

# *Mašinsko učenje.*

Danijela Petrović

May 31, 2016

- Bavi se *izgradnjom računarskih sistema* koji su sposobni da *poboljšavaju* svoje performanse koristeći informacije iz *iskustva*.

- Bavi se *izgradnjom računarskih sistema* koji su sposobni da *poboljšavaju* svoje performanse koristeći informacije iz *iskustva*.
- Pokušava da se približi i objasni *čovekovo učenje*.

- Bavi se *izgradnjom računarskih sistema* koji su sposobni da *poboljšavaju* svoje performanse koristeći informacije iz *iskustva*.
- Pokušava da se približi i objasni *čovekovo učenje*.
- **Glavna pitanja su:**
  - Šta se može naučiti i pod kojim uslovima?
  - Kako se povećava efikasnost učenja u zavisnosti od obima iskustva?
  - Koji su algoritmi pogodni za koje vrste problema?

- Bavi se *izgradnjom računarskih sistema* koji su sposobni da *poboljšavaju* svoje performanse koristeći informacije iz *iskustva*.
- Pokušava da se približi i objasni *čovekovo učenje*.
- **Glavna pitanja su:**
  - Šta se može naučiti i pod kojim uslovima?
  - Kako se povećava efikasnost učenja u zavisnosti od obima iskustva?
  - Koji su algoritmi pogodni za koje vrste problema?
- Praktični rezultati su prethodili teorijskim.

- **Donošenje zaključaka** o *nepoznatim slučajevima*, na osnovu znanja o nekim drugim, poznatim, slučajevima.

# Generalizacija

- **Donošenje zaključaka** o *nepoznatim slučajevima*, na osnovu znanja o nekim drugim, poznatim, slučajevima.
- Proces u kome se znanje koje važi za neki skup slučajeva prenosi na neki njegov nadskup, naziva se *generalizacijom ili induktivnim zaključivanjem*.

# Generalizacija

- **Donošenje zaključaka** o *nepoznatim slučajevima*, na osnovu znanja o nekim drugim, poznatim, slučajevima.
- Proces u kome se znanje koje važi za neki skup slučajeva prenosi na neki njegov nadskup, naziva se *generalizacijom ili induktivnim zaključivanjem*.
- *Apstrakcija* – određeni aspekti entiteta moraju biti zanemareni da bi generalizacija bila uspešna.



# Generalizacija

- **Donošenje zaključaka** o *nepoznatim slučajevima*, na osnovu *znanja* o nekim drugim, poznatim, slučajevima.
- Proces u kome se znanje koje važi za neki skup slučajeva prenosi na neki njegov nadskup, naziva se *generalizacijom ili induktivnim zaključivanjem*.
- *Apstrakcija* – određeni aspekti entiteta moraju biti zanemareni da bi generalizacija bila uspešna.
- Ne pruža garanciju da je zaključivanje ispravno.

# Generalizacija

- **Donošenje zaključaka** o *nepoznatim slučajevima*, na osnovu *znanja* o nekim drugim, poznatim, slučajevima.
- Proces u kome se znanje koje važi za neki skup slučajeva prenosi na neki njegov nadskup, naziva se *generalizacijom ili induktivnim zaključivanjem*.
- *Apstrakcija* – određeni aspekti entiteta moraju biti zanemareni da bi generalizacija bila uspešna.
- Ne pruža garanciju da je zaključivanje ispravno.
- Manji primer sa 'spam' porukama.

## *Nadgledano i nenadgledano učenje*

- *Nadgledano učenje* – algoritmu se daju **podaci iz kojih uči i željeni izlazi**. Algoritam treba da nauči da za date podatke pruži odgovarajuće izlaze.

# Nadgledano i nenadgledano učenje

- *Nadgledano učenje* – algoritmu se daju **podaci iz kojih uči i željeni izlazi**. Algoritam treba da nauči da za date podatke pruži odgovarajuće izlaze.
- *Nenadgledano učenje* – algoritmu pružaju **samo podaci** bez izlaza. Algoritam treba sam da uoči neke zakonitosti u podacima koji su mu dati.

## *Ciljna funkcija i modeli podataka*

- *Ciljna funkcija* – ono što je potrebno naučiti.

## *Ciljna funkcija i modeli podataka*

- *Ciljna funkcija* – ono što je potrebno naučiti.
- Učenje predstavlja **približno određivanje ove funkcije**.

## *Ciljna funkcija i modeli podataka*

- *Ciljna funkcija* – ono što je potrebno naučiti.
- Učenje predstavlja **približno određivanje ove funkcije**.
- *Modeli podataka, hipoteze* – funkcija kojom aproksimiramo ciljnu funkciju.

## Ciljna funkcija i modeli podataka

- *Ciljna funkcija* – ono što je potrebno naučiti.
- Učenje predstavlja **približno određivanje ove funkcije**.
- *Modeli podataka, hipoteze* – funkcija kojom aproksimiramo ciljnu funkciju.
- *Prostor hipoteza* – skup svih dopustivih hipoteza (linearne funkcije, if-then-else itd..)



## Ciljna funkcija i modeli podataka

- *Ciljna funkcija* – ono što je potrebno naučiti.
- Učenje predstavlja **približno određivanje ove funkcije**.
- *Modeli podataka, hipoteze* – funkcija kojom aproksimiramo ciljnu funkciju.
- *Prostor hipoteza* – skup svih dopustivih hipoteza (linearne funkcije, if-then-else itd..)
- Algoritmi učenja == pretraga prostora hipoteza korišćenjem podataka iz iskustva

## Ciljna funkcija i modeli podataka

- *Ciljna funkcija* – ono što je potrebno naučiti.
- Učenje predstavlja **približno određivanje ove funkcije**.
- *Modeli podataka, hipoteze* – funkcija kojom aproksimiramo ciljnu funkciju.
- *Prostor hipoteza* – skup svih dopustivih hipoteza (linearne funkcije, if-then-else itd..)
- Algoritmi učenja == pretraga prostora hipoteza korišćenjem podataka iz iskustva
- Izbor prostora hipoteza je jako značajan za kvalitet učenja:

## Ciljna funkcija i modeli podataka

- *Ciljna funkcija* – ono što je potrebno naučiti.
- Učenje predstavlja **približno određivanje ove funkcije**.
- *Modeli podataka, hipoteze* – funkcija kojom aproksimiramo ciljnu funkciju.
- *Prostor hipoteza* – skup svih dopustivih hipoteza (linearne funkcije, if-then-else itd..)
- Algoritmi učenja == pretraga prostora hipoteza korišćenjem podataka iz iskustva
- Izbor prostora hipoteza je jako značajan za kvalitet učenja:
  - dobro bi bilo da **sadrži ciljnu funkciju**

## Ciljna funkcija i modeli podataka

- *Ciljna funkcija* – ono što je potrebno naučiti.
- Učenje predstavlja **približno određivanje ove funkcije**.
- *Modeli podataka, hipoteze* – funkcija kojom aproksimiramo ciljnu funkciju.
- *Prostor hipoteza* – skup svih dopustivih hipoteza (linearne funkcije, if-then-else itd..)
- Algoritmi učenja == pretraga prostora hipoteza korišćenjem podataka iz iskustva
- Izbor prostora hipoteza je jako značajan za kvalitet učenja:
  - dobro bi bilo da **sadrži ciljnu funkciju**
  - **ne sme biti preveliko**

- *Instanca* – objekat

- *Instanca* – objekat
- Reprezentacija mora biti pogodna za primene algoritama mašinskog učenja

- *Instanca* – objekat
- Reprezentacija mora biti pogodna za primene algoritama mašinskog učenja
- Najčeće se instance predstavljaju *atributima, svojstvima, karakteristikama*.

## *Podaci za trening i podaci za testiranje*

- *Podaci za trening* – podaci na osnovu kojih se vrši generalizacija



## *Podaci za trening i podaci za testiranje*

- *Podaci za trening* – podaci na osnovu kojih se vrši generalizacija
- *Trening skup* – skup podataka za trening

## *Podaci za trening i podaci za testiranje*

- *Podaci za trening* – podaci na osnovu kojih se vrši generalizacija
- *Trening skup* – skup podataka za trening
- *Podaci za testiranje* – služe za procenu kvaliteta naučenog znanja

## *Podaci za trening i podaci za testiranje*

- *Podaci za trening* – podaci na osnovu kojih se vrši generalizacija
- *Trening skup* – skup podataka za trening
- *Podaci za testiranje* – služe za procenu kvaliteta naučenog znanja
- *Test skup* – skupo podataka za testiranje

## *Podaci za trening i podaci za testiranje*

- *Podaci za trening* – podaci na osnovu kojih se vrši generalizacija
- *Trening skup* – skup podataka za trening
- *Podaci za testiranje* – služe za procenu kvaliteta naučenog znanja
- *Test skup* – skupo podataka za testiranje
- Test skup bi trebalo da bude **disjuktan** sa trening skupom.

- Nadgledano ili nenadgledano učenje

## *Dizaj sistema koji uči*

- Nadgledano ili nenadgledano učenje
- Zapis podataka

## *Dizaj sistema koji uči*

- Nadgledano ili nenadgledano učenje
- Zapis podataka
- Izbor ciljne funkcije

- Nadgledano ili nenadgledano učenje
- Zapis podataka
- Izbor ciljne funkcije
- Izbor prostora hipoteza



# *Dizaj sistema koji uči*

- Nadgledano ili nenadgledano učenje
- Zapis podataka
- Izbor ciljne funkcije
- Izbor prostora hipoteza
- Izbor algoritma

- Nadgledano ili nenadgledano učenje
- Zapis podataka
- Izbor ciljne funkcije
- Izbor prostora hipoteza
- Izbor algoritma
- Izbor mera kvaliteta učenja

- Razvrstavanje instanci po klasama (autori, spam, slike, članci itd...).

- Razvrstavanje instanci po klasama (autori, spam, slike, članci itd...).
- **Određivanju vrednosti atributa** klase na osnovu preostalih atributa instance.

- Razvrstavanje instanci po klasama (autori, spam, slike, članci itd...).
- **Određivanju vrednosti atributa** klase na osnovu preostalih atributa instance.
- **Aproksimacija ciljne funkcije** koja svakoj instanci dodeljuje **oznaku klase** kojoj ta instanca pripada.

- Razvrstavanje instanci po klasama (autori, spam, slike, članci itd...).
- **Određivanju vrednosti atributa** klase na osnovu preostalih atributa instance.
- **Aproksimacija ciljne funkcije** koja svakoj instanci dodeljuje **oznaku klase** kojoj ta instanca pripada.
- Ciljna funkcija u ovom problemu je *diskretna*.

- Razvrstavanje instanci po klasama (autori, spam, slike, članci itd...).
- **Određivanju vrednosti atributa** klase na osnovu preostalih atributa instance.
- **Aproksimacija ciljne funkcije** koja svakoj instanci dodeljuje **oznaku klase** kojoj ta instanca pripada.
- Ciljna funkcija u ovom problemu je *diskretna*.
- Atribut klase, čiju je vrednost potrebno odrediti, je *kategorički atribut*.

- Razvrstavanje instanci po klasama (autori, spam, slike, članci itd...).
- **Određivanju vrednosti atributa** klase na osnovu preostalih atributa instance.
- **Aproksimacija ciljne funkcije** koja svakoj instanci dodeljuje **oznaku klase** kojoj ta instanca pripada.
- Ciljna funkcija u ovom problemu je *diskretna*.
- Atribut klase, čiju je vrednost potrebno odrediti, je *kategorički atribut*.
- Postoji više metoda klasifikacije: metode zasnovane na instancama, učenje stabla odlučivanja, metoda potpornih vektora i metode bayesovske klasifikacije zasnovane na verovatnoći.



## *Metode klasifikacije zasnovane na instancama*

- Ne gradi se model

## *Metode klasifikacije zasnove na instancama*

- Ne gradi se model
- Instance se čuvaju i bivaju upotrebljene kada je potrebno klasifikovati nepoznatu instancu

## *Metoda n-najbližih suseda*

- Nepoznatu instancu treba klasifikovati u klasu čije su instance najbližnje nepoznatoj.

## Metoda $n$ -najbližih suseda

- Nepoznatu instancu treba klasifikovati u klasu čije su instance najbližije nepoznatoj.
- Za sličnost koristimo *funkciju rastojanja*.

## Metoda $n$ -najbližih suseda

- Nepoznatu instancu treba klasifikovati u klasu čije su instance najbližije nepoznatoj.
- Za sličnost koristimo *funkciju rastojanja*.
- Nalaženju  $n$  instanci iz trening skupa koje su najbliže nepoznatoj instanci.

## Metoda $n$ -najbližih suseda

- Nepoznatu instancu treba klasifikovati u klasu čije su instance najbližije nepoznatoj.
- Za sličnost koristimo *funkciju rastojanja*.
- Nalaženju  $n$  instanci iz trening skupa koje su najbliže nepoznatoj instanci.
- U slučaju **izjednačenog ishoda** između više klasa nije moguće doneti odluku (u praksi se to nekako razrešava).

## Metoda $n$ -najbližih suseda

- Nepoznatu instancu treba klasifikovati u klasu čije su instance najbližije nepoznatoj.
- Za sličnost koristimo *funkciju rastojanja*.
- Nalaženju  $n$  instanci iz trening skupa koje su najbliže nepoznatoj instanci.
- U slučaju **izjednačenog ishoda** između više klasa nije moguće doneti odluku (u praksi se to nekako razrešava).
- Primer sa grafikom i tačkom u unutrašnjosti i na obodu.

## Metoda $n$ -najbližih suseda

- Nepoznatu instancu treba klasifikovati u klasu čije su instance najbližije nepoznatoj.
- Za sličnost koristimo *funkciju rastojanja*.
- Nalaženju  $n$  instanci iz trening skupa koje su najbliže nepoznatoj instanci.
- U slučaju **izjednačenog ishoda** između više klasa nije moguće doneti odluku (u praksi se to nekako razrešava).
- Primer sa grafikom i tačkom u unutrašnjosti i na obodu.
- Određivanje  $n$  se vrši se empirijski – tako da bude najefikasnije za klasifikaciju.



- Često se koriste zajedno sa metodom  $n$ -najbližih suseda

- Često se koriste zajedno sa metodom  $n$ -najbližih suseda
- Podatak se deli na delove dužine  $n$

- Često se koriste zajedno sa metodom  $n$ -najbližih suseda
- Podatak se deli na delove dužine  $n$
- Predstavljaju reprezentaciju pogodnu za metode klasifikacije

- Često se koriste zajedno sa metodom  $n$ -najbližih suseda
- Podatak se deli na delove dužine  $n$
- Predstavljaju reprezentaciju pogodnu za metode klasifikacije
- **Prednosti:** robusnost, jednostavnost, nezavisnost od domena koji se analizira, efikasnost

- Često se koriste zajedno sa metodom  $n$ -najbližih suseda
- Podatak se deli na delove dužine  $n$
- Predstavljaju reprezentaciju pogodnu za metode klasifikacije
- **Prednosti:** robusnost, jednostavnost, nezavisnost od domena koji se analizira, efikasnost
- **Problem:** Eksponencijalna zavisnost broja  $n$ -grama u odnosu na dužinu  $n$ -grama

Algoritmom 3 najbliža suseda klasifikovati instance iz trening skupa. Pri tom, koristiti Menhetn rastojanje. Izračunati preciznost, i udele tačno i lažno pozitivnih i tačno i lažno negativnih.

Trening skup

| $X_1$ | $X_2$ | $X_3$ | Klasa |
|-------|-------|-------|-------|
| 1     | 1     | 0     | A     |
| 1     | 0     | 2     | A     |
| 2     | 2     | 3     | A     |
| 3     | 2     | 4     | B     |
| 1     | 4     | 3     | B     |
| 4     | 3     | 3     | B     |

Test skup

| $X_1$ | $X_2$ | $X_3$ | Klasa |
|-------|-------|-------|-------|
| 0     | 0     | 0     | A     |
| 3     | 3     | 3     | A     |
| 1     | 3     | 4     | B     |
| 4     | 5     | 3     | B     |

Date su instance  $(0,0,A)$ ,  $(1,1,A)$ ,  $(1,2,A)$ ,  $(0,2,A)$ ,  $(1,5,B)$ ,  $(4,5,B)$ ,  $(5,6,B)$ ,  $(5,2,C)$ ,  $(4,0,C)$ , pri čemu prve dve koordinate predstavljaju koordinate tačke, a poslednja koordinata predstavlja oznaku klase. Algoritmom 3 najbliža suseda odrediti kojoj klasi pripadaju instance  $(0,1,A)$ ,  $(4,3,B)$ ,  $(3,1,C)$ ? Kao meru rastojanja koristiti Euklidovo rastojanje u ravni. Odrediti preciznost.

- Ciljna funkcija  $\Rightarrow$  predstavlja se uz pomoć stabla odlučivanja



## Učenje stabla odlučivanja

- Ciljna funkcija  $\Rightarrow$  predstavlja se uz pomoć stabla odlučivanja
- Instance se predstavljaju pomoću *fiksno*g broja atributa.

## Učenje stabla odlučivanja

- Ciljna funkcija == predstavlja se uz pomoć stabla odlučivanja
- Instance se predstavljaju pomoću *fiksno*g broja atributa.
- Skup vrednosti – **diskretan i mali**

## Učenje stabla odlučivanja

- Ciljna funkcija == predstavlja se uz pomoć stabla odlučivanja
- Instance se predstavljaju pomoću *fiksno*g broja atributa.
- Skup vrednosti – **diskretan i mali**
- Pogodni su kada podaci za trening sadrže greške ili fali vrednost nekog atributa.

## Učenje stabla odlučivanja

- Ciljna funkcija == predstavlja se uz pomoć stabla odlučivanja
- Instance se predstavljaju pomoću *fiksnog broja atributa*.
- Skup vrednosti – **diskretan i mali**
- Pogodni su kada podaci za trening sadrže greške ili fali vrednost nekog atributa.
- *ID3* – najpoznatiji algoritam za učenje stabla odlučivanja

## Učenje stabla odlučivanja

- Ciljna funkcija == predstavlja se uz pomoć stabla odlučivanja
- Instance se predstavljaju pomoću *fiksnog broja atributa*.
- Skup vrednosti – **diskretan i mali**
- Pogodni su kada podaci za trening sadrže greške ili fali vrednost nekog atributa.
- *ID3* – najpoznatiji algoritam za učenje stabla odlučivanja
- Bira se atribut koji maksimizuje dobitak informacije.

## *Učenje stabla odlučivanja*

- $p_i$  – verovatnoća da instanca pripada  $i$ -toj klasi

## Učenje stabla odlučivanja

- $p_i$  – verovatnoća da instanca pripada  $i$ -toj klasi
- Potrebno je odrediti *meru neuređivosti nekog skupa*

## Učenje stabla odlučivanja

- $p_i$  – verovatnoća da instanca pripada  $i$ -toj klasi
- Potrebno je odrediti *meru neure d enosti nekog skupa*
- *Greška klasifikacije:  $Err(S) = 1 - \max_i p_i$*



## Učenje stabla odlučivanja

- $p_i$  – verovatnoća da instanca pripada  $i$ -toj klasi
- Potrebno je odrediti *meru neure d enosti nekog skupa*
- *Greška klasifikacije:*  $Err(S) = 1 - \max_i p_i$
- *Entropija:*  $Entropija(S) = - \sum_{i=1}^c p_i \log_2 p_i$

## Učenje stabla odlučivanja

- $p_i$  – verovatnoća da instanca pripada  $i$ -toj klasi
- Potrebno je odrediti *meru neure d enosti nekog skupa*
- *Greška klasifikacije:*  $Err(S) = 1 - \max_i p_i$
- *Entropija:*  $Entropija(S) = - \sum_{i=1}^c p_i \log_2 p_i$
- *Dobitak:*

$$Dobitak(S, A) = Err(S) - \sum_{v \in Vred(A)} \frac{|S_v|}{|S|} \cdot Err(S_v)$$

## Učenje stabla odlučivanja

- $p_i$  – verovatnoća da instanca pripada  $i$ -toj klasi
- Potrebno je odrediti *meru neure d enosti nekog skupa*
- *Greška klasifikacije:*  $Err(S) = 1 - \max_i p_i$
- *Entropija:*  $Entropija(S) = - \sum_{i=1}^c p_i \log_2 p_i$
- *Dobitak:*

$$Dobitak(S, A) = Err(S) - \sum_{v \in Vred(A)} \frac{|S_v|}{|S|} \cdot Err(S_v)$$

- $|S|$  – veličina skupa

## Učenje stabla odlučivanja

- $p_i$  – verovatnoća da instanca pripada  $i$ -toj klasi
- Potrebno je odrediti *meru neure d enosti nekog skupa*
- *Greška klasifikacije:*  $Err(S) = 1 - \max_i p_i$
- *Entropija:*  $Entropija(S) = - \sum_{i=1}^c p_i \log_2 p_i$
- *Dobitak:*

$$Dobitak(S, A) = Err(S) - \sum_{v \in Vred(A)} \frac{|S_v|}{|S|} \cdot Err(S_v)$$

- $|S|$  – veličina skupa
- $|S_v|$  – veličina skupa sa vrednosti atributa  $Vred(A) = v$

| <b>Životinja</b> | <b>Veličina</b> | <b>Ishrana</b> | <b>Otr</b> | <b>Br Nogu</b> | <b>PV</b> | <b>Opasna</b> |
|------------------|-----------------|----------------|------------|----------------|-----------|---------------|
| Lav              | Velika          | Meso           | Ne         | 4              | N         | DA            |
| Mačka            | Mala            | Meso           | Ne         | 4              | N         | NE            |
| Koza             | Mala            | Biljke         | Ne         | 4              | P         | NE            |
| Zec              | Mala            | Biljke         | Ne         | 4              | N         | NE            |
| Zmaj             | Velika          | Meso           | Da         | 4              | N         | DA            |

# Primer

| <b>Životinja</b> | <b>Veličina</b> | <b>Ishrana</b> | <b>Otr</b> | <b>Br Nogu</b> | <b>PV</b> | <b>Opasna</b> |
|------------------|-----------------|----------------|------------|----------------|-----------|---------------|
| Lav              | Velika          | Meso           | Ne         | 4              | N         | DA            |
| Mačka            | Mala            | Meso           | Ne         | 4              | N         | NE            |
| Koza             | Mala            | Biljke         | Ne         | 4              | P         | NE            |
| Zec              | Mala            | Biljke         | Ne         | 4              | N         | NE            |
| Zmaj             | Velika          | Meso           | Da         | 4              | N         | DA            |
| Zmija            | Mala            | Meso           | Da         | 0              | P         | DA            |
| Pčela ubica      | Mala            | Biljke         | Da         | 6              | P         | DA            |
| Morska krava     | Velika          | Biljke         | Ne         | 0              | P         | NE            |

Otrv = Ne

| Životinja    | Veličina | Ishrana | Br Nogu | PV | Opasna |
|--------------|----------|---------|---------|----|--------|
| Lav          | Velika   | Meso    | 4       | N  | DA     |
| Mačka        | Mala     | Meso    | 4       | N  | NE     |
| Koza         | Mala     | Biljke  | 4       | P  | NE     |
| Zec          | Mala     | Biljke  | 4       | N  | NE     |
| Morska krava | Velika   | Biljke  | 0       | P  | NE     |

- Ne može se desiti da ciljna funkcija nije u prostoru hipoteza (jer sve se može prikazati stablom).



- Ne može se desiti da ciljna funkcija nije u prostoru hipoteza (jer sve se može prikazati stablom).
- Preferira stabla sa manjom dubinom.

## Mere kvaliteta i tehnike evaluacije klasifikacije

- *preciznost* =  $\frac{\text{broj\_dobro\_klasifikovanih}}{\text{ukupan\_broj\_instanci}}$

## Mere kvaliteta i tehnike evaluacije klasifikacije

- *preciznost* =  $\frac{\text{broj\_dobro\_klasifikovanih}}{\text{ukupan\_broj\_instanci}}$
- Prethodna mera nije uvek dovoljna.

## Mere kvaliteta i tehnike evaluacije klasifikacije

- *preciznost* =  $\frac{\text{broj\_dobro\_klasifikovanih}}{\text{ukupan\_broj\_instanci}}$
- Prethodna mera nije uvek dovoljna.
- *osetljivost* klasifikatora

## Mere kvaliteta i tehnike evaluacije klasifikacije

- *preciznost* =  $\frac{\text{broj\_dobro\_klasifikovanih}}{\text{ukupan\_broj\_instanci}}$
- Prethodna mera nije uvek dovoljna.
- *osetljivost* klasifikatora
- *SP – stvarno pozitivne* (pozitivne i klasifikovane kao pozitivne)

## Mere kvaliteta i tehnike evaluacije klasifikacije

- *preciznost* =  $\frac{\text{broj\_dobro\_klasifikovanih}}{\text{ukupan\_broj\_instanci}}$
- Prethodna mera nije uvek dovoljna.
- *osetljivost* klasifikatora
- *SP – stvarno pozitivne* (pozitivne i klasifikovane kao pozitivne)
- *SN - stvarno negativne* (negativne i klasifikovane kao negativne)

## Mere kvaliteta i tehnike evaluacije klasifikacije

- *preciznost* =  $\frac{\text{broj\_dobro\_klasifikovanih}}{\text{ukupan\_broj\_instanci}}$
- Prethodna mera nije uvek dovoljna.
- *osetljivost* klasifikatora
- *SP – stvarno pozitivne* (pozitivne i klasifikovane kao pozitivne)
- *SN - stvarno negativne* (negativne i klasifikovane kao negativne)
- *LP - lažno pozitivne* (negativne, ali greškom klasifikovane kao pozitivne)

## Mere kvaliteta i tehnike evaluacije klasifikacije

- *preciznost* =  $\frac{\text{broj\_dobro\_klasifikovanih}}{\text{ukupan\_broj\_instanci}}$
- Prethodna mera nije uvek dovoljna.
- *osetljivost* klasifikatora
- *SP – stvarno pozitivne* (pozitivne i klasifikovane kao pozitivne)
- *SN - stvarno negativne* (negativne i klasifikovane kao negativne)
- *LP - lažno pozitivne* (negativne, ali greškom klasifikovane kao pozitivne)
- *LN - lažno negativne* (pozitivne, ali greškom klasifikovane kao negativne)



## Mere kvaliteta i tehnike evaluacije klasifikacije

- *USP - udeo stvarno pozitivnih*:  $USP = \frac{SP}{SP+LN}$   
(SP + LN – ukupan broj pozitivnih)

## Mere kvaliteta i tehnike evaluacije klasifikacije

- *USP - udeo stvarno pozitivnih*:  $USP = \frac{SP}{SP+LN}$   
(SP + LN – ukupan broj pozitivnih)
- *USN - udeo stvarno negativnih*:  $USN = \frac{SN}{SN+LP}$

## Mere kvaliteta i tehnike evaluacije klasifikacije

- *USP - udeo stvarno pozitivnih*:  $USP = \frac{SP}{SP+LN}$   
(SP + LN – ukupan broj pozitivnih)
- *USN - udeo stvarno negativnih*:  $USN = \frac{SN}{SN+LP}$
- *ULP - udeo lažno pozitivnih*:  $ULP = \frac{LP}{LP+SN}$

## Mere kvaliteta i tehnike evaluacije klasifikacije

- *USP - udeo stvarno pozitivnih*:  $USP = \frac{SP}{SP+LN}$   
(SP + LN – ukupan broj pozitivnih)
- *USN - udeo stvarno negativnih*:  $USN = \frac{SN}{SN+LP}$
- *ULP - udeo lažno pozitivnih*:  $ULP = \frac{LP}{LP+SN}$
- *ULN - udeo lažno negativnih*:  $ULN = \frac{LN}{LN+SP}$

## Mere kvaliteta i tehnike evaluacije klasifikacije

- *USP - udeo stvarno pozitivnih*:  $USP = \frac{SP}{SP+LN}$   
(SP + LN – ukupan broj pozitivnih)
- *USN - udeo stvarno negativnih*:  $USN = \frac{SN}{SN+LP}$
- *ULP - udeo lažno pozitivnih*:  $ULP = \frac{LP}{LP+SN}$
- *ULN - udeo lažno negativnih*:  $ULN = \frac{LN}{LN+SP}$
- *Preciznost*:  $Preciznost = \frac{SP+SN}{SP+SN+LP+LN}$

Primer 4.2.png

Na osnovu atributa "ima krila", "leže jaja", "leti" konstruisati stablo odlučivanja koje prepoznaje ptice. Za trening koristiti sledeće životinje: roda, krava, vrabac, slepi miš, noj, zebra, gavran. Kolika je preciznost predviđanja tog stabla na sledećem skupu: kokoška, kornjača, konj, lav?

Na osnovu mere "greška klasifikacije" i datih podataka, odabrati najbolji atribut za izgradnju stabla odlučivanja.

| $X_1$ | $X_2$ | $X_3$ | Klasa |
|-------|-------|-------|-------|
| T     | T     | T     | A     |
| F     | T     | T     | A     |
| F     | T     | T     | A     |
| F     | F     | T     | A     |
| F     | F     | F     | A     |
| F     | F     | F     | B     |
| T     | F     | F     | B     |
| T     | F     | F     | B     |
| T     | T     | F     | B     |
| T     | T     | F     | B     |



## Primer

Na osnovu sledećih podataka, konstruisati stablo odlučivanja dubine 1 korišćenjem mere "greška klasifikacije".

|       |   |   |   |   |   |   |   |   |
|-------|---|---|---|---|---|---|---|---|
| A     | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 |
| B     | 1 | 1 | 2 | 3 | 3 | 3 | 1 | 2 |
| C     | 2 | 2 | 1 | 2 | 1 | 3 | 3 | 3 |
| Klasa | + | - | + | + | - | + | - | + |

Izračunati preciznost dobijenog drveta odlučivanja na sledećem test skupu.

|       |   |   |   |   |
|-------|---|---|---|---|
| A     | 1 | 1 | 1 | 3 |
| B     | 1 | 2 | 2 | 3 |
| C     | 1 | 1 | 3 | 1 |
| Klasa | - | + | - | + |

Na osnovu sledećih podataka, konstruisati stablo odlučivanja dubine 1 korišćenjem mere "greška klasifikacije". Trening skup:

| Obrazovan | Visok | Marljiv | Odgovoran | Dobar |
|-----------|-------|---------|-----------|-------|
| Da        | Ne    | Da      | Da        | Da    |
| Ne        | Ne    | Da      | Da        | Da    |
| Ne        | Da    | Ne      | Ne        | Ne    |
| Da        | Ne    | Ne      | Da        | Da    |

Izračunati preciznost dobijenog drveta odlučivanja na sledećem test skupu.

Test skup:

| Obrazovan | Visok | Marljiv | Odgovoran |
|-----------|-------|---------|-----------|
| Da        | Ne    | Da      | Ne        |
| Da        | Da    | Da      | Da        |
| Da        | Ne    | Ne      | Da        |

Na osnovu sledećih podataka, konstruisati stablo odlučivanja dubine 1 korišćenjem mere "greška klasifikacije".

| Temperatura | Kašalj | Glavobolja | Osip | Grip |
|-------------|--------|------------|------|------|
| Visoka      | Da     | Ne         | Ne   | Da   |
| Visoka      | Da     | Da         | Da   | Ne   |
| Normalna    | Da     | Da         | Ne   | Da   |
| Visoka      | Da     | Da         | Ne   | Da   |

Konstruisati stablo odlučivanja potrebne dubine koje prepoznaje parnost 4-bitnih brojeva na osnovu njihovih binarnih reprezentacija. Neka se trening skup sastoji od brojeva 1, 3, 6, 9, 12 i 14. Kolika je preciznost ovog stabla na brojevima 2,4,5 i 7?