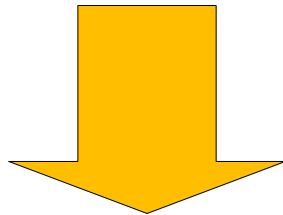


Apprendimento

Rote learning (apprendimento meccanico)

Strategia: memorizzare le istanze

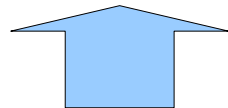


att1	att2	att3	att4	classe
a	b	b	d	A
a	f	b	b	B
b	b	c	d	A
c	c	d	f	C

MODELLO: non esiste in senso proprio
Il sistema ricorda dei “casi”, non generalizza

att1	att2	att3	att4	classe
a	b	b	d	A
a	f	b	b	B
b	b	c	d	A
c	c	d	f	C

MEMORIA



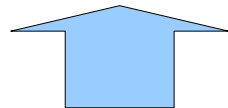
< a, f, b, b >

ISTANZA

Tramite confronti cerca un'istanza identica, se la trova restituisce la classe corrispondente

att1	att2	att3	att4	classe
a	b	b	d	A
a	f	b	b	B
b	b	c	d	A
c	c	d	f	C

MEMORIA



< a, f, b, d >

ISTANZA

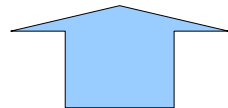
Se non la trova prova a indovinare:

Cerca istanze simili

Misura di distanza come misura di similitudine
(vicino = simile)

att1	att2	att3	att4	classe
a	b	b	d	A
a	f	b	b	B
b	b	c	d	A
c	c	d	f	C

MEMORIA



< a, f, b, d >

ISTANZA

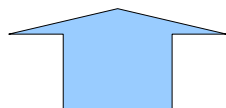
Se non la trova prova a indovinare:

Cerca istanze **simili** ...:

se le loro classi sono identiche, quello è l'output

att1	att2	att3	att4	classe
a	b	b	d	A
a	f	b	b	B
b	b	c	d	A
c	c	d	f	C

MEMORIA



< a, f, b, d >

ISTANZA

Se non la trova prova a indovinare:

Se **sono diverse** occorre attuare una strategia per comporre il risultato, es.

- **Votazione a maggioranza:**
corrispondenza = voto, la classe più votata vince
- **Votazione pesata:**
ogni voto ha un peso maggiore/minore a seconda della “distanza” fra le istanze considerate

- In caso di votazione pesata, i pesi vengono calcolati, usati e poi dimenticati

Generalizzazione di un modello

- Algoritmi di apprendimento diversi producono **modelli di tipo diverso**:
 - Alberi di decisione: albero
 - Sistemi a regole: insiemi di regole if-then
 - Reti neurali: matrici di numeri
 - Apprendimento per rinforzo: distribuzioni di probabilità e matrici di numeri
 - ...

Avvertenze prima dell'uso

- Un algoritmo di apprendimento impara **quello che c'è nei dati per come sono etichettati**, non quello che chi ha costruito i dati immagina ci sia
- **RICORDARE SEMPRE** che le classi **esistono solo nella mente** di chi usa il sistema
- Il sistema non ha altra conoscenza oltre ai dati
- Attenzione ai numeri! Richiedono competenza

Il mio classificatore è fantastico!!!!

Ha un'accuratezza del 99.9% !!!
(error rate pari a 0.1%)



WOW?

Attenti al LS e ai numeri ...

LEARNING SET

Istanze di classe A: 9990

Istanze di classe B: 10

Il classificatore ha costruito la regola “**if (TRUE) then classe A**” quindi restituisce sempre: *l'istanza è di classe A*

Accuratezza: $9990/10000 = 0.999$!!
(su LS)

Error rate: $10 / 10000 = 0.001$
(su LS)



Non ha
imparato
nulla

- **Esempio:** riconoscere “frutta” da altre cose

Esempio

- Servono dei dati

- Servono dei dati



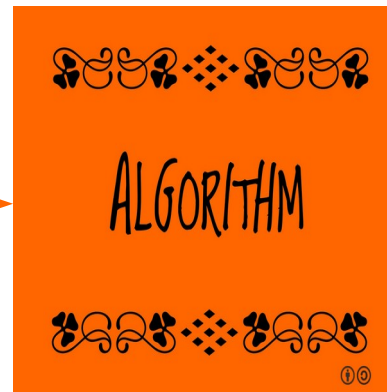
Eccoli !

Esempio

- Servono dei dati



Esempi di frutti

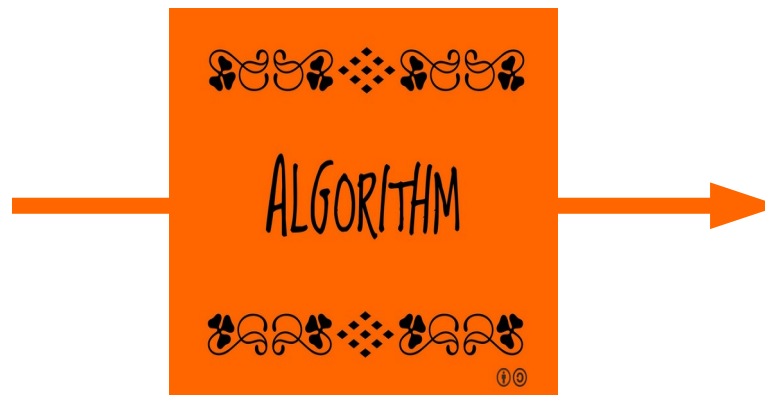


Esempio

- Servono dei dati



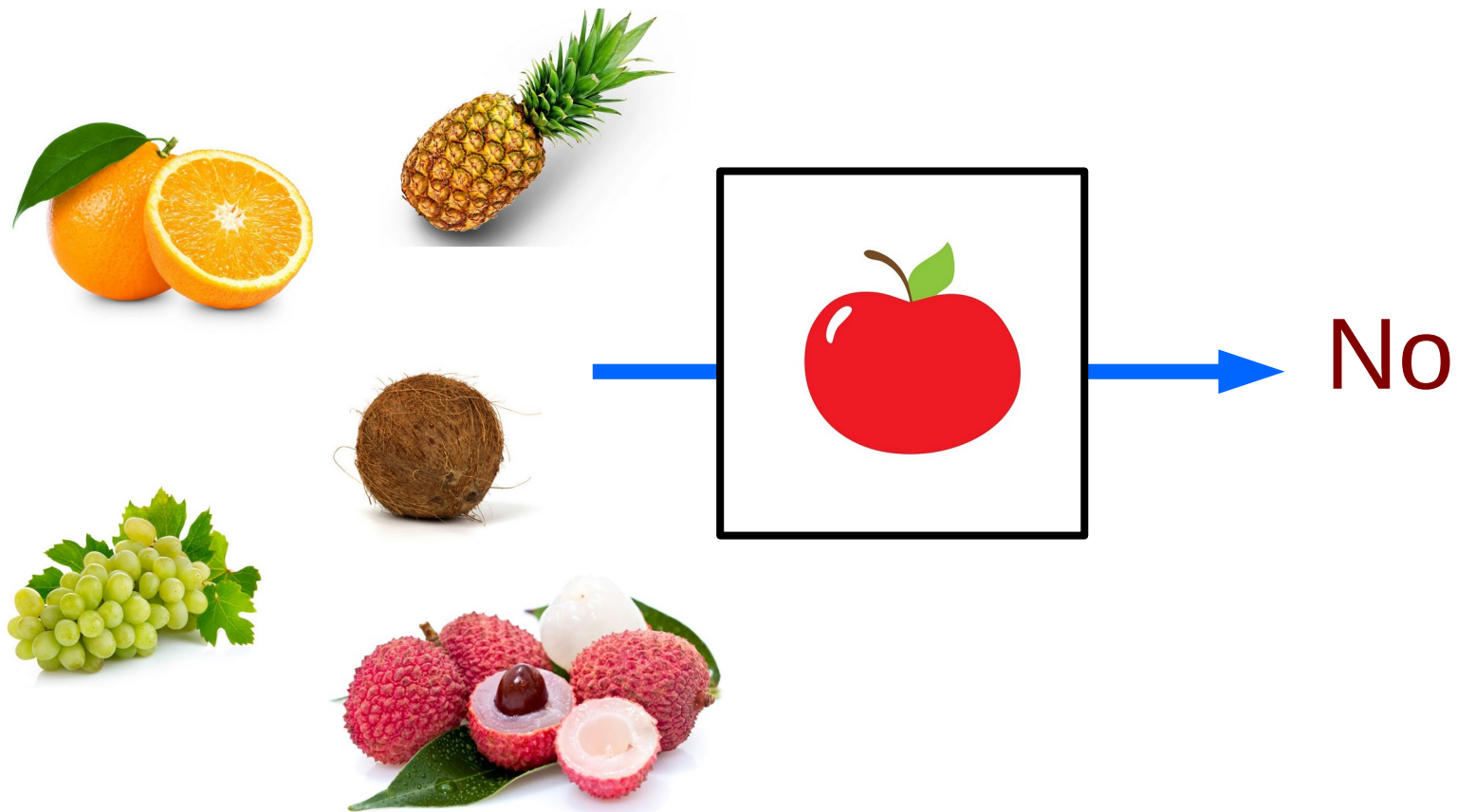
Esempi di frutti



Modello di frutta (?)

Esempio

- Modello troppo povero
- Insufficiente, riconosce solo le mele



Perché si fanno questi errori?

- I dati sono difficili da reperire
- Risorse disponibili + fretta (es. archivio fotografico coldiretti di una certa area, es. Trentino Alto Adige)
- Bias mentale (se vivo circondato dai meleti non mi sembra strano usare foto di mele, non mi vengono in mente uva o ciliegie)
- Attribuire all'algoritmo i meccanismi umani: noi siamo bravi a generalizzare, più di un software, e usiamo moltissima conoscenza senza rendercene conto (es. è prodotto da una pianta dall'impollinazione di un fiore, allora è un frutto)

Un passo oltre la frutta ...

- Fotografia: riconoscere i **volti delle persone**
- Decidere se **concedere un mutuo** (un rischio per la banca)
- Decidere se **concedere un'assicurazione sanitaria**
- **Applicazioni militari** (es. riconoscimento automatico di un obiettivo)
- Applicazioni civili, **tecnologie see and spray**: identificare le piante con parassiti per limitare irrorazione di insetticidi

... applicazioni reali (esempi)

- Volti: <https://www.luxand.com/facesdk/>
- Mutui: <https://www.app0.io/blog/ai-in-lending>
- Salute: <https://research.aimultiple.com/ai-health-insurance/>
- Militari: https://en.wikipedia.org/wiki/Automatic_target_recognition
- Agricoltura: <https://www.deere.com/en/sprayers/see-spray-ultimate/>

(aggiornati maggio 2024)

Responsabilità? Esempio

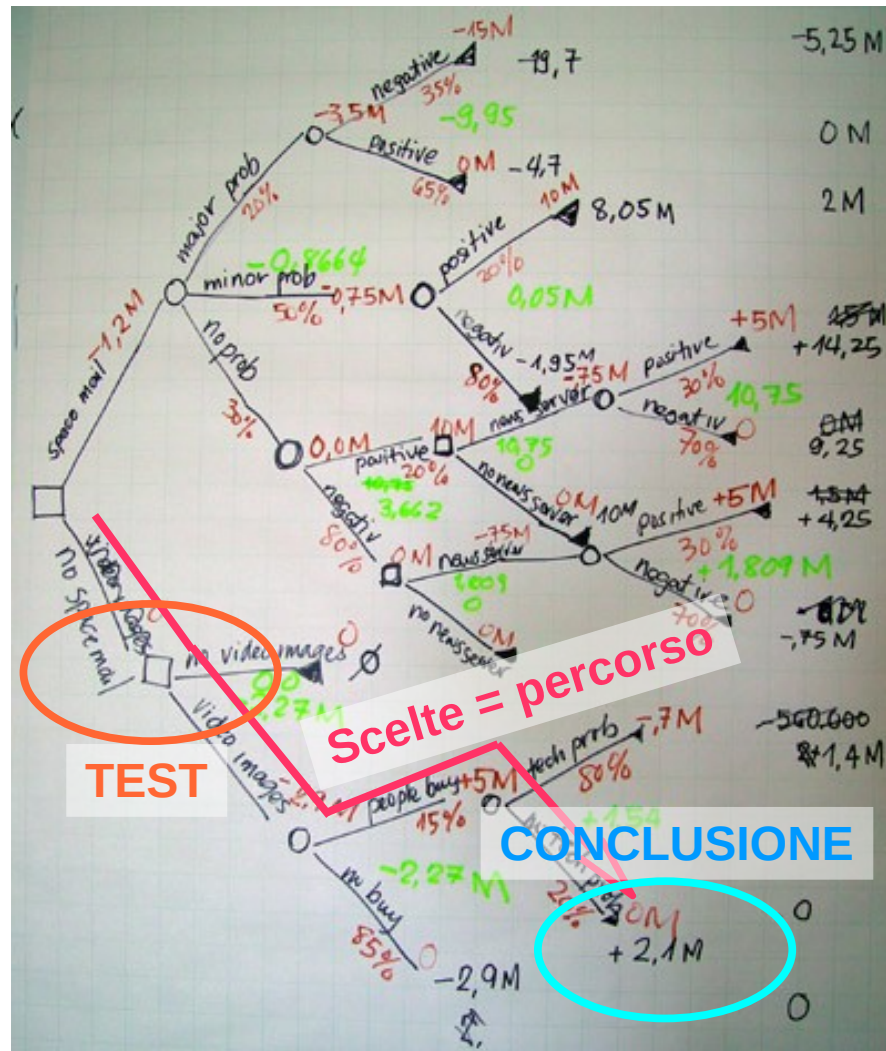
- **Trolley problem** adattato alla guida autonoma
 - Un bambino attraversa la strada all'improvviso
 - Se l'auto non frena lo investe
 - Se frena, data la velocità, l'incidente può mettere a rischio l'incolumità dei viaggiatori
 - Se sterza finisce sul marciapiede di cui non ha percezione (c'è qualcuno?)
- **Di chi è la responsabilità della decisione** di un agente autonomo?
- Problem of the many hands

Esempio

<https://ndpr.nd.edu/reviews/moral-responsibility-and-the-problem-of-many-hands/>

Induzione di alberi di decisione

Alberi di decisione



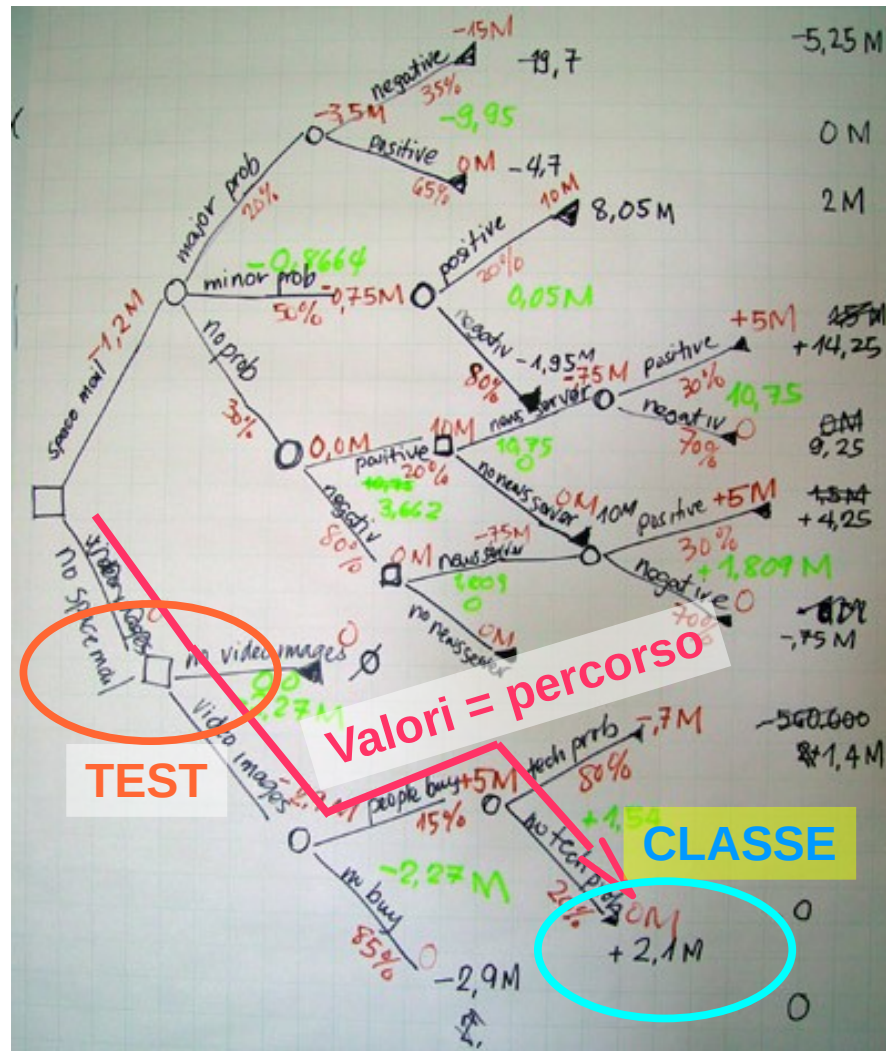
Sono strumenti di supporto alle decisioni che usano modelli strutturati ad albero, comunemente utilizzati per esempio per la definizione di strategie mirate al conseguimento di un goal

Esistono da tempi immemorabili

Hanno moltissime applicazioni. Per esempio la struttura di navigazione di molti negozi on-line implementa un albero di decisione (ti interessa un libro, un film o un gioco?)

Fonte: wikipedia

Alberi di decisione

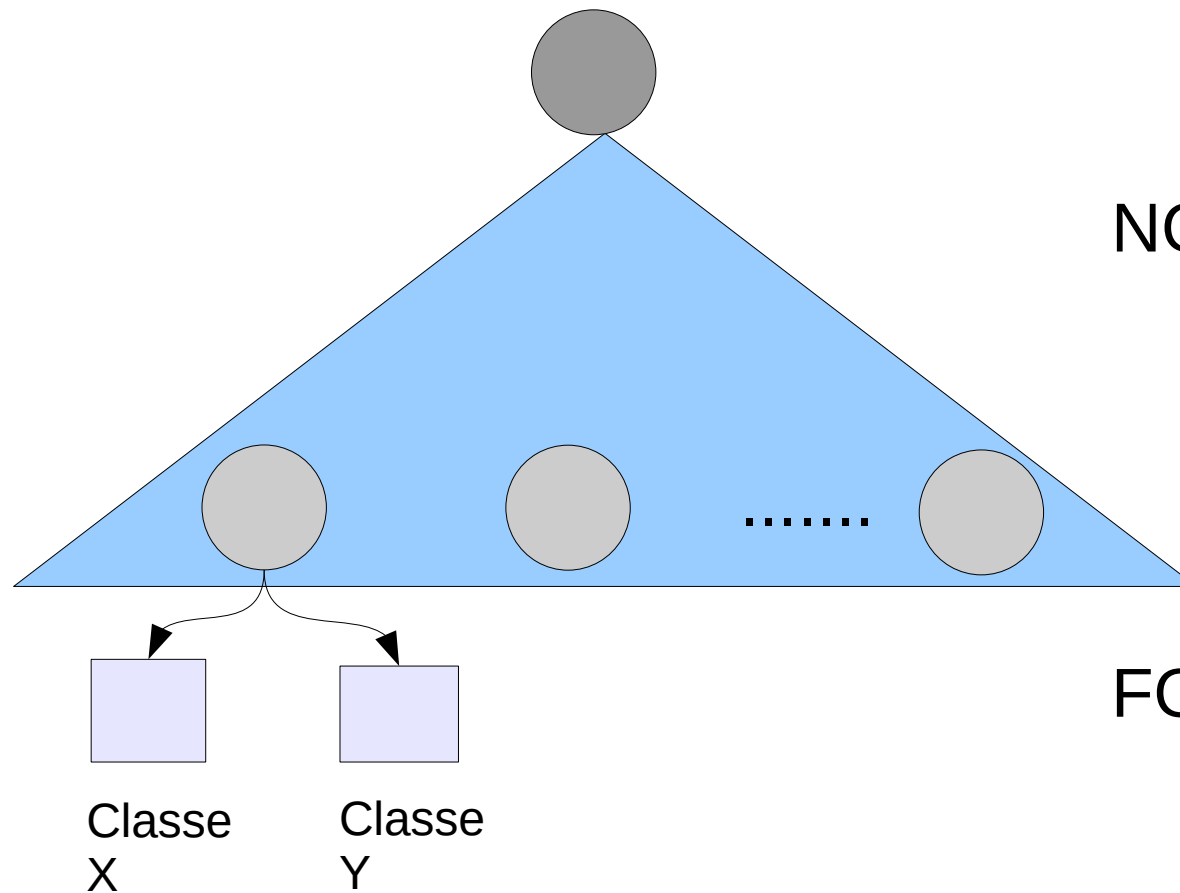


Noi studieremo un metodo che consente di indurre da esempi degli alberi di decisione da utilizzare in compiti di classificazione

MODELLO = albero

Fonte: wikipedia

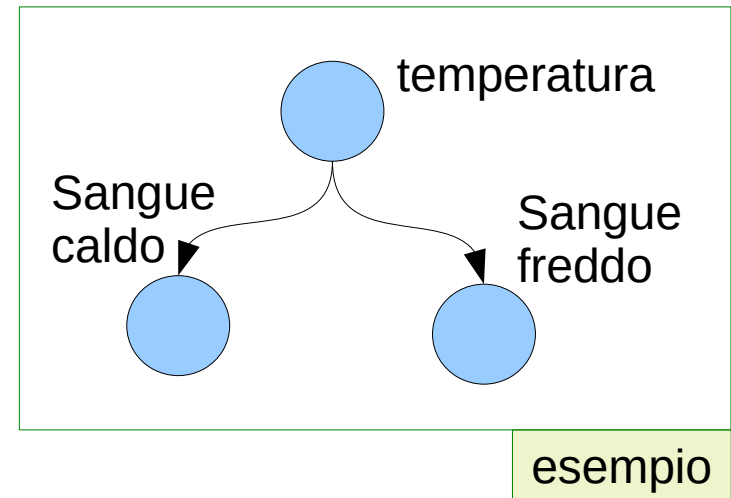
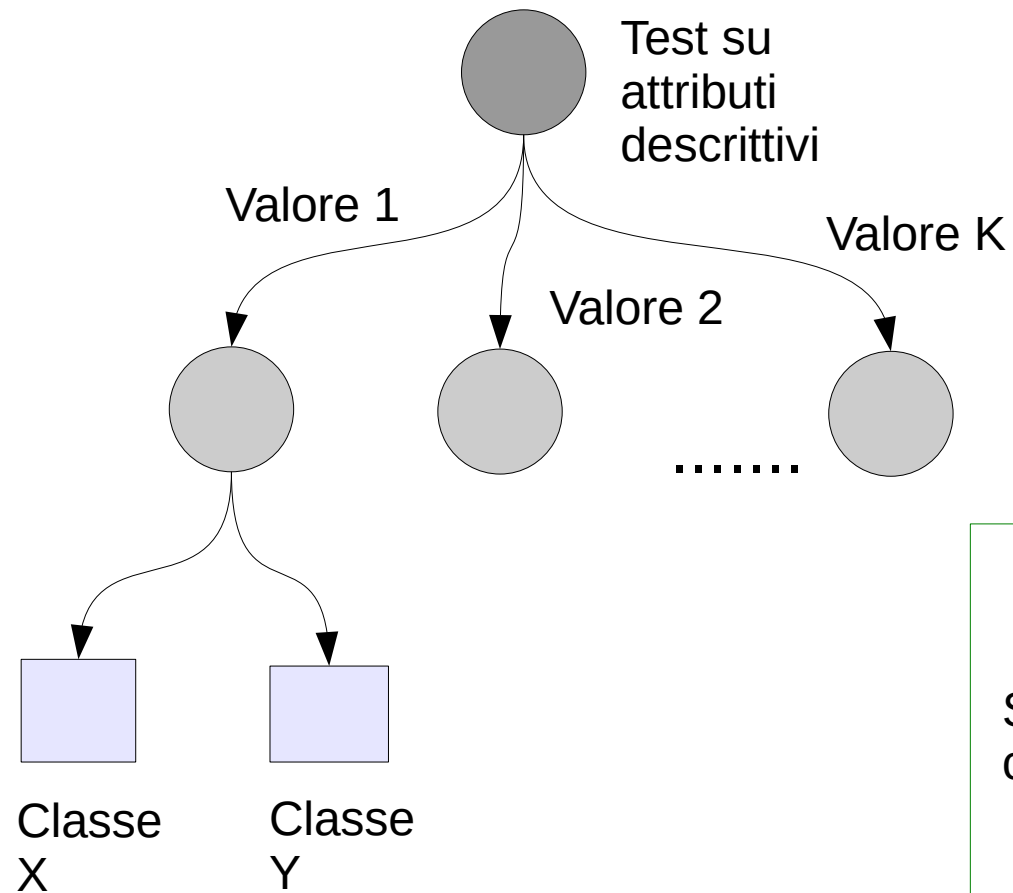
Struttura degli alberi di decisione



NODI INTERNI: test

FOGLIE: decisioni

Struttura degli alberi di decisione



Iris: 3 classi



Iris virginica (1)



Iris setosa (2)



Iris versicolor (3)

ATTRIBUTI: petal width, petal length, sepal width, sepal length (di tipo continuo)

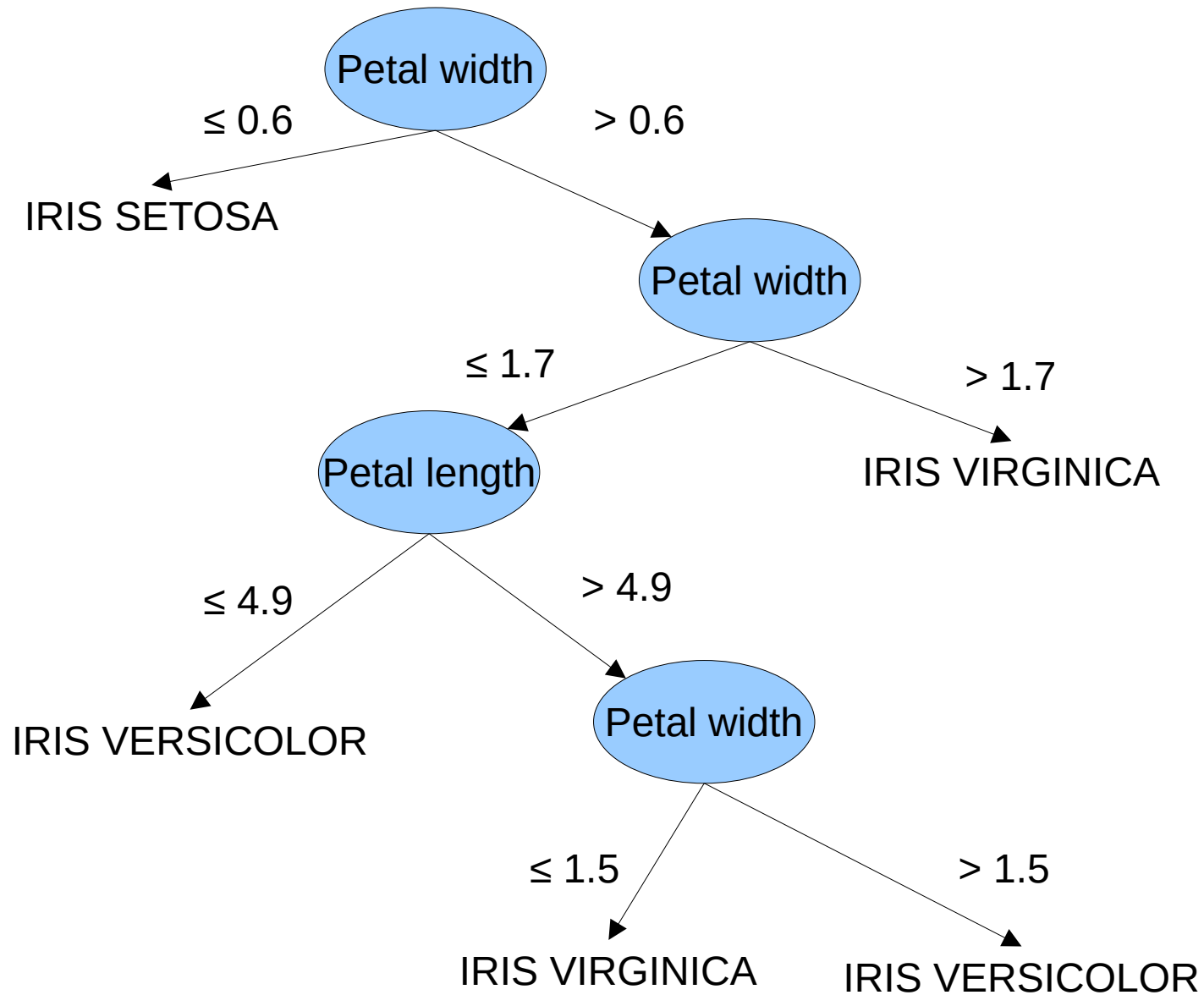
Dataset: <https://archive.ics.uci.edu/dataset/53/iris>

(1) By Eric Hunt - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=72555262>

(2) By w:ru:Денис Анисимов (talk | contribs) - Own work by the original uploader, Public Domain, <https://commons.wikimedia.org/w/index.php?curid=57593826>

(3) By D. Gordon E. Robertson - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=10227368>

Esempio albero di decisione



Algoritmi di apprendimento

- Alcuni algoritmi per l'apprendimento di alberi di decisione:
 - **Algoritmo di Hunt**
 - ID3, C4.5
 - CART
 - ...

Algoritmo di Hunt

L'albero viene costruito procedendo *ricorsivamente* e suddividendo il learning set in sottoinsiemi *via via più "puri"*. Il generico passo di suddivisione di un nodo dell'albero esegue questi passi:

Dati:

D_t = sottoinsieme del learning set associato al nodo t

$y = \{y_1, y_2, \dots, y_c\}$ = insieme delle etichette che identificano le classi

passo 1: se tutte le istanze in D_t appartengono alla stessa classe y_t allora il nodo è una foglia etichettata dalla classe y_t delle sue istanze

passo 2: si sceglie un attributo fra quelli che descrivono le istanze, si produce un nodo figlio per ogni possibile valore dell'attributo (il range è dato dal learning set?).

A ciascun nodo figlio si associa uno specifico valore e si associano ad esso anche quelle istanze, già associate al padre, per le quali l'attributo assume il valore corrispondente al nodo

Algoritmo di Hunt

nota 1: se una certa combinazione di valori non è rappresentata da nessuna istanza, questa sarà associata alla classe di default (se esiste)

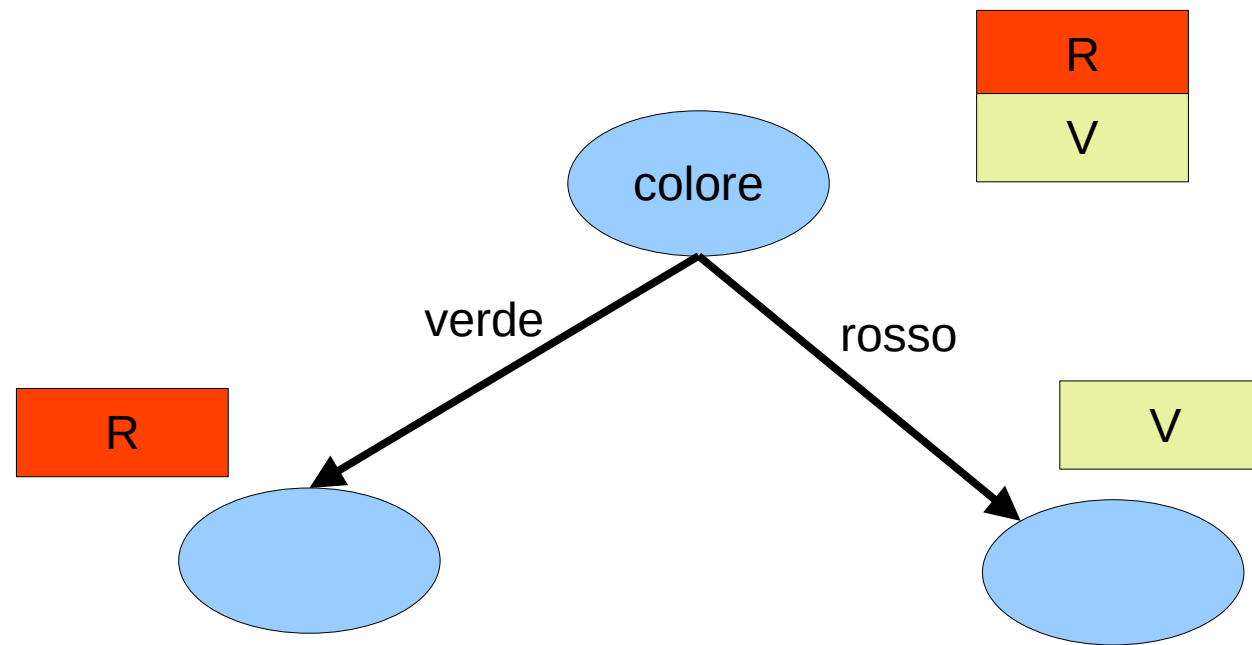
nota 2: se tutte le istanze associate a un nodo sono identiche come tuple ma corrispondono a classi differenti (non-determinismo), il nodo non può essere scisso.

Diventa una foglia che ha associata la classe più rappresentata

nota 3: quando si termina la costruzione dell'albero?

nota 4: come si sceglie l'attributo di split?

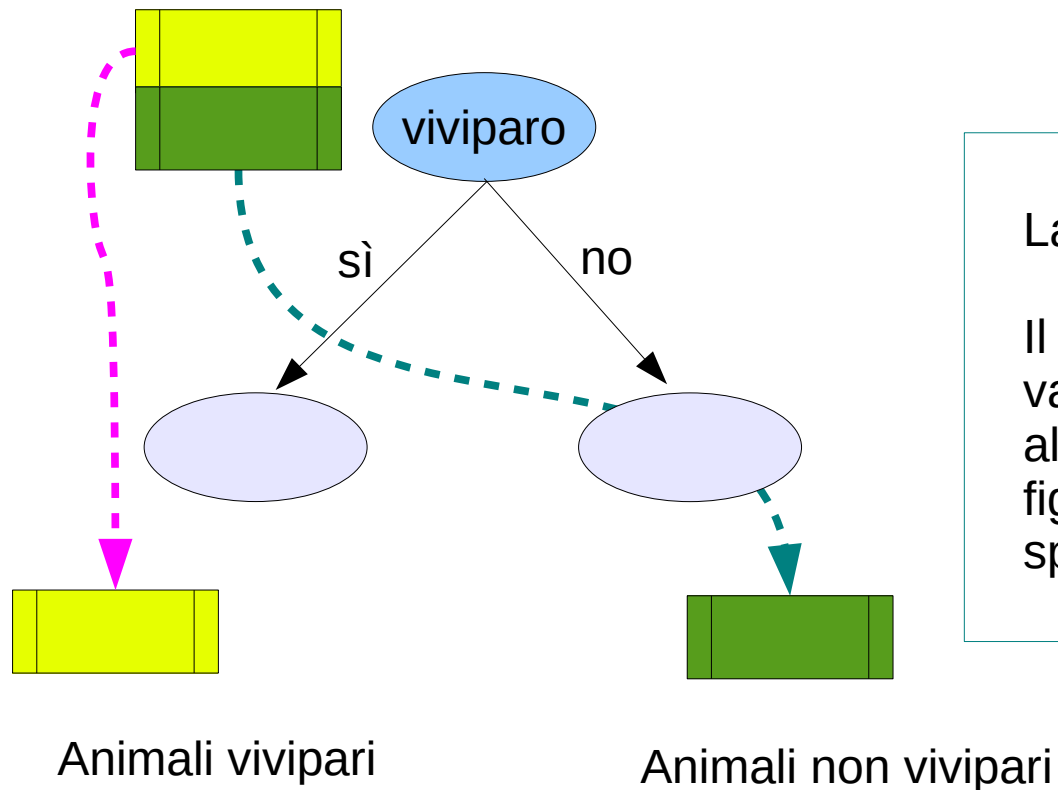
Split = suddividere il learning set



In generale ...

- **Strategia greedy**
 - Gli attributi su cui effettuare gli split sono selezionati in modo da massimizzare una qualche misura di riferimento
- **Problemi:**
 - Come **specificare la condizione di test** sugli attributi scelti?
 - *Attributi binari*
 - *Attributi nominali*
 - *Attributi ordinali*
 - *Attributi continui*
 - Come **determinare lo split migliore?**
 - **Quando fermarsi** nella costruzione dell'albero?

Split su attributi binari



La risposta non può che essere sì oppure no

Il nodo corrente avrà due figli a seconda del valore rappresentato. Gli esempi associati al nodo radice verranno suddivisi fra i due figli a seconda del valore riportato in corrispondenza dell'attributo

Altri esempi: animale acquatico, volatile, a sangue caldo, ...

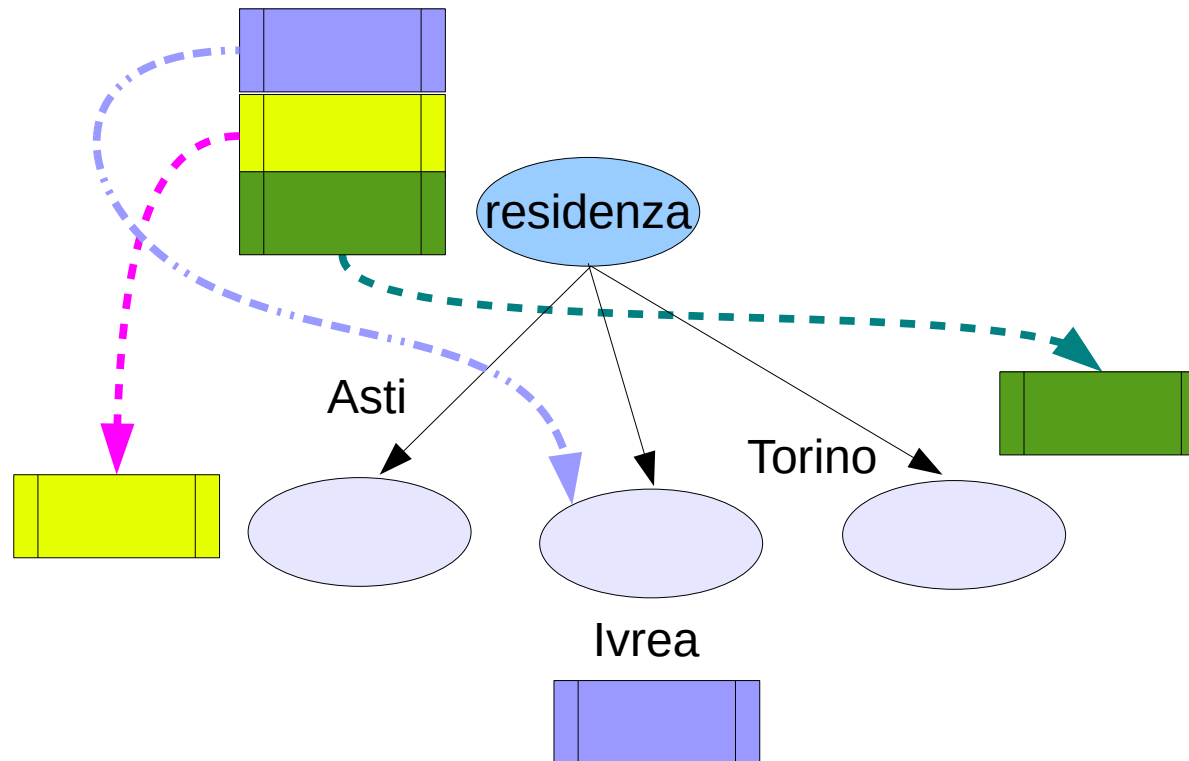
Split su attributi nominali

L'attributo assume valori su un insieme (finito) di etichette $\{L1, L2, \dots, L_n\}$

Gli split possono essere **binari** oppure **multivalore**

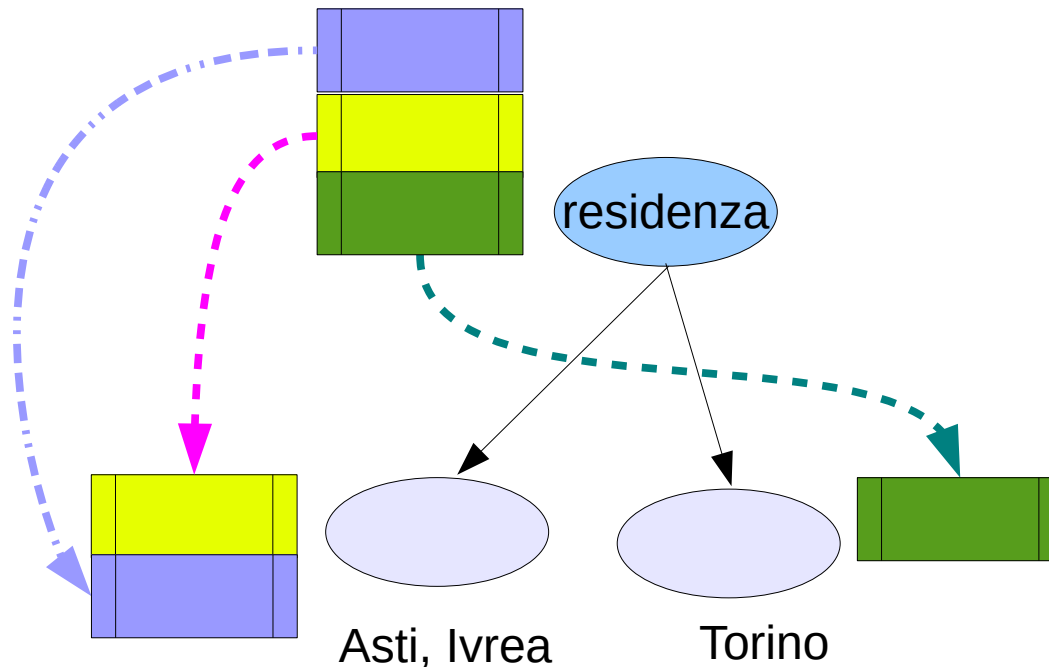
Split su attributi nominali

Split multivalore: il nodo avrà tanti figli quanti sono i possibili valori dell'attributo



Split su attributi nominali

Split binari: il nodo avrà due figli, uno corrisponde a un valore, l'altro all'insieme dei rimanenti valori

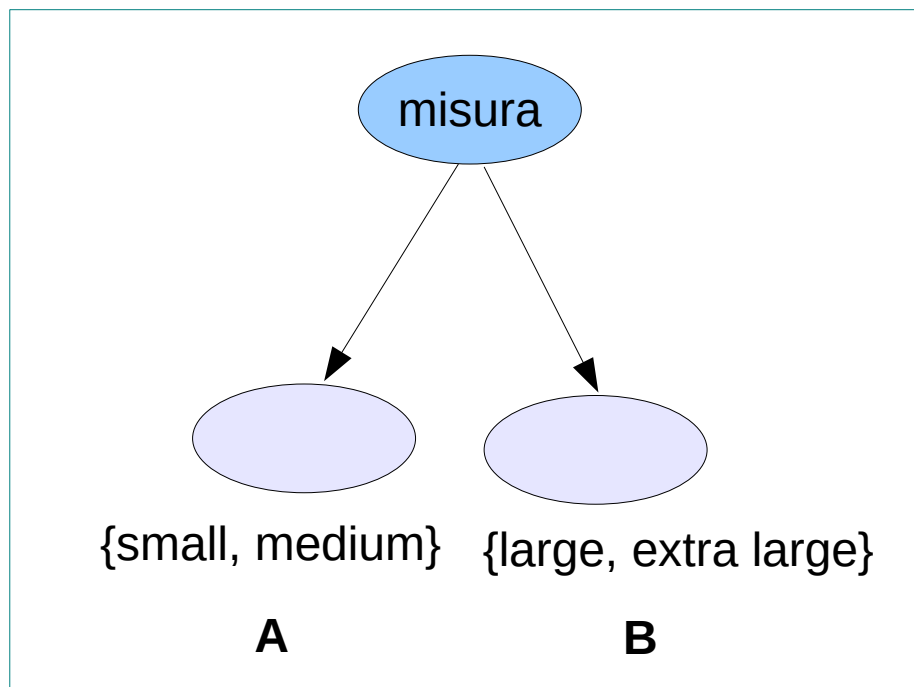


Ci sono $2^{k-1} - 1$ possibili alternative se l'attributo ha k valori alternativi possibili

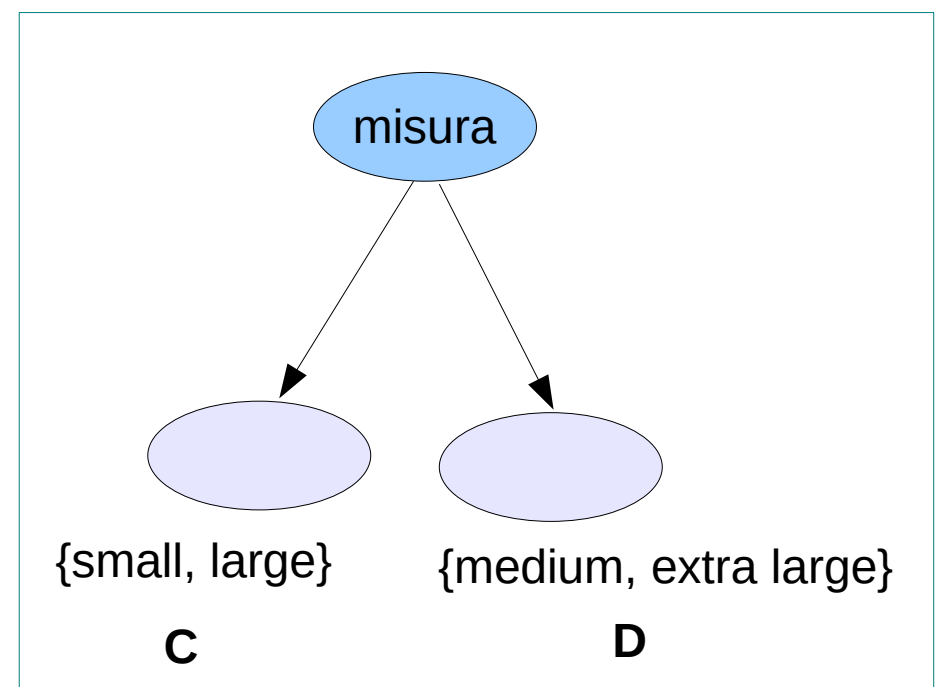
Split su attributi ordinali

Anche in questo caso si possono avere split binari o multivalore con un vincolo: il raggruppamento dei valori deve rispettare l'ordinamento

Esempio: supponiamo di avere le misure *small* < *medium* < *large* < *extralarge*



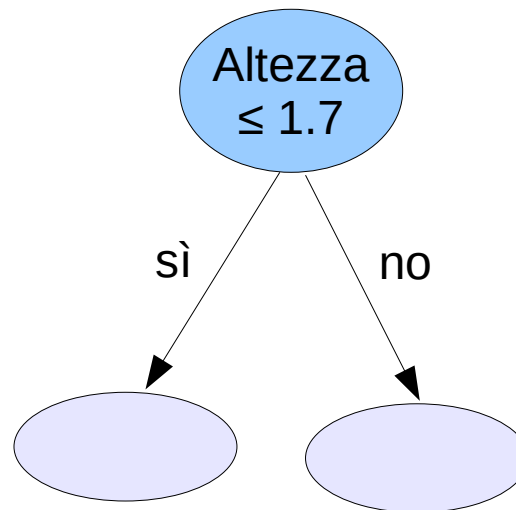
Corretto!! $A < B$



Errato!! C ?? D

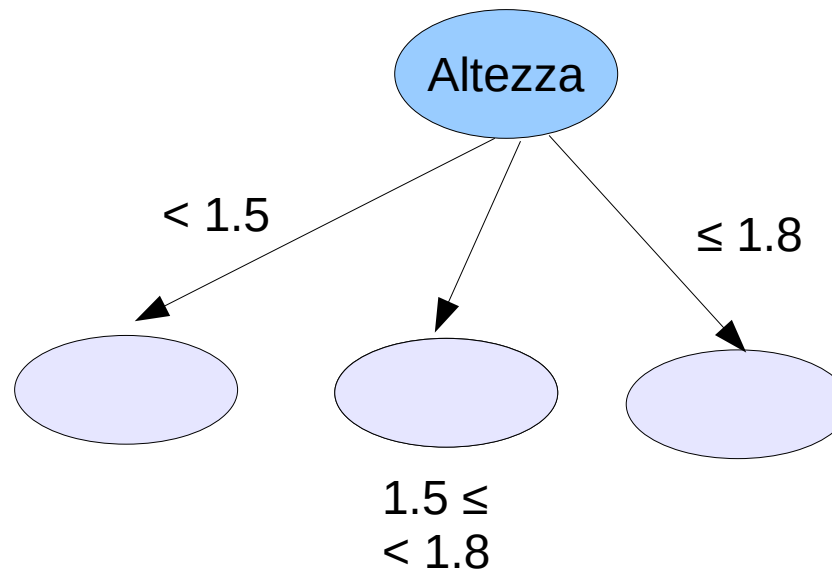
Split binari di attributi continui

In questo caso il test prevedono l'identificazione di un valore possibile v per l'attributo A in questione



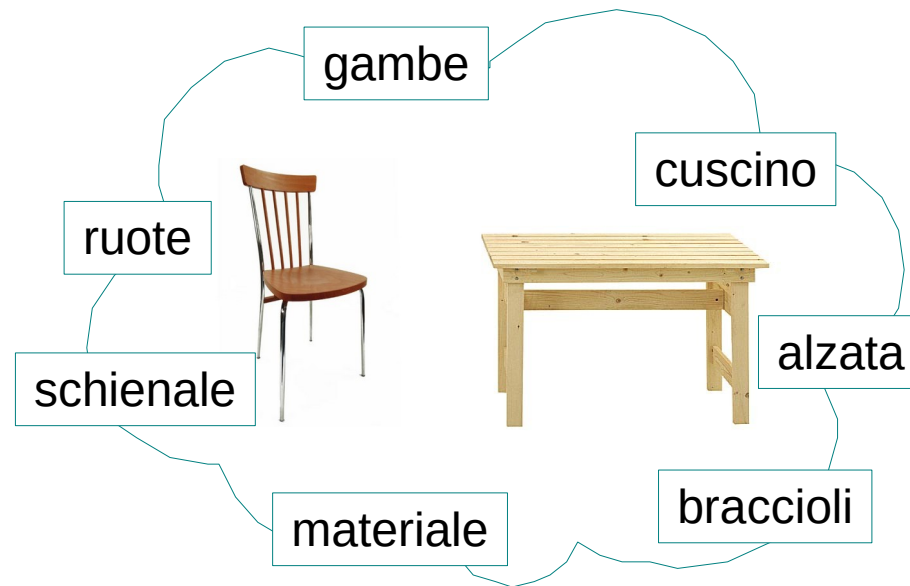
Split multivalore di attributi continui

In questo caso il test prevedono l'identificazione di un insieme di valori v_i per l'attributo A in questione e la produzione di una serie di test $v_i \leq A < v_{i+1}$



Occorre **discretizzare** la variabile continua identificando un numero finito di intervalli significativi di valori

Quale split?



Supponiamo di dover costruire un DT che consenta di distinguere sedie da tavoli.
È indifferente l'ordine con il quale scegliamo di effettuare gli split?

Partire da *materiale* da *ruote* oppure da *schienale* fa qualche differenza?

Bontà degli split

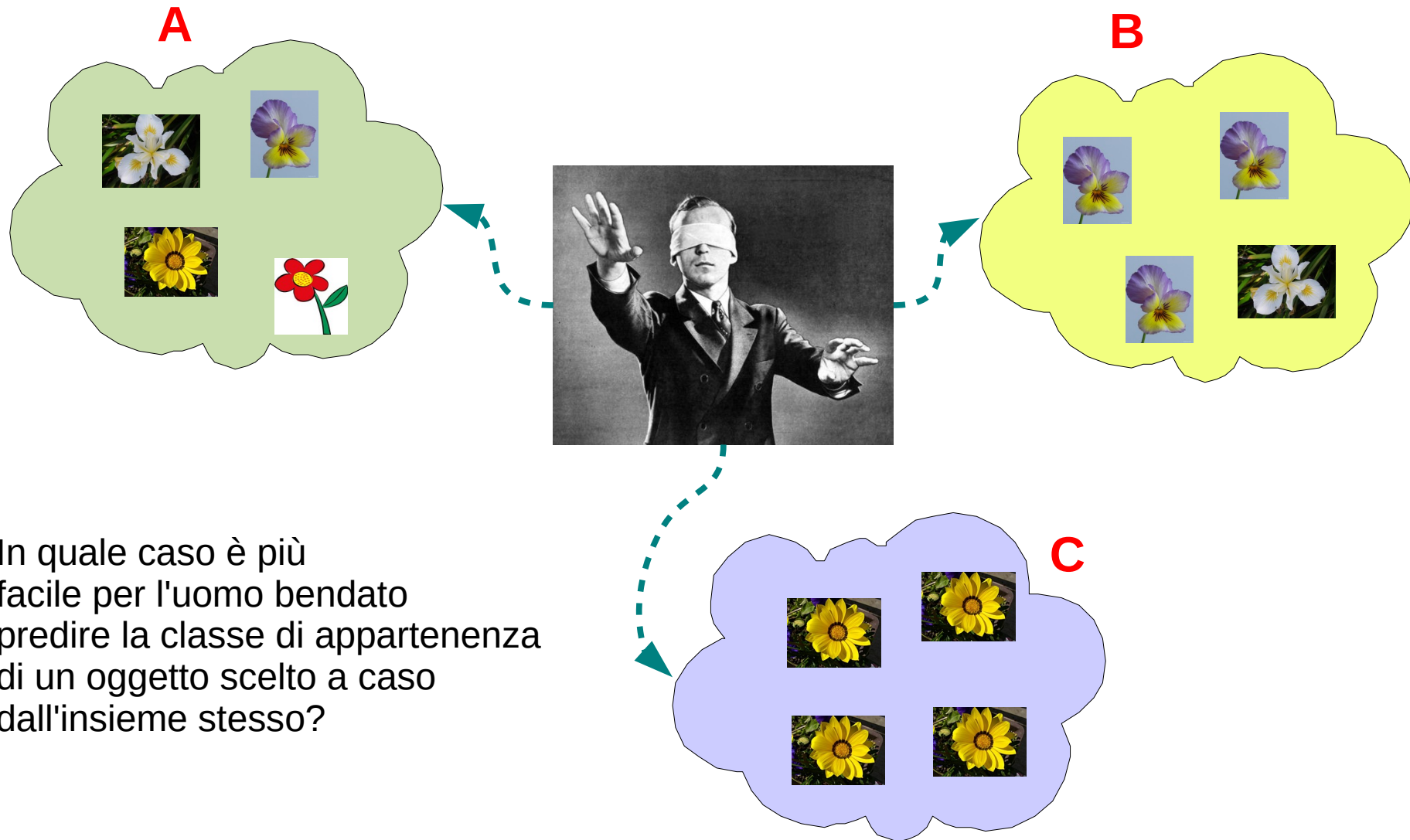
Criterio generale: alberi compatti sono preferiti ad alberi che consentono di raggiungere lo stesso grado di accuratezza (e di error rate) usando un maggior numero di test

Rasoio di Occam: a parità di assunzioni, la spiegazione più semplice è da preferire

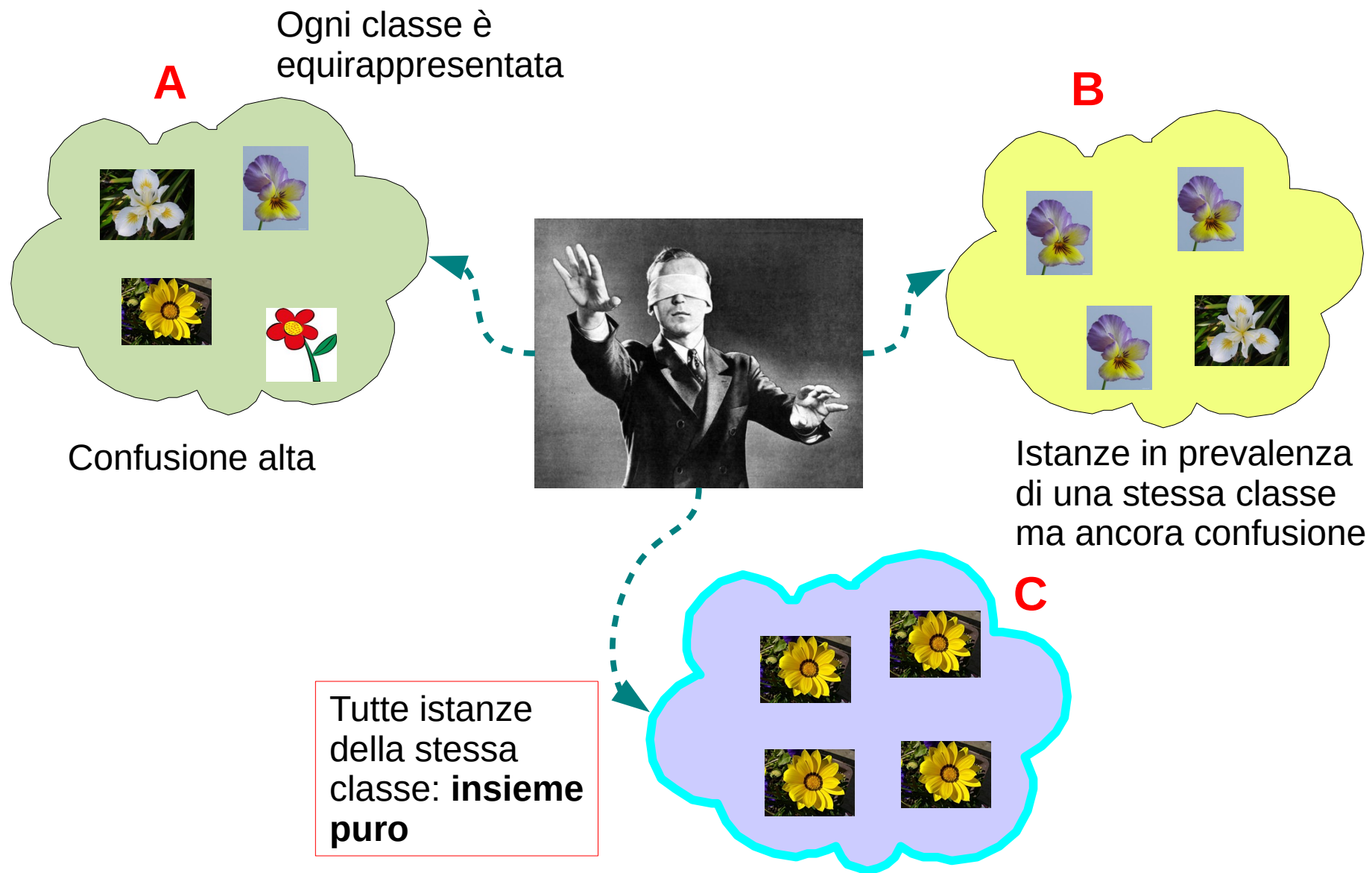


William of Ockham

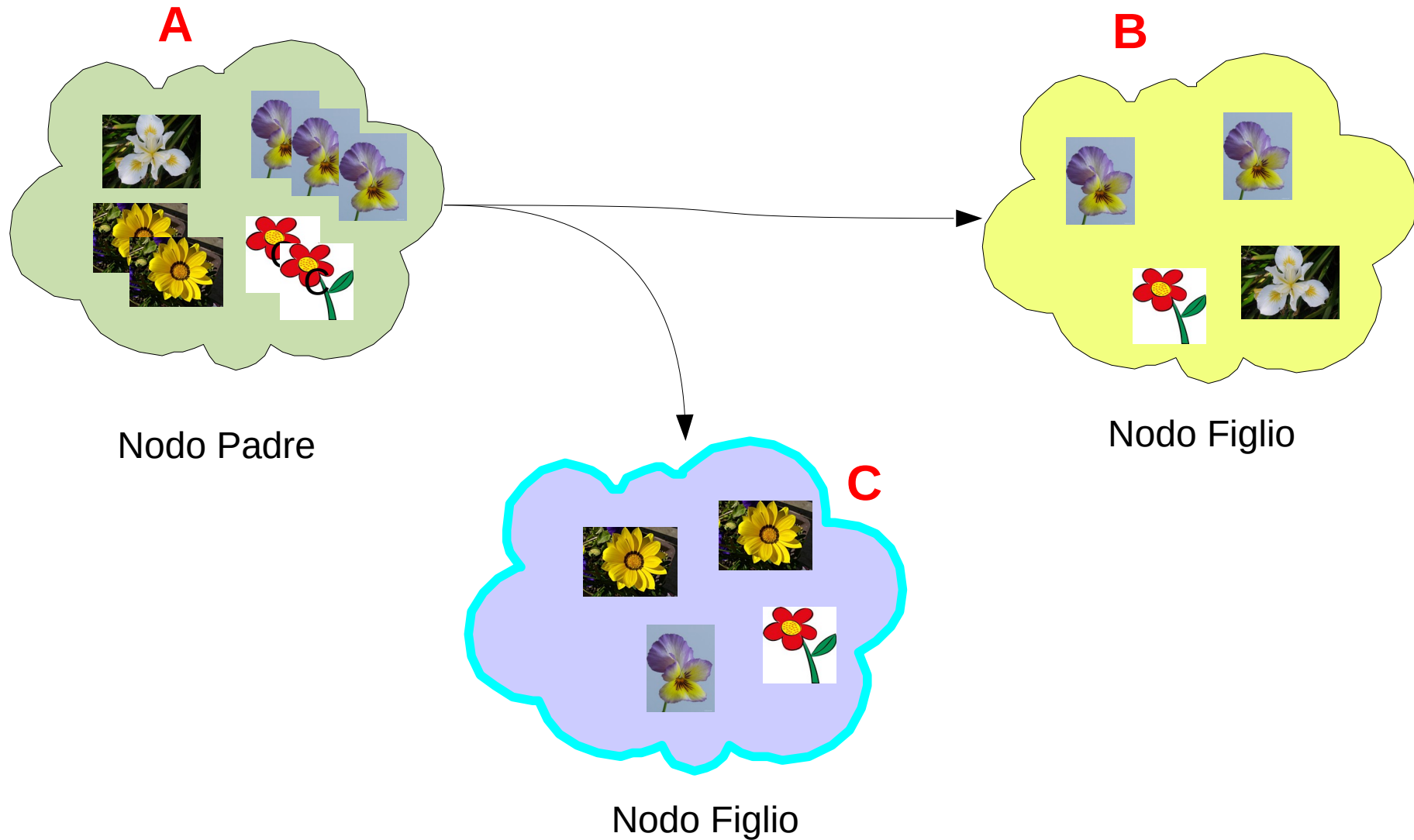
Misure per determinare la bontà di uno split



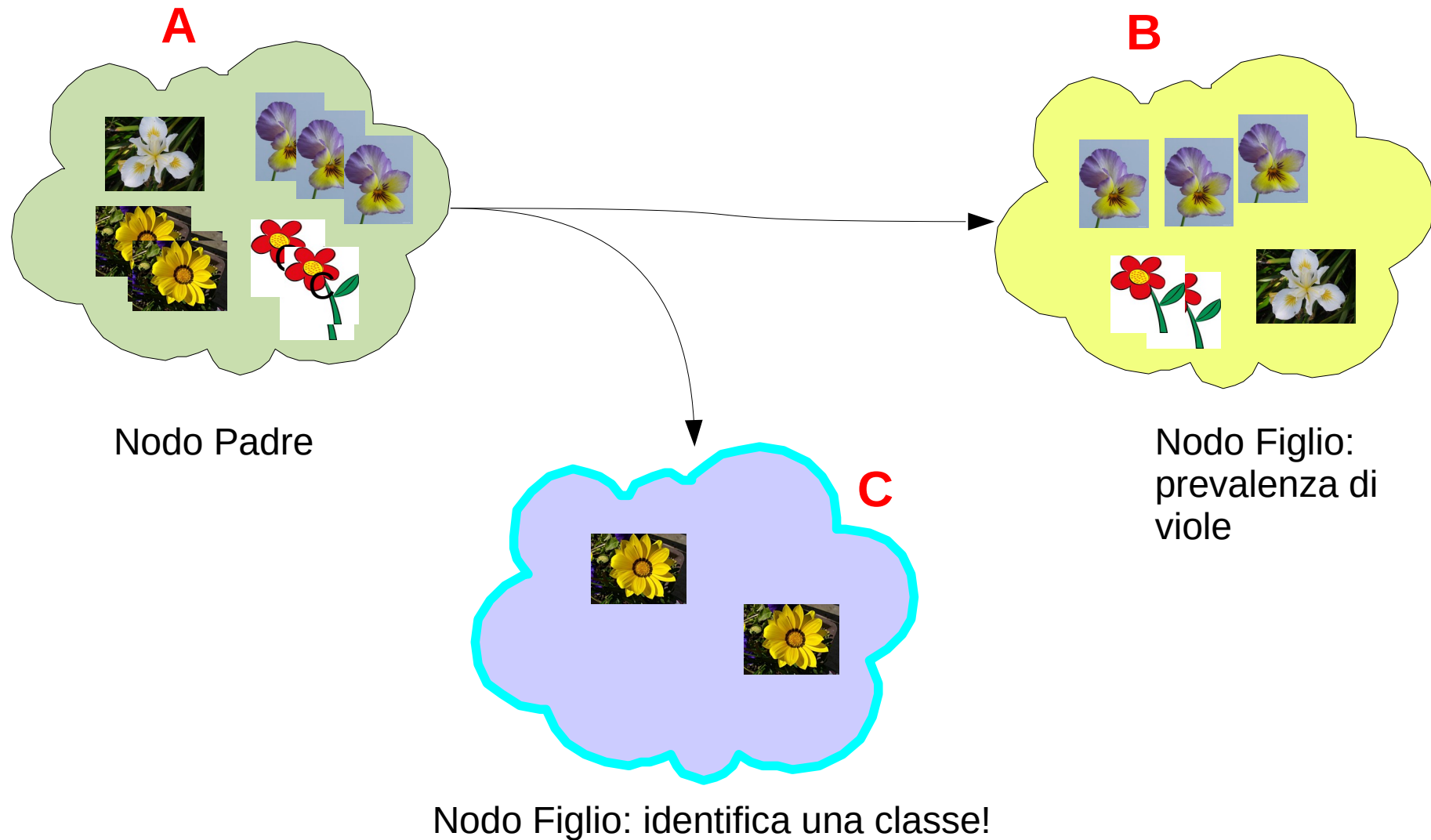
Misure per determinare la bontà di uno split



È meglio questo split ...



... o questo split?



Criterio generale

- Sono preferiti gli split che producono nodi figli la cui estensione prevede minore confusione (il cui grado di purezza è maggiore)
- Misure alternative:
 - *Entropia*
 - *Gini*
 - *Errore di classificazione*

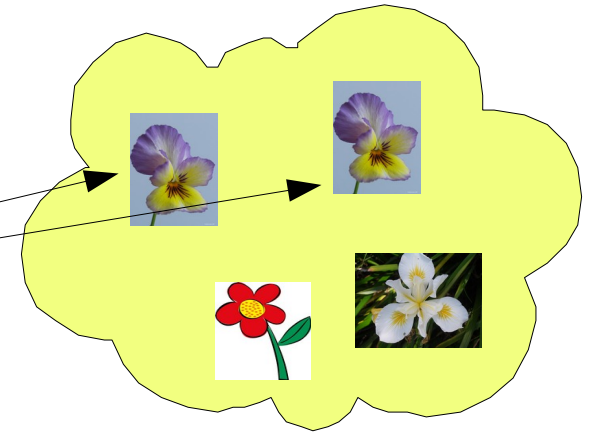
Base delle misure di selezione

Dato un nodo t , sia $p(i | t)$ la probabilità che un elemento estratto casualmente dall'insieme sia di classe i

$$p(\text{viola} | t) = 0.5 \quad (2 \text{ su } 4)$$

$$p(\text{iris} | t) = 0.25 \quad (1 \text{ su } 4)$$

$$p(\text{finto} | t) = 0.25 \quad (1 \text{ su } 4)$$



(0.5, 0.25, 0.25) è la **distribuzione di probabilità** di appartenenza di un record estratto a caso dall'estensione associata al nodo t a una delle classi in questione

Entropia

$$Entropia(t) = - \sum_{i=0}^{c-1} p(i|t) \log_2 p(i|t)$$

Per ogni classe

Probabilità che l'elemento appartenga alla classe i-ma

Assunto nel calcolo dell'entropia: $0 \log_2 0 = 0$

Entropia

$$Entropia(t) = - \sum_{i=0}^{c-1} p(i|t) \log_2 p(i|t)$$

Assunto nel calcolo dell'entropia: $0 \log_2 0 = 0$

Supponiamo di avere due sole classi:

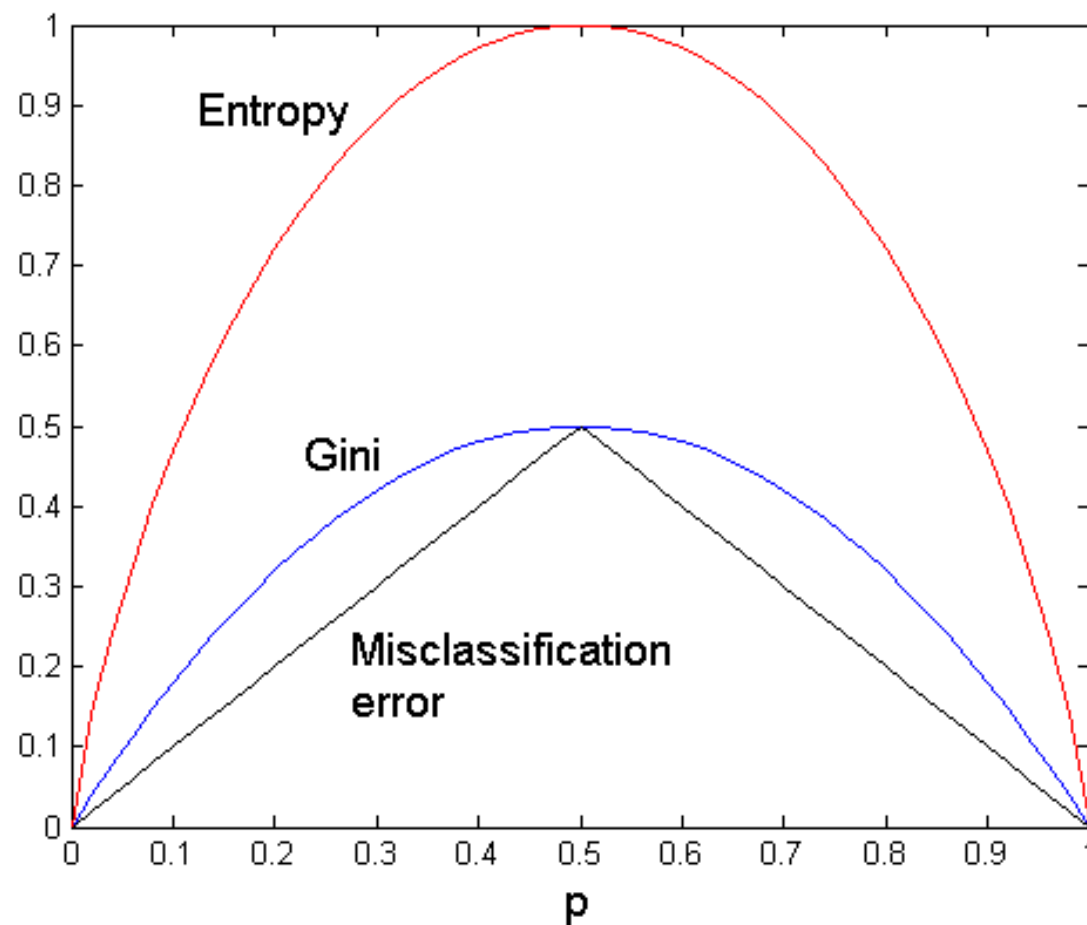
le distribuzioni (0, 1) e (1, 0) sono le migliori, purezza massima dell'insieme, nessuna confusione. Calcoliamo per il caso **(0, 1)**

Entropia: $-0 \log_2 0 - 1 \log_2 1 = 0$

La distribuzione **(0.5, 0.5)** è la peggiore, massimo grado di confusione:

Entropia: $-0.5 \log_2 0.5 - 0.5 \log_2 0.5 = -\log_2 2^{-1} = 1$

Misure: confronto

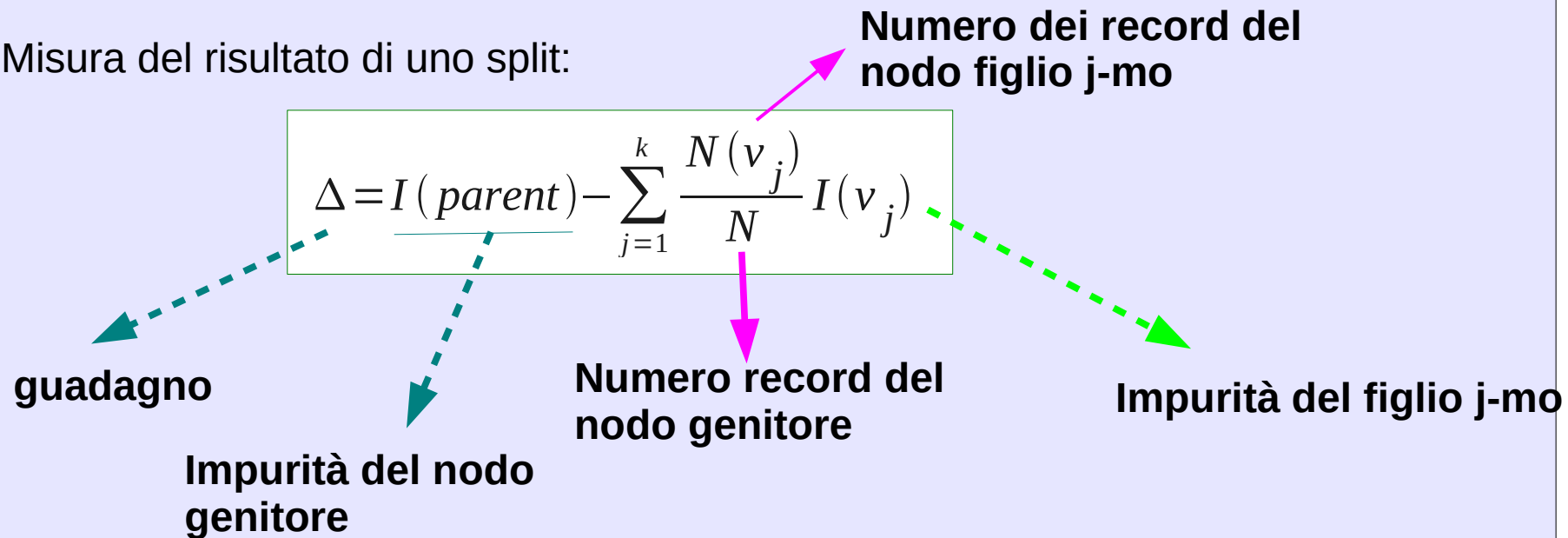


Valore dell'entropia per una sorgente binaria

Calcolo del guadagno

Problema della scelta dello split: valuto i diversi split che posso fare, usando attributi diversi, tramite una delle misure viste e scelgo quello che mi restituisce il risultato col minor grado di confusione

Misura del risultato di uno split:



Dall'impurità del nodo genitore viene sottratta la media pesata delle impurità dei nodi figli. Di solito la misura dell'impurità è scelta in modo tale da minimizzare l'impurità / massimizzare il guadagno

Information gain

Per **information gain** si intende una misura del guadagno ottenuta usando l'**entropia** come valore dell'impurità dei nodi:

$$\Delta = \text{entropia}(\text{parent}) - \sum_{j=1}^k \frac{N(v_j)}{N} \text{entropia}(v_j)$$