

Si consideri il seguente scenario di classificazione binaria in cui si vuole predire se il capitale annuo supera i 50K

Age	Workclass	Education	Marital-status	Sex	Capital
39	State-gov	Bachelors	Never-married	Male	<=50K
50	Self-emp-not-inc	Bachelors	Married-civ-spouse	Male	<=50K
38	Private	HS-grad	Divorced	Male	<=50K
53	Private	11th	Married-civ-spouse	Male	<=50K
31	Private	Masters	Never-married	Female	>50K
42	Private	Bachelors	Married-civ-spouse	Male	>50K
37	Private	Some-college	Married-civ-spouse	Male	>50K

Esercizio 1

Trova almeno 2 criteri di split che ci forniscano una decisione certa in almeno un ramo.

Esercizio 2

Se volessimo definire uno split sull'attributo "Education" che generi due raggruppamenti quale potrebbe essere il miglior criterio da utilizzare secondo la metrica di *Gini Impurity*?

$$Gini\ Impurity = 1 - \sum_{k=1}^K f_k^2$$

Esercizio 3

Se volessimo definire uno split sull'attributo "Marital-Status" che generi due raggruppamenti quale potrebbe essere il miglior criterio da utilizzare secondo la metrica di *Information Gain*?

$$Entropy = -\sum_{k=1}^K f_k * \log_2 f_k$$

$$InformationGain = Entropy(before) - Entropy(after) = Entropy(parent) - Entropy(children)$$

Esercizio 4

Si considerino i seguenti due criteri di split (mostrati come raggruppamenti nelle rispettive tabelle) e valutare quale dei due criteri risulta essere la scelta migliore. Applicare almeno una metrica di valutazione del criterio di split.

Criterio #1: *Education = 'Bachelors' OR 'Masters' OR 'some-college'?*

Age	Workclass	Education	Marital-status	Sex	Capital	
39	State-gov	Bachelors	Never-married	Male	<=50K	Group1
50	Self-emp-not-inc	Bachelors	Married-civ-spouse	Male	<=50K	Group1
38	Private	HS-grad	Divorced	Male	<=50K	Group2
53	Private	11th	Married-civ-spouse	Male	<=50K	Group2
31	Private	Masters	Never-married	Female	>50K	Group1
42	Private	Bachelors	Married-civ-spouse	Male	>50K	Group1
37	Private	Some-college	Married-civ-spouse	Male	>50K	Group1

Criterio #2: *Marital-status* = '*Married-civ-spouse*'?

Age	Workclass	Education	Marital-status	Sex	Capital	
39	State-gov	Bachelors	Never-married	Male	<=50K	Group2
50	Self-emp-not-inc	Bachelors	Married-civ-spouse	Male	<=50K	Group1
38	Private	HS-grad	Divorced	Male	<=50K	Group2
53	Private	11th	Married-civ-spouse	Male	<=50K	Group1
31	Private	Masters	Never-married	Female	>50K	Group2
42	Private	Bachelors	Married-civ-spouse	Male	>50K	Group1
37	Private	Some-college	Married-civ-spouse	Male	>50K	Group1