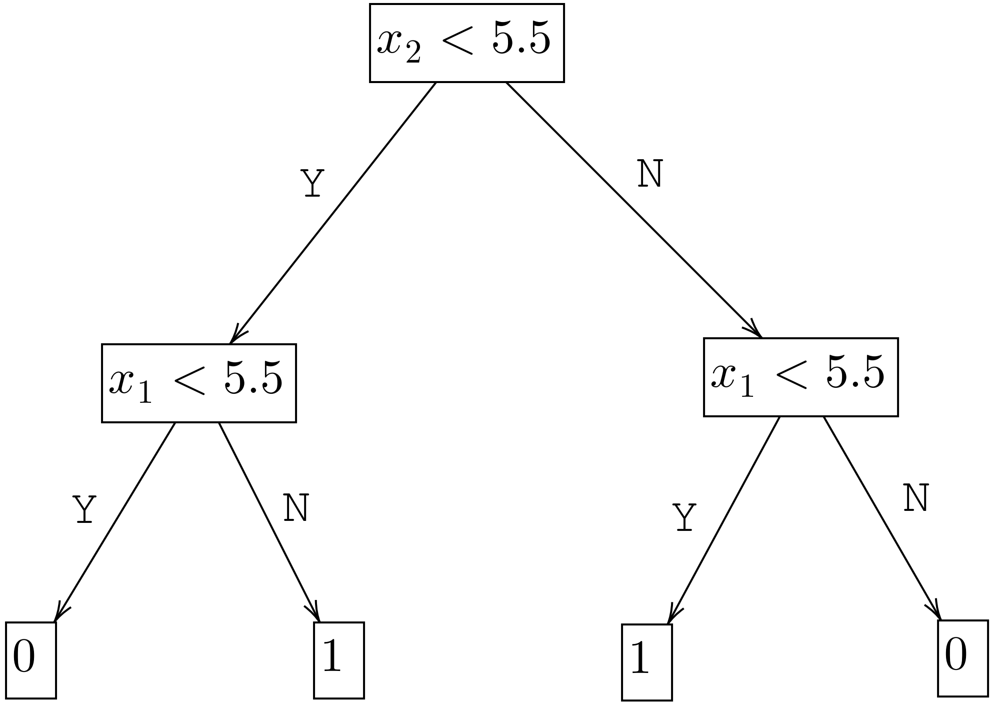
Growing a Tree



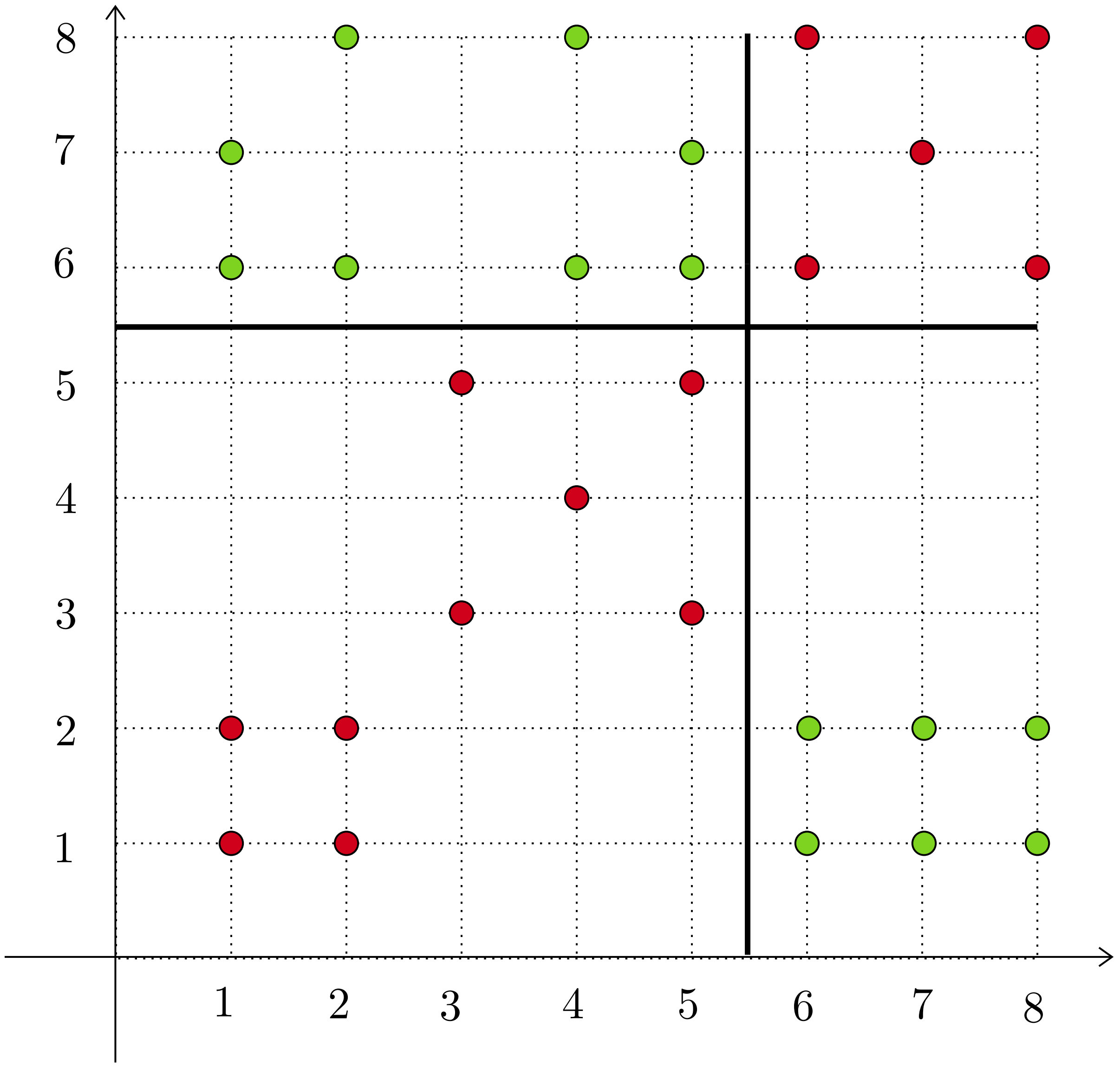


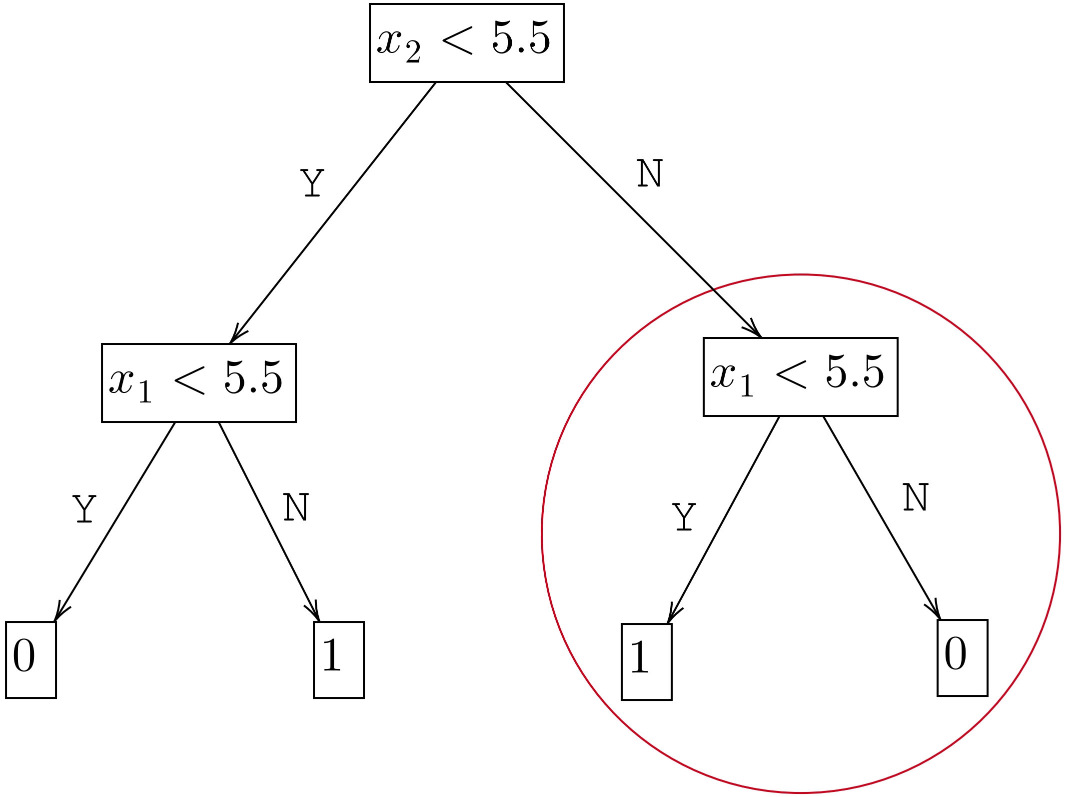
Growing a Tree



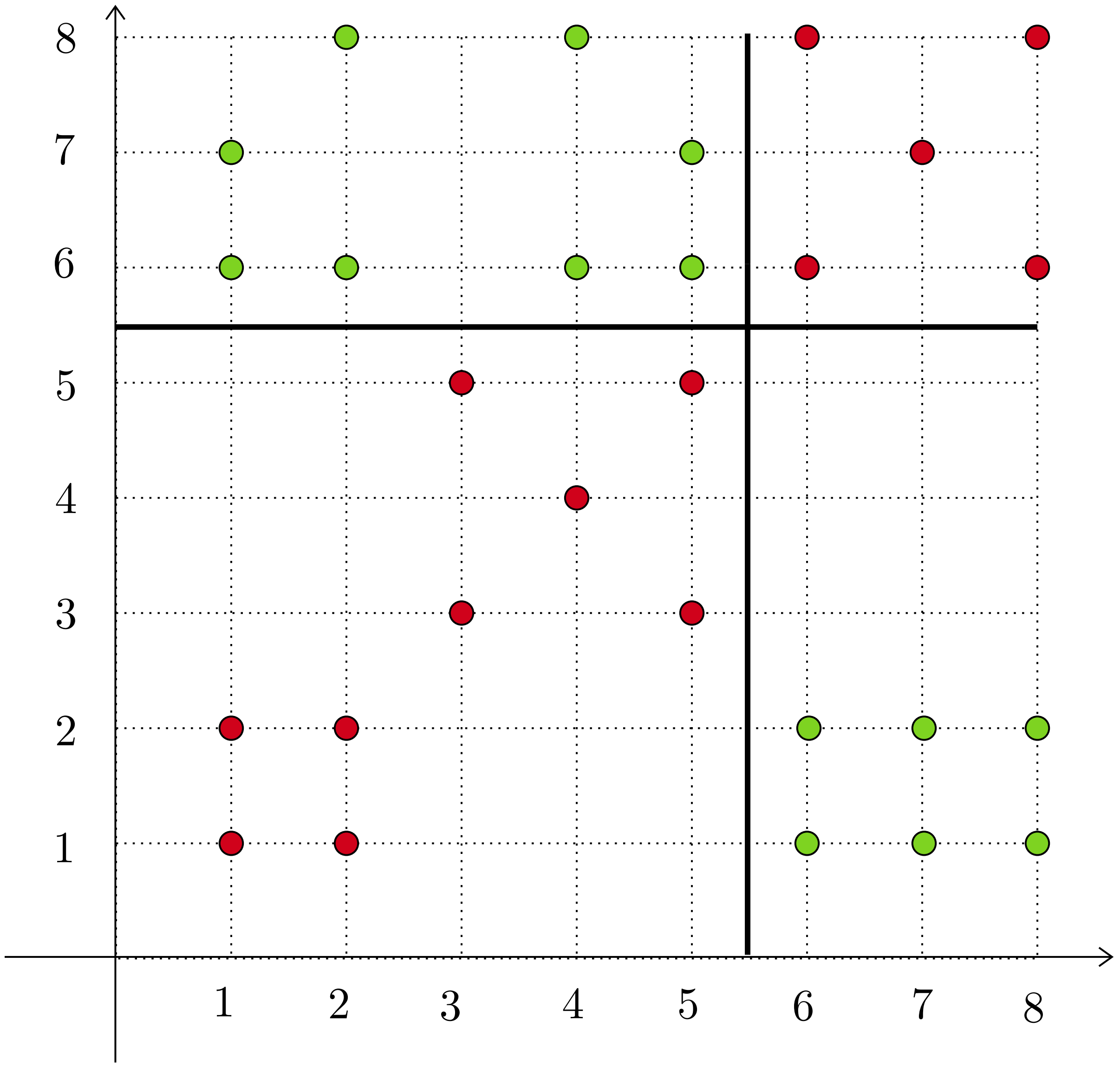


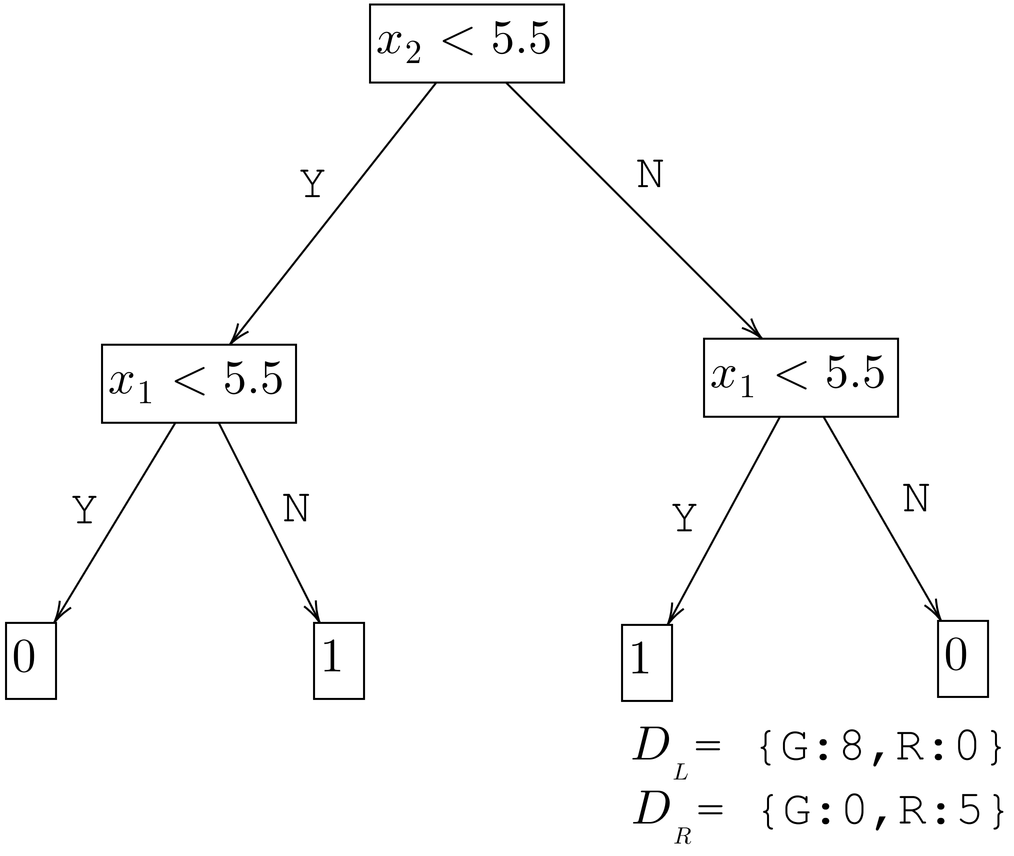
Growing a Tree



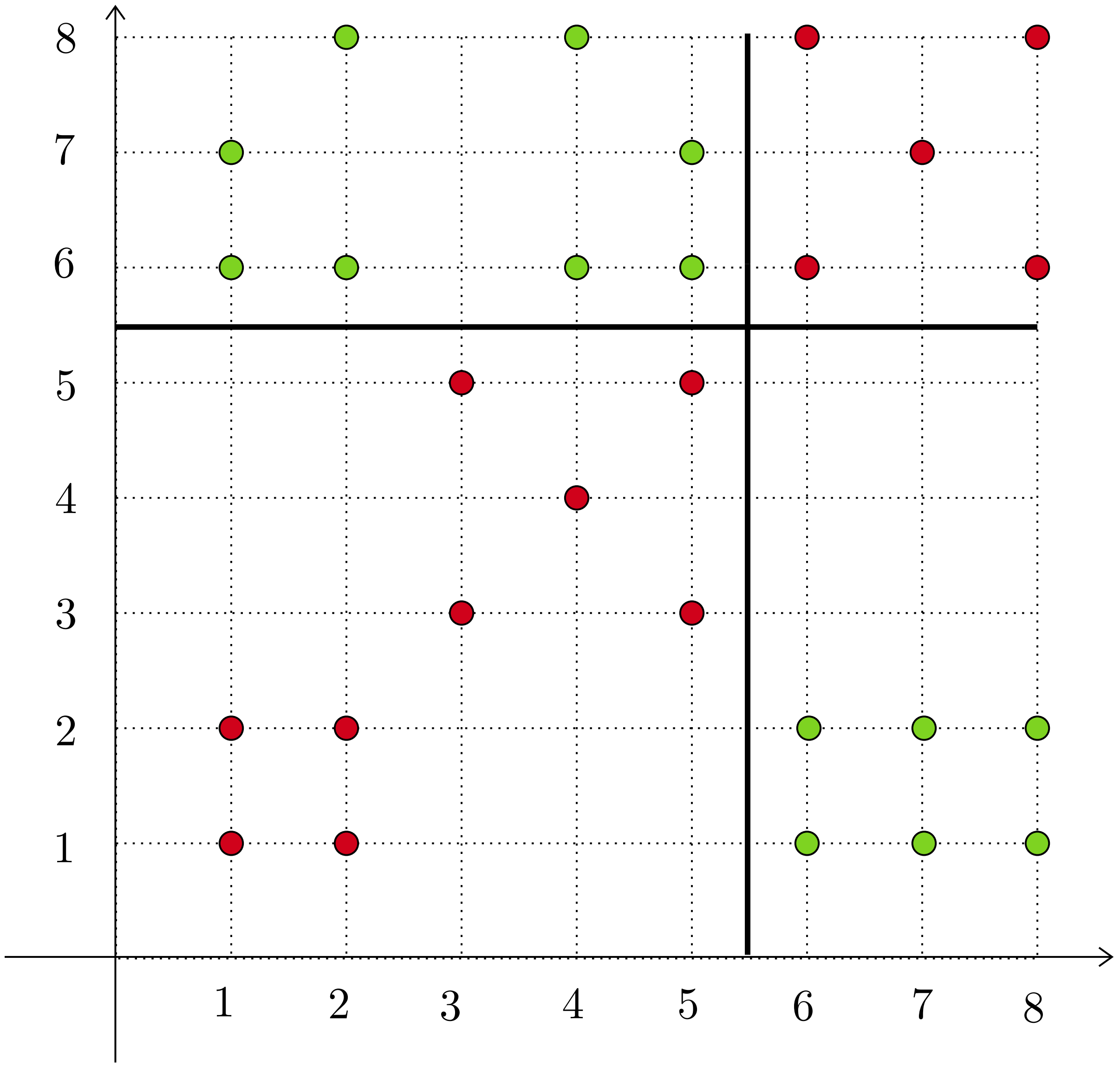


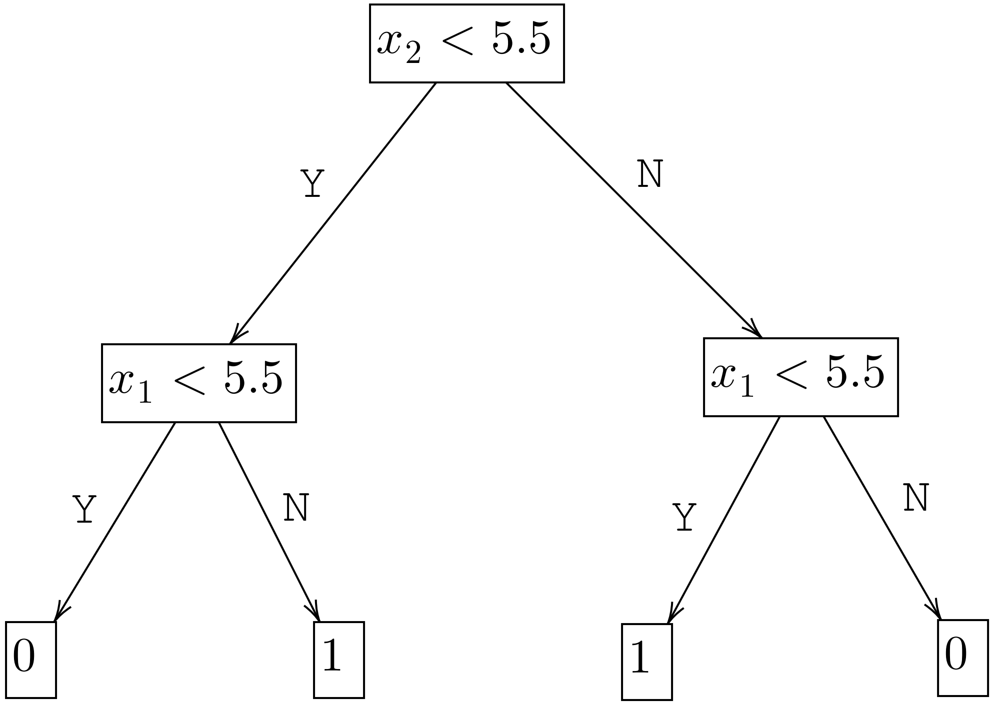
Growing a Tree





Growing a Tree





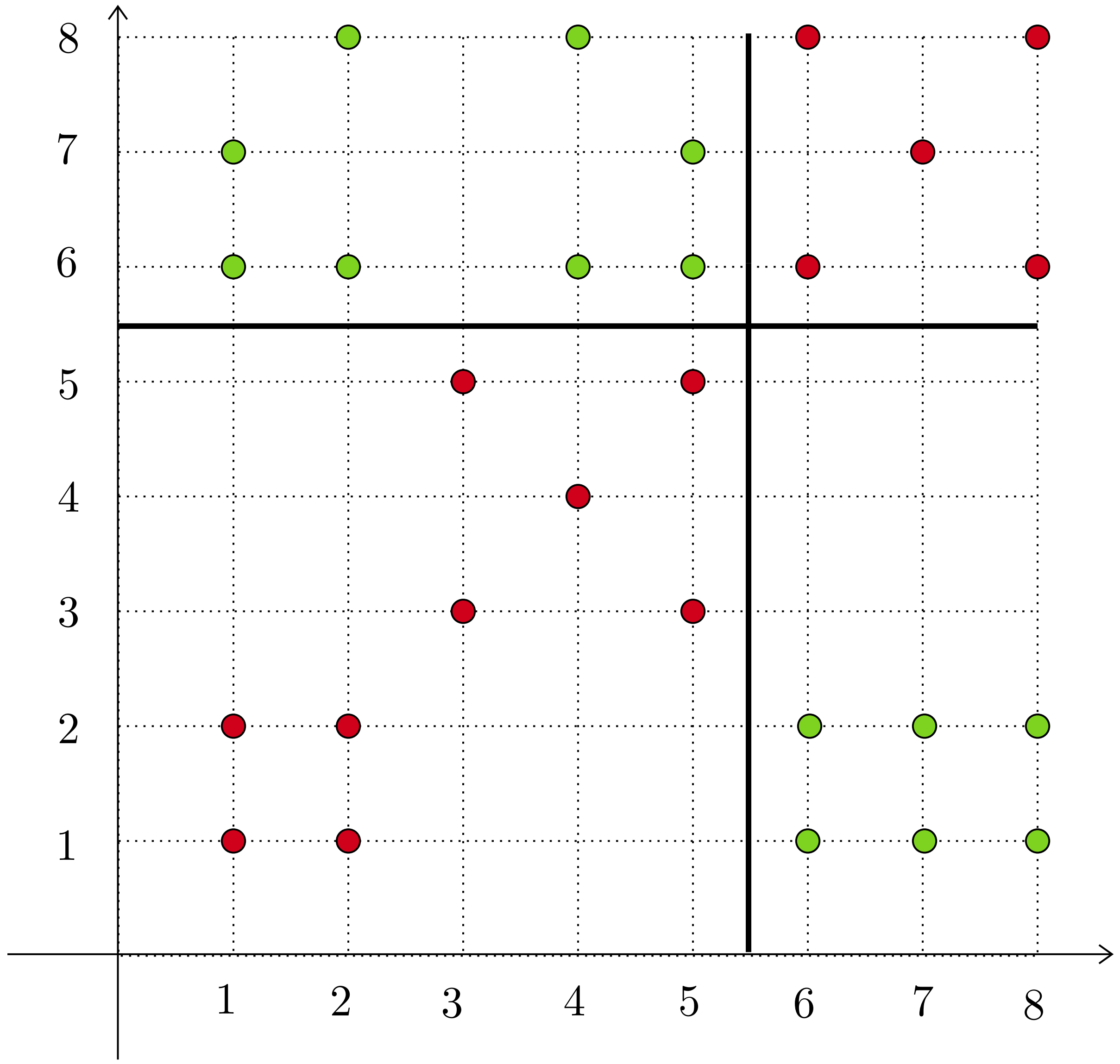
$D_L = \{G:8, R:0\}$
$D_R = \{G:0, R:5\}$

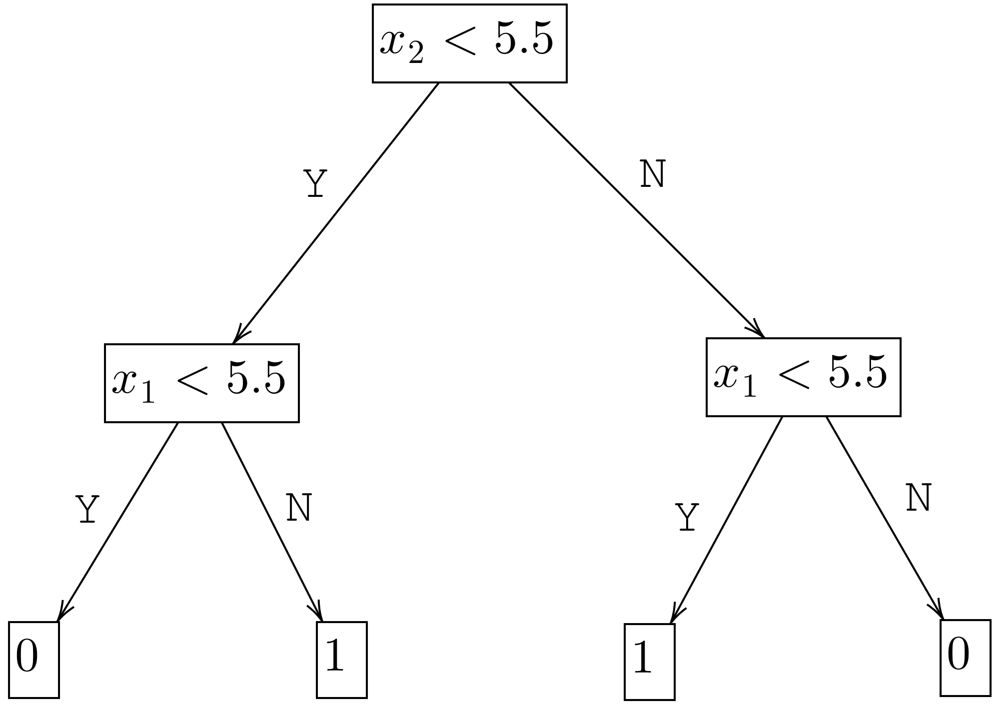
$p_P = (8 \,/\, 13) = 0.615$

$p_L = (8 \,/\, 8) = 1$

$p_R = (0 \,/\, 5) = 0$

Growing a Tree





$D_L =$ {G:8,R:0}

$D_R =$ {G:0,R:5}

$p_P = (8\,/\,13)$

$\quad = \ 0.615$

$E_P = 0.96$

$p_L = (8\,/\,8)$

$\quad = 1$

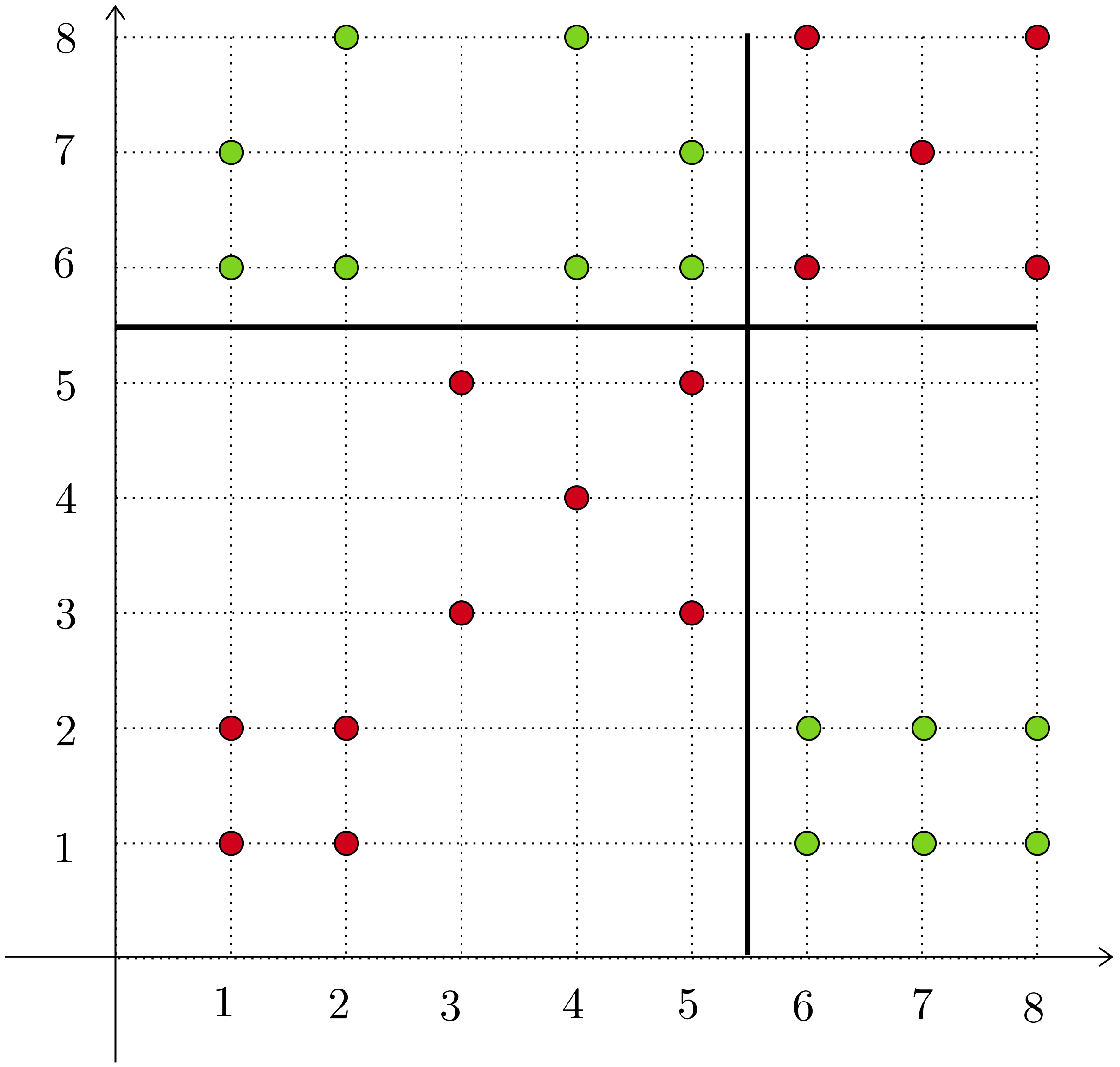
$E_L = 0$

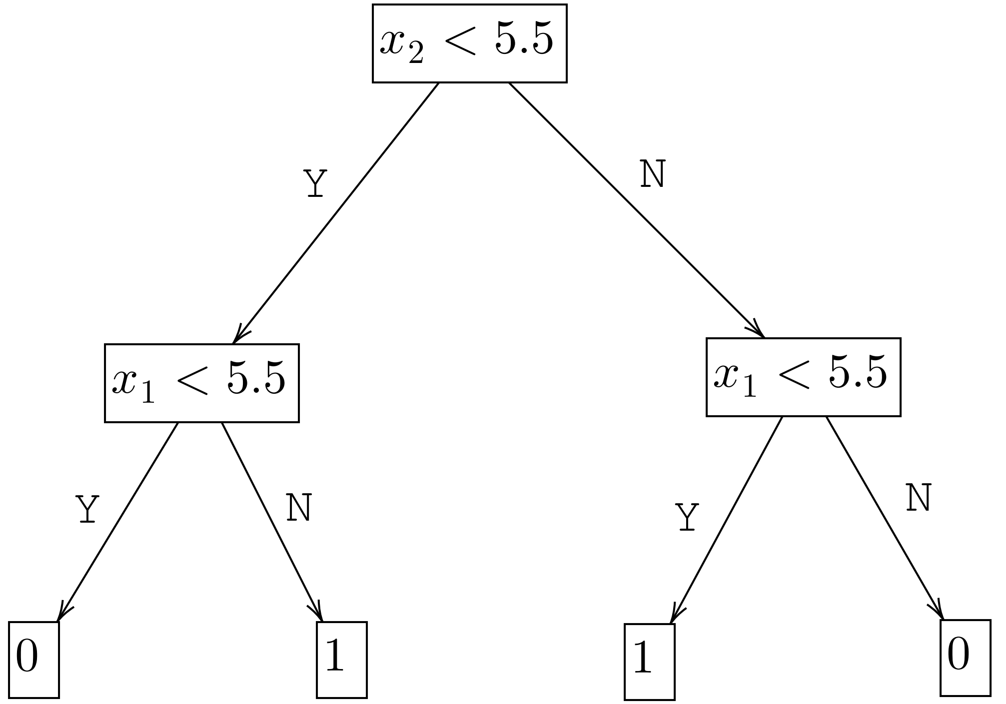
$p_R = (0\,/\,5)$

$\quad = 0$

$E_R = 0$

Growing a Tree





$D_L =$ {G:8,R:0}
$D_R =$ {G:0,R:5}

$$p_P = (8\,/\,13) = 0.615 \qquad p_L = (8\,/\,8) = 1 \qquad p_R = (0\,/\,5) = 0$$

$$E_P = 0.96 \qquad E_L = 0 \qquad E_R = 0$$

$$\begin{aligned} IG &= E_P - [\gamma E_L + (1-\gamma)E_R] \\ &= 0.96 - 0.615(0) + (0.385)(0) \\ &= 0.96 \end{aligned}$$