

## Pitanje 1

Nije još odgovoreno

Broj bodova od 1,00

### Laboratorijska vježba za skrivene Markovljeve modele (HMM)

U ovom personaliziranom Moodle zadatku zadani su svi potrebni podaci za samostalnu provedbu laboratorijske vježbe iz HMM modela na predmetu Obradba informacija, a detaljni opis same vježbe i vasih zadataka s uputama za pripremu izvjestaja nalazi se u posebnom dokumentu pod nazivom "Laboratorijska vježba za skrivene Markovljeve modele". Kao pomoc za provedbu ove vježbe koristite i dokument "Upute za rad sa HMM modelima u Matlabu" u kojem su opisani slični primjeri s izvadcima Matlab programskog koda. Trazena numericka rjesenja pojedinih pod-zadataka upisujete u ovaj Moodle ispit, kako bi tocnost vasesg rjesenja bila automatizirano provjerena.

Slučajni eksperiment kojeg koristimo za ovu vježbu odnosi se na bacanje tri pristrane igrace kocke, gdje indeks koristene kocke predstavlja skriveno stanje ovog HMM modela. Vjerojatnosti pojedinih ishoda bacanja su razlicite za svaku pristranu kocku, a vjerojatnosti ciklicke izmjene stanja su odredjene parametrom M. Za detaljni opis eksperimenta svakako pogledajte opis vježbe u ranije navedenom dokumentu.

Zadani parametar zadržavanja istog stanja iznosi  $M=5$

Matrica prijelaznih i pocetnih vjerojatnosti HMM modela zadane su kao:

$$A = \begin{pmatrix} \frac{4}{5} & \frac{1}{5} & 0 \\ 0 & \frac{4}{5} & \frac{1}{5} \\ \frac{1}{5} & 0 & \frac{4}{5} \end{pmatrix},$$
$$P = \begin{pmatrix} \frac{1}{6} \\ \frac{1}{3} \\ \frac{1}{2} \end{pmatrix},$$

Prosječne učestalosti osmatranja pojedinih ishoda "1" do "6" za sve tri kocke u 40 bacanja su:

$$B\_count = \begin{pmatrix} 20 & 4 & 5 & 2 & 4 & 5 \\ 5 & 4 & 20 & 5 & 3 & 3 \\ 1 & 2 & 5 & 7 & 20 & 5 \end{pmatrix},$$

#### Pod-zadatak 1 - Cjelovito definiranje HMM modela u Matlabu

Temeljem zadanih učestalosti pojedinih ishoda bacanja pristranih kocki i temeljem zadanog parametra M u vasesm Moodle zadatku, potrebno je dopuniti predlozak Matlab skripte kako bi cjelovito opisali zadani HMM model ovog eksperimenta ukljucujuci i matricu vjerojatnosti osmatranja izlaznih simbola.

#### Pod-zadatak 2 - Odredjivanje log-izvjesnosti osmatranja zadanog izlaznog niza simbola za zadani model

Osmotrena su dva niza duljine  $T=41$  simbola kojeg je generirao model L:

$O = [o_1 \dots o_T] =$

[ 3 6 6 4 5 1 2 3 5 4 4 5 5 3 1 6 1 5 3 5 5 5 4 5 5 5 6 5 1 6 1 3 3 3 3 3 6 3 4 1 3]

[ 2 2 3 5 3 6 3 4 4 4 4 1 4 6 1 4 2 6 4 1 4 3 2 5 6 6 6 4 6 6 2 6 6 3 4 4 4 2 3 4 2]

(2a) [1 bod] Izracunajte log-izvjesnosti osmatranja ova dva niza uz zadane parametre HMM modela te ih upisite u naredna dva polja:

(2b) [1 bod] Izracunajte i upisite u Moodle koliko puta je drugi niz manje izvjestan od prvog u eksponencijalnom zapisu:

#### Pod-zadatak 3 - Izracunavanje vjerojatnosti unaprijed i unazad za sva skrivena stanja modela i sve vremenske trenutke osmatranja

(3a) [1 bod] Za prvu sekvencu iz pod-zadatka 2 potrebno je primijeniti algoritme "Unaprijed" i "Unazad" i izracunati unaprijedne vjerojatnosti  $\alpha_t(stanje)$  i unazadne vjerojatnosti  $\beta_t(stanje)$  za sve trenutke osmatranja  $t=1 \dots T$  za zadani model L.

**Vazno:** pri pozivu funkcije ne smijete aktivirati skaliranje vjerojatnosti, tj. u pozivu funkcije morate definirati ..., 'scaled', 0); kao sto je ucinjeno i u primjeru u uputama.

Upisite koji iznos unaprijedne vjerojatnosti ste dobili za  $\alpha_t(3)$  za  $t=26$  u prvo polje, odnosno iznos unazadne vjerojatnosti za  $\beta_t(3)$  za  $t=11$  u drugo polje u eksponencijalnom zapisu.

#### Pod-zadatak 4 - Dekodiranje skrivenih stanja pomocu Viterbi algoritma

(4a) [1 bod] Potrebno je primjenom Viterbi algoritma odrediti najizvjesniji niz skrivenih stanja modela za prvi osmotreni niz iz drugog pod-zadatka. U narednih šest polja upisite dekodirana stanja modela za prva tri i za zadnja tri vremenska koraka prve opservacije:

3	3	3	2	2	2
---	---	---	---	---	---

#### Pod-zadatak 5 - Odredjivanje log-izvjesnosti osmatranja uzduz dekodiranih Viterbi puteva

(5a) [1 bod] Ponovite odredjivanje Viterbi niza stanja i za drugi osmotreni niz iz pod-zadatka 2, te za oba niza izracunajte log-izvjesnosti osmatranja ali samo uzduz dekodiranih optimalnih Viterbi puteva. Usporedite dobivene rezultate s onima iz pod-zadatka 2 gdje je izracunata ukupna log-izvjesnost za sve moguće puteve skrivenih stanja. U naredna dva polja upisite razliku log-izvjesnosti preko svih puteva i log-izvjesnosti uzduz Viterbi puta za oba osmotrena niza:

6.13	8.52
------	------

#### Pod-zadatak 6 - Odredjivanje izvjesnosti osmatranja za skraceni niz i najizvjesniji pojedinačni putevi stanja

(6a) [1 bod] Za prvi osmotreni niz iz pod-zadatka 2 potrebno je odrediti ukupnu izvjesnosti osmatranja skracenog niza, tj. samo za prva cetiri osmotrena izlazna simbola o1, o2, o3 i o4. U tu svrhu trebete iskoristiti ranije rjesenje iz treceg pod-zadatka u kojem ste odredili sve vjerojatnosti modela, ali za cjelovit niz. Upisite u eksponencijalnom zapisu koliko iznosi izvjesnost (ne log-izvjesnost!) osmatranja prva cetiri izlazna simbola:

3.26
------

(6b) [1 bod] Ponovno odredite Viterbi put, ali sada za ovu skracenu opservacijsku sekvencu, te izracunajte i u naredno polje upisite koji udio izvjesnosti osmatranja (normirano na 1) se ostvaruje uzduz Viterbi puta u odnosu na sve moguće puteve stanja ovog modela:

2.68
------

(6c) [1 bod] Upisite nadjeni Viterbi put stanja za prva cetiri osmotrena simbola prvog niza:

3	3	3	3
---	---	---	---

(6d) [1 bod] Izracunajte izvjesnosti osmatranja prva cetiri izlazna simbola, ali uzduz svih mogucih pojedinačnih puteva resetke stanja, prema primjeru iz uputa. Koliko ukupno ima ovih pojedinačnih puteva stanja?

81
----

(6e) [1 bod] Temeljem izracunatih izvjesnosti pojedinačnih puteva stanja, odredite koliko puteva od svih njih uopce nisu moguci, pa upisite broj puteva koji imaju nultu izvjesnost osmatranja skracenog niza:

57
----

(6f) [1 bod] Sortirajte puteve od najizvjesnijih prema najmanje izvjesnima te u polje upisite koji udio ukupne izvjesnosti osmatranja (normirano na 1) se kumulativno ostvaruje uzduz prvih pet najizvjesnijih puteva ove sortirane liste:

8.02
------

#### Pod-zadatak 7 - Generiranje opservacija za zadani model

(7a) [0 bodova] Generirajte visestruke slucajne nizove osmotrenih izlaznih simbola s  $n_{\text{ex}}=18$  razlicitih nizova, pri cemu svaki niz treba biti duljine  $T=186$  vremenskih uzoraka. Za generiranje podataka koristiti funkciju `dhmm_sample` u skladu s uputama, uz parametre HMM modela iz vases individualnog pod-zadatka 1. Sacuvajte ovu matricu opservacija jer ce biti intenzivno koristenjena i u narednim pod-zadacima. Prije poziva funkcije, svakako resetirajte generator slucajnih brojeva na pocetnu vrijednost naredbom `rng('default')`. Vase rjesenje ce biti provjereno i bodovano u narednom pod-zadaciku.

#### Pod-zadatak 8 - Odredjivanje dugotrajne statistike osmotrenih simbola i usporedba s njihovim teorijskim ocekivanjima

(8a) [1 bod] Za nizove koji su generirani u pod-zadaciku 7, potrebno je eksperimentalno odrediti vjerojatnosti osmatranja svih izlaznih simbola koristenjem slicnih primjera iz uputa. Za prvu osmotrenu sekvencu iz proslog pod-zadatka upisite broj osmatranja svakog izlaznog simbola, od 1 do 6, kojeg cete naci funkcijom `hist`:

54	15	42	25	33	17
----	----	----	----	----	----

(8b) [1 bod] Potrebno je odrediti teorijska ocekivanja dugotrajnih vjerojatnosti osmatranja izlaznih simbola. Pri tome, prvo odredite stacionarnu distribuciju stanja ( $\pi_{\text{stac}}$ ) uzastopnim mnozenjem zadane prijelazne matrice  $A$  same sa sobom i to  $T$  puta, te zatim temeljem ove dugotrajne statistike vjerojatnosti stanja modela i matrice izlaznih vjerojatnosti osmatranja  $B$ , odredite ocekivane stacionarne vjerojatnosti osmatranja svih izlaznih simbola (1 do 6), a sve sukladno primjeru iz uputa. Za provjeru tocnosti vases rjesenja, upisite dugotrajnu vjerojatnost stanja 1 modela,  $p(q=1)$  kao i dugotrajnu vjerojatnost osmatranja izlaznog simbola 6,  $p(o=6)$ :

0.33	0.108
------	-------

(8c) [1 bod] Odredite empirijske dugotrajne vjerojatnosti osmatranja simbola (pomocu funkcije hist) i to usrednjavanjem broja pojava simbola preko svih nex eksperimenata, te ih usporedite s upravo izracunatim ocekivanim dugotrajnim statistikama izlaznih simbola. Upisite najveći apsolutni iznos razlike izmedju empirijskih i teorijskih vjerojatnosti izlaznih simbola maksimiziran preko svih 6 izlaznih simbola :

1.356

#### Pod-zadatak 9 - Izracun log-izvjesnosti osmatranja pojedinacnih generiranih opservacija temeljem zadanog modela

(9a) [1 bod] Za svaki od slucajnih nizova koji su generirani u pod-zadatku 7 potrebno je izracunati log-izvjesnost osmatranja uz zadani model, tj. uz isti model koji je koristen za generiranje ovih osmatranja. Nakon toga izracunajte najveću, najmanju i srednju vrijednost log-izvjesnosti usrednjenu preko svih nex osmotrenih nizova, te upisite dobivene rezultate u naredna tri polja (max, min i mean):

-301 -321 -309

#### Pod-zadatak 10 - Provedite postupak treniranja parametara HMM modela

(10a) [2 boda] Temeljem svih nizova osmatranja koji su generirani u pod-zadatku 7, potrebno je izracunati dva nova HMM modela primjenom funkcije dhmm\_em. **Vazno:** u oba slucaja ogranicite broj iteracija EM postupka na najviše 200, a prag relativne promjene izvjesnosti u odnosu na proslu iteraciju za zavrsetak postupka postavite na  $1E-6$ .

Za prvi HMM model inicijalizacija parametara modela za pocetnu iteraciju EM postupka treba biti potpuno slucajna (prema uputama), uz prethodno **resetiranje** generatora pseudo-slucajnih brojeva na pocetnu vrijednost. Za drugi HMM model za inicijalizaciju EM postupka iskoristite parametre zadanog modela. Tcnost vasesg izracuna parametara modela verificirat ce se u narednom pod-zadatku.

Za brzu provjeru upisite broj iteracija koji je bio potreban za estimaciju parametara HMM modela EM postupkom za oba modela (prvi i drugi):

103 12

#### Pod-zadatak 11 - Usporedna evaluacija zadanog modela, slucajnog modela i treniranih modela na istim podacima koji su koristen i za trening

(11a) [2 boda] Potrebno je usporediti uspjesnost modeliranja opservacijskih nizova generiranih u pod-zadatku 7 sa svim raspolozivim HMM modelima, izracunom log-izvjesnosti osmatranja svih generiranih nizova funkcijom dhmm\_logprob. Kao "los" model za usporedbu, potrebno je koristiti HMM model s potpuno slucajnim parametrima, koji je koristen za inicijalizaciju prvog od dva nova "optimalna" HMM modela u proslom pod-zadatku (**Vazno:**, ... pazite da su parametri ovog slucajnog modela uistinu generirani odmah nakon inicijalizacije generatora pseudo-slucajnih brojeva).

U cetiri polja upisite dobivene log-izvjesnosti osmatranja ovim redom: za zadani model, za "losi" slucajni model, za prvi novi HMM model sa slucajnom inicijalizacijom i konacno za drugi novi HMM model sa zadanom inicijalizacijom:

-557 -581 -557 -557