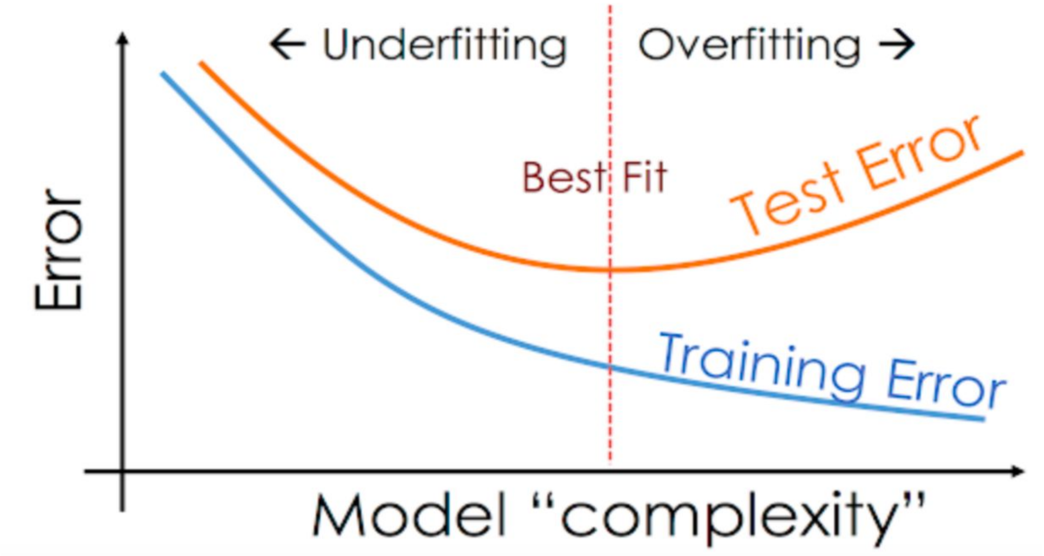
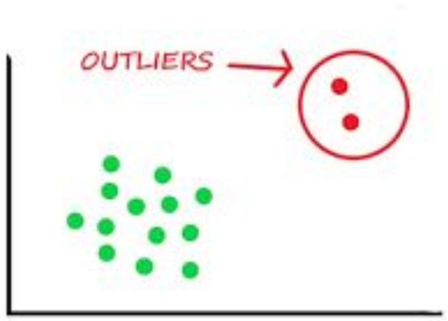
Machine Learning

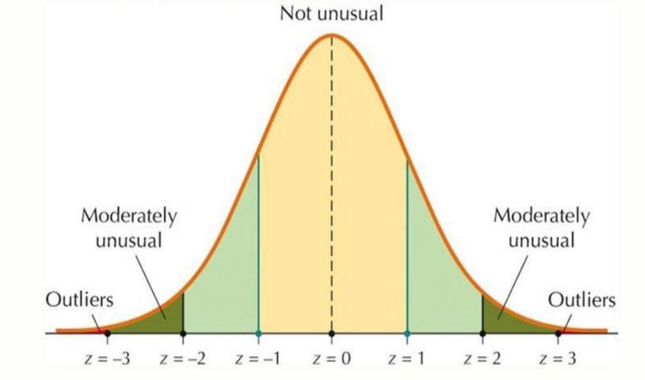
Overfitting - Algorithm perform well with training dataset but not with testing dataset

Underfitting - Algorithm does not perform well with training dataset and testing dataset



Outliers - Observation is either very low value or very high value in comparison to other observed values.





z-score

$$Z = \frac{\overset{\text{Score}}{x} - \overset{\text{Mean}}{\mu}}{\underset{\text{SD}}{\sigma}}$$

Gradient descent

Repeat until convergence{

$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$ (for $j = 0$ and $j = 1$)

}

α: Learning rate (step size)

$\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$: derivative (rate of change)

Gradient descent

Correct: simultaneous update

$$\text{temp0} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$\text{temp1} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

$$\theta_0 := \text{temp0}$$

$$\theta_1 := \text{temp1}$$

Incorrect:

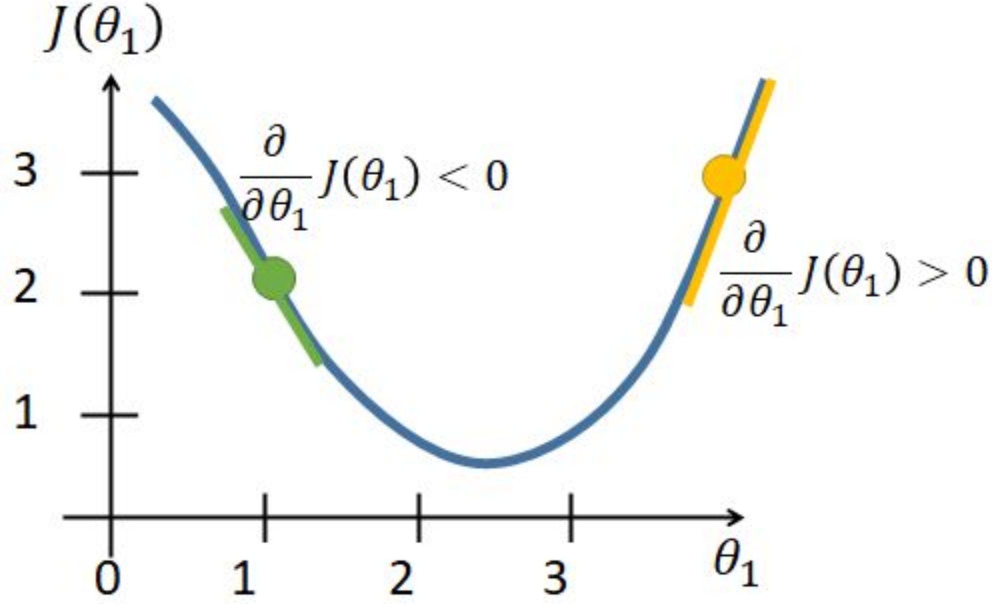
$$\text{temp0} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$\theta_0 := \text{temp0}$$

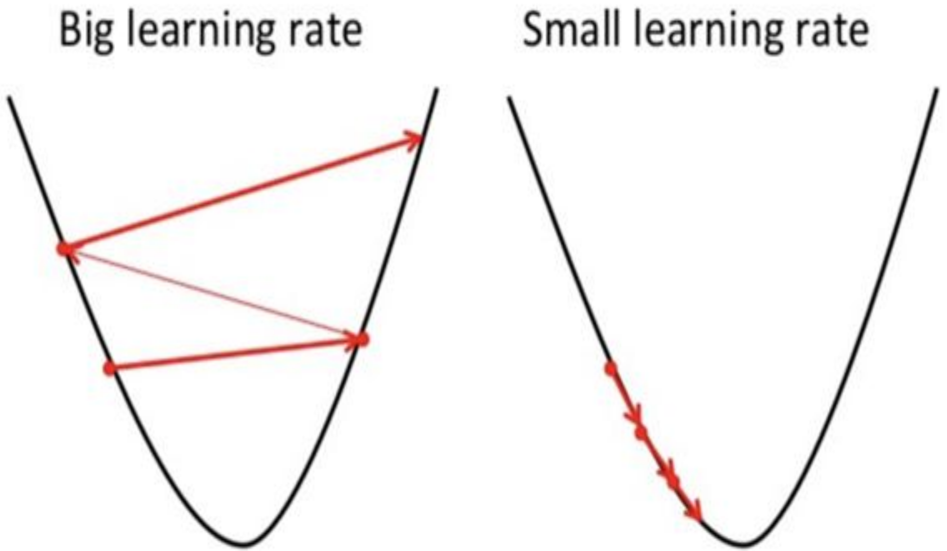
$$\text{temp1} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

$$\theta_1 := \text{temp1}$$

$$\theta_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_1)$$

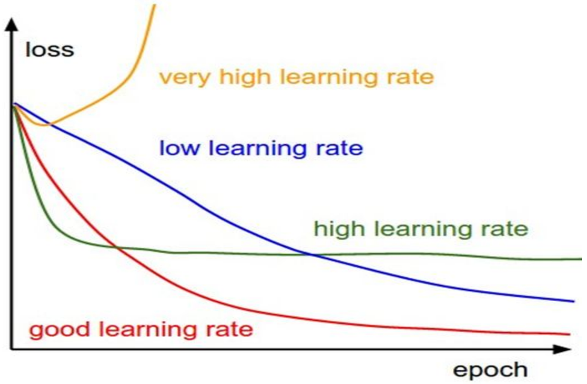


Learning rate



It cause drastic updates which leads to divergent behaviors

It swiftly reaches the minimum point



- α -> 0.001, … 0.01, …, 0.1, … , 1

Gradient descent for linear regression

Repeat until convergence{

$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$ (for $j = 0$ and $j = 1$)

}

- Linear regression model

$h_\theta(x) = \theta_0 + \theta_1 x$

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

Computing partial derivative

- $\frac{\partial}{\partial\theta_j} J(\theta_0, \theta_1) = \frac{\partial}{\partial\theta_j} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$

  $= \frac{\partial}{\partial\theta_j} \frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)^2$

- $j = 0$: $\frac{\partial}{\partial\theta_0} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)$
- $j = 1$: $\frac{\partial}{\partial\theta_1} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) x^{(i)}$

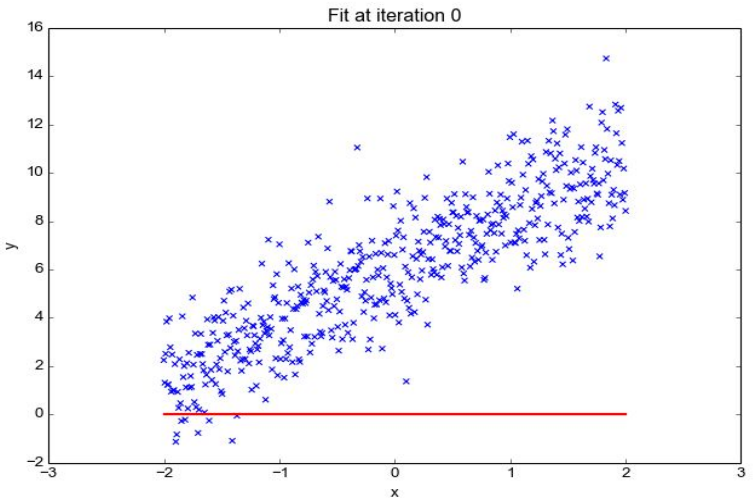
Gradient descent for linear regression

Repeat until convergence{

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) x^{(i)}$$

}



Update θ_0 and θ_1 simultaneously

Batch gradient descent

- "Batch": Each step of gradient descent uses all the training examples

Repeat until convergence{

m: Number of training examples

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{\boxed{m}} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right)$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{\boxed{m}} \left(h_\theta\left(x^{(i)}\right) - y^{(i)} \right) x^{(i)}$$

}

m training examples, n features

Gradient Descent

- Need to choose α
- Need many iterations
- Works well even when n is large

Normal Equation

- No need to choose α
- Don't need to iterate
- Need to compute $(X^\top X)^{-1}$
- Slow if n is very large

(x_0)	Size in feet^2 (x_1)	Number of bedrooms (x_2)	Number of floors (x_3)	Age of home (years) (x_4)	Price ($) in 1000's (y)
1	2104	5	1	45	460
1	1416	3	2	40	232
1	1534	3	2	30	315
1	852	2	1	36	178
	...				...

$$X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \end{bmatrix} \qquad y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$$

$$\theta = (X^\top X)^{-1} X^\top y$$

Model Performance Evaluation -

R2 Score/ R squared/coefficient of determination -

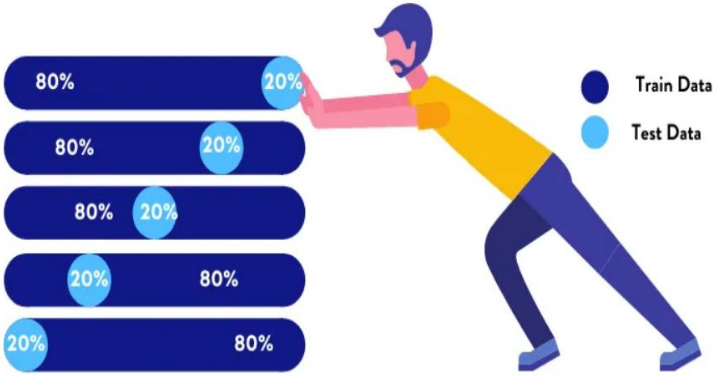
- If the value of the R squared score is 1, it means that the model is perfect and if its value is 0, it means that the model will perform badly on an unseen dataset.
- This also implies that the closer the value of the R squared score is to 1, the more perfectly the model is trained.

Cross Validation -

Technique in which we train our model using the subset of the data-set and then evaluate using the complementary subset of the data-set.

K-Fold Cross Validation -

Split the data-set into k number of subsets(known as folds) then we perform training on the all the subsets but leave one(k-1) subset for the evaluation of the trained model.



K=5

Regularization -

Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.

1. L1 regularization - **LASSO(Least Absolute Shrinkage and Selection Operator)**

$$Loss = Error(y, \hat{y}) + \lambda \sum_{i=1}^{N} |w_i|$$

$$||\mathbf{w}||_1 = |w_1| + |w_2| + ... + |w_N|$$

2. L2 regularization - **Ridge regression**

$$Loss = Error(y, \hat{y}) + \lambda \sum_{i=1}^{N} w_i^2$$

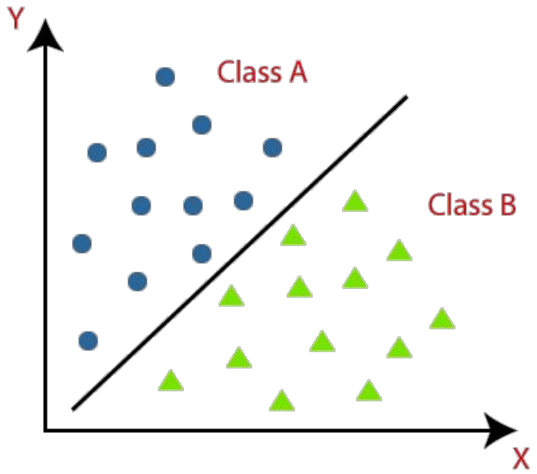
$$\|\mathbf{w}\|_2 = \left(|w_1|^2 + |w_2|^2 + \ldots + |w_N|^2\right)^{\frac{1}{2}}$$

lambda is Known as regularization constant

Regression coding rest

Classification -

Classification algorithms are used when the output variable is categorical, not value which means there are classes such as Yes-No, Male-Female, True-false, etc.



y=f(x)

where y = categorical output
x = input feature

- **Binary Classifier**
- **Multi-class Classifier**

Lazy Learners: Firstly stores the training dataset and wait until it receives the test dataset. e.g. K-NN algorithm

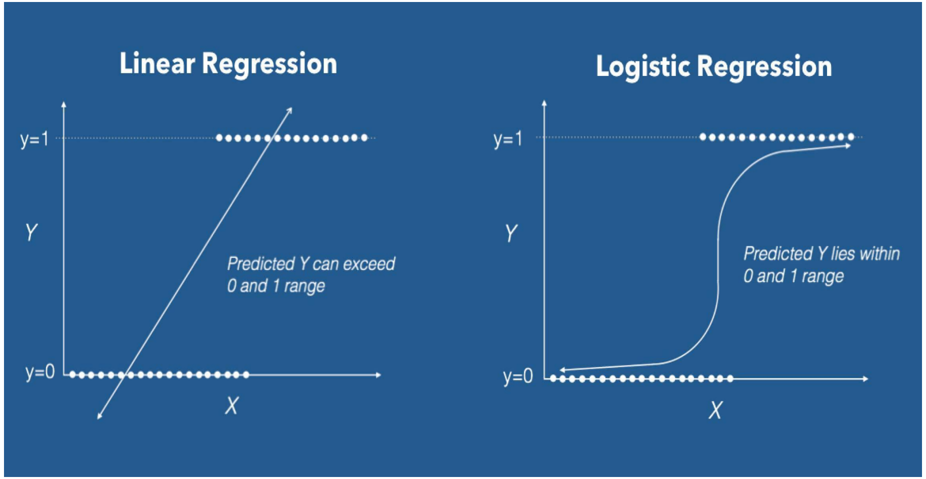
Eager Learners: Develop a classification model based on a training dataset before receiving a test dataset. e.g. Decision Trees

Classification -

- **Logistic Regression**
- **Decision Trees**
- **Random Forest**
- **Support vector Machines**
- **K-NN**
- **Naïve Bayes**

Logistic Regression -

Uses the concept of predictive modeling as regression, but it is used to classify samples. It predicts the output of a categorical dependent variable using a given set of independent variables.



$$log\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

We need result as 0 and 1 only, so divide the equation by (1-y).

Sigmoid function

$$f(x)= \frac{1}{1+e^{-x}}$$

Confusion Matrix -

A matrix used to determine the performance of the classification models for a given set of test data.

- Prediction of 2 classes classifiers, the matrix is of 2*2 table, for 3 classes, it is 3*3 table, and so on.

		Predicted Class		
		Positive	**Negative**	
Actual Class	**Positive**	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{(TP+FN)}$
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{(TN+FP)}$
		Precision $\frac{TP}{(TP+FP)}$	**Negative Predictive Value** $\frac{TN}{(TN+FN)}$	**Accuracy** $\frac{TP+TN}{(TP+TN+FP+FN)}$

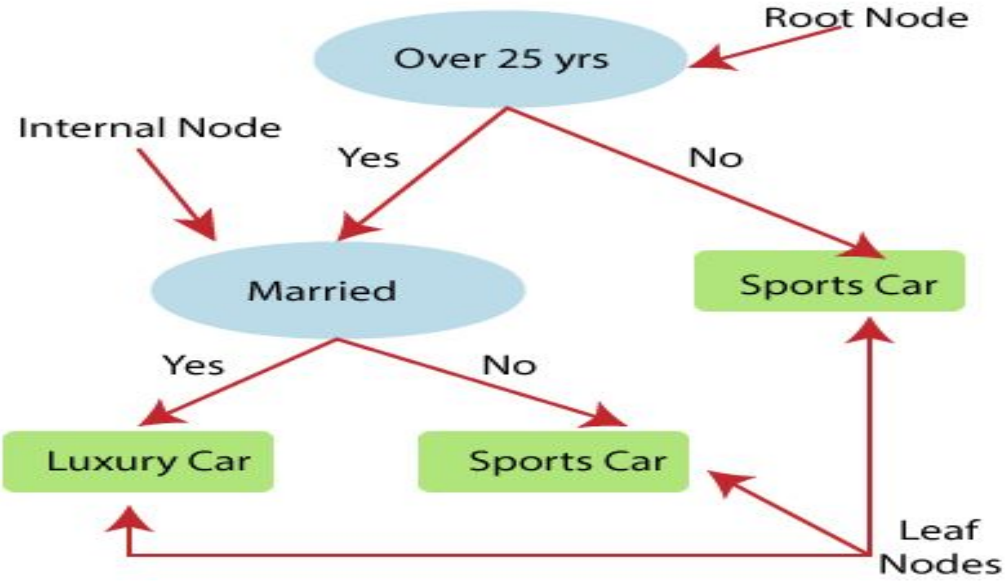
$$\text{Error rate} = \frac{FP+FN}{TP+FP+FN+TN}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{F-measure} = \frac{2*Recall*Precision}{Recall+Precision}$$

Decision Tree -

- It can solve problems for both categorical and numerical data



Root Node

Leaf Node

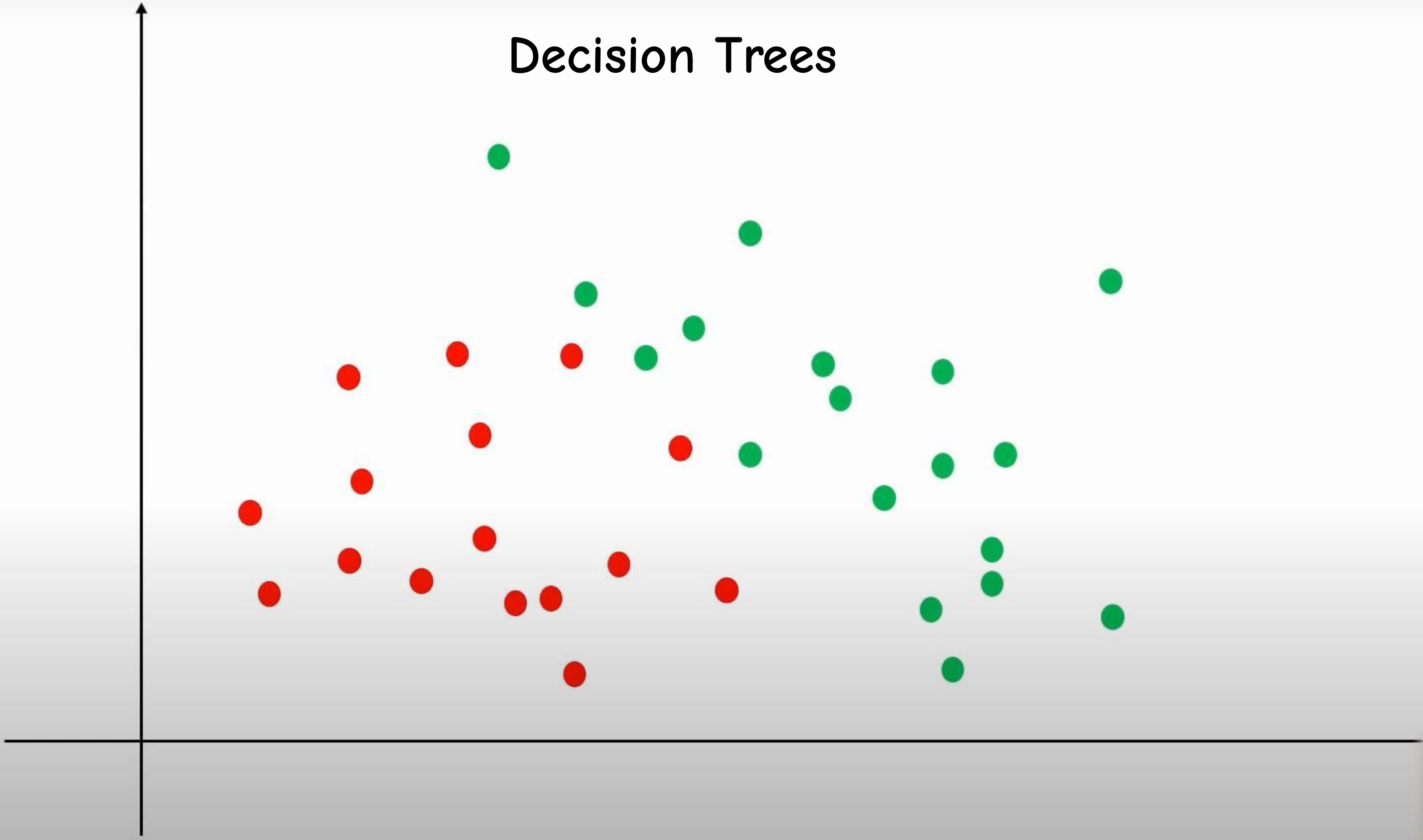
Splitting

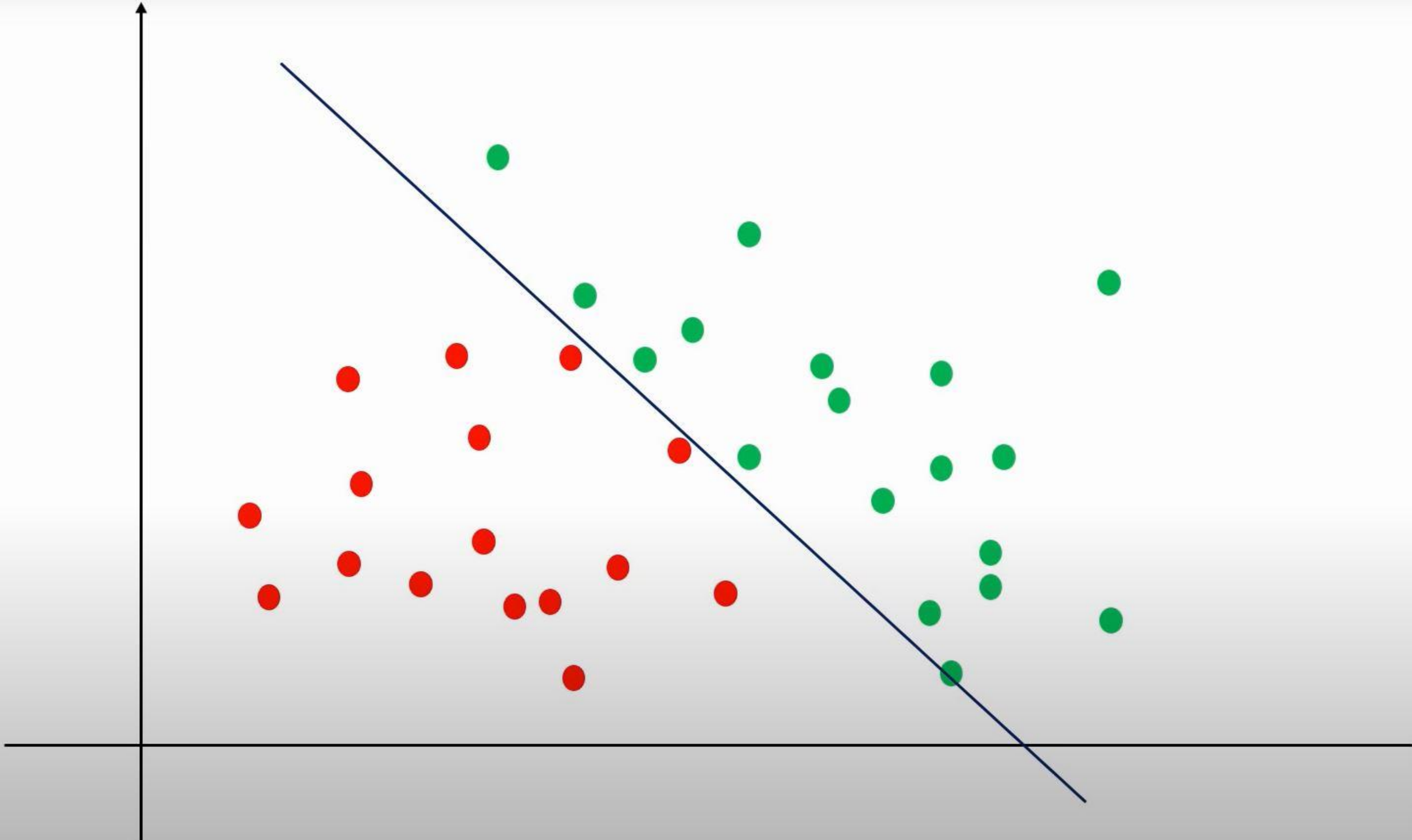
Branch/Subtree

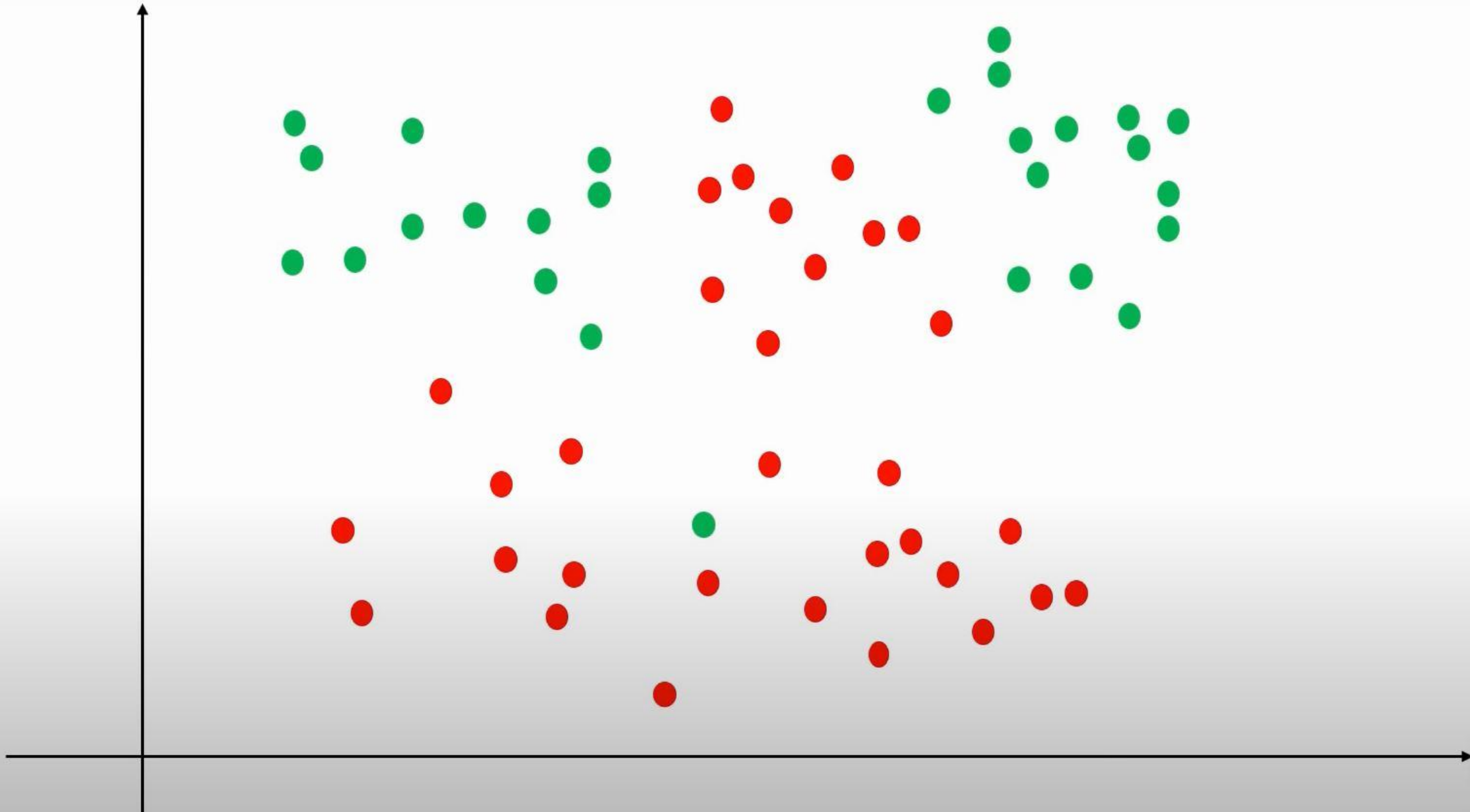
Pruning

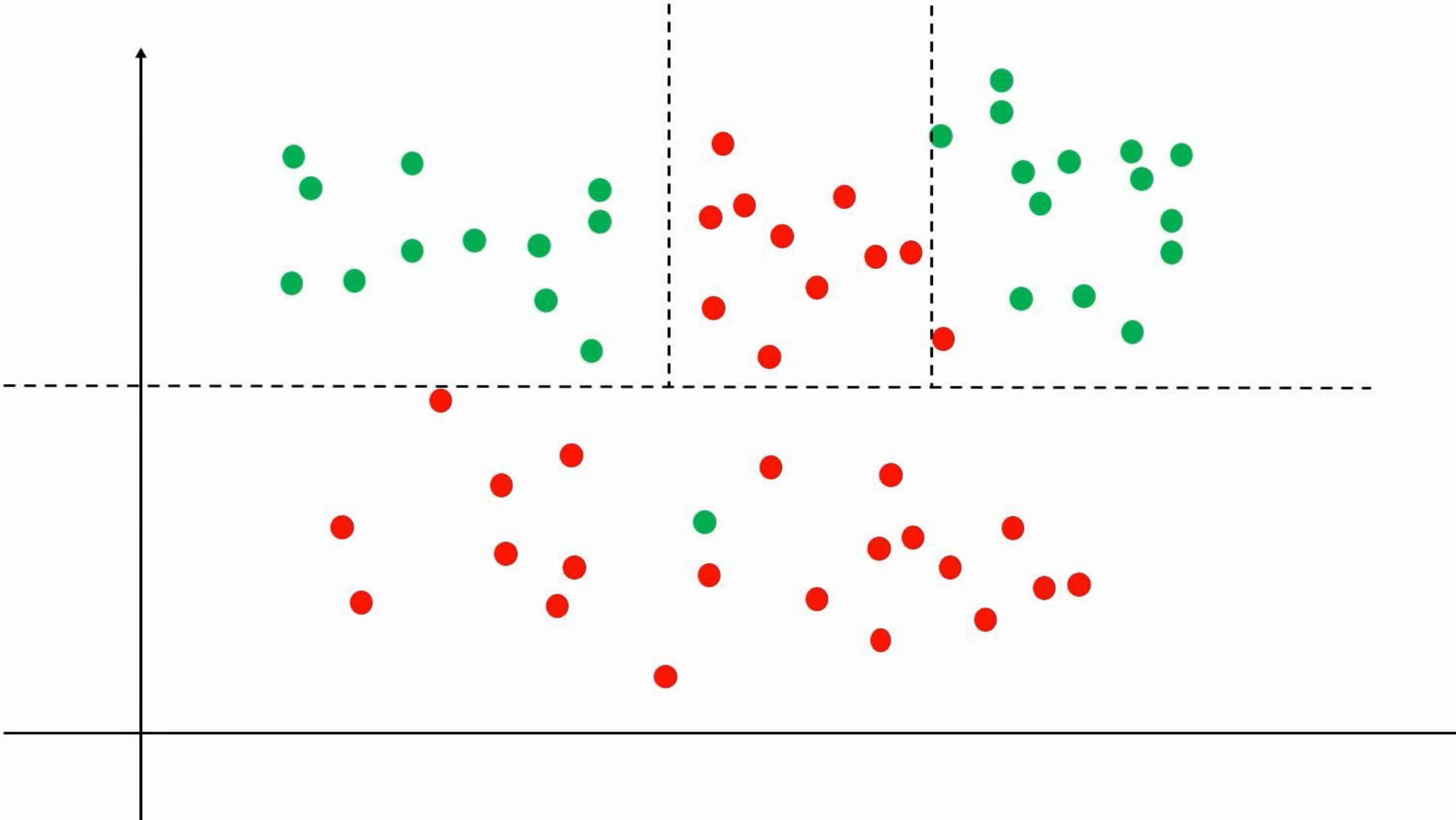
Parent/Child node

Decision Trees



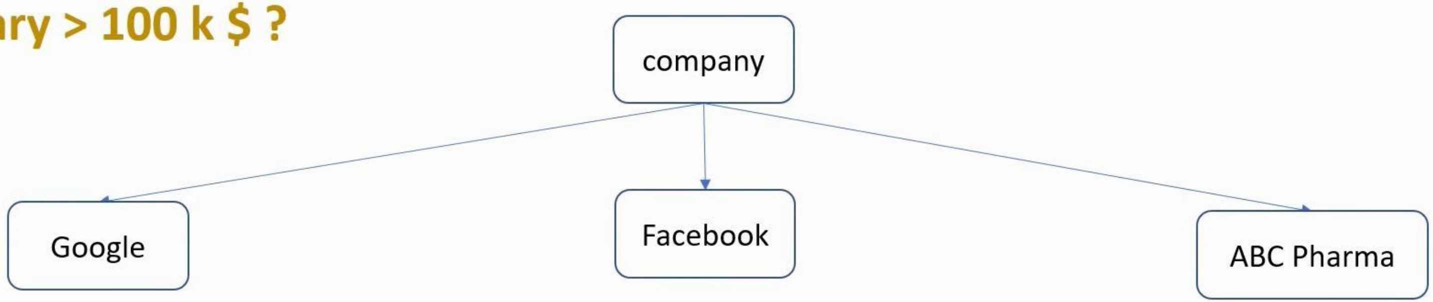






Company	Job	Degree	Salary_more_then_100k
google	sales executive	bachelors	0
google	sales executive	masters	0
google	business manager	bachelors	1
google	business manager	masters	1
google	computer programmer	bachelors	0
google	computer programmer	masters	1
abc pharma	sales executive	masters	0
abc pharma	computer programmer	bachelors	0
abc pharma	business manager	bachelors	0
abc pharma	business manager	masters	1
facebook	sales executive	bachelors	1
facebook	sales executive	masters	1
facebook	business manager	bachelors	1
facebook	business manager	masters	1
facebook	computer programmer	bachelors	1
facebook	computer programmer	masters	1

Salary > 100 k $?



google	sales executive	bachelors
google	sales executive	masters
google	business manager	bachelors
google	business manager	masters
google	computer programmer	bachelors
google	computer programmer	masters

?

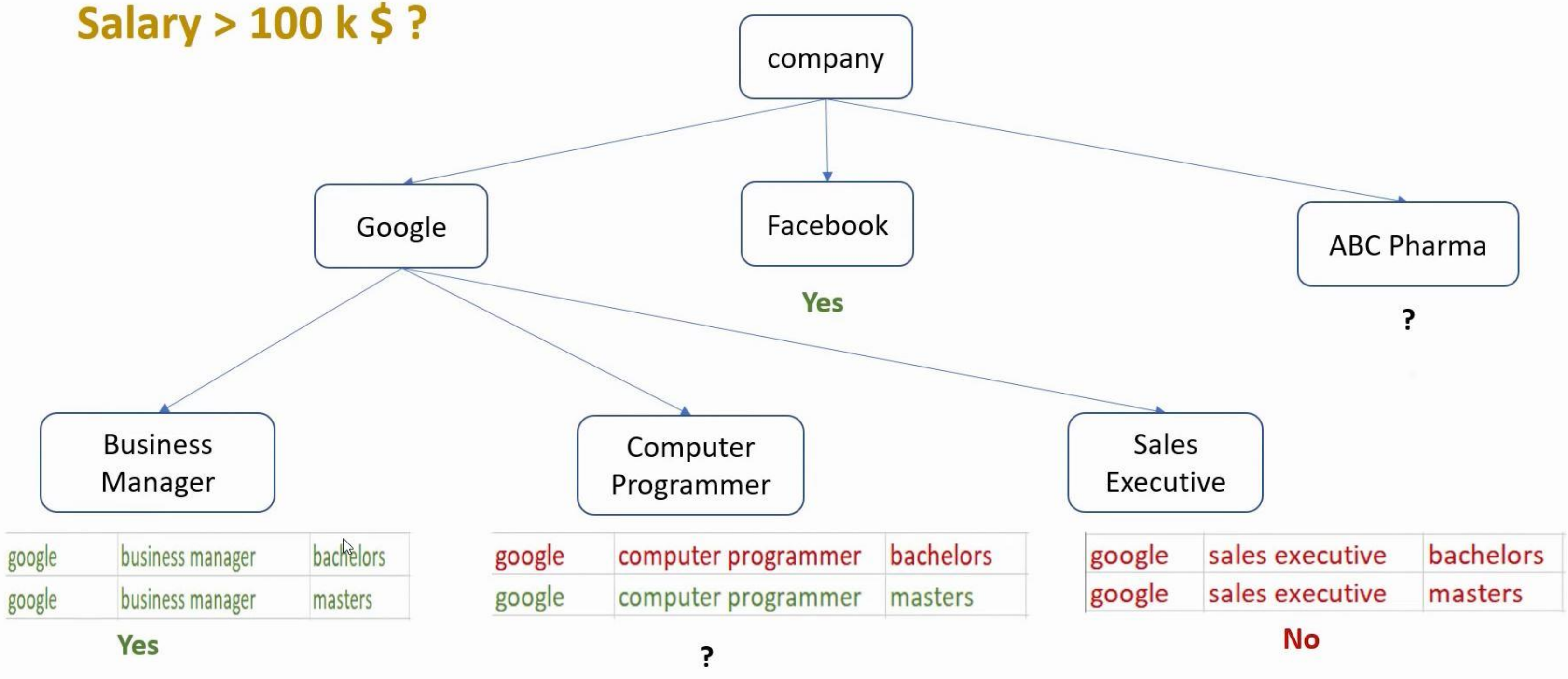
facebook	sales executive	bachelors
facebook	sales executive	masters
facebook	business manager	bachelors
facebook	business manager	masters
facebook	computer programmer	bachelors
facebook	computer programmer	masters

Yes

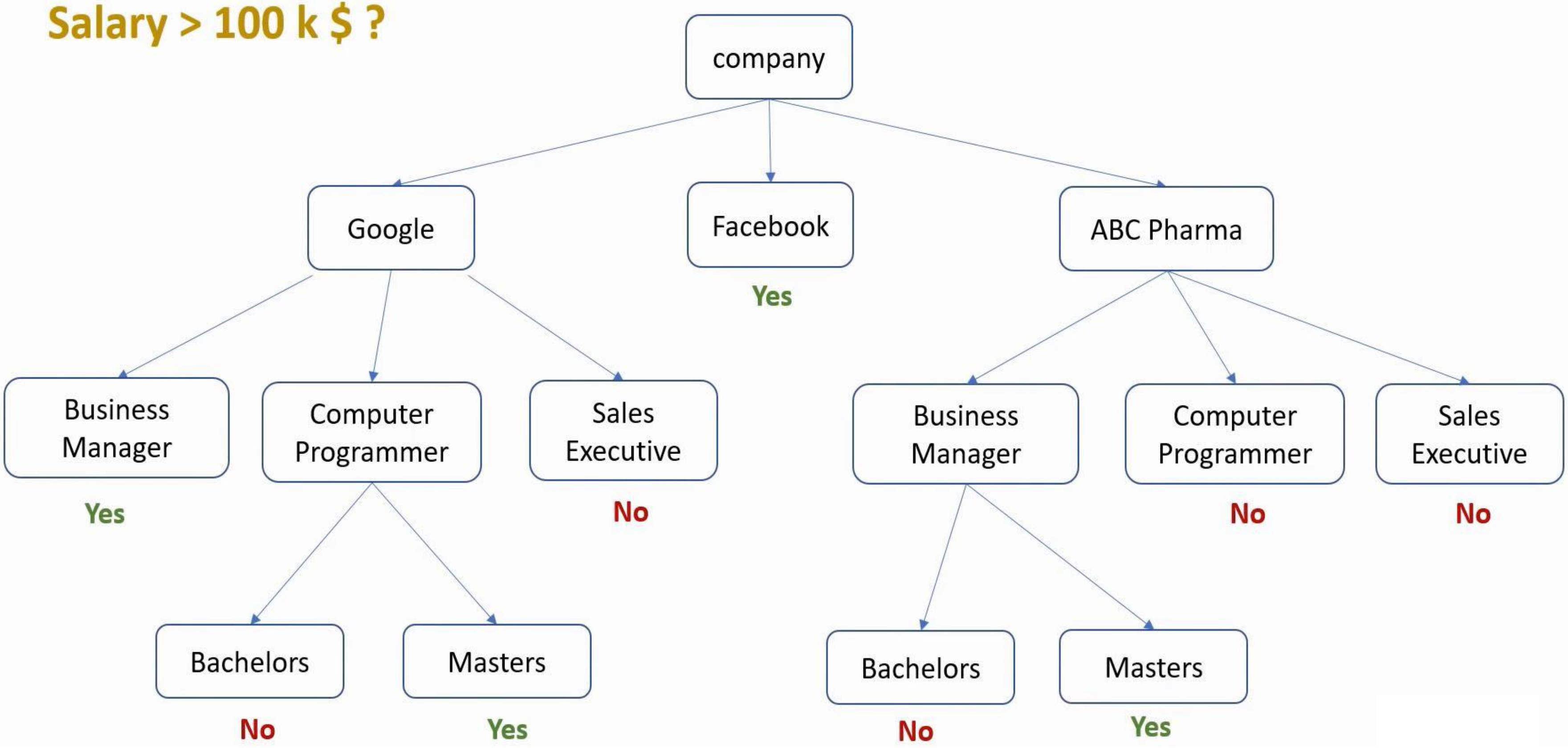
abc pharma	sales executive	masters
abc pharma	computer programmer	bachelors
abc pharma	business manager	bachelors
abc pharma	business manager	masters

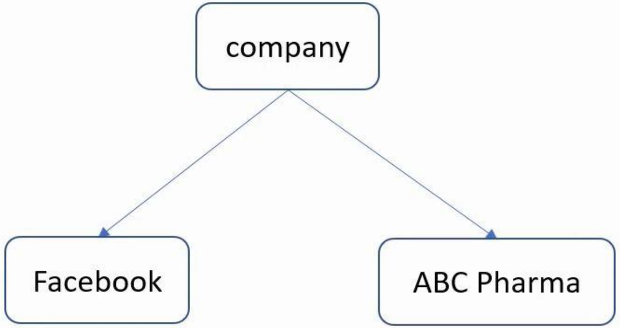
?

Salary > 100 k $?


company
Google
Facebook
ABC Pharma
Yes
?
Business Manager
Computer Programmer
Sales Executive
google business manager bachelors
google business manager masters
Yes
google computer programmer bachelors
google computer programmer masters
?
google sales executive bachelors
google sales executive masters
No

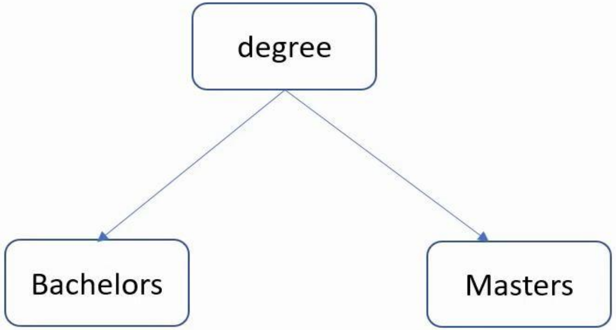
Salary > 100 k $?


company
Google
Facebook
ABC Pharma
Yes
Business Manager
Computer Programmer
Sales Executive
Business Manager
Computer Programmer
Sales Executive
Yes
No
No
No
Bachelors
Masters
Bachelors
Masters
No
Yes
No
Yes



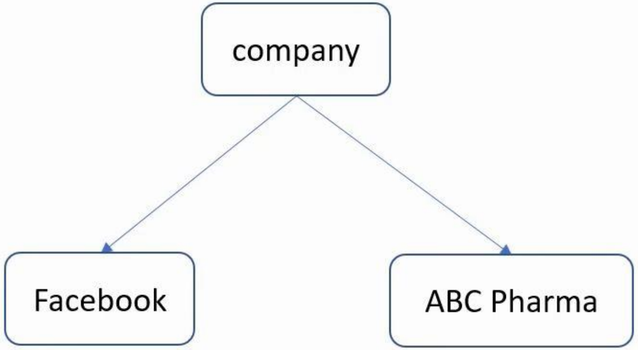
facebook	sales executive	bachelors
facebook	sales executive	masters
facebook	business manager	bachelors
facebook	business manager	masters
facebook	computer programmer	bachelors
facebook	computer programmer	masters

abc pharma	sales executive	masters
abc pharma	computer programmer	bachelors
abc pharma	business manager	bachelors
abc pharma	business manager	masters



google	sales executive	bachelors
google	business manager	bachelors
google	computer programmer	bachelors
abc pharma	computer programmer	bachelors
abc pharma	business manager	bachelors
facebook	sales executive	bachelors
facebook	business manager	bachelors
facebook	computer programmer	bachelors

google	sales executive	masters
google	business manager	masters
google	computer programmer	masters
abc pharma	sales executive	masters
abc pharma	business manager	masters
facebook	sales executive	masters
facebook	business manager	masters
facebook	computer programmer	masters



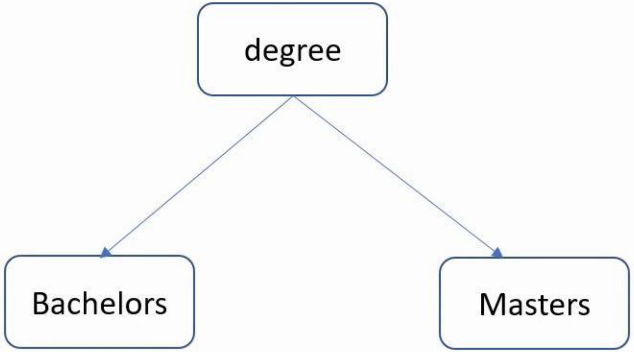
facebook	sales executive	bachelors
facebook	sales executive	masters
facebook	business manager	bachelors
facebook	business manager	masters
facebook	computer programmer	bachelors
facebook	computer programmer	masters

6 / 0 (low entropy)

abc pharma	sales executive	masters
abc pharma	computer programmer	bachelors
abc pharma	business manager	bachelors
abc pharma	business manager	masters

1 / 3

High Information Gain



google	sales executive	bachelors
google	business manager	bachelors
google	computer programmer	bachelors
abc pharma	computer programmer	bachelors
abc pharma	business manager	bachelors
facebook	sales executive	bachelors
facebook	business manager	bachelors
facebook	computer programmer	bachelors

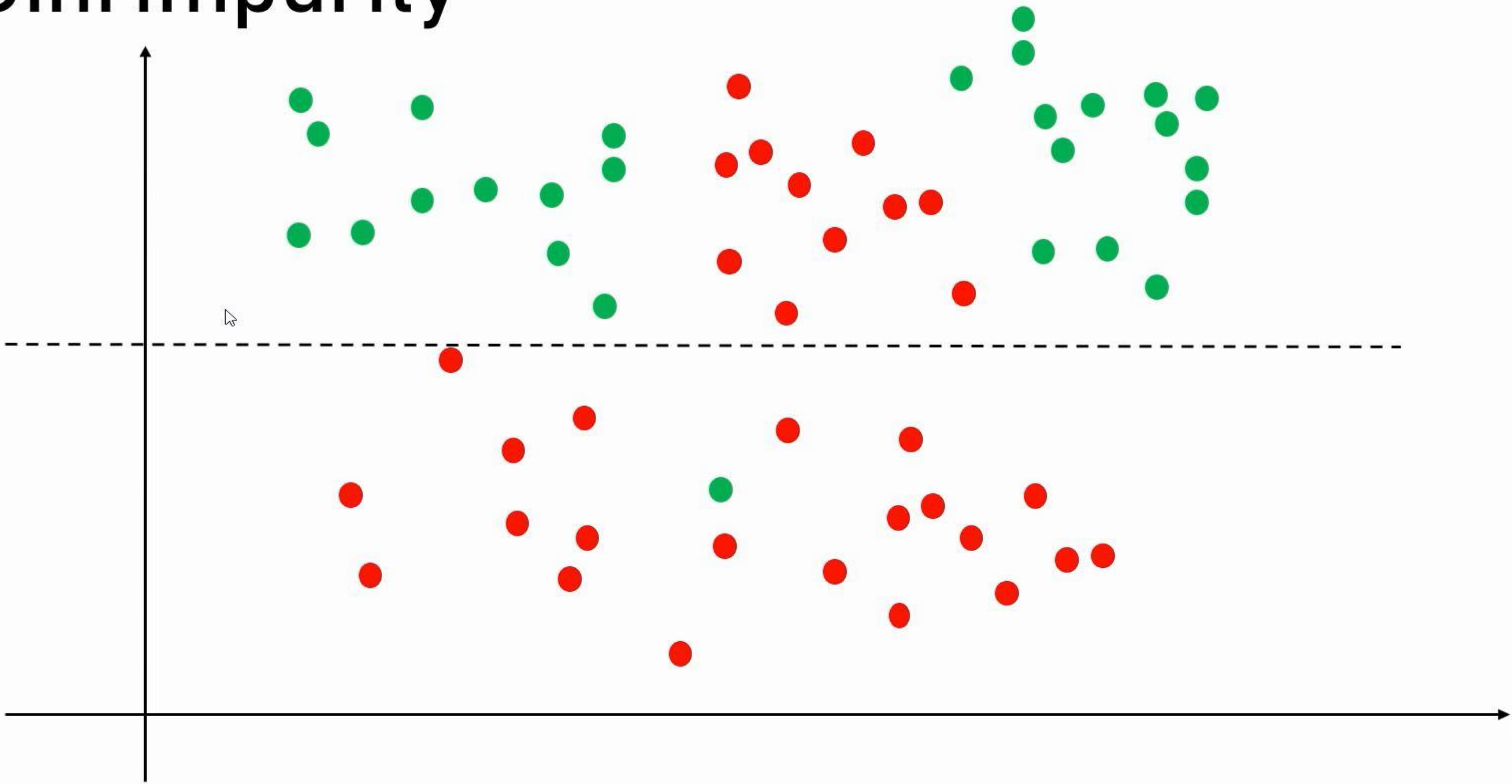
4 / 4 (high entropy)

google	sales executive	masters
google	business manager	masters
google	computer programmer	masters
abc pharma	sales executive	masters
abc pharma	business manager	masters
facebook	sales executive	masters
facebook	business manager	masters
facebook	computer programmer	masters

6 / 2

Low Information Gain

Gini Impurity

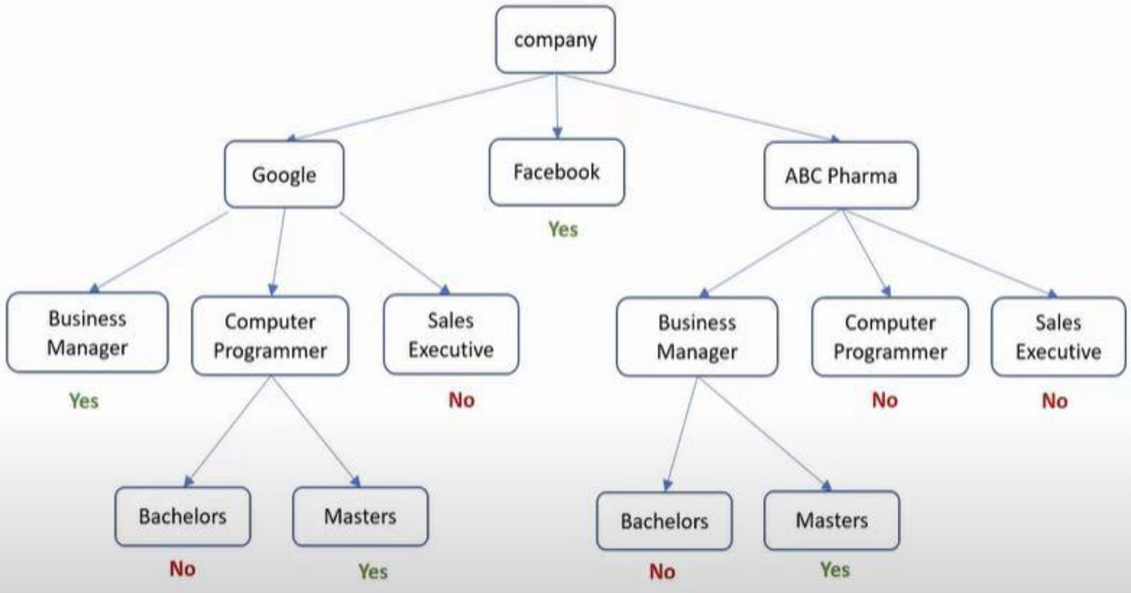


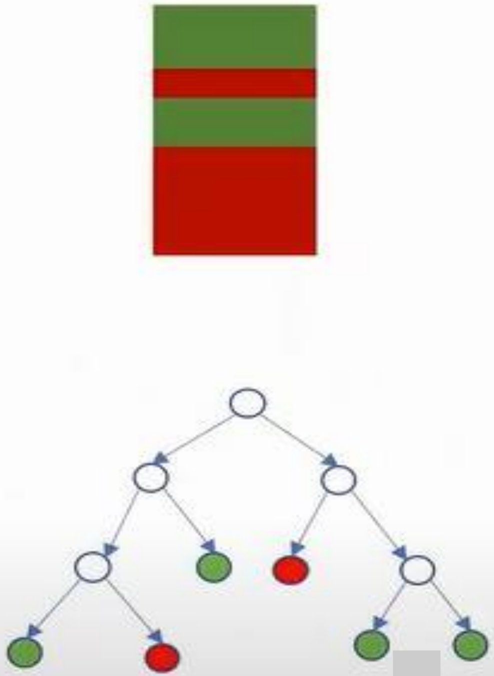
A	B	C	D	E	F	G	H	I	J	K	L
Passenger	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 2117	7.25		S
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2.	7.925		S
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1	C123	S
5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.05		S
6	0	3	Moran, Mr. James	male		0	0	330877	8.4583		Q
7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46	S
8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909	21.075		S
9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742	11.1333		S
10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736	30.0708		C
11	1	3	Sandstrom, Miss. Marguerite Rut	female	4	1	1	PP 9549	16.7	G6	S
12	1	1	Bonnell, Miss. Elizabeth	female	58	0	0	113783	26.55	C103	S
13	0	3	Saundercock, Mr. William Henry	male	20	0	0	A/5. 2151	8.05		S
14	0	3	Andersson, Mr. Anders Johan	male	39	1	5	347082	31.275		S
15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14	0	0	350406	7.8542		S
16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55	0	0	248706	16		S
17	0	3	Rice, Master. Eugene	male	2	4	1	382652	29.125		Q
18	1	2	Williams, Mr. Charles Eugene	male		0	0	244373	13		S
19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vandemoortele)	female	31	1	0	345763	18		S
20	1	3	Masselmani, Mrs. Fatima	female		0	0	2649	7.225		C
21	0	2	Fynney, Mr. Joseph J	male	35	0	0	239865	26		S
22	1	2	Beesley, Mr. Lawrence	male	34	0	0	248698	13	D56	S
23	1	3	McGowan, Miss. Anna "Annie"	female	15	0	0	330923	8.0292		Q
24	1	1	Sloper, Mr. William Thompson	male	28	0	0	113788	35.5	A6	S
25	0	3	Palsson, Miss. Torborg Danira	female	8	3	1	349909	21.075		S
26	1	3	Asplund, Mrs. Carl Oscar (Selma Augusta Emilia Johansson)	female	38	1	5	347077	31.3875		S
27	0	3	Emir, Mr. Farred Chehab	male		0	0	2631	7.225		C
28	0	1	Fortune, Mr. Charles Alexander	male	19	3	2	19950	263	C23 C25 C	S
29	1	3	O'Dwyer, Miss. Ellen "Nellie"	female		0	0	330959	7.8792		Q
30	0	3	Todoroff, Mr. Lalio	male		0	0	349216	7.8958		S
31	0	1	Uruchurtu, Don. Manuel E	male	40	0	0	PC 17601	27.7208		C
32	1	1	Spencer, Mrs. William Augustus (Marie Eugenie)	female		1	0	PC 17569	146.521	B78	C

Random Forest

Salary > 100 k $?

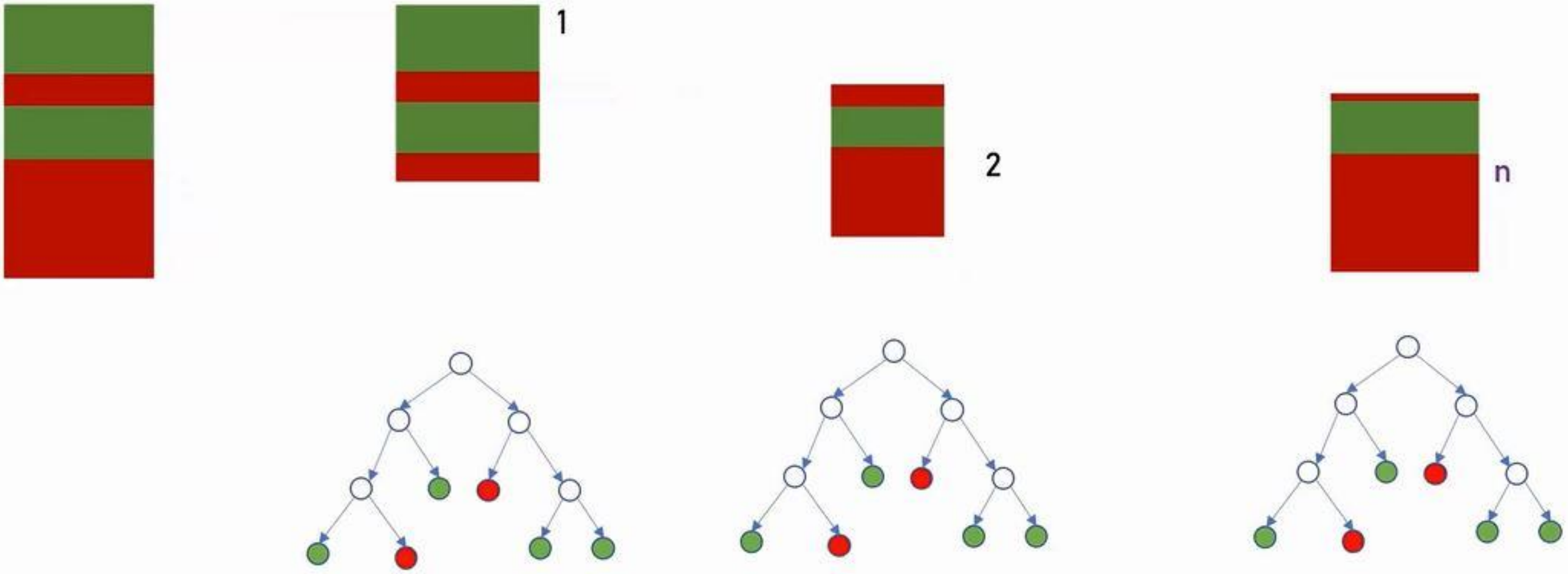
Company	Job	Degree	Salary_more_then_100k
google	sales executive	bachelors	0
google	sales executive	masters	0
google	business manager	bachelors	1
google	business manager	masters	1
google	computer programmer	bachelors	0
google	computer programmer	masters	1
abc pharma	sales executive	masters	0
abc pharma	computer programmer	bachelors	0
abc pharma	business manager	bachelors	0
abc pharma	business manager	masters	1
facebook	sales executive	bachelors	1
facebook	sales executive	masters	1
facebook	business manager	bachelors	1
facebook	business	masters	1

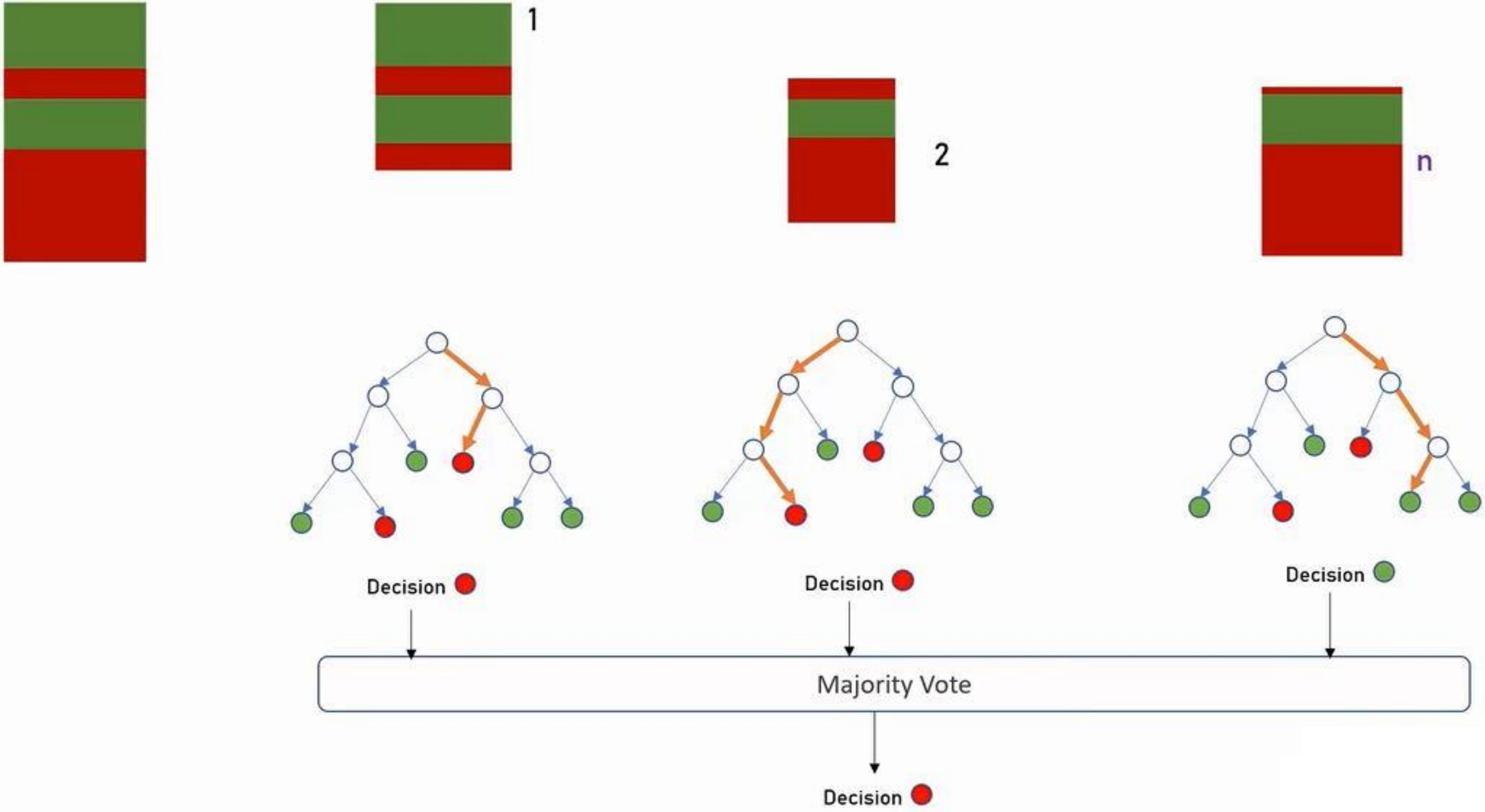




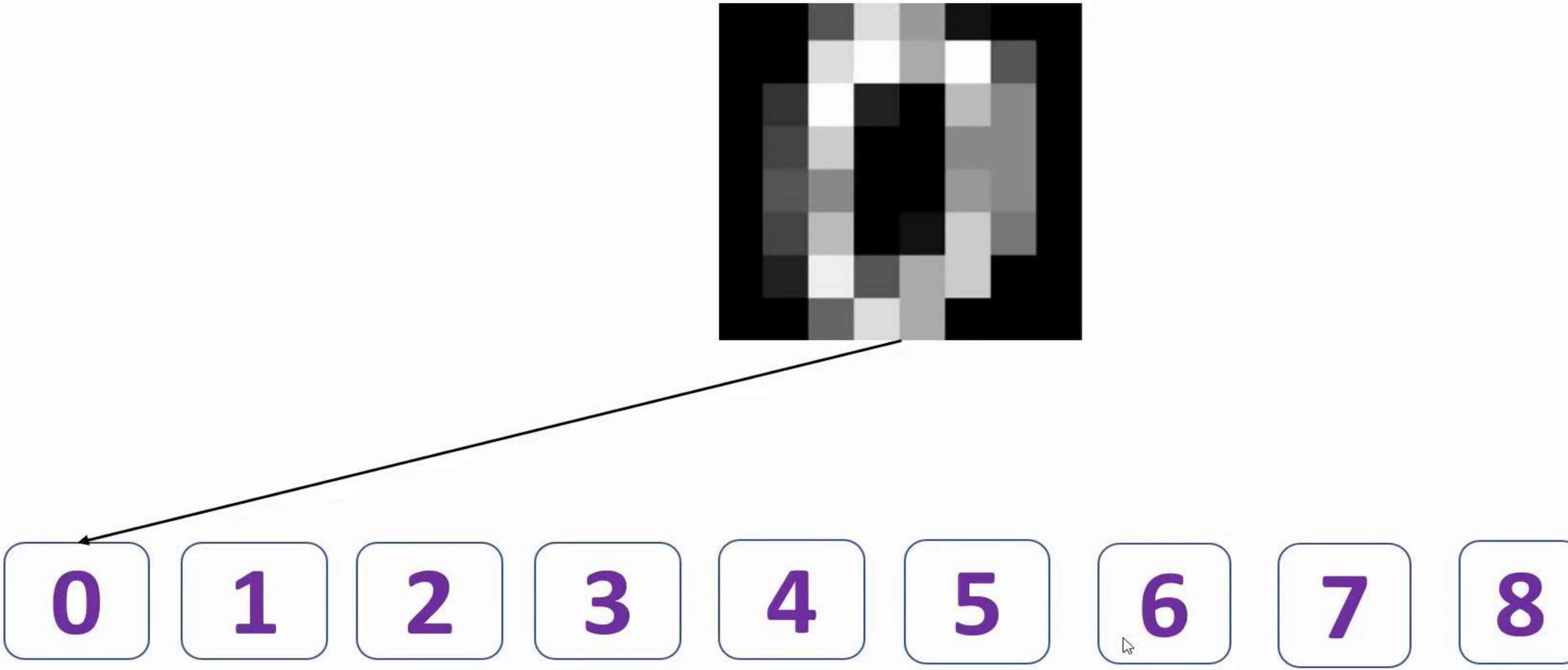
Decision tree based on green and red nodes


1
2
n


1
2
n

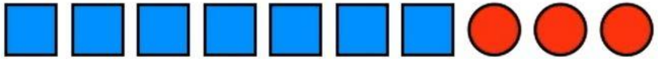

1
2
n
Decision
Decision
Decision
Majority Vote
Decision

Identify hand written digits recognition

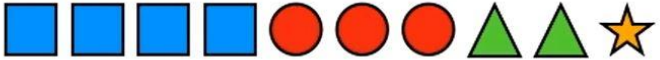


Which set is more diverse?

Gini = 0.42



Gini = 0.7



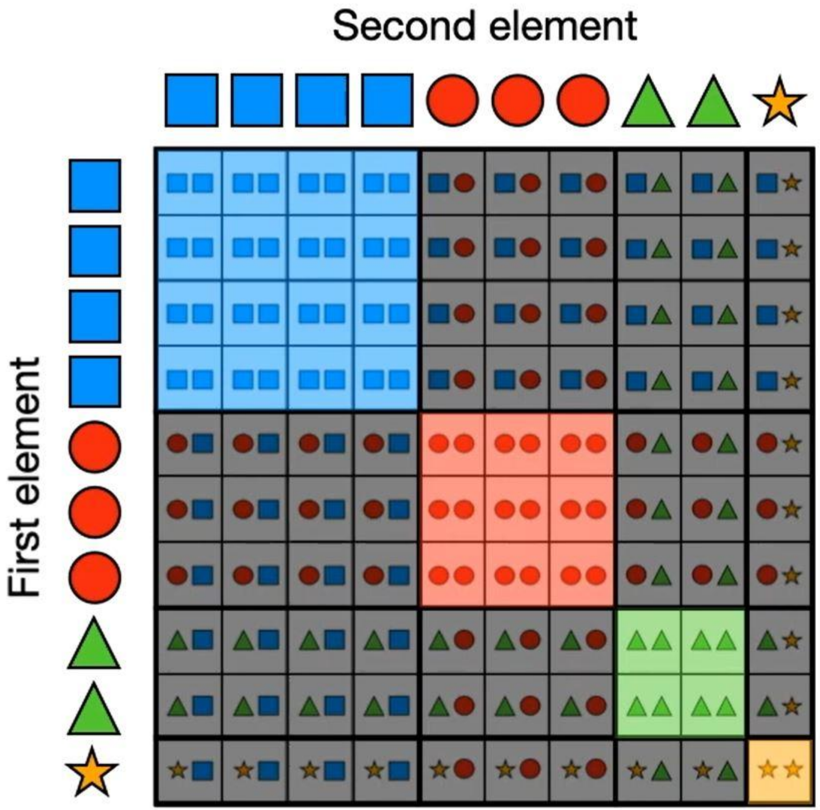
Gini Index = Probability of picking two distinct elements

■	■	Same
■	●	**Different**
●	■	**Different**
■	■	Same
●	●	Same
■	●	**Different**
●	●	Same
■	■	Same
●	■	**Different**
■	■	Same

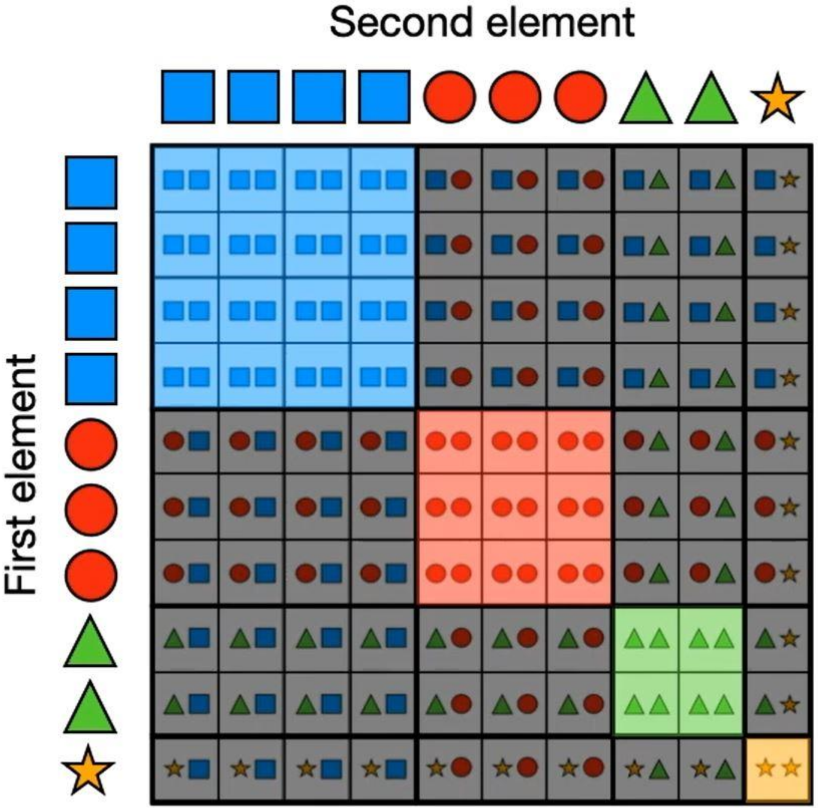
Different:
4 out of 10

●	▲	**Different**
■	■	Same
▲	■	**Different**
★	●	**Different**
■	▲	**Different**
■	■	Same
●	●	Same
▲	●	**Different**
■	★	**Different**
●	■	**Different**

Different:
7 out of 10

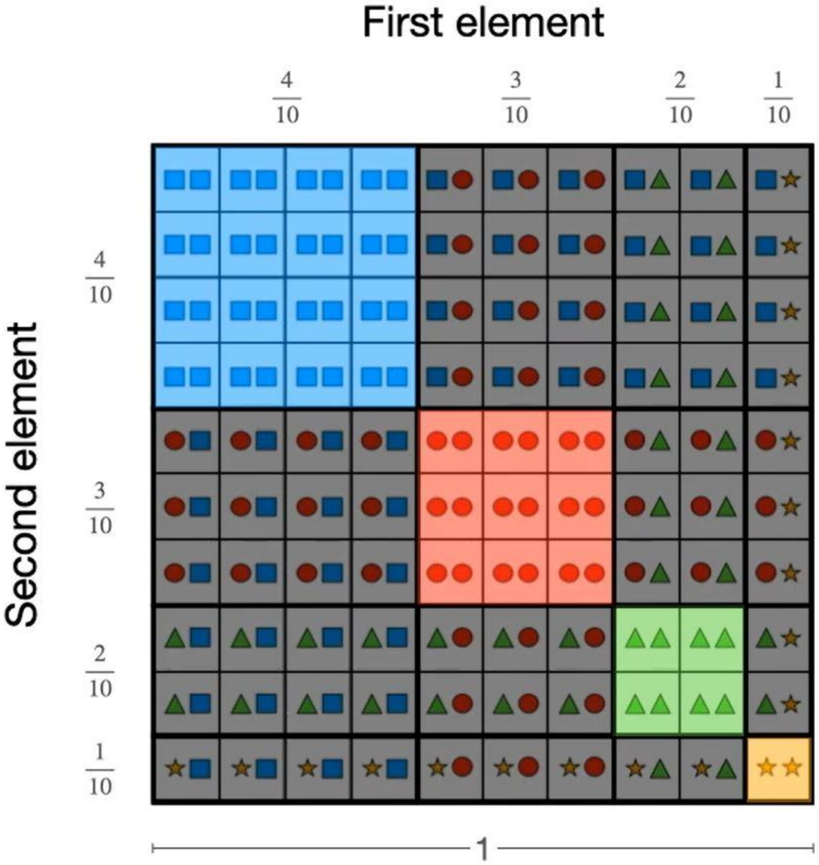

Second element
First element


=
-



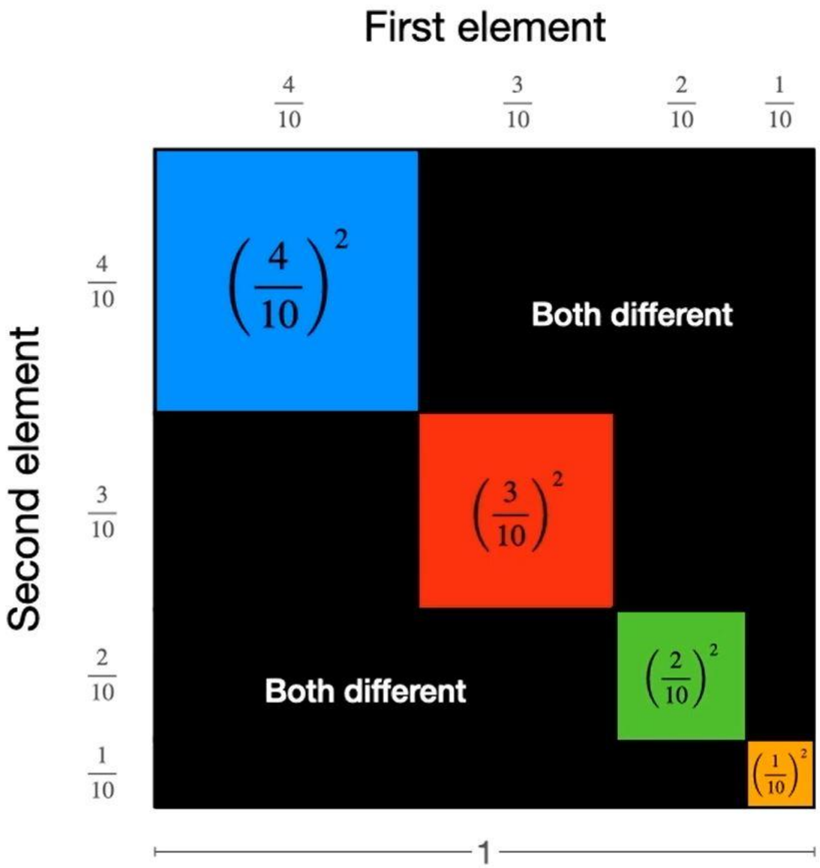
P(Both different) = P(Anything) - P(Both equal)

= 1 - P(Both blue) - P(Both red) - P(Both green) - P(Both yellow)



P(Both different) = P(Anything) - P(Both equal)

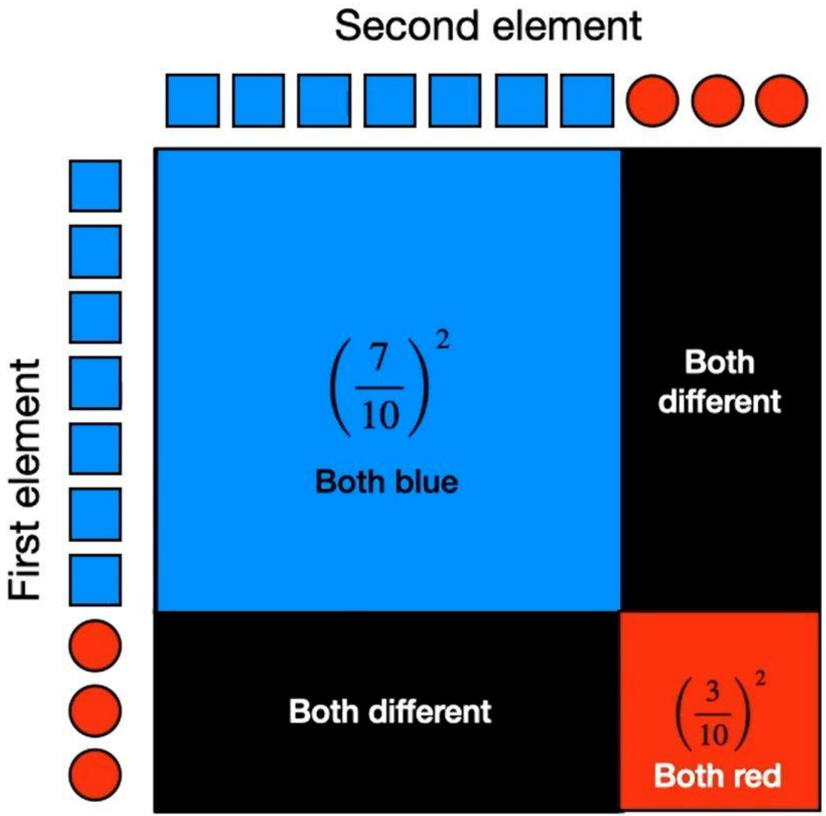
= 1 - P(Both blue) - P(Both red) - P(Both green) - P(Both yellow)



P(Both different) = P(Anything) - P(Both equal)

= 1 - P(Both blue) - P(Both red) - P(Both green) - P(Both yellow)

$$= 1 - \left(\frac{4}{10}\right)^2 - \left(\frac{3}{10}\right)^2 - \left(\frac{2}{10}\right)^2 - \left(\frac{1}{10}\right)^2$$



$$P(\text{Both different}) = P(\text{Anything}) - P(\text{Both equal})$$

$$= 1 - P(\text{Both blue}) - P(\text{Both red})$$

$$= 1 - \left(\frac{7}{10}\right)^2 - \left(\frac{3}{10}\right)^2$$

$$= 1 - 0.7^2 - 0.3^2$$

$$= 1 - 0.49 - 0.09$$

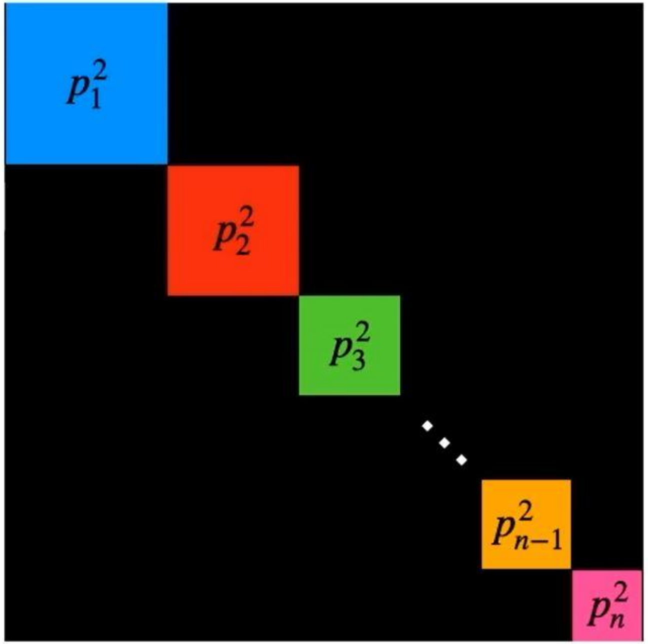
$$= 0.42$$

General formula

n classes

Proportions: $p_1, p_2, \ldots, p_n$

Gini impurity index: $1 - p_1^2 - p_2^2 - \cdots - p_n^2$



General formula

n classes

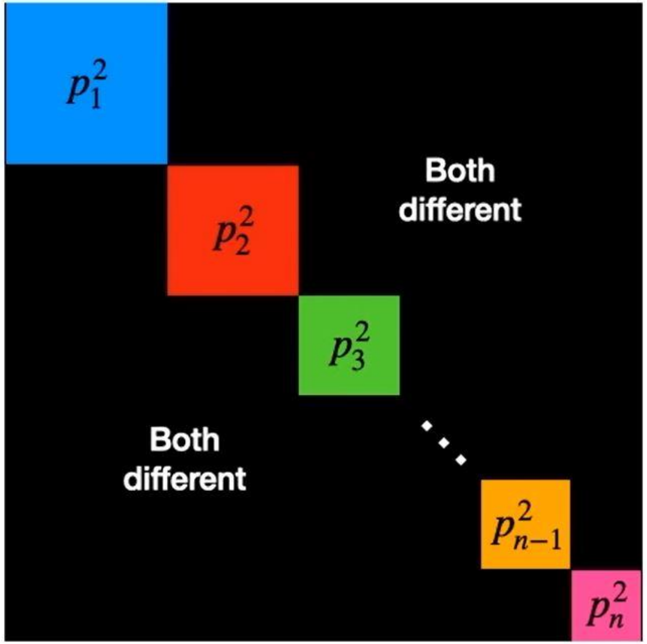
Proportions: $p_1, p_2, \ldots, p_n$

Gini impurity index: $1 - p_1^2 - p_2^2 - \cdots - p_n^2$

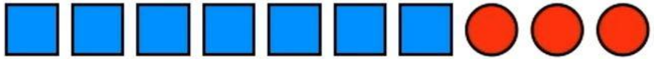
P(Both different)

P(Anything)

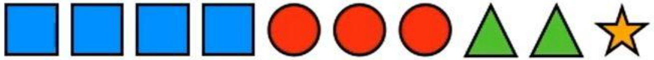
P(Both equal)



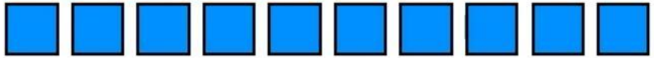
Which set is more diverse?



Gini = 0.42



Gini = 0.7



Gini = 0

$$1 - 1^2 = 0$$

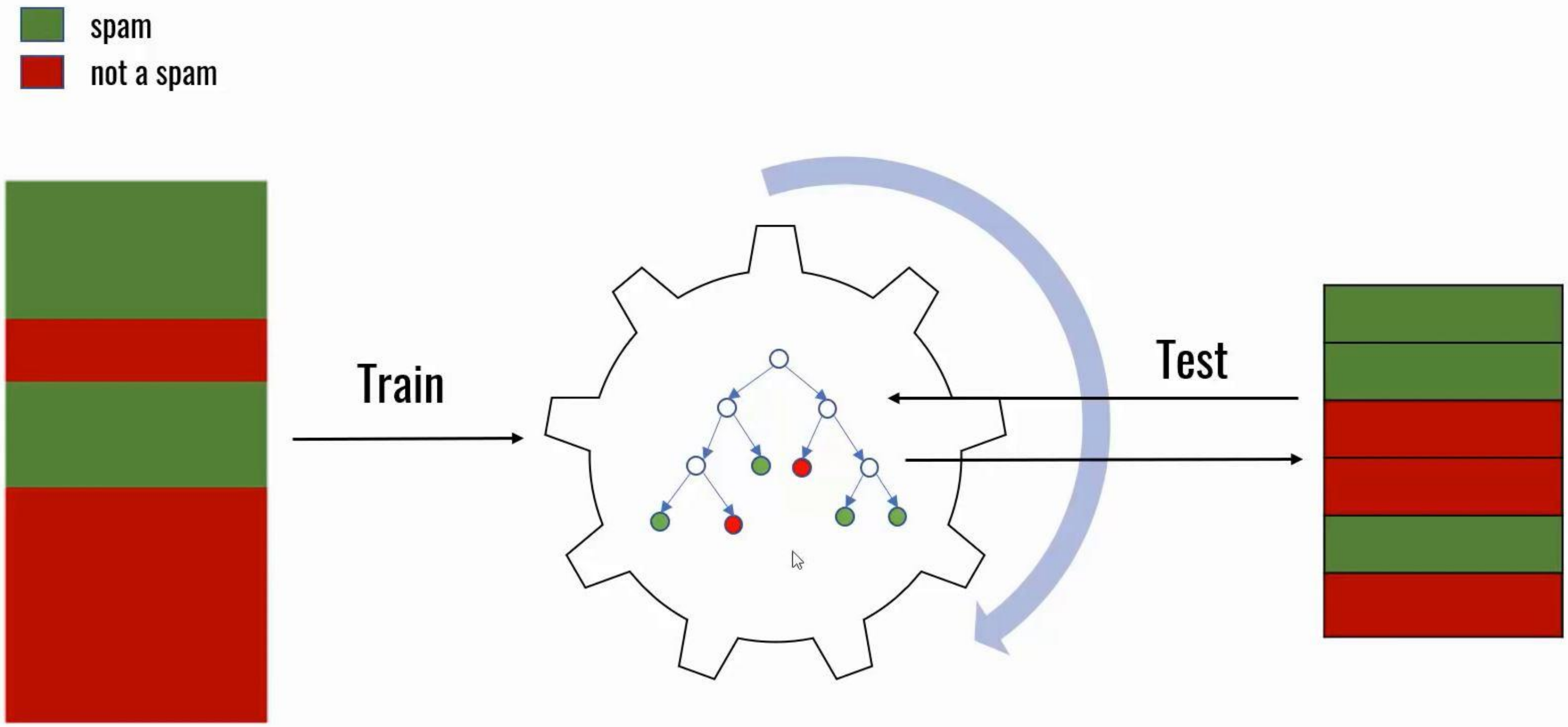
Gini = ?

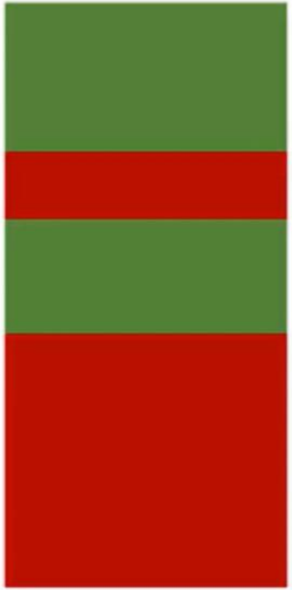
$$1 \underbrace{- 0.1^2 - 0.1^2 - \cdots - 0.1^2}_{\text{10 times}} = 0.9$$

Introduction to K Fold Cross Validation….

need?

Option 1:





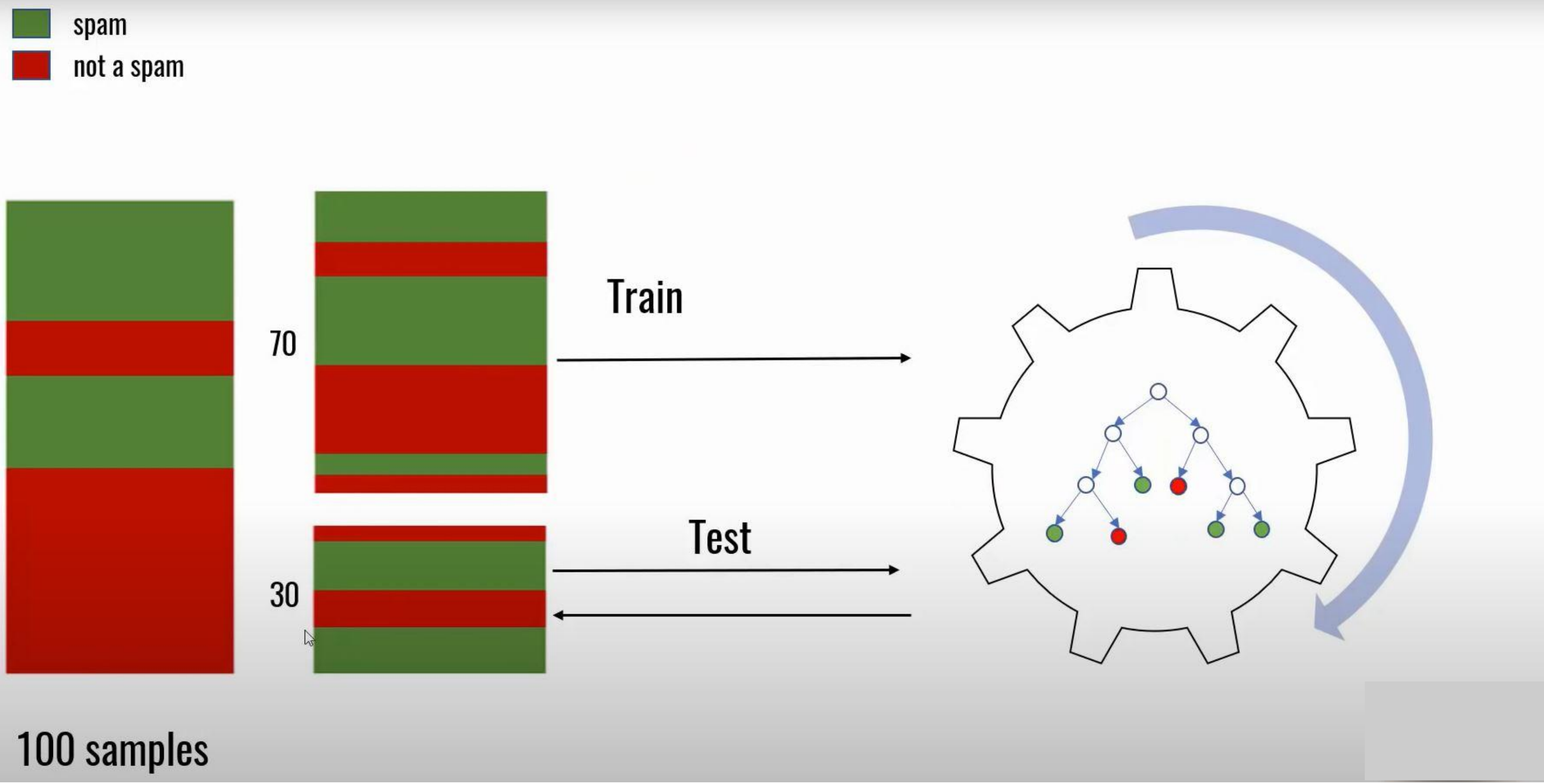
100 math questions

Train

Test

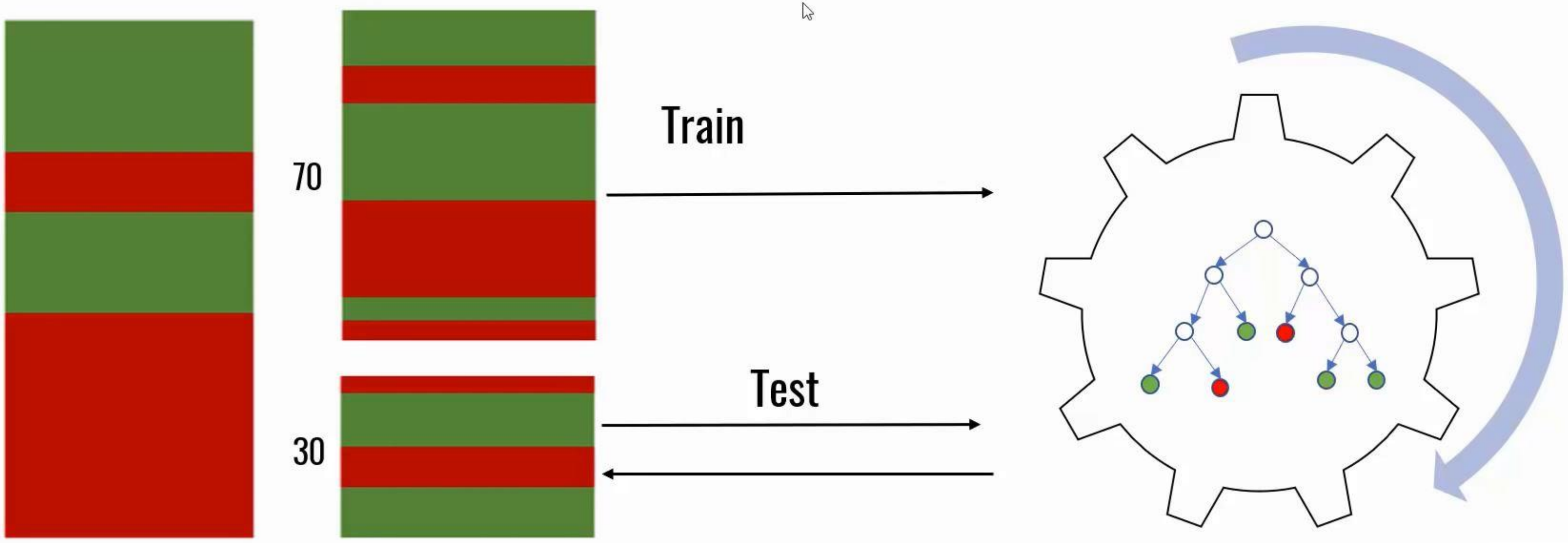
Same 100 math questions
as test set

Option 2:

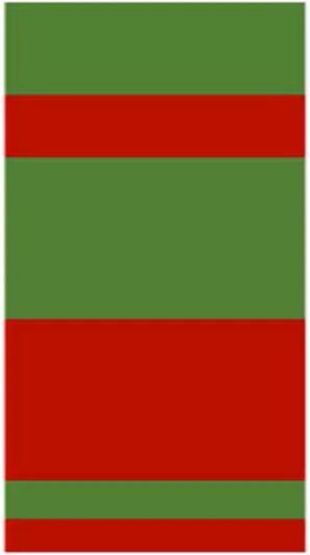


```python
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.3)
```

spam
not a spam



100 samples



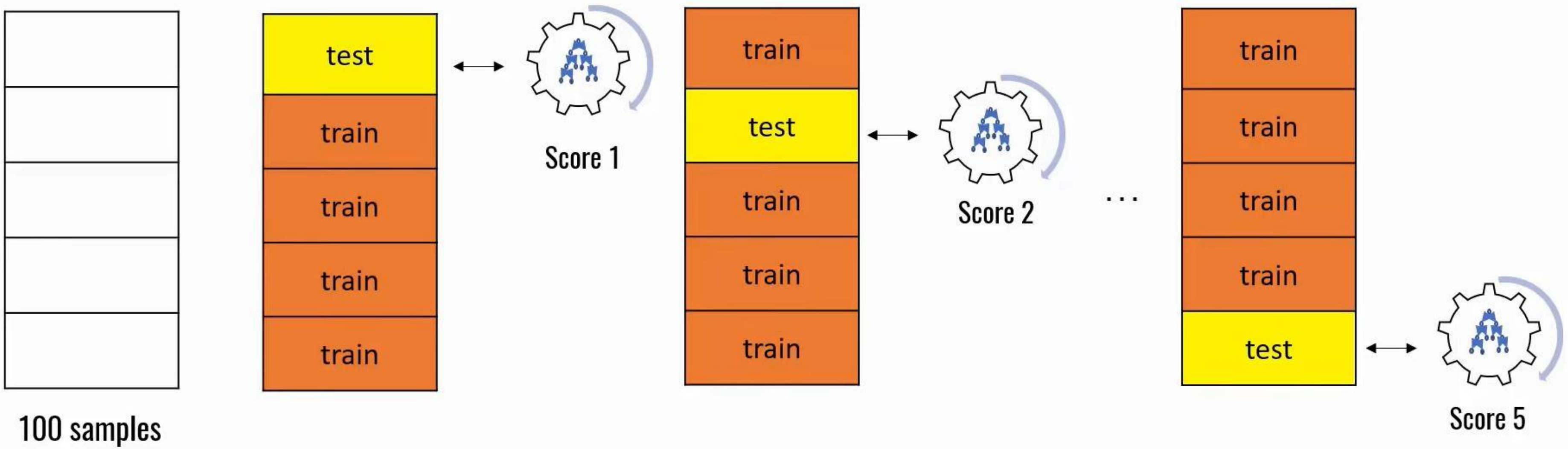
Train

Test

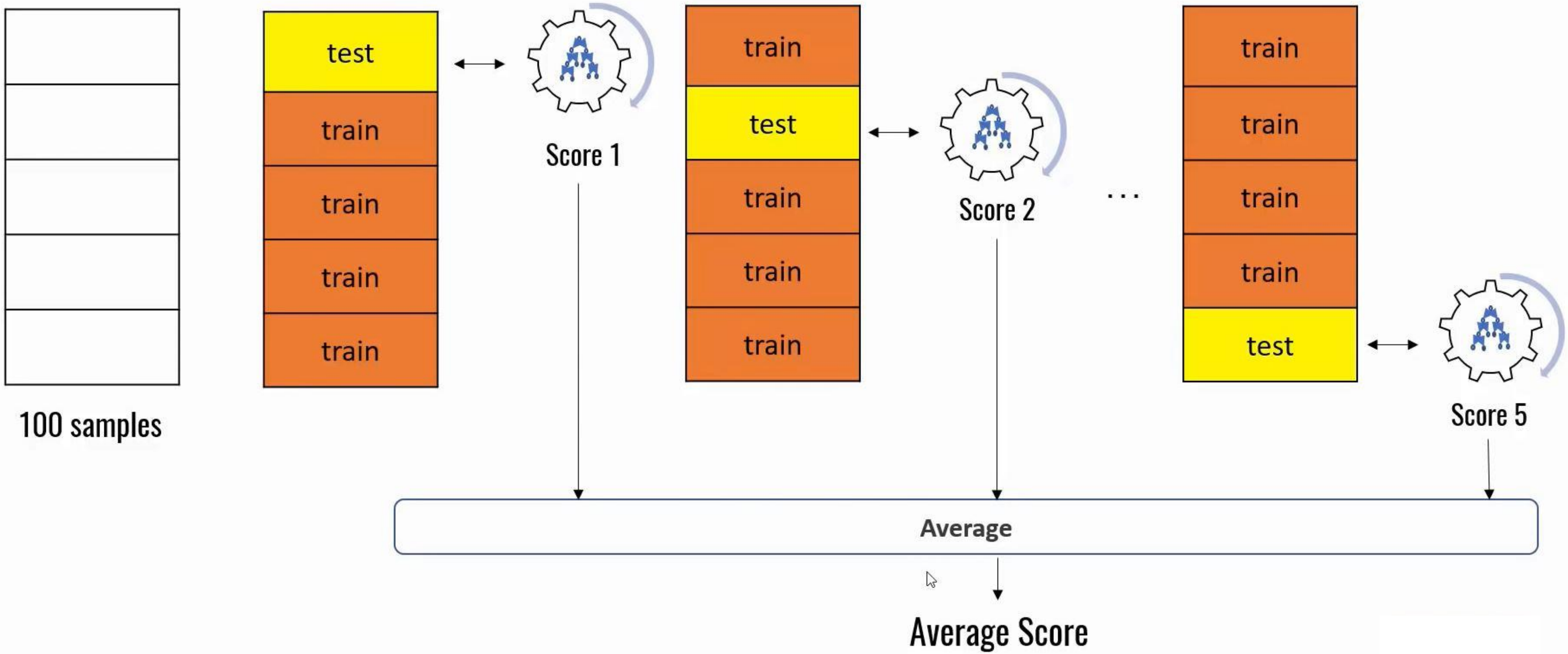
Remaining 30 questions

70 math questions

Option 3:



100 samples
test
train
train
train
train
Score 1
train
test
train
train
train
Score 2
. . .
train
train
train
train
test
Score 5



100 samples
test
train
train
train
train
Score 1
train
test
train
train
train
Score 2
. . .
train
train
train
train
test
Score 5
Average
Average Score