

Ansampli

Mašinsko učenje 2020/21.

Matematički fakultet
Univerzitet u Beogradu

Ansambli

- ▶ Ansambl - skup većeg broja modela koji zajednički donose odluke

Ansampli

- ▶ Ansambl - skup većeg broja modela koji zajednički donose odluke
- ▶ Anekdota o poznatom naučniku Frensisu Galtonu

Ansampli

- ▶ Ansambl - skup većeg broja modela koji zajednički donose odluke
- ▶ Anegdota o poznatom naučniku Frensisu Galtonu
- ▶ Veći broj modela koji samostalno možda i nisu dovoljno dobri, ali prave relativno smisljena predviđanja i greše na nezavisne načine

Ansambl

- ▶ Ansambl - skup većeg broja modela koji zajednički donose odluke
- ▶ Anegdota o poznatom naučniku Frensisu Galtonu
- ▶ Veći broj modela koji samostalno možda i nisu dovoljno dobri, ali prave relativno smisljena predviđanja i greške na nezavisne načine
- ▶ Uprosečavanjem, nezavisne greške se poništavaju, što značajno popravlja preciznost predviđanja

Ansampli

- ▶ Postoje različiti pristupi konstrukciji ansambla, ali vredi naglasiti da su neki od njih po preciznosti koju nude dosledno među najboljim metodama za rešavanje različitih problema

Ansampli

- ▶ Postoje različiti pristupi konstrukciji ansambla, ali vredi naglasiti da su neki od njih po preciznosti koju nude dosledno među najboljim metodama za rešavanje različitih problema
- ▶ Tipično u slučaju da su podaci predstavljeni u vektorskom obliku, ansampli predstavljaju najpouzdaniji pristup postizanju visoke preciznosti u predviđanju

Ansambli

- ▶ Postoje različiti pristupi konstrukciji ansambla, ali вреди naglasiti da su neki od njih po preciznosti koju nude dosledno među najboljim metodama za rešavanje različitih problema
- ▶ Tipično u slučaju da su podaci predstavljeni u vektorskom obliku, ansambli predstavljaju najpouzdaniji pristup postizanju visoke preciznosti u predviđanju
- ▶ U takvom slučaju obično su i precizniji i efikasniji i jednostavniji za obučavanje od neuronskih mreža

Ansambli

- ▶ Postoje različiti pristupi konstrukciji ansambla, ali vredi naglasiti da su neki od njih po preciznosti koju nude dosledno među najboljim metodama za rešavanje različitih problema
- ▶ Tipično u slučaju da su podaci predstavljeni u vektorskom obliku, ansambli predstavljaju najpouzdaniji pristup postizanju visoke preciznosti u predviđanju
- ▶ U takvom slučaju obično su i precizniji i efikasniji i jednostavniji za obučavanje od neuronskih mreža
- ▶ Naravno, njihov kvalitet i efikasnost zavise i od vrste modela koji sačinjavaju ansambl

Pregled

Prosta agregacija

Stabla odlučivanja

Slučajne šume

Pojačavanje

Prosta agregacija

- ▶ *Prosta agregacija* (eng. *bootstrap aggregation, bagging*) podrazumeva obučavanje većeg broja modela koji pojedinačno ne moraju imati visoku preciznost, ali čije su greške *nezavisne*

Prosta agregacija

- ▶ *Prosta agregacija* (eng. *bootstrap aggregation, bagging*) podrazumeva obučavanje većeg broja modela koji pojedinačno ne moraju imati visoku preciznost, ali čije su greške *nezavisne*
- ▶ Prilikom predviđanja, svi modeli nude svoja predviđanja, koja se agregiraju kako bi se dobilo predviđanje ansambla

Prosta agregacija

- ▶ *Prosta agregacija* (eng. *bootstrap aggregation, bagging*) podrazumeva obučavanje većeg broja modela koji pojedinačno ne moraju imati visoku preciznost, ali čije su greške *nezavisne*
- ▶ Prilikom predviđanja, svi modeli nude svoja predviđanja, koja se agregiraju kako bi se dobilo predviđanje ansambla
- ▶ Ukoliko se radi o regresiji, agregacija se tipično vrši uprosečavanjem (a moguće je koristiti i medijanu), a u slučaju klasifikacije glasanjem

Prosta agregacija

- ▶ Snaga ovakvog modela počiva na ideji da će se prilikom agregacije greške koje modeli nezavisno prave *poništiti*

Prosta agregacija

- ▶ Snaga ovakvog modela počiva na ideji da će se prilikom agregacije greške koje modeli nezavisno prave *poništiti*
- ▶ Na primer, ukoliko se rešava problem regresije i ako se modeli posmatraju kao nezavisne slučajne promenljive $X_1, \dots, X_m$ sa istom raspodelom koja ima prosek μ i varijansu σ^2 , na osnovu centralne granične teoreme sledi

$$\sqrt{m} \left(\frac{1}{m} \sum_{i=1}^m X_i - \mu \right) \rightarrow \mathcal{N}(0, \sigma^2)$$

Prosta agregacija

- ▶ Snaga ovakvog modela počiva na ideji da će se prilikom agregacije greške koje modeli nezavisno prave *poništiti*
- ▶ Na primer, ukoliko se rešava problem regresije i ako se modeli posmatraju kao nezavisne slučajne promenljive $X_1, \dots, X_m$ sa istom raspodelom koja ima prosek μ i varijansu σ^2 , na osnovu centralne granične teoreme sledi

$$\sqrt{m} \left(\frac{1}{m} \sum_{i=1}^m X_i - \mu \right) \rightarrow \mathcal{N}(0, \sigma^2)$$

- ▶ Prosek modela približno ima očekivanje μ (isto kao svaki model zasebno) i varijansu σ^2/m (manje nego svaki model zasebno!)

Prosta agregacija

- ▶ Snaga ovakvog modela počiva na ideji da će se prilikom agregacije greške koje modeli nezavisno prave *poništiti*
- ▶ Na primer, ukoliko se rešava problem regresije i ako se modeli posmatraju kao nezavisne slučajne promenljive $X_1, \dots, X_m$ sa istom raspodelom koja ima prosek μ i varijansu σ^2 , na osnovu centralne granične teoreme sledi

$$\sqrt{m} \left(\frac{1}{m} \sum_{i=1}^m X_i - \mu \right) \rightarrow \mathcal{N}(0, \sigma^2)$$

- ▶ Prosek modela približno ima očekivanje μ (isto kao svaki model zasebno) i varijansu σ^2/m (manje nego svaki model zasebno!)
- ▶ Prosta agregacija nudi mogućnost smanjenja varijanse bez povećanja sistematskog odstupanja (očekivanje je isto pa je time i sistematsko odstupanje isto)

Pregled

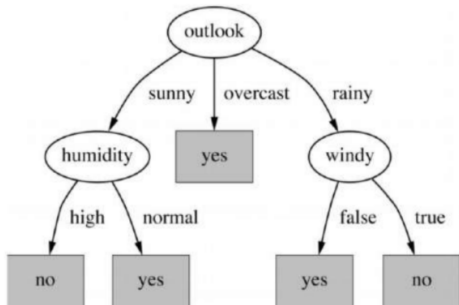
Prosta agregacija

Stabla odlučivanja

Slučajne šume

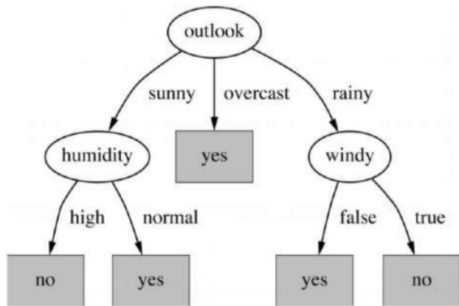
Pojačavanje

Stabla odlučivanja



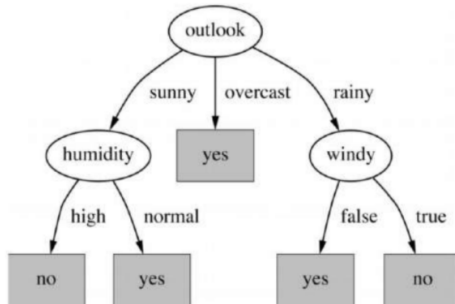
- U svakom čvoru stabla, osim listova, nalazi se po jedan test

Stabla odlučivanja



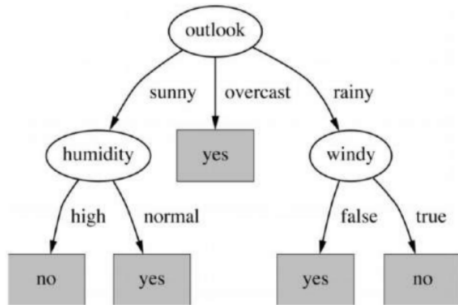
- ▶ U svakom čvoru stabla, osim listova, nalazi se po jedan test
- ▶ Svaki test ima više od jednog ishoda

Stabla odlučivanja



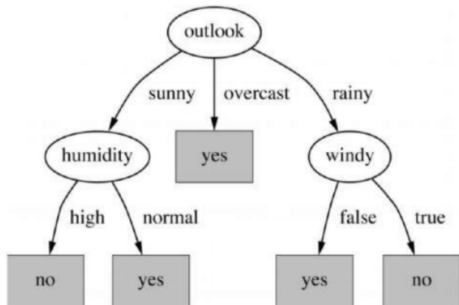
- ▶ U svakom čvoru stabla, osim listova, nalazi se po jedan test
- ▶ Svaki test ima više od jednog ishoda
- ▶ Svakom ishodu odgovara jedna grana stabla koja vodi do sledećeg čvora

Stabla odlučivanja



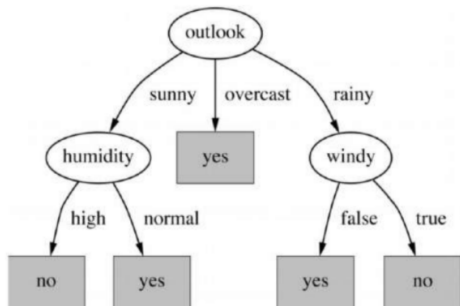
- ▶ U svakom čvoru stabla, osim listova, nalazi se po jedan test
- ▶ Svaki test ima više od jednog ishoda
- ▶ Svakom ishodu odgovara jedna grana stabla koja vodi do sledećeg čvora
- ▶ Listovi su označeni vrednostima koje predstavljaju predviđanja stabla (na primeru, da li je vreme pogodno za igranje golfa)

Stabla odlučivanja



- Na instancu za koju je potrebno dati predviđanje prvo se primenjuje test u korenu stabla

Stabla odlučivanja



- ▶ Na instancu za koju je potrebno dati predviđanje prvo se primenjuje test u korenu stabla
- ▶ U zavisnosti od ishoda, instanca se prosleđuje duž grane koja odgovara tom ishodu ka sledećem čvoru i postupak se rekurzivno nastavlja dok instanca ne stigne do lista, čija vrednost predstavlja traženo predviđanje

Stabla odlučivanja

- ▶ Postoji ogroman broj varijanti stabala odlučivanja

Stabla odlučivanja

- ▶ Postoji ogroman broj varijanti stabala odlučivanja
- ▶ Testovi koji se vrše u svakom od čvorova su tipično provjere vrednosti *pojedinačnih* atributa:

Stabla odlučivanja

- ▶ Postoji ogroman broj varijanti stabala odlučivanja
- ▶ Testovi koji se vrše u svakom od čvorova su tipično provere vrednosti *pojedinačnih* atributa:
 - ▶ *kategorički atributi*: provera jednakosti sa nekom konkretnom vrednošću atributa, provera pripadnosti nekom skupu vrednosti, itd.

Stabla odlučivanja

- ▶ Postoji ogroman broj varijanti stabala odlučivanja
- ▶ Testovi koji se vrše u svakom od čvorova su tipično provere vrednosti *pojedinačnih* atributa:
 - ▶ *kategorički atributi*: provera jednakosti sa nekom konkretnom vrednošću atributa, provera pripadnosti nekom skupu vrednosti, itd.
 - ▶ *neprekidni atributi*: provera da li je vrednost atributa veća ili manja od neke vrednosti

Stabla odlučivanja

- ▶ Postoji ogroman broj varijanti stabala odlučivanja
- ▶ Testovi koji se vrše u svakom od čvorova su tipično provere vrednosti *pojedinačnih* atributa:
 - ▶ *kategorički atributi*: provera jednakosti sa nekom konkretnom vrednošću atributa, provera pripadnosti nekom skupu vrednosti, itd.
 - ▶ *neprekidni atributi*: provera da li je vrednost atributa veća ili manja od neke vrednosti
- ▶ Neke varijante stabala odlučivanja sadrže testove *nad više promenljivih odjednom*, na primer ispitivanje vrednosti neke njihove linearne kombinacije.

Stabla odlučivanja

- ▶ Važna razlika u odnosu na većinu prethodno pomenutih modela je *diskretna struktura stabala odlučivanja*, koja čini gradijentne metode optimizacije neprimenljivim

Stabla odlučivanja

- ▶ Važna razlika u odnosu na većinu prethodno pomenutih modela je *diskretna struktura stabala odlučivanja*, koja čini gradijentne metode optimizacije neprimenljivim
- ▶ Alternativu predstavljaju kombinatorne metode optimizacije

Stabla odlučivanja

- ▶ Kombinatorne metode optimizacije su tipično previše zahtevne za praktičnu upotrebu, pa se pribegava pohlepnim pristupima optimizacije, koji u skladu sa nekom heuristikom rade sledeće:

Stabla odlučivanja

- ▶ Kombinatorne metode optimizacije su tipično previše zahtevne za praktičnu upotrebu, pa se pribegava pohlepnim pristupima optimizacije, koji u skladu sa nekom heuristikom rade sledeće:
 1. formulišu najbolji test u korenu stabla

Stabla odlučivanja

- ▶ Kombinatorne metode optimizacije su tipično previše zahtevne za praktičnu upotrebu, pa se pribegava pohlepnim pristupima optimizacije, koji u skladu sa nekom heuristikom rade sledeće:
 1. formulišu najbolji test u korenu stabla
 2. dele instance skupa za obučavanje na više grupa kojima odgovaraju isti ishodi tog testa

Stabla odlučivanja

- ▶ Kombinatorne metode optimizacije su tipično previše zahtevne za praktičnu upotrebu, pa se pribegava pohlepnim pristupima optimizacije, koji u skladu sa nekom heuristikom rade sledeće:
 1. formulišu najbolji test u korenu stabla
 2. dele instance skupa za obučavanje na više grupa kojima odgovaraju isti ishodi tog testa
 3. rekurzivno konstruišu stabla nad tim podskupovima

Stabla odlučivanja

- ▶ Kombinatorne metode optimizacije su tipično previše zahtevne za praktičnu upotrebu, pa se pribegava pohlepnim pristupima optimizacije, koji u skladu sa nekom heuristikom rade sledeće:
 1. formulišu najbolji test u korenu stabla
 2. dele instance skupa za obučavanje na više grupa kojima odgovaraju isti ishodi tog testa
 3. rekurzivno konstruišu stabla nad tim podskupovima
 4. povezuju tako dobijena stabla na koren ukupnog stabla

Stabla odlučivanja

- ▶ Kako bi se ovakav postupak implementirao, potrebno je precizirati:

Stabla odlučivanja

- ▶ Kako bi se ovakav postupak implementirao, potrebno je precizirati:
 - ▶ način na koji se formuliše test u nekom čvoru

Stabla odlučivanja

- ▶ Kako bi se ovakav postupak implementirao, potrebno je precizirati:
 - ▶ način na koji se formuliše test u nekom čvoru
 - ▶ kriterijum zaustavljanja

Stabla odlučivanja

- ▶ Kako bi se ovakav postupak implementirao, potrebno je precizirati:
 - ▶ način na koji se formuliše test u nekom čvoru
 - ▶ kriterijum zaustavljanja
 - ▶ kako se određuju vrednosti u listovima

Testovi

- ▶ Pretpostavimo da su svi atributi kategorički, da *test* predstavlja ispitivanje vrednosti nekog atributa i da svakoj vrednosti atributa odgovara po jedna grana

Testovi

- ▶ Pretpostavimo da su svi atributi kategorički, da *test* predstavlja ispitivanje vrednosti nekog atributa i da svakoj vrednosti atributa odgovara po jedna grana
- ▶ Kako najbolje izabrati atribut koji će biti testiran u korenu stabla za dati skup instanci?

Testovi

- ▶ Pretpostavimo da su svi atributi kategorički, da *test* predstavlja ispitivanje vrednosti nekog atributa i da svakoj vrednosti atributa odgovara po jedna grana
- ▶ Kako najbolje izabrati atribut koji će biti testiran u korenu stabla za dati skup instanci?
- ▶ Idealno, izabrani atribut bi razvrstao instance tako da su klase instanci u svakoj podgrupi iste

Testovi

- ▶ Pretpostavimo da su svi atributi kategorički, da *test* predstavlja ispitivanje vrednosti nekog atributa i da svakoj vrednosti atributa odgovara po jedna grana
- ▶ Kako najbolje izabrati atribut koji će biti testiran u korenu stabla za dati skup instanci?
- ▶ Idealno, izabrani atribut bi razvrstao instance tako da su klase instanci u svakoj podgrupi iste
- ▶ Ako to nije moguće, onda je ideja izabrati atribut koji *u nekom smislu* najbolje razvrstava instance, tako da dobijene podgrupe budu što homogenije u odnosu na klase

Testovi

- ▶ Pretpostavimo da su svi atributi kategorički, da *test* predstavlja ispitivanje vrednosti nekog atributa i da svakoj vrednosti atributa odgovara po jedna grana
- ▶ Kako najbolje izabrati atribut koji će biti testiran u korenu stabla za dati skup instanci?
- ▶ Idealno, izabrani atribut bi razvrstao instance tako da su klase instanci u svakoj podgrupi iste
- ▶ Ako to nije moguće, onda je ideja izabrati atribut koji *u nekom smislu* najbolje razvrstava instance, tako da dobijene podgrupe budu što homogenije u odnosu na klase
- ▶ Potrebno je definisati neku meru kvaliteta atributa u odnosu na razvrstavanje koje nudi, odnosno meru čistoće (eng. *purity*) ili homogenosti dobijenih podgrupa

Testovi

- ▶ Pretpostavimo da se u skupu od 10 instanci javljaju 4 instance klase A , tri instance klase B i tri instance klase C

Testovi

- ▶ Pretpostavimo da se u skupu od 10 instanci javljaju 4 instance klase A , tri instance klase B i tri instance klase C
- ▶ Udeli instanci svake klase u ukupnom skupu su $p_A = 0.4$, $p_B = 0.3$ i $p_C = 0.3$.

Testovi

- ▶ Pretpostavimo da se u skupu od 10 instanci javljaju 4 instance klase A , tri instance klase B i tri instance klase C
- ▶ Udeli instanci svake klase u ukupnom skupu su $p_A = 0.4$, $p_B = 0.3$ i $p_C = 0.3$.
- ▶ U opštem slučaju, za C klasa, *entropija* i *Ginijev indeks* se definišu kao:

$$H(p_1, \dots, p_C) = - \sum_{i=1}^C p_i \log p_i \qquad G(p_1, \dots, p_C) = 1 - \sum_{i=1}^C p_i^2$$

Testovi

- ▶ Pretpostavimo da se u skupu od 10 instanci javljaju 4 instance klase A , tri instance klase B i tri instance klase C
- ▶ Udeli instanci svake klase u ukupnom skupu su $p_A = 0.4$, $p_B = 0.3$ i $p_C = 0.3$.
- ▶ U opštem slučaju, za C klasa, *entropija* i *Ginijev indeks* se definišu kao:

$$H(p_1, \dots, p_C) = - \sum_{i=1}^C p_i \log p_i \qquad G(p_1, \dots, p_C) = 1 - \sum_{i=1}^C p_i^2$$

- ▶ Kada oba izraza dostižu maksimalnu vrednost?

Testovi

- ▶ Pretpostavimo da se u skupu od 10 instanci javljaju 4 instance klase A , tri instance klase B i tri instance klase C
- ▶ Udeli instanci svake klase u ukupnom skupu su $p_A = 0.4$, $p_B = 0.3$ i $p_C = 0.3$.
- ▶ U opštem slučaju, za C klasa, *entropija* i *Ginijev indeks* se definišu kao:

$$H(p_1, \dots, p_C) = - \sum_{i=1}^C p_i \log p_i \qquad G(p_1, \dots, p_C) = 1 - \sum_{i=1}^C p_i^2$$

- ▶ Kada oba izraza dostižu maksimalnu vrednost?
 - ▶ ako za svako i važi $p_i = 1/C$

Testovi

- ▶ Pretpostavimo da se u skupu od 10 instanci javljaju 4 instance klase A , tri instance klase B i tri instance klase C
- ▶ Udeli instanci svake klase u ukupnom skupu su $p_A = 0.4$, $p_B = 0.3$ i $p_C = 0.3$.
- ▶ U opštem slučaju, za C klasa, *entropija* i *Ginijev indeks* se definišu kao:

$$H(p_1, \dots, p_C) = - \sum_{i=1}^C p_i \log p_i \qquad G(p_1, \dots, p_C) = 1 - \sum_{i=1}^C p_i^2$$

- ▶ Kada oba izraza dostižu maksimalnu vrednost?
 - ▶ ako za svako i važi $p_i = 1/C$
- ▶ Kada oba izraza dostižu minimalnu vrednost?

Testovi

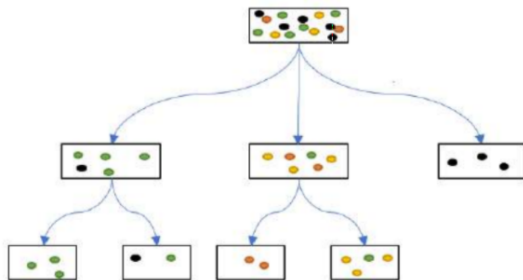
- ▶ Pretpostavimo da se u skupu od 10 instanci javljaju 4 instance klase A , tri instance klase B i tri instance klase C
- ▶ Udeli instanci svake klase u ukupnom skupu su $p_A = 0.4$, $p_B = 0.3$ i $p_C = 0.3$.
- ▶ U opštem slučaju, za C klasa, *entropija* i *Ginijev indeks* se definišu kao:

$$H(p_1, \dots, p_C) = - \sum_{i=1}^C p_i \log p_i \qquad G(p_1, \dots, p_C) = 1 - \sum_{i=1}^C p_i^2$$

- ▶ Kada oba izraza dostižu maksimalnu vrednost?
 - ▶ ako za svako i važi $p_i = 1/C$
- ▶ Kada oba izraza dostižu minimalnu vrednost?
 - ▶ ako je $p_i = 1$ za neko i

Testovi

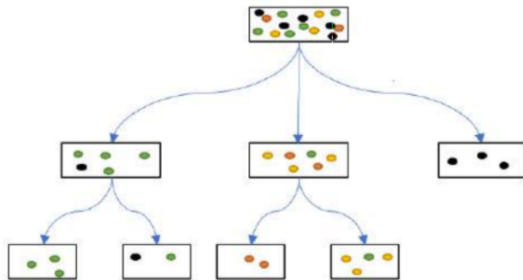
- ▶ Na osnovu neke od ovih ili neke slične mere, atribut se obično evaluira tako što se izračuna smanjenje nehomogenosti nakon particionisanja skupa instanci po vrednostima tog atributa



Testovi

- ▶ Na osnovu neke od ovih ili neke slične mere, atribut se obično evaluira tako što se izračuna smanjenje nehomogenosti nakon partitionisanja skupa instanci po vrednostima tog atributa
- ▶ Ako je $\mathcal{D}_i$ skup svih instanci za koje je vrednost atributa A jednaka i , smanjenje nehomogenosti se računa sledećim izrazom (u slučaju korišćenja entropije):

$$H(\mathcal{D}) - \sum_{i=1}^C \frac{|\mathcal{D}_i|}{|\mathcal{D}|} H(\mathcal{D}_i)$$



Kriterijum zaustavljanja

- ▶ Pretpostavimo da su svi atributi kategorički

Kriterijum zaustavljanja

- ▶ Pretpostavimo da su svi atributi kategorički
- ▶ Postoje razni kriterijumi zaustavljanja

Kriterijum zaustavljanja

- ▶ Pretpostavimo da su svi atributi kategorički
- ▶ Postoje razni kriterijumi zaustavljanja
- ▶ Kada dubina stabla dostigne broj atributa

Kriterijum zaustavljanja

- ▶ Pretpostavimo da su svi atributi kategorički
- ▶ Postoje razni kriterijumi zaustavljanja
- ▶ Kada dubina stabla dostigne broj atributa
- ▶ Kada sve instance u čvoru imaju identične vrednosti atributa ili identične vrednosti ciljne promenljive

Kriterijum zaustavljanja

- ▶ Pretpostavimo da su svi atributi kategorički
- ▶ Postoje razni kriterijumi zaustavljanja
- ▶ Kada dubina stabla dostigne broj atributa
- ▶ Kada sve instance u čvoru imaju identične vrednosti atributa ili identične vrednosti ciljne promenljive
- ▶ Kada je dostignuta unapred određena dubina stabla

Kriterijum zaustavljanja

- ▶ Pretpostavimo da su svi atributi kategorički
- ▶ Postoje razni kriterijumi zaustavljanja
- ▶ Kada dubina stabla dostigne broj atributa
- ▶ Kada sve instance u čvoru imaju identične vrednosti atributa ili identične vrednosti ciljne promenljive
- ▶ Kada je dostignuta unapred određena dubina stabla
- ▶ Kada mera nehomogenosti podskupa instanci koje je pridružen tekućem čvoru dostigne neku nisku, unapred definisanu vrednost

Kriterijum zaustavljanja

- ▶ Pretpostavimo da su svi atributi kategorički
- ▶ Postoje razni kriterijumi zaustavljanja
- ▶ Kada dubina stabla dostigne broj atributa
- ▶ Kada sve instance u čvoru imaju identične vrednosti atributa ili identične vrednosti ciljne promenljive
- ▶ Kada je dostignuta unapred određena dubina stabla
- ▶ Kada mera nehomogenosti podskupa instanci koje je pridružen tekućem čvoru dostigne neku nisku, unapred definisanu vrednost
- ▶ ...

Određivanje vrednosti u listovima

- ▶ Kod regresije: prosek vrednosti ciljnih promenljivih za instance pridružene listu

Određivanje vrednosti u listovima

- ▶ Kod regresije: prosek vrednosti ciljnih promenljivih za instance pridružene listu
- ▶ Kod klasifikacije: većinska klasa među instancama koje su pridružene listu

- ▶ Stabla odlučivanja se tipično ne smatraju modelima visoke preciznosti

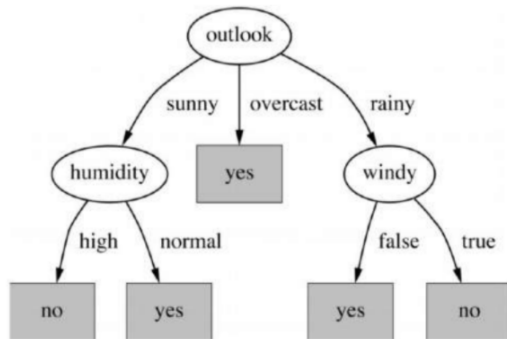
- ▶ Stabla odlučivanja se tipično ne smatraju modelima visoke preciznosti
- ▶ Duboka stabla su sklona preprilagođavanju

- ▶ Stabla odlučivanja se tipično ne smatraju modelima visoke preciznosti
- ▶ Duboka stabla su sklona preprilagođavanju
- ▶ Plitka stabla su sklona potprilagođavanju

- ▶ Stabla odlučivanja se tipično ne smatraju modelima visoke preciznosti
- ▶ Duboka stabla su sklona preprilagođavanju
- ▶ Plitka stabla su sklona potprilagođavanju
- ▶ Dubina stabla predstavlja regularizacioni metaparametar, ali uprkos mogućnosti regularizacije, u praksi ispoljavaju visoku varijansu

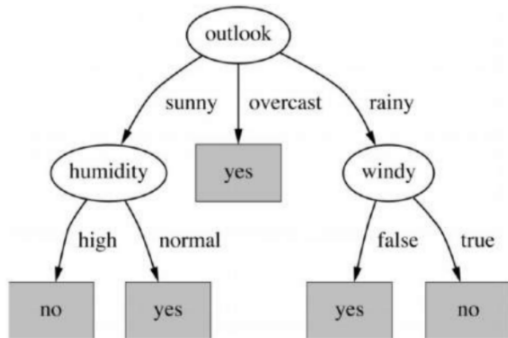
Prednosti

► Interpretabilnost



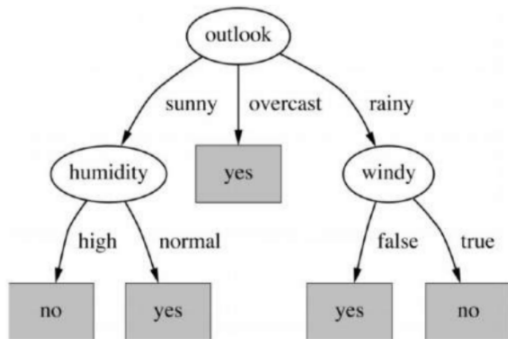
Prednosti

- Interpretabilnost
- Ukoliko je stablo malo, testovi koje vrši jasno ukazuju na to na osnovu čega stablo donosi odluke



Prednosti

- ▶ Interpretabilnost
- ▶ Ukoliko je stablo malo, testovi koje vrši jasno ukazuju na to na osnovu čega stablo donosi odluke
- ▶ Svaka putanja kroz stablo odlučivanja od korena do lista može se videti kao jedno `if ... then` pravilo, u kojem uslov predstavlja konjunkciju svih ishoda testova duž putanje, dok je odluka vrednost u listu.



Prednosti

- ▶ Prirodno kombinuju attribute različitih vrsta (kategoričke i neprekidne)

Prednosti

- ▶ Prirodno kombinuju attribute različitih vrsta (kategoričke i neprekidne)
- ▶ Neosetljiva su na monotone transformacije atributa

Prednosti

- ▶ Prirodno kombinuju attribute različitih vrsta (kategoričke i neprekidne)
- ▶ Neosetljiva su na monotone transformacije atributa
- ▶ Vrlo su efikasna u vreme predviđanja

Pregled

Prosta agregacija

Stabla odlučivanja

Slučajne šume

Pojačavanje

Slučajne šume

- ▶ Zasniva se na prostoj agregaciji stabala odlučivanja

Slučajne šume

- ▶ Zasniva se na prostoj agregaciji stabala odlučivanja
- ▶ Ansambl se sastoji od m stabala treniranih na različitim podskupovima skupa za obučavanje

Slučajne šume

- ▶ Zasniva se na prosto agregaciji stabala odlučivanja
- ▶ Ansambl se sastoji od m stabala treniranih na različitim podskupovima skupa za obučavanje
- ▶ Jedno stablo se obučava tako što se izabere podskup skupa za obučavanje određene veličine, pri čemu je moguće koristiti i samo podskup ukupnog skupa atributa

Slučajne šume

- ▶ Zasniva se na prosto agregaciji stabala odlučivanja
- ▶ Ansambl se sastoji od m stabala treniranih na različitim podskupovima skupa za obučavanje
- ▶ Jedno stablo se obučava tako što se izabere podskup skupa za obučavanje određene veličine, pri čemu je moguće koristiti i samo podskup ukupnog skupa atributa
- ▶ Stabla se obučavaju na različitim podskupovima kako bi njihove greške bile što slabije korelisane, što ostavlja prostor za popravku agregacijom

Slučajne šume

- ▶ Zasniva se na prosto agregaciji stabala odlučivanja
- ▶ Ansambl se sastoji od m stabala treniranih na različitim podskupovima skupa za obučavanje
- ▶ Jedno stablo se obučava tako što se izabere podskup skupa za obučavanje određene veličine, pri čemu je moguće koristiti i samo podskup ukupnog skupa atributa
- ▶ Stabla se obučavaju na različitim podskupovima kako bi njihove greške bile što slabije korelisane, što ostavlja prostor za popravku agregacijom
- ▶ Metaparametri su očito broj stabala m i veličine podskupova instanci i atributa

Slučajne šume

- ▶ Broj stabala m se može posmatrati i kao regularizacioni parametar

Slučajne šume

- ▶ Broj stabala m se može posmatrati i kao regularizacioni parametar
- ▶ Visoke vrednosti su tipično bolje

Slučajne šume

- ▶ Broj stabala m se može posmatrati i kao regularizacioni parametar
- ▶ Visoke vrednosti su tipično bolje
- ▶ Korišćenje većeg broja stabala tipično smanjuje preprilagođavanje i daje bolje rezultate

Slučajne šume

- ▶ Stabla odlučivanja su jedan od najprimenjenijih algoritama mašinskog učenja

Slučajne šume

- ▶ Stabla odlučivanja su jedan od najprimenjenijih algoritama mašinskog učenja
- ▶ Njihovo obučavanje je relativno efikasno, a preciznost predviđanja obično među najboljim za vektorski predstavljene podatke

Slučajne šume

- ▶ Stabla odlučivanja su jedan od najprimenjenijih algoritama mašinskog učenja
- ▶ Njihovo obučavanje je relativno efikasno, a preciznost predviđanja obično među najboljim za vektorski predstavljene podatke
- ▶ Slučajna šuma, kao i drugi ansambli nije interpretabilna.

Pregled

Prosta agregacija

Stabla odlučivanja

Slučajne šume

Pojačavanje

Pojačavanje

- ▶ Prilikom proste agregacije, modeli se konstruišu potpuno nezavisno

Pojačavanje

- ▶ Prilikom proste agregacije, modeli se konstruišu potpuno nezavisno
- ▶ Drugačiji pristup: modeli nekako uzimaju u obzir ponašanje drugih modela

Pojačavanje

- ▶ Prilikom proste agregacije, modeli se konstruišu potpuno nezavisno
- ▶ Drugačiji pristup: modeli nekako uzimaju u obzir ponašanje drugih modela
- ▶ Osnovna ideja *pojačavanja* (eng. *boosting*) je da se modeli prave inkrementalno, jedan po jedan, i dodavati u ansambl

Pojačavanje

- ▶ Prilikom proste agregacije, modeli se konstruišu potpuno nezavisno
- ▶ Drugačiji pristup: modeli nekako uzimaju u obzir ponašanje drugih modela
- ▶ Osnovna ideja *pojačavanja* (eng. *boosting*) je da se modeli prave inkrementalno, jedan po jedan, i dodavati u ansambl
- ▶ Na koji način se vrši izgradnja sledećeg modela kada već imamo skup modela?

Pojačavanje

- ▶ Prilikom proste agregacije, modeli se konstruišu potpuno nezavisno
- ▶ Drugačiji pristup: modeli nekako uzimaju u obzir ponašanje drugih modela
- ▶ Osnovna ideja *pojačavanja* (eng. *boosting*) je da se modeli prave inkrementalno, jedan po jedan, i dodavati u ansambl
- ▶ Na koji način se vrši izgradnja sledećeg modela kada već imamo skup modela?
- ▶ Vrš se tako da novi model najbolje kompenzuje greške modela koji već jesu u ansamblu

Pojačavanje

- ▶ Prilikom proste agregacije, modeli se konstruišu potpuno nezavisno
- ▶ Drugačiji pristup: modeli nekako uzimaju u obzir ponašanje drugih modela
- ▶ Osnovna ideja *pojačavanja* (eng. *boosting*) je da se modeli prave inkrementalno, jedan po jedan, i dodavati u ansambl
- ▶ Na koji način se vrši izgradnja sledećeg modela kada već imamo skup modela?
- ▶ Vršiti se tako da novi model najbolje kompenzuje greške modela koji već jesu u ansamblu
- ▶ Za razliku od proste agregacije gde se modeli prave nasumice, ovde se prave ciljano, u odnosu na učinak već postojećih modela