

How to select best split point in Decision Trees?

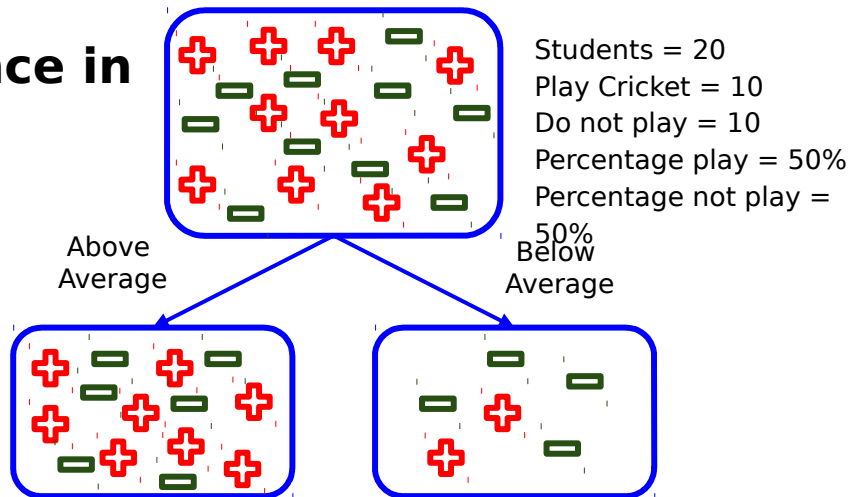
Chi-Square

- Statistical significance of differences between child nodes and their parent node
- It is measured as sum of squared standardized differences between actual and expected frequencies of target variable for each node.

$$\text{Chi-Square} = \sum [(Actual - Expected)^2 / Expected]$$

Steps to calculate Chi-Square for a split

Split on Performance in Class



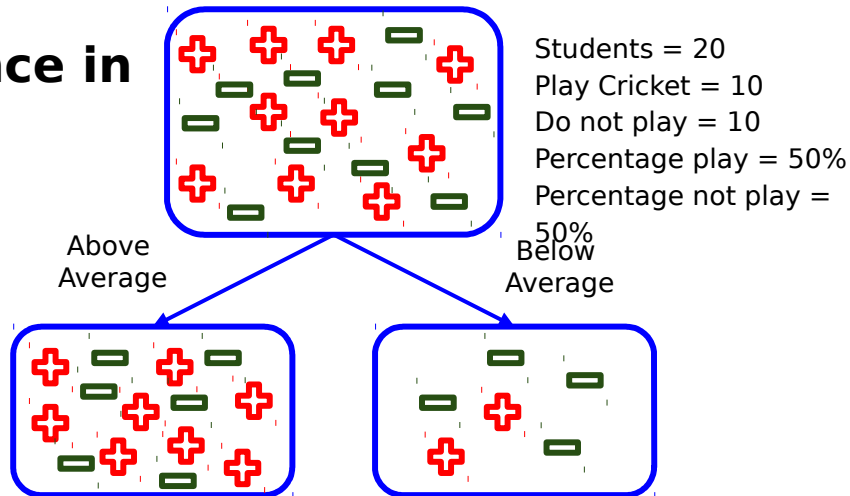
Steps to calculate Chi-Square for a split

Split on Performance in Class

Students = 14

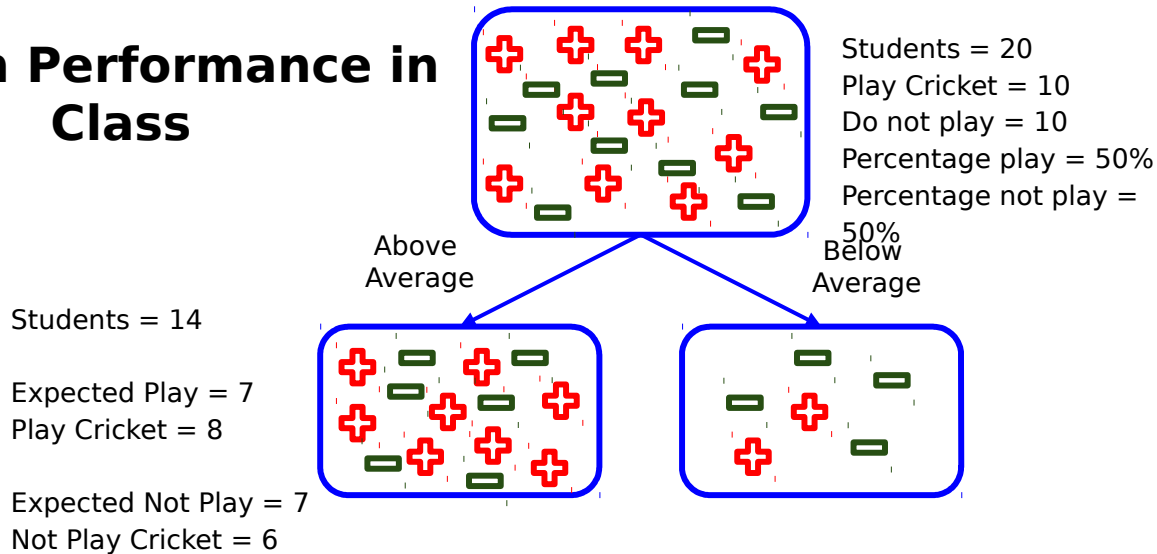
Expected Play = 7

Play Cricket = 8



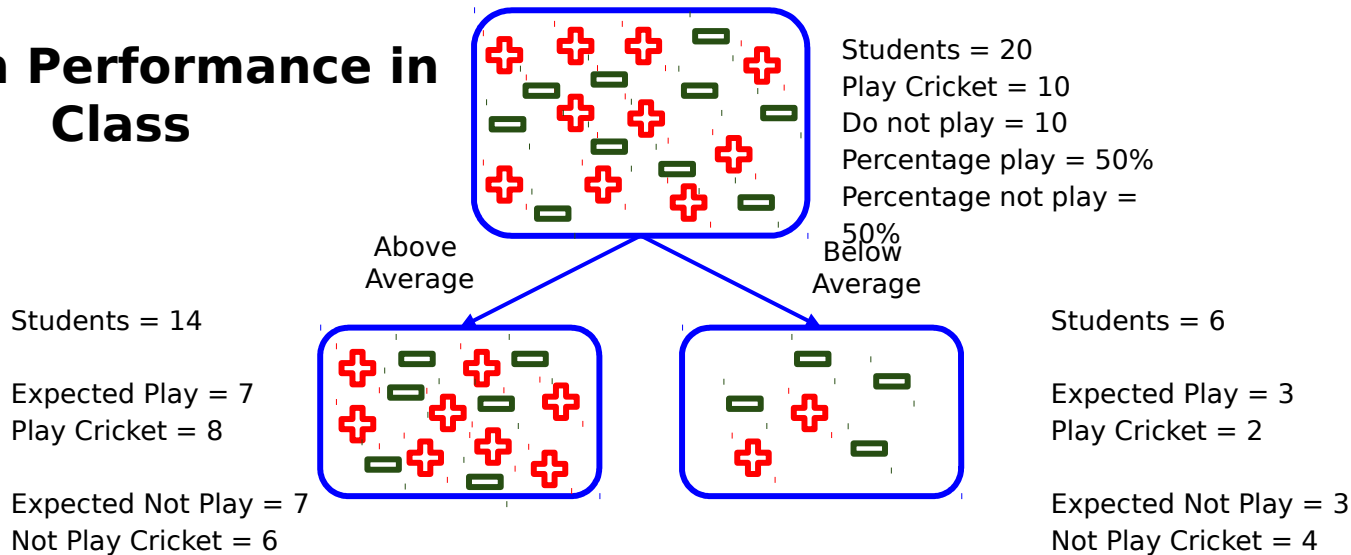
Steps to calculate Chi-Square for a split

Split on Performance in Class



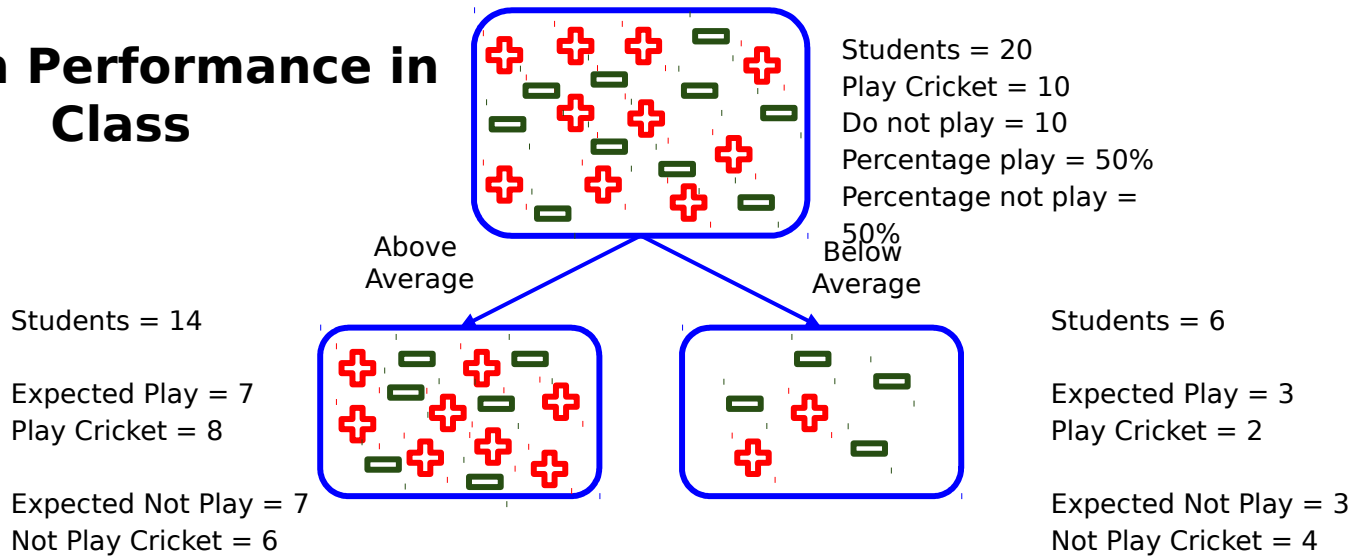
Steps to calculate Chi-Square for a split

Split on Performance in Class



Steps to calculate Chi-Square for a split

Split on Performance in Class



$$\text{Chi-Square} = \sqrt{[(\text{Actual} - \text{Expected})^2 / \text{Expected}]}$$

Properties of Chi-Square

- Works only with categorical target variable
- Can perform two or more splits
- Higher the Chi-Square value, higher the homogeneity of nodes

Steps to calculate Chi-Square for a split

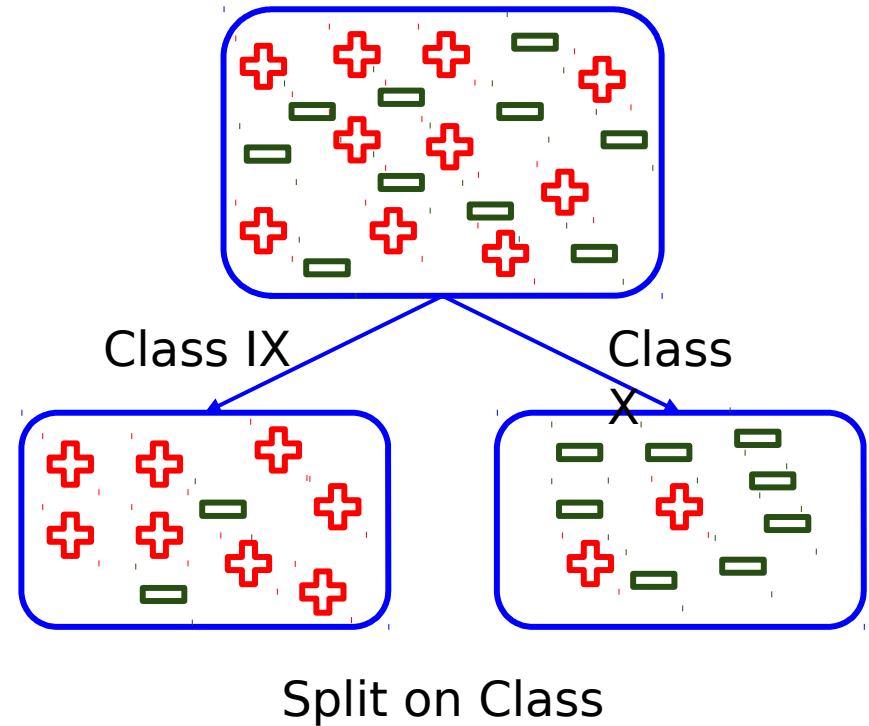
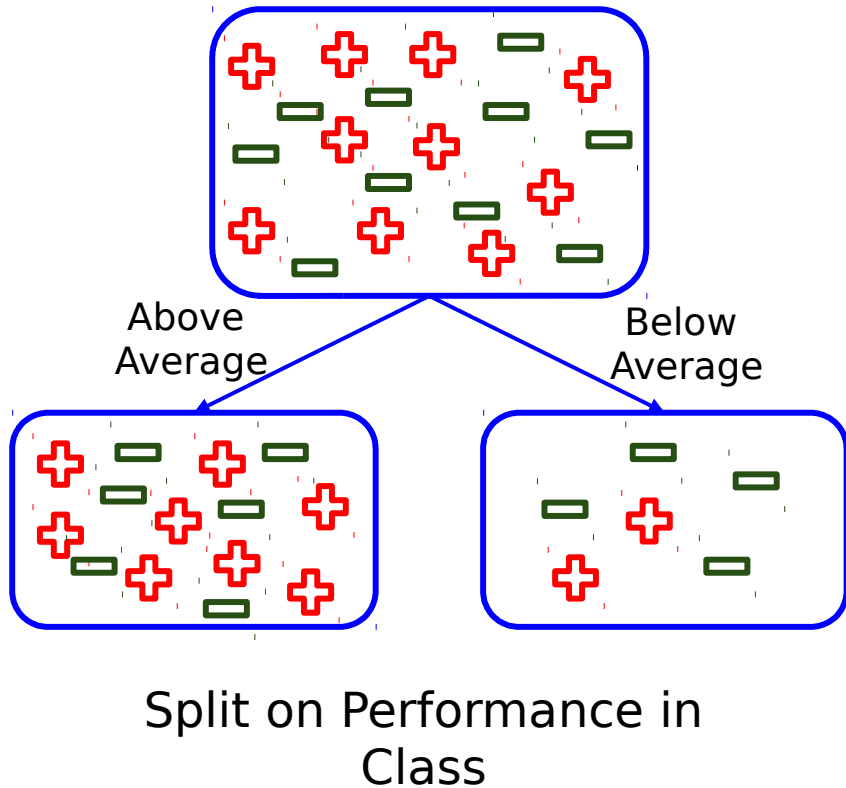
- Calculate the expected values for each class for every child nodes

- Calculate the Chi-Square for every child node

$$\text{Chi-Square} = \sum [(Actual - Expected)^2 / Expected]$$

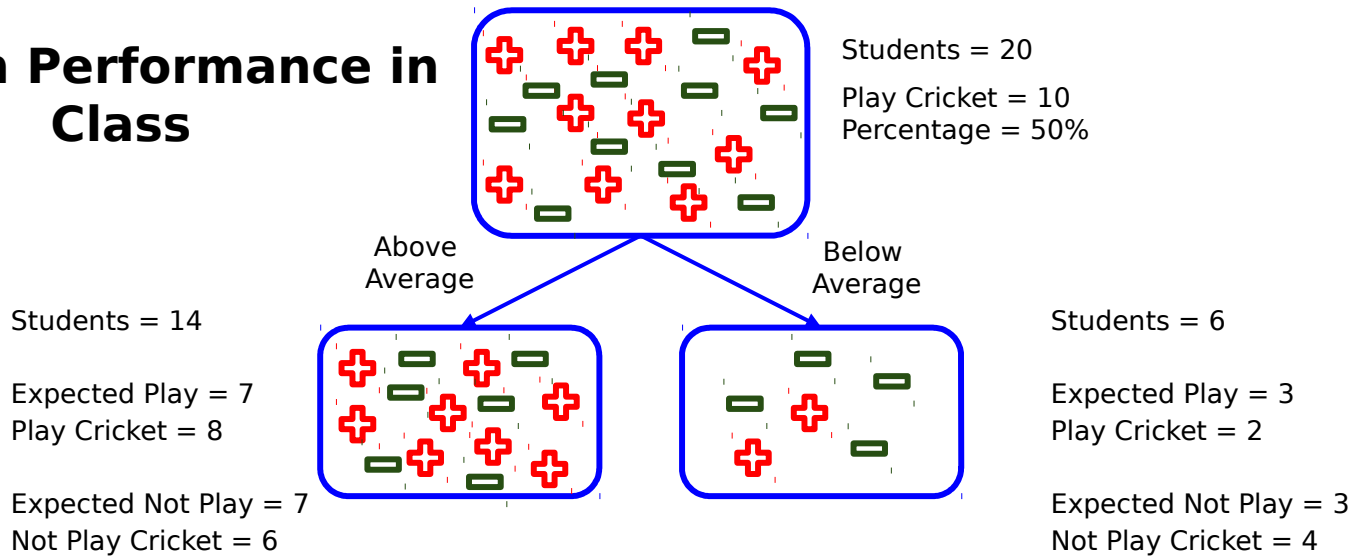
- Calculate Chi-Square for split using sum of Chi-Square of each child node of that split

Steps to calculate Chi-Square for a split



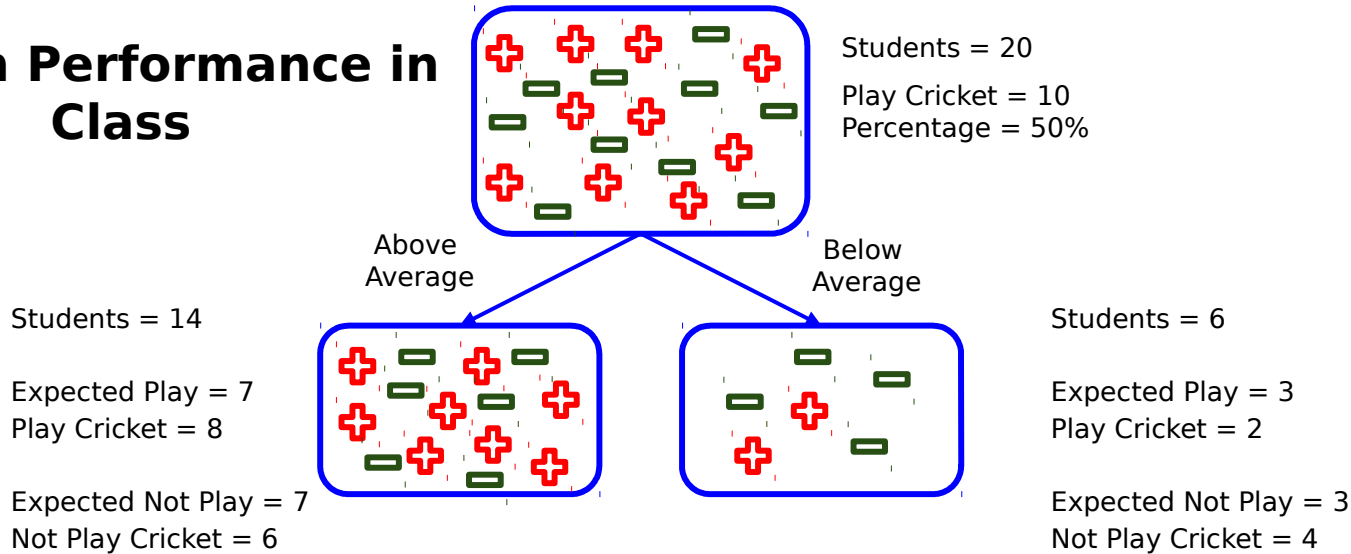
Steps to calculate Chi-Square for a split

Split on Performance in Class



Steps to calculate Chi-Square for a split

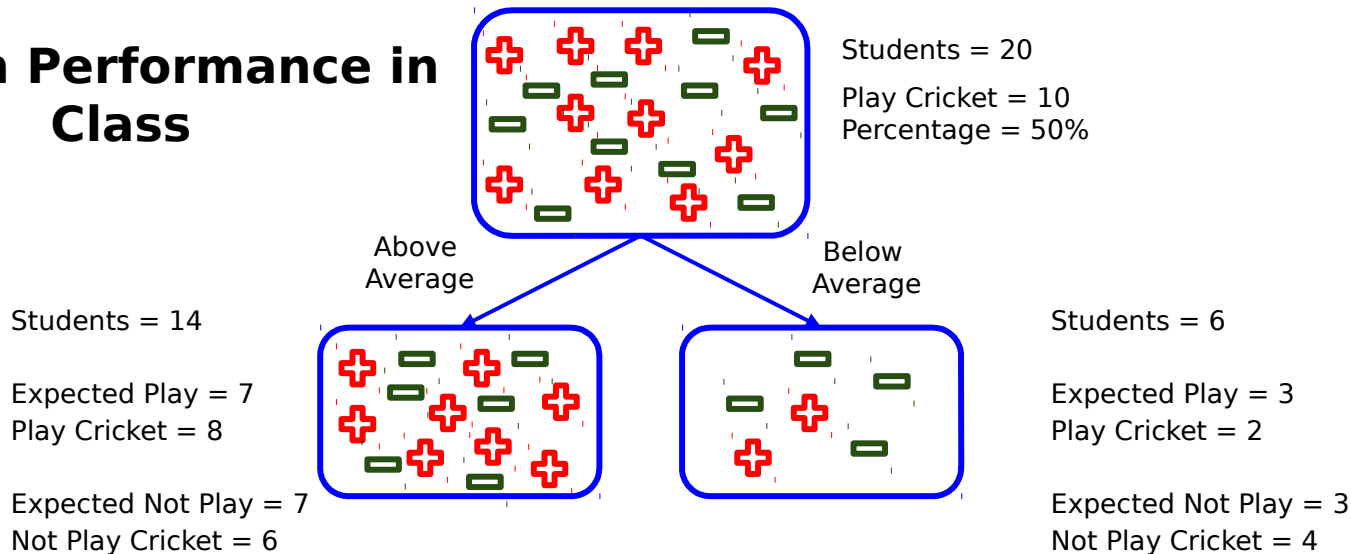
Split on Performance in Class



Node	Actual Play	Actual Not Play	Expected Play	Expected Not Play	Deviation Play	Deviation Not Play	Chi-Square (Play)	Chi-Square (Not Play)
Above Average								
Below Average								

Steps to calculate Chi-Square for a split

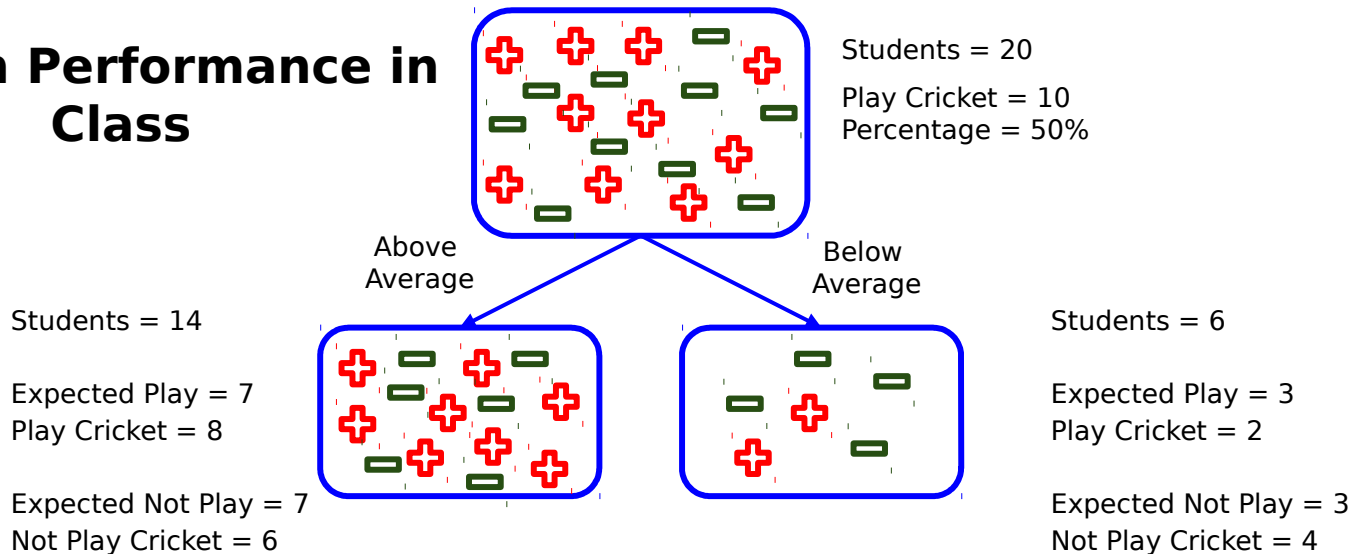
Split on Performance in Class



Node	Actual Play	Actual Not Play	Expected Play	Expected Not Play	Deviation Play	Deviation Not Play	Chi-Square (Play)	Chi-Square (Not Play)
Above Average	8	6						
Below Average								

Steps to calculate Chi-Square for a split

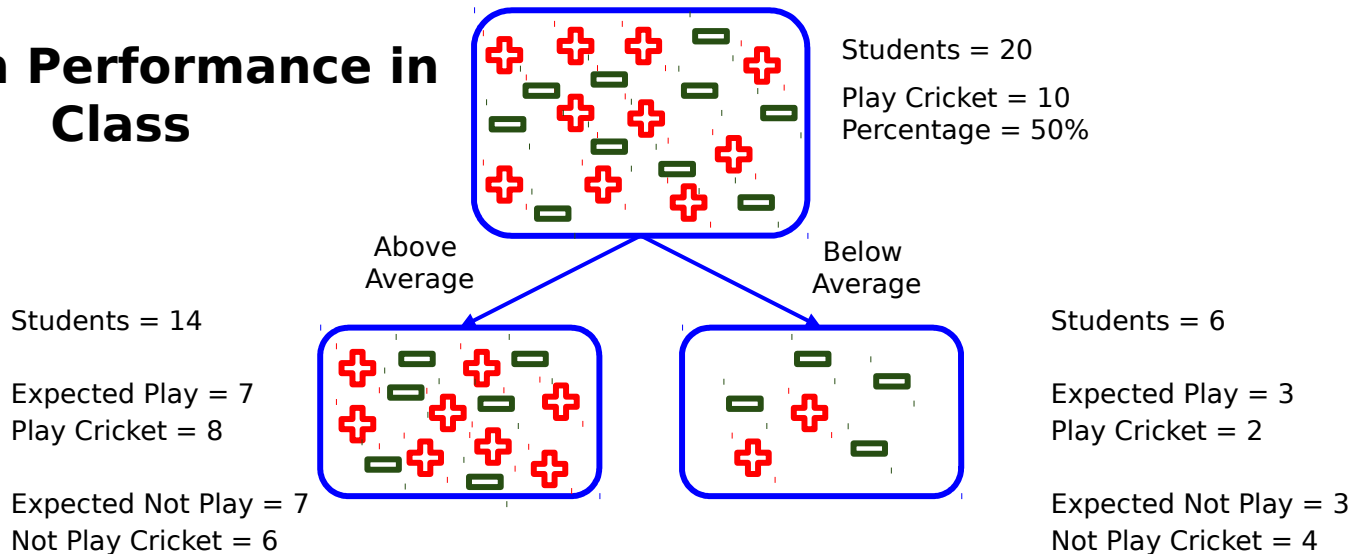
Split on Performance in Class



Node	Actual Play	Actual Not Play	Expected Play	Expected Not Play	Deviation Play	Deviation Not Play	Chi-Square (Play)	Chi-Square (Not Play)
Above Average	8	6	7	7				
Below Average								

Steps to calculate Chi-Square for a split

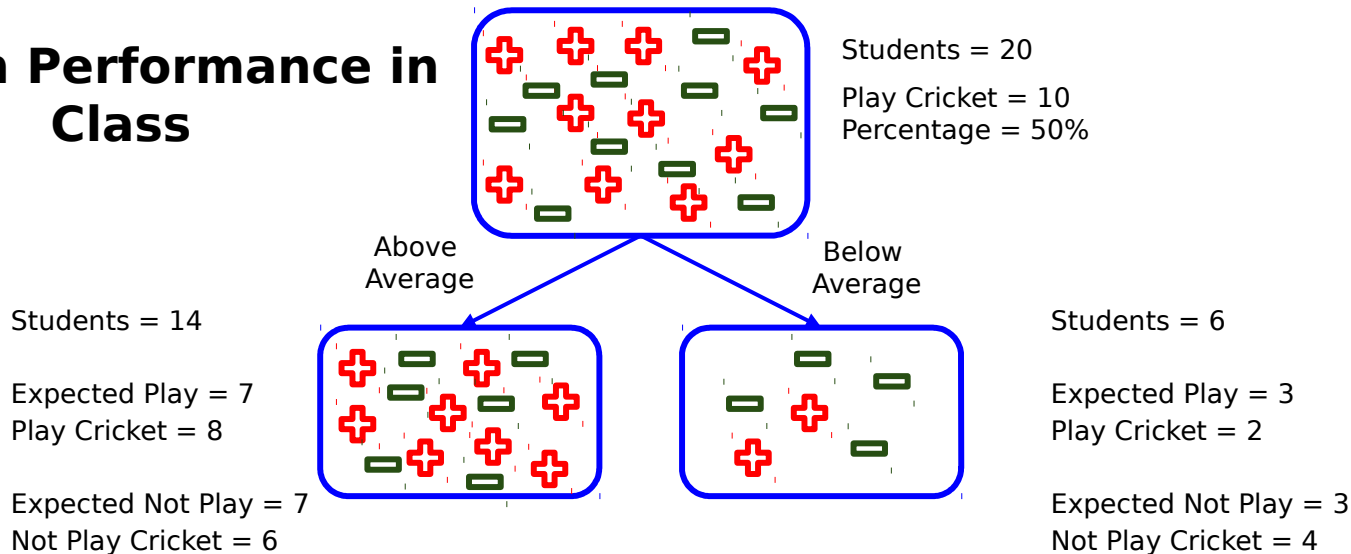
Split on Performance in Class



Node	Actual Play	Actual Not Play	Expected Play	Expected Not Play	Deviation Play	Deviation Not Play	Chi-Square (Play)	Chi-Square (Not Play)
Above Average	8	6	7	7				
Below Average	2	4	3	3				

Steps to calculate Chi-Square for a split

Split on Performance in Class



Node	Actual Play	Actual Not Play	Expected Play	Expected Not Play	Deviation Play	Deviation Not Play	Chi-Square (Play)	Chi-Square (Not Play)
Above Average	8	6	7	7	1	-1		
Below Average	2	4	3	3	-1	1		

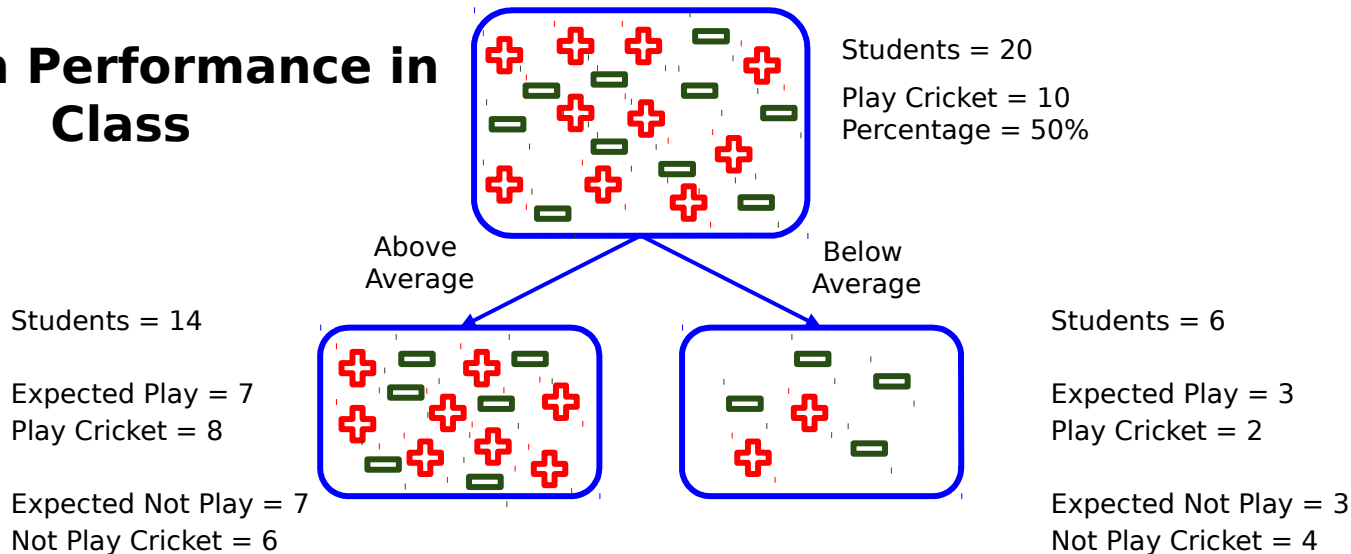
Steps to calculate Chi-Square for a split

Split on Performance in Class

$$\text{Chi-Square} = \sum \frac{(\text{Actual} - \text{Expected})^2}{\text{Expected}}$$

Steps to calculate Chi-Square for a split

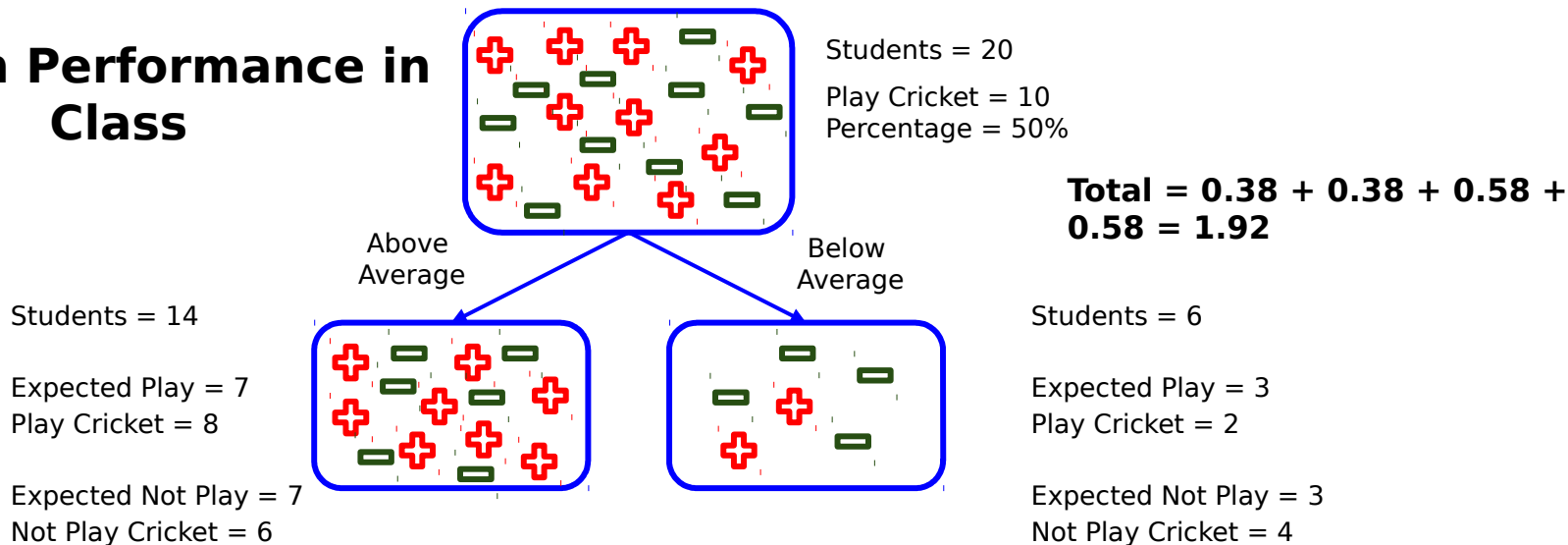
Split on Performance in Class



Node	Actual Play	Actual Not Play	Expected Play	Expected Not Play	Deviation Play	Deviation Not Play	Chi-Square (Play)	Chi-Square (Not Play)
Above Average	8	6	7	7	1	-1	0.38	0.38
Below Average	2	4	3	3	-1	1	0.58	0.58

Steps to calculate Chi-Square for a split

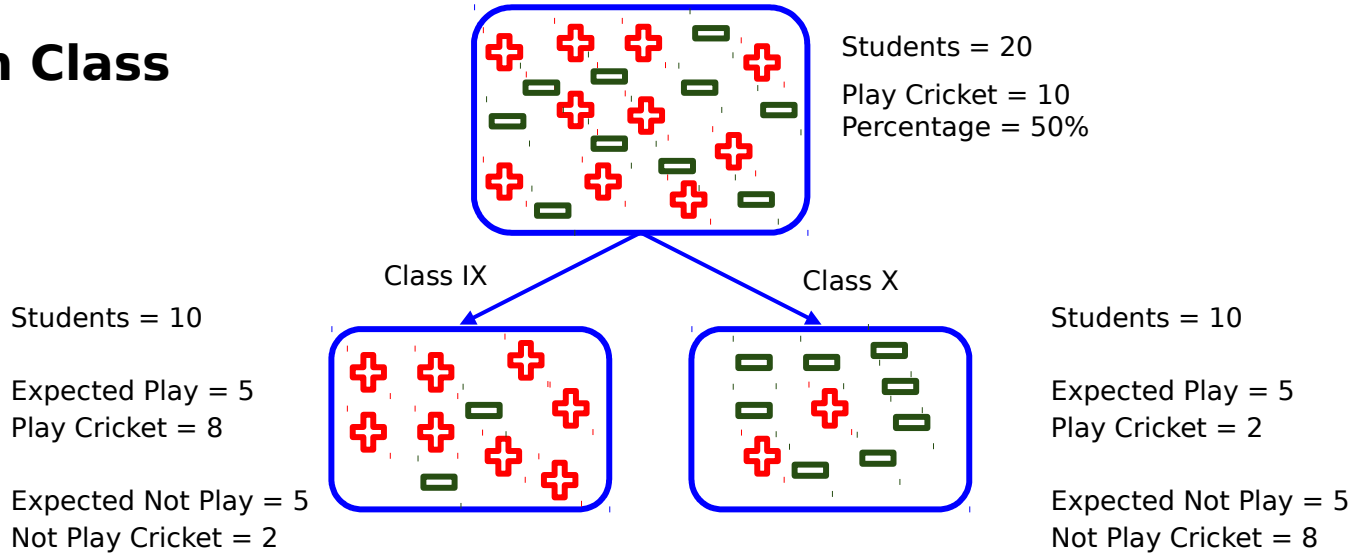
Split on Performance in Class



Node	Actual Play	Actual Not Play	Expected Play	Expected Not Play	Deviation Play	Deviation Not Play	Chi-Square (Play)	Chi-Square (Not Play)
Above Average	8	6	7	7	1	-1	0.38	0.38
Below Average	2	4	3	3	-1	1	0.58	0.58

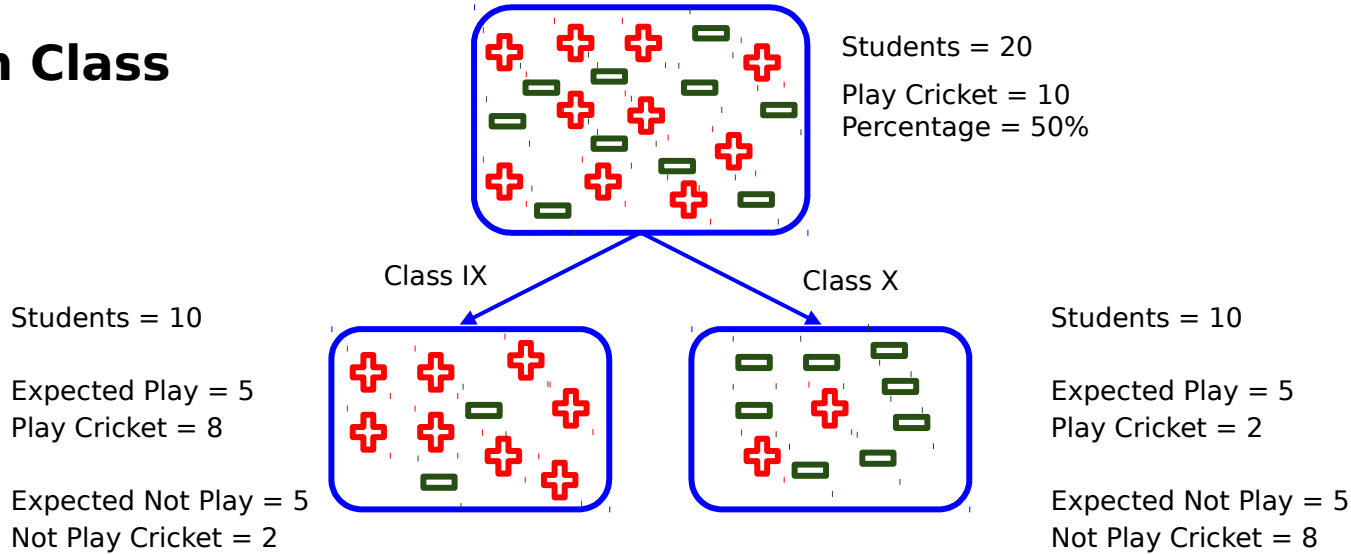
Steps to calculate Chi-Square for a split

Split on Class



Steps to calculate Chi-Square for a split

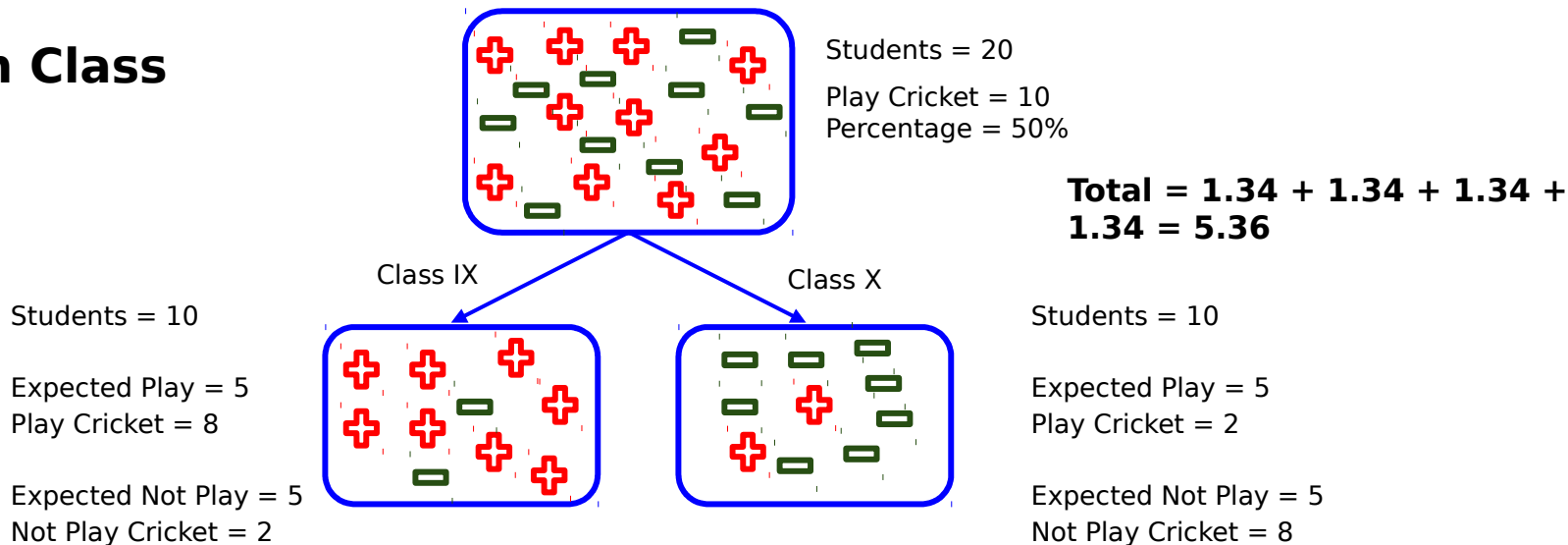
Split on Class



Node	Actual Play	Actual Not Play	Expected Play	Expected Not Play	Deviation Play	Deviation Not Play	Chi-Square (Play)	Chi-Square (Not Play)
IX	8	2	5	5	3	-3	1.34	1.34
X	2	8	5	5	-3	3	1.34	1.34

Steps to calculate Chi-Square for a split

Split on Class



Node	Actual Play	Actual Not Play	Expected Play	Expected Not Play	Deviation Play	Deviation Not Play	Chi-Square (Play)	Chi-Square (Not Play)
IX	8	2	5	5	3	-3	1.34	1.34
X	2	8	5	5	-3	3	1.34	1.34

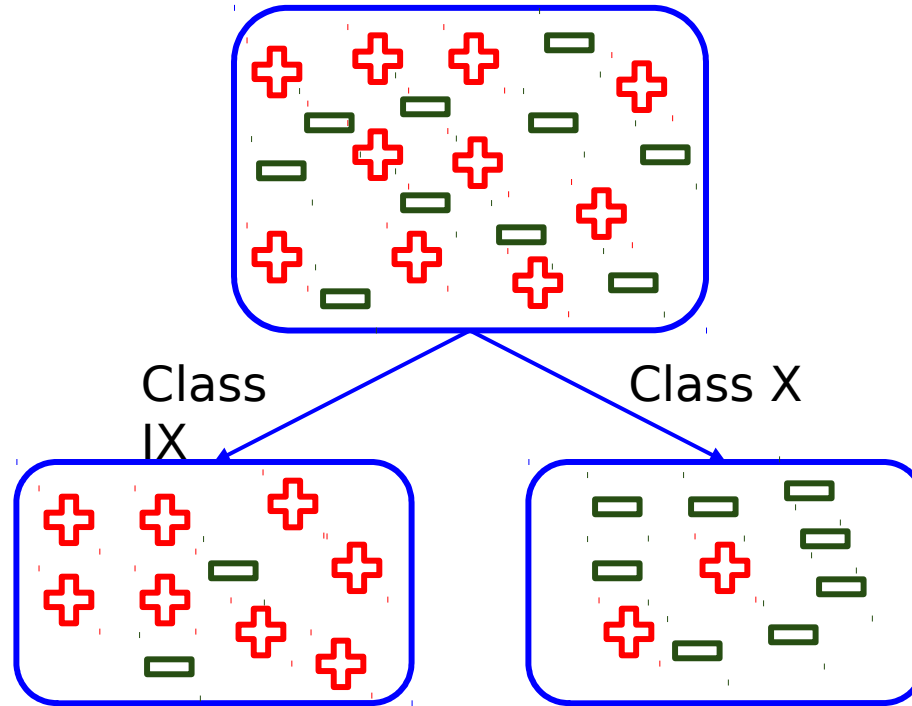
Steps to calculate Chi-Square for a split

Split	Chi-Square
Performance in Class	1.92
Class	5.36

Steps to calculate Chi-Square for a split

Split	Chi-Square
Performance in Class	1.92
Class	5.36

Steps to calculate Chi-Square for a split



Split on Class

Thank
You!