

①

9/1/2023

Monday

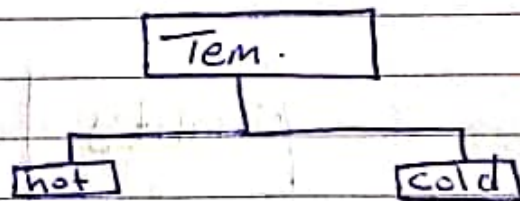
Types of Decision Tress

It can be spit at each level to be a two-way split. Then we can split any variable type using split criteria as:-

1- Dichotomy

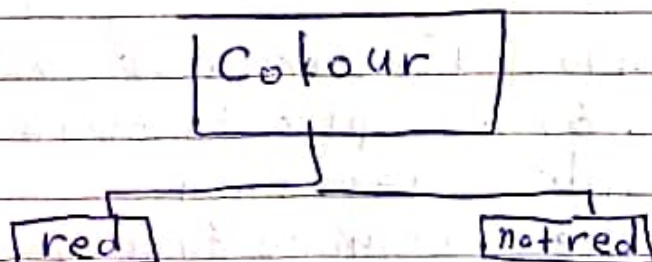
Variables with values are simple to split since each branch represent a specific value

Example:- Temperature can be classified to Hot, cold



2. Nominal

Since nominal values are discrete values with a two-way can be shown as follows



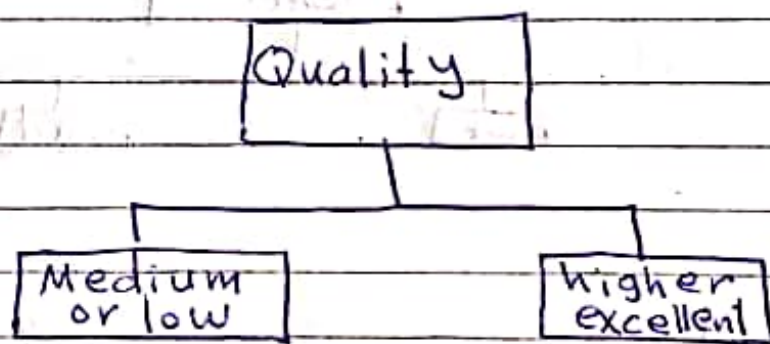
المتغير المتألف من شجرة القرار يتم استنتاجه مع المتغيرات الوصفية (الاسمية)

② المتغير المنقطع يكون اسمياً وغير قابل للترتيب أي لا توجد قاعدة لتعيينه لذا لا يمكن أن يبدأ معين في الترتيب فلو كان لدينا لون معين فليس التمييز على أساس درجة اللون أو إذا كان لون آخر

③ Ordinal

In this case the variable has discrete values with orders, then the subsets have more than one values

Example:- In one collage there is a department which gives Bich, M.sc., D.ph and has specified degree



هذا مثال وجود أربع أنواع ولكن تم تجزئتها فقط إلى جزئين وكل جزء من قسم

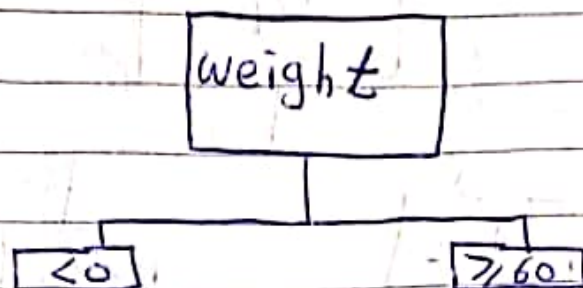
④ Continuous: For variables with continuous values to be split to two-way

① one side is less than

② The other is (more than or equal)

(3)

يجب توضيح انه الماوان في احد الاطراف



Rank of Split

There are two approaches for split based on whether the response is categorical or continuous

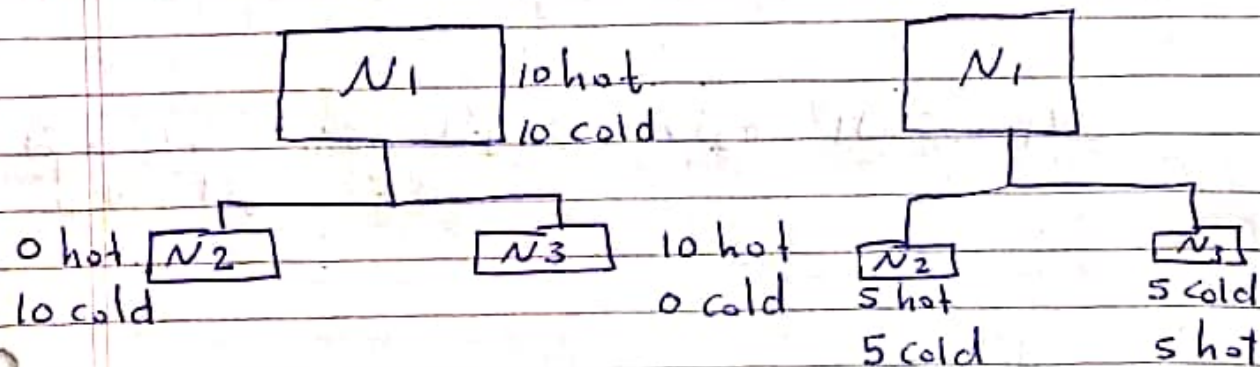
a. Scoring Split for Categorical Variables:-

To illustrate how the score splits when the response is categorical variables these splits (a, b, c) shown in this examples:-

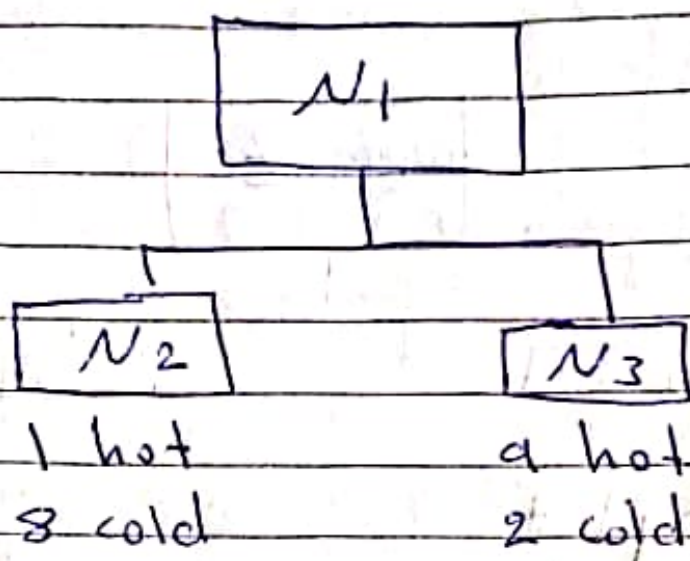
Example:- Twenty areas: ten areas are hot and another are cold so we split them depending on the temperature degree

split a

split b



split c



Scoring split For continuous Response:
For the continuous variables we find the score of every node by using (SSE) (sum of squares of error)

$$SSE = \sum_{i=1}^n (y_i - \bar{y})^2$$

$\bar{y}$ = average value for the subset

y_i = is the individual values for the response

نمنا SSE لكل عقده ونقوده العقده التي
تنتجها أقل SSE

5

Example: - The following table containing information about weight (x) and MpG (miles per Gallon) (y)

Weight	1850	1770	1610	1830	4610	4730	4950	4740
x								
MpG	26	31	35	27	10	9	12	13
y								

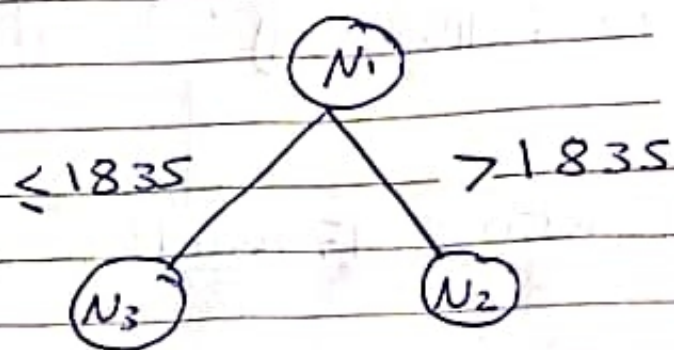
If the splitting values for variable weight (x) are (1835, 3250, 4650). Calculate the score of the MpG response (x) and find the best split

Solution

① ترتيب الوزن لتباينها	
weight	MpG
1610	35
1770	31
1830	27
1850	26
4610	10
4730	9
4740	13
4950	12

② تكون العقد

Split 1 : 1835



node 2

$$\bar{y} = \frac{26 + 10 + 9 + 13 + 12}{5} = 14$$

$$SSE_1 = \sum_{i=1}^5 (y_i - \bar{y})^2$$

$$= (26 - 14)^2 + (10 - 14)^2 + (9 - 14)^2 + (13 - 14)^2 + (12 - 14)^2$$

$$= 190$$

node 3

$$\bar{y} = \frac{35 + 31 + 27}{3} = 31$$

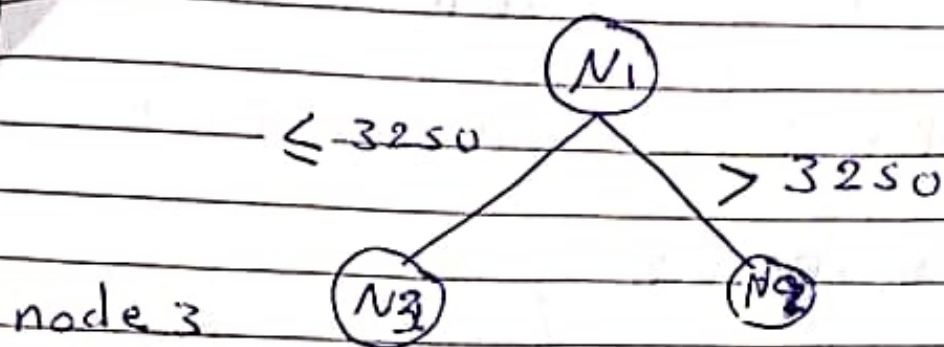
$$SSE = (35 - 31)^2 + (31 - 31)^2 + (27 - 31)^2$$

$$= 32$$

$$\text{Split score of split 1} = 190 + 32 = 222$$

(7)

split 2 : 3250



$$\bar{y} = \frac{35 + 31 + 27 + 26}{4} = 29.75$$

$$SSE_1 = (35 - 29.75)^2 + (31 - 29.75)^2 + (27 - 29.75)^2 + (26 - 29.75)^2 = 50.75$$

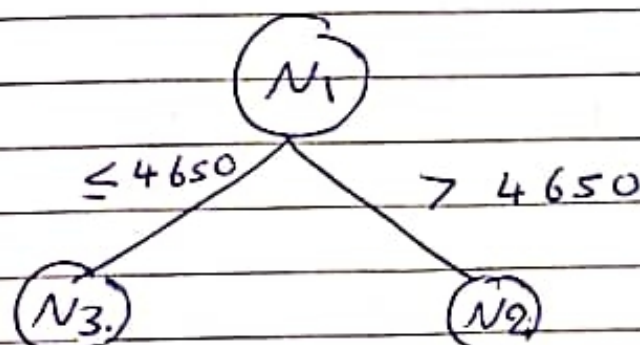
Node 2

$$\bar{y} = 11$$

$$SSE_2 = 10$$

$$SSE(\text{split 2}) = 50.75 + 10 = 60.75$$

split 3



node 3

$$\bar{y} = 25.8$$

$$SSE = 362.8$$

node 2

$$\bar{y} = 11.35$$

$$SSE = 8.67$$

(5)

$$\text{Score for split 3} = 362.8 + 8.67 \\ = 371.4$$

Since Split 2 has smaller SSE

∴ we take split 2

ملاحظات
1- عليه التجزئة تبين عملية التمييز
أي يتم وضع المتاحات المتأخرة مع بعضها

2- إذا كانت التجزئة كـ L1 أو MP6 هنا يفترض
الوزن هو المتاحات (أي يتم أخذ قيم الأوزان في
حساب المعدل SSE لكل عقدة)

3- إذا لم يتم إظهار أوزان قياسية فهذا يكون عند
عدد الأوزان