

OSNOVE UMETNE INTELIGENCE

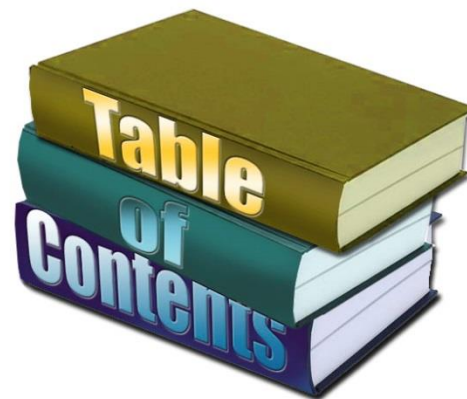
2022/23

uvod v strojno učenje
učenje odločitvenih dreves

Pridobljeno znanje s prejšnjih predavanj

- različne **definicije** umetne inteligence
- filozofske in psihološke **implikacije** umetne inteligence
- **Turingov test**
- **zgodovina** umetne inteligence
- sodobne **aplikacije in področja** umetne inteligence
- **strojno učenje**
 - namen in cilji učenja
 - vrste učenja: nadzorovano, nenadzorovano, spodbujevano
 - nadzorovano učenje:
 - atributna predstavitev podatkov
 - klasifikacijski/regresijski problemi
 - prostor hipotez
 - evalvacija hipotez: konsistentnost, splošnost, preprostost/razumljivost/interpretabilnost

Pregled

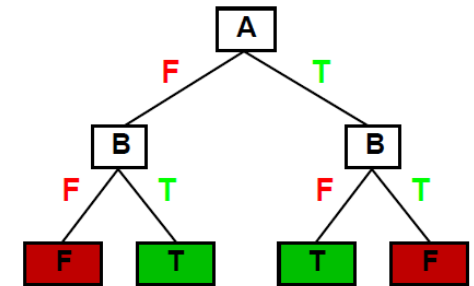


- strojno učenje
 - uvod v strojno učenje
 - vrste strojnega učenja
 - predstavitev podatkov, hipoteze
 - učenje odločitvenih dreves

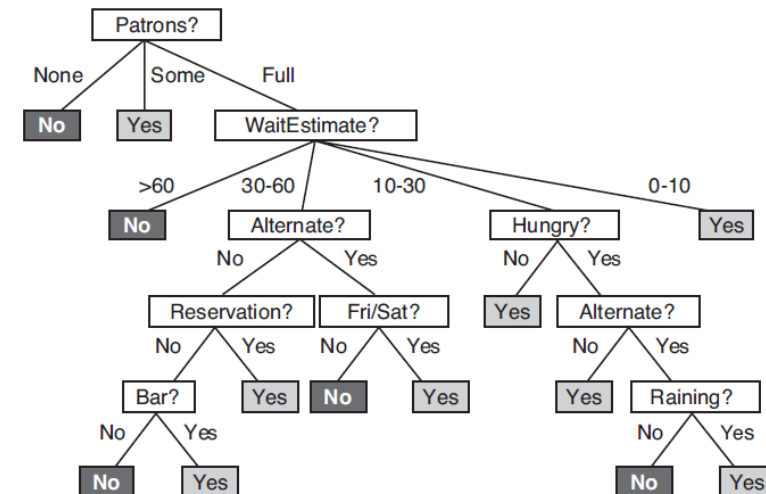
Odločitveno drevo

- ponazarja relacijo med vhodnimi vrednostmi (atributi) in odločitvijo (ciljna spremenljivka – razred ali označba)
 - notranja vozlišča: test glede na vrednost posameznega atributa
 - listi: odločitev (vrednost ciljne spremenljivke)
 - pot: konjunkcija pogojev v notranjih vozliščih na poti, ki vodi do lista
- poseben primer: binarna klasifikacija (razred ima dve možni vrednosti - npr. pozitivni/negativni, strupen/užiten itd.)

A	B	A xor B
F	F	F
F	T	T
T	F	T
T	T	F

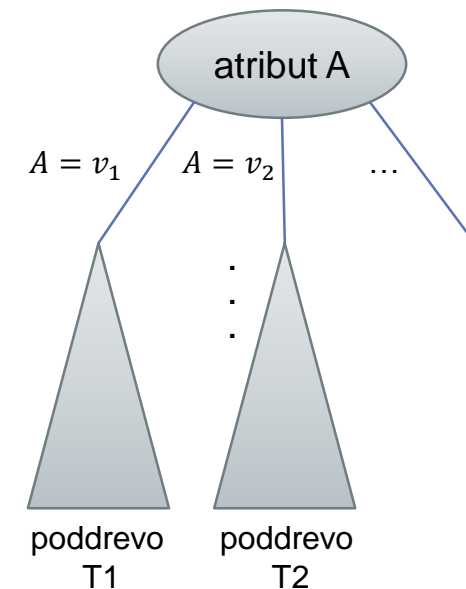


Example	Attributes										Target WillWait
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	
X ₁	T	F	F	T	Some	\$\$\$	F	T	French	0-10	T
X ₂	T	F	F	T	Full	\$	F	F	Thai	30-60	F
X ₃	F	T	F	F	Some	\$	F	F	Burger	0-10	T
X ₄	T	F	T	T	Full	\$	F	F	Thai	10-30	T
X ₅	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X ₆	F	T	F	T	Some	\$\$	T	T	Italian	0-10	T
X ₇	F	T	F	F	None	\$	T	F	Burger	0-10	F
X ₈	F	F	F	T	Some	\$\$	T	T	Thai	0-10	T
X ₉	F	T	T	F	Full	\$	T	F	Burger	>60	F
X ₁₀	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30	F
X ₁₁	F	F	F	F	None	\$	F	F	Thai	0-10	F
X ₁₂	T	T	T	T	Full	\$	F	F	Burger	30-60	T



Gradnja odločitvenega drevesa

- cilj: zgradi **čim manjše** drevo, ki je **konsistentno** z učnimi podatki
- prostor iskanja: kombinatoričen, vsa možna drevesa (neučinkovito!)
- **hevristični požrešni algoritem** s strategijo **razveji in omeji**:
 - izberi najbolj pomemben atribut – tisti, ki najbolj odločilno vpliva na klasifikacijo primera – in razdeli primere v poddrevesa glede na njegove vrednosti,
 - rekurzivno ponovi za poddrevesa,
 - če vsi elementi v listu pripadajo istemu razredu ali vozlišča ni možno deliti naprej (ni razpoložljivih atributov), ustavi gradnjo.
- imenovano tudi **Top Down Induction of Decision Trees (TDIDT)**
- primeri implementacij: ID3, CART, Assistant, C4.5, C5, ...



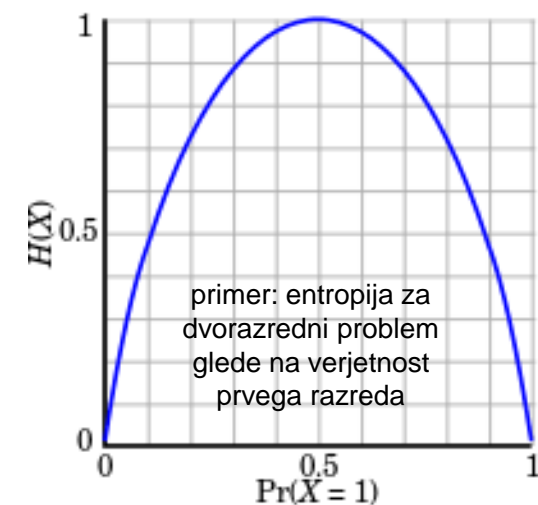
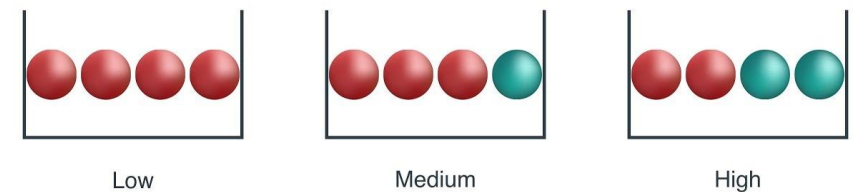
Izbor najbolj pomembnega atributa

- najboljši atribut je tisti, ki razdeli učno množico v najbolj "čiste" podmnožice (glede na razred)
- uporabimo lahko **mero entropije**:

$$H = - \sum_k p_k \log_2 p_k$$

- mera nečistoče oz. mera nedoločenosti naključne spremenljivke (Shannon in Weaver, 1949)
- enota: količina informacije v bitih, ki jo pridobimo
- primeri:
 - met kovanca: 1 bit informacije
 - poskus s štirimi enako verjetnimi možnimi izidi: 2 bita informacije
 - poskus z dvema izidoma, od katerih je eden 99%: ~ 0 bitov informacije

Entropy



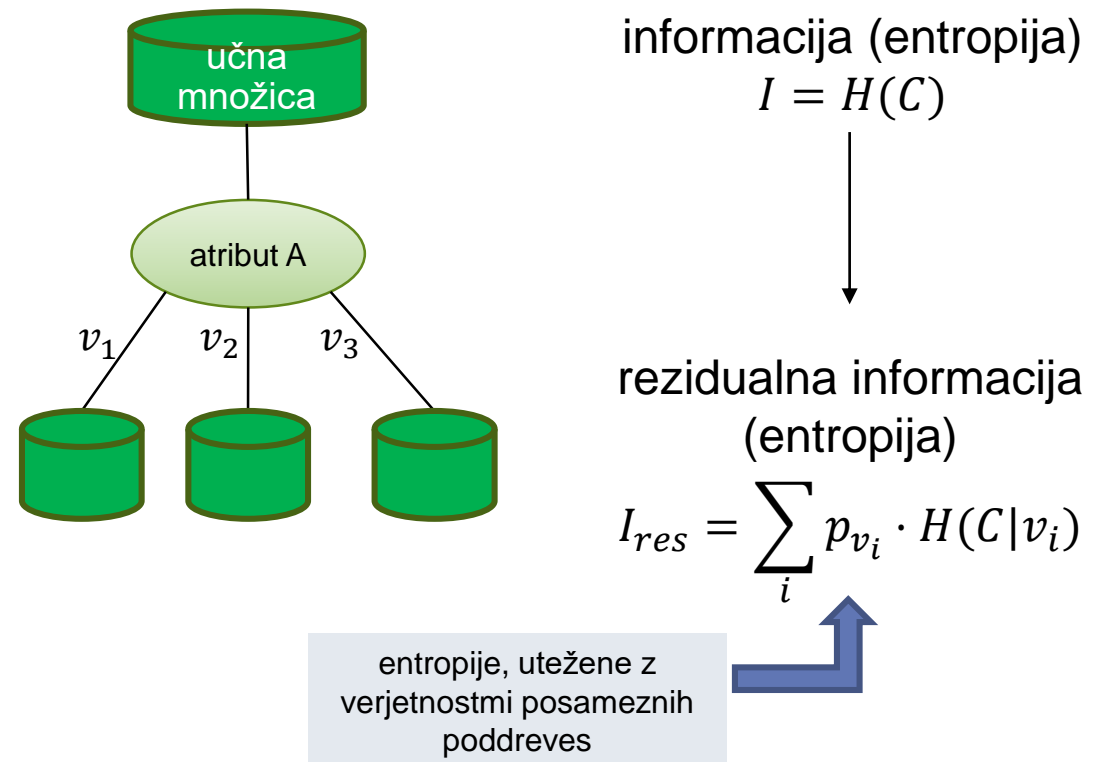
Informacijski prispevek

- dejansko nas zanima **znižanje entropije** (nedoločenosti) ob delitvi učne množice glede na vrednosti atributa A
- znižanje entropije ob delitvi učne množice glede na vrednosti atributa A
- **informacijski prispevek:**

$$Gain(A) = I - I_{res}(A)$$

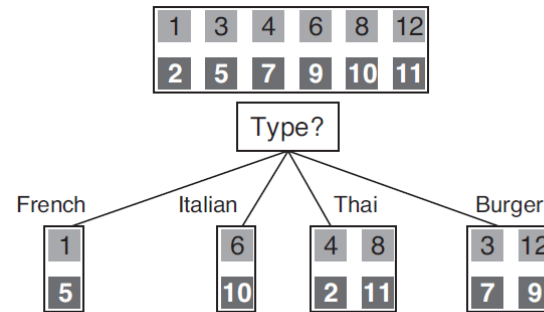
$$I_{res} = - \sum_{v_i \in A} p_{v_i} \sum_c p(c|v_i) \log_2 p(c|v_i)$$

- najbolj informativni atribut **maksimizira informacijski prispevek** (minimizira I_{res})

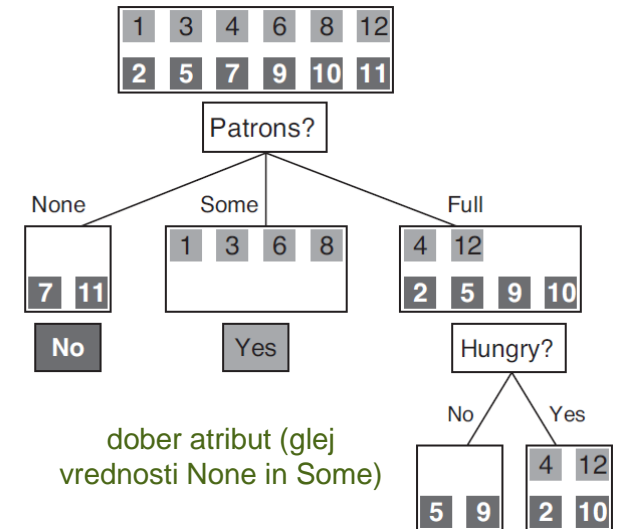


Izbor najbolj pomembnega atributa

Example	Attributes										Target
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	WillWait
X ₁	T	F	F	T	Some	\$\$\$	F	T	French	0-10	T
X ₂	T	F	F	T	Full	\$	F	F	Thai	30-60	F
X ₃	F	T	F	F	Some	\$	F	F	Burger	0-10	T
X ₄	T	F	T	T	Full	\$	F	F	Thai	10-30	T
X ₅	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X ₆	F	T	F	T	Some	\$\$	T	T	Italian	0-10	T
X ₇	F	T	F	F	None	\$	T	F	Burger	0-10	F
X ₈	F	F	F	T	Some	\$\$	T	T	Thai	0-10	T
X ₉	F	T	T	F	Full	\$	T	F	Burger	>60	F
X ₁₀	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30	F
X ₁₁	F	F	F	F	None	\$	F	F	Thai	0-10	F
X ₁₂	T	T	T	T	Full	\$	F	F	Burger	30-60	T



slab atribut (slabo loči pozitivne in negativne primere)



dober atribut (glej vrednosti None in Some)

- znižanje entropije ob delitvi učne množice glede na vrednosti atributa A

- $Gain(A) = I - I_{res}(A)$

$$I = -p(T) \log_2 p(T) - p(F) \log_2 p(F) = -\frac{6}{12} \log_2 \frac{6}{12} - \frac{6}{12} \log_2 \frac{6}{12} = -\log_2 \frac{1}{2} = 1$$

$$I_{res}(Type) = -\frac{2}{12} \left[\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right] - \frac{2}{12} \left[\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right] - \frac{4}{12} \left[\frac{2}{4} \log_2 \frac{2}{4} + \frac{2}{4} \log_2 \frac{2}{4} \right] - \frac{4}{12} \left[\frac{2}{4} \log_2 \frac{2}{4} + \frac{2}{4} \log_2 \frac{2}{4} \right] = 1$$

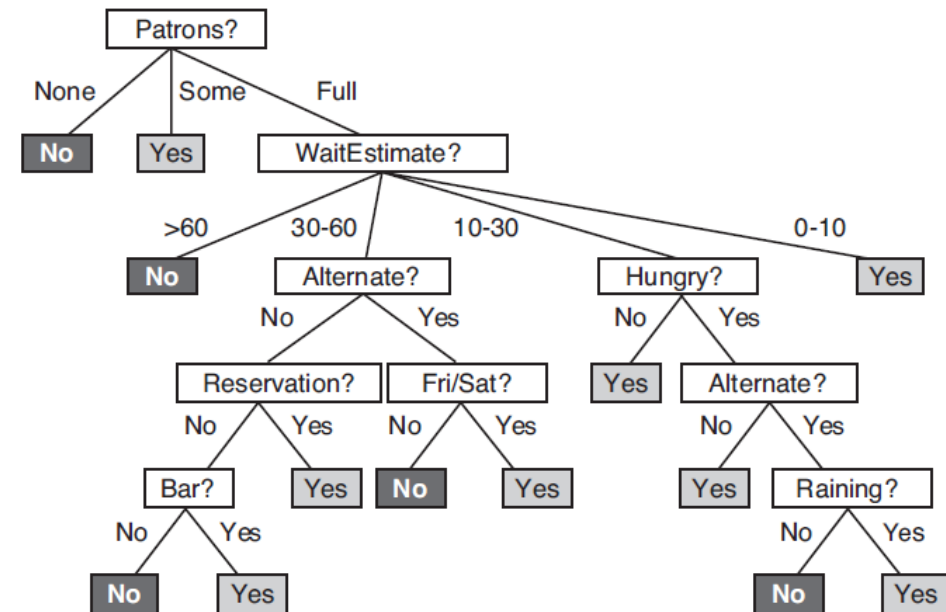
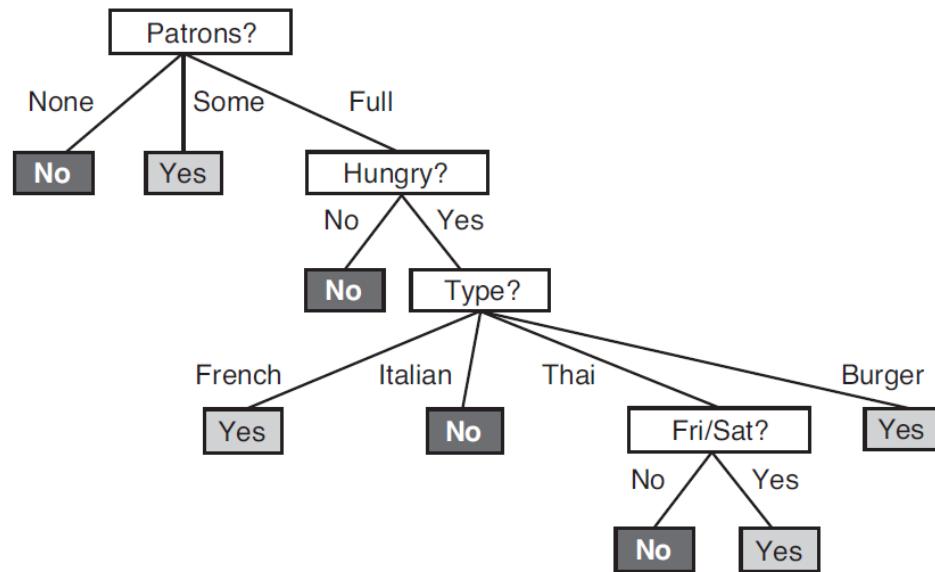
$$I_{res}(Patrons) = -\frac{2}{12} \cdot 0 - \frac{4}{12} \cdot 0 - \frac{6}{12} \left[\frac{2}{6} \log_2 \frac{2}{6} + \frac{4}{6} \log_2 \frac{4}{6} \right] \approx 0,46$$

$$Gain(Type) = 1 - 1 = 0$$

$$Gain(Patrons) = 1 - 0,46 = 0,54$$

Primer

- naučeno odločitveno drevo (levo) je krajše od ročno zgrajenega drevesa (desno)



- obe drevesi sta konsistentni s primeri
- v zgrajenem drevesu ne nastopajo vsi atributi (npr. *Raining* in *Reservation*), zakaj?

Večvrednostni atributi

- težava z atributi, ki imajo več kot dve vrednosti: informacijski prispevek precenjuje njihovo kakovost (entropija je višja na račun večjega števila vrednosti in ne na račun kakovosti atributa)
- primer: dve situaciji z enako nedoločenostjo razredov:
 - dve vrednosti: **111222**, $H = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = -\log_2\frac{1}{2} = 1$
 - tri vrednosti: **112233**, $H = -\frac{1}{3}\log_2\frac{1}{3} - \frac{1}{3}\log_2\frac{1}{3} - \frac{1}{3}\log_2\frac{1}{3} = -\log_2\frac{1}{3} = 1,58$
- rešitve:
 1. normalizacija informacijskega prispevka (**relativni informacijski prispevek**)
 2. uporaba **alternativnih mer** (informacijskih, ocene verjetnosti itd.)
 3. **binarizacija** atributov

Relativni informacijski prispevek in Gini

- Normalizacija informacijskega prispevka: **information gain ratio** (sistem ID3, Quinlan, 1986)

$$Gain(a) = I - I_{res}(A)$$

$$I(A) = - \sum_v p_v \log_2 p_v$$

$$GainRatio(A) = \frac{Gain(A)}{I(A)} = \frac{I - I_{res}(A)}{I(A)}$$

v – vrednost atributa

c – razred

informacija, ki jo potrebujemo
za določitev vrednosti atributa
A (entropija atributa)

- Alternativna mera za kakovost atributa: **Gini index**
 - ocena pričakovane klasifikacijske napake (vsota produktov verjetnosti razredov)

$$Gini = \sum_{c_1 \neq c_2} p(c_1)p(c_2)$$

$$Gini(A) = \sum_v p(v) \sum_{c_1 \neq c_2} p(c_1|v)p(c_2|v)$$

Izpitna naloga

- 2. izpitni rok, 15. 2. 2018 (prilagojena naloga)

Podana je učna množica primerov, ki je prikazana v tabeli (*vreme* in *pritisk* sta atributa, *glavobol* pa je razred). Naloge:

- a) Zgradi odločitveno drevo, pri čemer za ocenjevanje atributov uporabi informacijski prispevek. V primeru enakega števila primerov – predstavnikov obeh razredov – naj vozlišče klasificira v večinski razred iz učne množice.
- b) Ali bi dobljeno drevo bilo drugačno, če bi uporabili razmerje | informacijskega prispevka? Utemelji.
- c) V kateri razred bi drevo klasificiralo učni primer z vrednostmi atributov *vreme=deževno*, *pritisk=srednji*?

vreme	pritisk	glavobol
sončno	nizek	ne
sončno	nizek	ne
sončno	srednji	da
sončno	visok	ne
sončno	nizek	ne
sončno	nizek	da
deževno	srednji	ne
deževno	srednji	da
deževno	visok	da

Binarizacija atributov

- alternativa za reševanje problematike z večvrednostnimi atributi
- zalogo vrednosti atributa lahko razbijemo v dve množici
- primer: atribut $barva \in \{rdeča, rumena, zelena, modra\}$
- strategije:
 - $\{\{rdeča\}, \{rumena, zelena, modra\}\}$ (one-vs-all)
 - $\{\{rdeča, rumena\}, \{zelena, modra\}\}$
 - vpeljava binarnih atributov za vsako barvo
 - itd.
- prednost: manjše vejanje drevesa (statistično bolj zanesljivo, možna višja klasifikacijska točnost)
 - različne načine binarizacije atributa lahko nastopajo kot samostojni atributi, ki se v drevesu pojavijo večkrat

Kratkovidnost algoritma TDIDT

- TDIDT je požrešni algoritem, ki "lokalno" izbira najboljši atribut in ne upošteva, kako dobro drugi algoritmi dopolnjujejo izbrani atribut
- prednosti in slabosti zgornjega pristopa?
- kratkovidnost (angl. myopy) izbora atributa
- primer: problem XOR



A_1	A_2	Razred
0	0	0
0	1	1
1	0	1
1	1	0

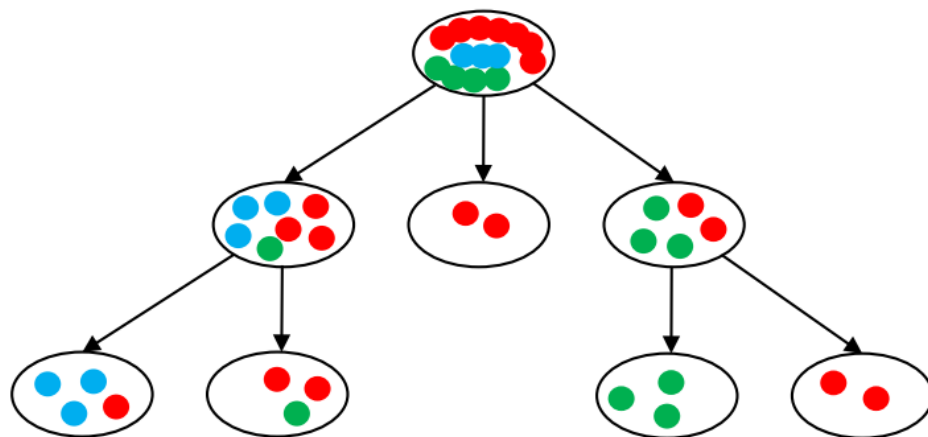
$Gain(A_1) = ?$

$Gain(A_2) = ?$

$Gain(A_1 A_2) = ?$

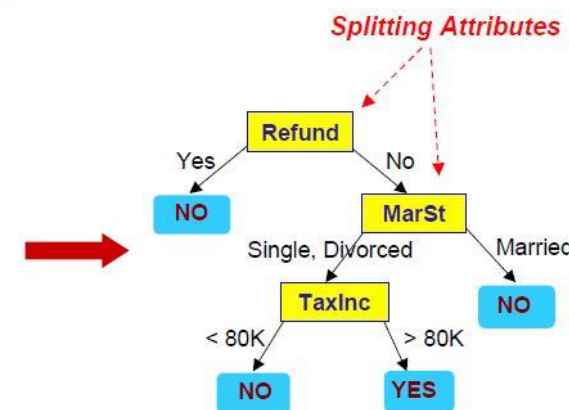
Prostor hipotez odločitvenih dreves

- **diskretni atributi** - odločitvena drevesa delijo prvotno učno množico na vse manjše podmnožice (cilj: maksimizirati čistost podmnožic)
- **zvezni atributi** – delitev podmnožice glede na smiselno mejo izbranega atributa
- primera:



Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

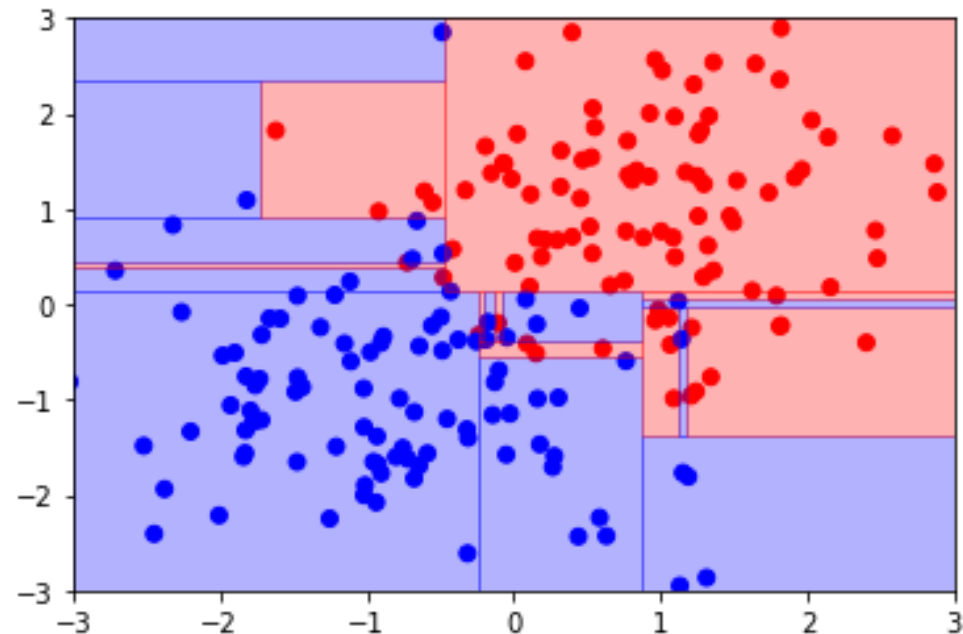
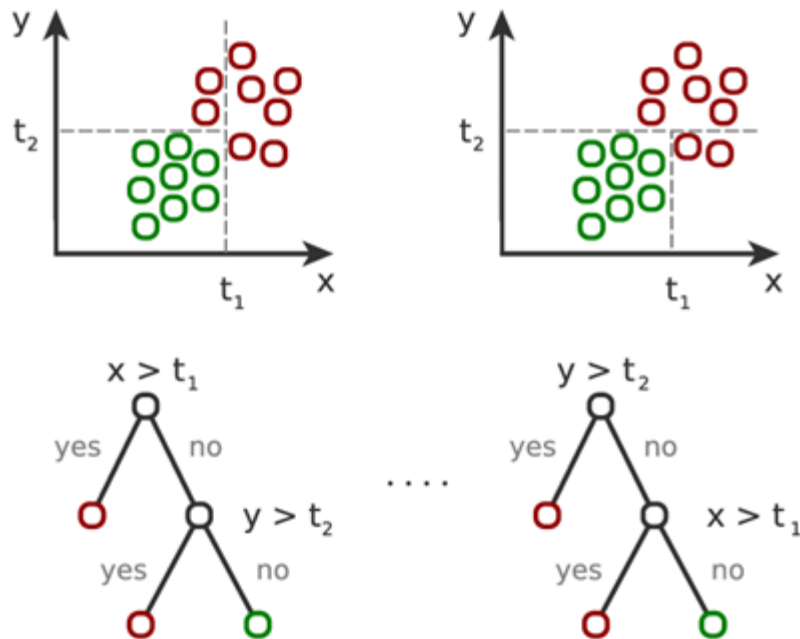
Training Data



Model: Decision Tree

Prostor hipotez odločitvenih dreves

- **zvezni atributi** (npr. višina, dolžina, IQ, koncentracija ozona, poraba el. energije, ipd.)
- v vozliščih običajno testiramo primerjavo zveznega atributa z izbrano mejo (večje/manjše)
- takšna odločitvena drevesa delijo prostor na particije (hiper-kvadre), katerih meje so vzporedne koordinatnim osem
- dva primera:



Privzeta točnost

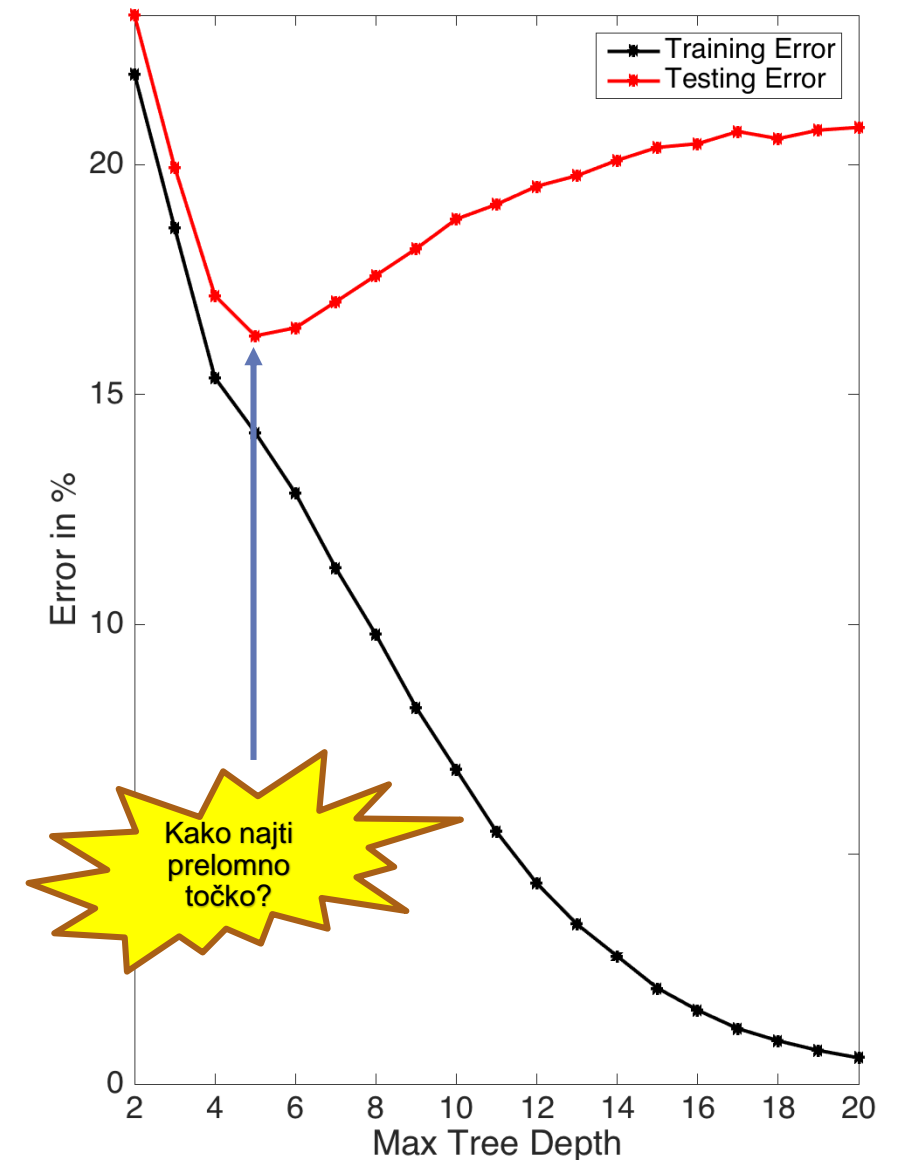
- smiselna mera za **privzeto točnost** (minimalno pričakovano točnost) odločitvenega drevesa je **verjetnost večinskega razreda** v učni množici
- drevo je koristno/uporabno, če je njegova točnost **višja od privzete točnosti**

Primer:

- $[\text{\#Yes}, \text{\#No}] = [3, 7]$
- pričakovana točnost (verjetnost večinskega razreda) je 0,7
- želimo, da ima zgrajeno drevo na testnih podatkih višjo točnost

Pristranost na učni množici

- cilj: maksimiziraj pričakovano točnost drevesa (vendar ne na učnih podatkih - **pretirano prilagajanje?**)
- alternativa – uporaba **nevidenih primerov**:
 - izvzamemo posebno množico **testnih primerov**, če imamo dovolj podatkov (ostane manj podatkov za gradnjo)
 - tipična delitev podatkov: **učna množica** (70%), **testna množica** (30%)





**Učenje dreves, rezanje,
šumni podatki**