

## Laboratorijska vježba za skrivene Markovljeve modele (HMM)

Zadani parametar zadržavanja istog stanja iznosi  $M=5$

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

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

Kod:

```
clear;
addpath(genpath('C:\Users\Luka\Desktop\FER\OBRADAINFO\treci lab\HMM\HMMall'));
prior0=[
    1 % Prva kocka
    2 % Druga kocka
    3 % Treća kocka
]/6;
% Broj stanja HMM modela
Q=size(prior0,1);
% -----
M=5;
transmat0=[
M-1 1 0 % P(1|1) P(2|1) P(3|1)
0 M-1 1 % P(1|2) P(2|2) P(3|2)
1 0 M-1 % P(1|3) P(2|3) P(3|3)
]/M;
bcount = [20 4 5 5 1 6
          6 20 2 3 3 5
          4 6 20 2 6 2]/40;
```

Spremljene varijable:

Workspace	
Name ^	Value
bcount	3x6 double
M	5
prior0	[0.1667;0.3333;0....
Q	3
transmat0	[0.8000,0.2000,0;...

Podijelili smo matricu vjerojatnosti osmatranja izlaznih simbola ukupnim brojem bacanja koji iznosi 40 da bismo dobili vjerojatnost izlaza s obzirom na navedeni uzorak.

## Pod-zadatak 2 - Određjivanje 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 5 4 3 3 1 3 5 3 6 1 4 5 5 2 1 4 5 4 5 6 5 2 3 2 5 5 2 1 1 1 1 3 5 4 4 1 1 1 1 2]

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

(2a) Izračunajte log-izvjesnosti osmatranja ova dva niza uz zadane parametre HMM modela te ih upisite u naredna dva polja:

```
sample1 = [ 3 5 4 3 3 1 3 5 3 6 1 4 5 5 2 1 4 5 4 5 6 5 2 3 2 5 5 2 1 1 1 1 3 5 4 4 1 1 1 1 2];  
sample2 = [ 5 3 1 4 4 2 6 5 6 2 5 3 4 4 2 3 5 2 3 2 3 5 6 2 6 4 4 4 2 4 4 4 4 6 6 6 4 4 2 5 2];  
ll1 = dhmm_logprob(sample1, prior0, transmat0, obsmat0);  
ll2 = dhmm_logprob(sample2, prior0, transmat0, obsmat0);
```

Workspace		
Name ^	Value	
ll1	-75.8129	
ll2	-82.8061	
M	5	
obsmat0	3x6 double	
prior0	[0.1667;0.3333;0....	
Q	3	
sample1	1x41 double	
sample2	1x41 double	
transmat0	[0.8000,0.2000,0;...	

**IZVJEŠTAJ:** Ime matrice bcount mijenjamo u obsmat. Sample1 i sample2, koji nam predstavljaju dva uzorka, ubacujemo s ostalim parametrima u funkciju, već implementiranu u HMMall „knjižnici alata“, dhmm\_logprob.

(2b) Izračunajte i upisite u Moodle koliko puta je drugi niz manje izvjestan od prvog u eksponencijalnom zapisu:

```
llDiff = exp(ll1)/exp((ll2)); %razlika u izvjesnosti u eksponencijalnom zapisu
```

llDiff	1.0892e+03
--------	------------

**IZVJEŠTAJ:** Drugi je niz 1089.2 puta manje izvjestan od prvog kad se računaju u eksponencijalnom zapisu.

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

(3a) Za prvu sekvencu iz pod-zadatka 2 potrebno je primijeniti algoritme "Unaprijed" i "Unazad" i izracunati unaprijedne vjerojatnosti  $\alpha_t(\text{stanje})$  i unazadne vjerojatnosti  $\beta_t(\text{stanje})$  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  $\alpha(1)$  za  $t=15$  u prvo polje, odnosno iznos unazadne vjerojatnosti za  $\beta(1)$  za  $t=23$  u drugo polje u eksponencijalnom zapisu.

```
[alpha, beta, gamma, ll] = ...  
  fwdback(prior0, transmat0, obslik0, 'scaled', 0);  
  
fwdvalue15 = (alpha(1,15));  
bwdvalue23 = (beta(1,23));
```

**IZVJEŠTAJ:** Brojevi koje dobivamo su poprilično mali zato što je mala vjerojatnost navedenih događaja.

fwdvalue15 =

6.1221e-14

bwdvalue23 =

3.3325e-14

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

(4a) 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:

```
vpath = viterbi_path(prior0, transmat0, obslik0); %najizvjesniji put  
  
firstThree = vpath(1:3);  
vpathR = flip(vpath);  
lastThree = flip(vpathR(1:3));
```

firstThree	[3,3,3]
fwdvalue15	6.1221e-14
gamma	3x41 double
lastThree	[1,1,2]

**IZVJEŠTAJ:** viterbi\_path metoda vraća cijeli niz vrijednosti koji predstavljaju najveću vjerojatnost odabira kocki da bi nam se niz ostvario

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

(5a) 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:

```
obslik2 = multinomial_prob(sample2, obsmat0); %izracunavanje vrijednosti osmatranja odabranog niza
vpath2 = viterbi_path(prior0, transmat0, obslik2); %najizvjesniji put za drugi niz

% vjerojatnost uzduz najizvjesnijeg puta, slicna funkcija kao za sve puteve
% samo jos dodajemo određen viterbi put kao parametar
[l11V, p1] = dhmm_logprob_path(prior0, transmat0, obslik0, vpath);
[l12V, p2] = dhmm_logprob_path(prior0, transmat0, obslik2, vpath2);

diff1 = l11 - l11V; %razlika log izvjesnosti za sve puteve i log izvjesnosti preko određenog viterbi puta
diff2 = l12 - l12V;
```

diff1	5.9074
diff2	6.2511

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

(6a) 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 trebate 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:

```
%izracunavanje skracenog niza, ali putem alphe tako da ćemo sumirat cijeli
%četrty stupac i tako saznat ukupnu izvjesnost
```

```
lR = sum(alpha(:,4)); % l - likelihood (izvjesnost) za R - reducirani niz
```

lR	1.0892e+03
IR	6.4081e-04

**IZVJEŠTAJ:** Izračunavanje smo mogli riješiti i na način da smo odrezali niz te ga ubacili u funkciju `dhmm_logprob`, ali na ovaj način nismo morali ništa „dirati“, a izvođenje funkcije za cijeli niz od 41 elemenata nam je dalo razne informacije u poljima s kojim možemo manipulirati te uzeti ono što nas zanima kao sad.

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

```
obslikR = multinomial_prob(sample1(1:4), obsmat0); %izracunavamo obslik polje za reducirani niz, potreban nam je za
%izracun viterbi puta

vpathR = viterbi_path(prior0, transmat0, obslikR);
[l1RV, p3] = dhmm_logprob_path(prior0, transmat0, obslikR, vpathR); %izracun izvjesnosti osmatranja uzduž Viterbi puta

likelihoodStake = exp(l1RV)/lR; %udio izvjesnosti
```

```
lastInree      [1,1,4]
likelihoodStake 0.7491
..
```

(6c) Upisite nadjeni Viterbi put stanja za prva četiri osmotrena simbola prvog niza:

```
firstFour = vpathR(1:4); %izracun prva četiri stanja Viterbi puta reduciranog niza
```

```
firstFour      0.2311
firstFour      [3,3,3,3]
..
```

(6d) Izračunajte izvjesnosti osmatranja prva četiri izlazna simbola, ali uzduž svih mogućih pojedinačnih puteva resetke stanja, prema primjeru iz uputa. Koliko ukupno ima ovih pojedinačnih puteva stanja?

```
numPaths      81
```

**IZVJEŠTAJ:** Ukupan broj puteva za određen model smo dobili alatom iz vjerojatnosti koji nam daje broj svih permutacija.

(6e) Temeljem izračunatih izvjesnosti pojedinačnih puteva stanja, odredite koliko puteva od svih njih uopće nisu mogući, pa upisite broj puteva koji imaju nultu izvjesnost osmatranja skraćenog niza:

```

%treba nam funkcija koja ce nam naci sve varijacije s ponavljanjem za polje
%"v"
llm=zeros(numPaths,1); % Stupac za log-izvjesnosti
mpath = [];
for i = 1:3 %dodavanje svih varijacija u polje mpath
    for j = 1:3
        for k = 1:3
            for l = 1:3
                mpath = [mpath; [i j k]];
            end
        end
    end
end

for i=1:numPaths %racunanje log-izvjesnosti za svaki put
    %te upisivanje vrijednosti u stupac
    [llm(i), p] = dhmm_logprob_path(prior0, transmat0, obslikR, mpath(i,:));
end

numOfNulls = 0;

for i=1:numPaths %racunanje ukupnog broja izvjesnosti koje su jednaki nuli
    if llm(i) == 0
        numOfNulls = numOfNulls + 1;
    end
end

```

## IZVJEŠTAJ:

### Neshvatljiv problem

```

>> skripta_treci_labos
Index exceeds the number of array elements. Index must not exceed 3.

Error in dhmm_logprob_path (line 12)
    p(t) = transmat(qs(t-1), qs(t)) * obsmat(qs(t),t);

Error in skripta_treci_labos (line 90)
    [llm(i), p] = dhmm_logprob_path(prior0, transmat0, obslikR, mpath(i,:));

```

(6f) 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:

```
% Sortiraj puteve prema izvjesnosti osmatranja od najizvjesnijeg do
% najmanje izvjesnog
[sllm,illm]=sort(-llm);

% Zbrajamo izvjesnosti
sumL = sum(exp(-sllm));
%kako bi mogli dijelit kumulativnu sumu da dobijemo udio
cumStake = cumsum(exp(-sllm)) / sumL;
```

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

(7a) Generirajte višestruke slučajne nizove osmotrenih izlaznih simbola s  $nex=14$  različitih nizova, pri čemu svaki niz treba biti duljine  $T=100$  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 koristena i u narednim pod-zadacima. Prije poziva funkcije, svakako resetirajte generator slučajnih brojeva na pocetnu vrijednost naredbom `rng('default')`. Vase rjesenje ce biti provjereno i bodovano u narednom pod-zadatku.

```
% generiraj višestruki opservacijski niz:
T = 100; % duljina svakog niza
nex = 14; % broj opservacijskih nizova
rng('default'); %resetiranje generatora slučajnih brojeva
data = dhmm_sample(prior0, transmat0, obsmat0, nex, T); %poziv funkcije
```

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	1
1	5	2	1	1	2	3	5	2	2	3	1	2	5	3	
2	2	4	6	2	3	1	6	1	5	2	1	2	2	2	
3	3	3	2	5	3	2	3	2	3	3	3	3	3	3	
4	3	2	3	2	1	1	2	5	3	1	1	5	2	5	
5	2	2	2	2	4	5	2	5	3	6	4	1	1	5	
6	3	4	3	3	3	4	2	5	2	4	2	2	4	2	
7	6	3	5	6	3	3	5	5	1	6	6	3	1	5	
8	3	3	5	2	4	4	4	1	2	2	6	1	3	2	
9	3	1	3	1	4	1	3	1	3	3	1	1	1	6	
10	5	1	3	2	1	3	1	2	2	2	6	2	4	6	
11	1	4	3	2	1	3	1	4	2	2	4	6	2	2	
12	3	2	1	4	3	2	1	1	1	1	3	1	1	4	
13	3	3	3	3	3	1	2	2	2	1	6	2	2	2	
14	2	2	3	3	3	3	5	5	2	2	3	3	3	2	

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

(8a) Za nizove koji su generirani u pod-zadatku 7, potrebno je eksperimentalno odrediti vjerojatnosti osmatranja svih izlaznih simbola korištenjem 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:

```
hm=hist(data(1,:), [1 2 3 4 5 6]);
```

```
gamma      34 22 16 7 15 6
hm          [34,22,16,7,15,6]
```

**IZVJEŠTAJ:** Uzeli smo samo prvu sekvencu te iskoristili metodu koja opisuje koliko se određenih izlaznih simbola pojavilo. Suma dobivenog polja ce biti 100.

(8b) Potrebno je odrediti teorijska ocekivanja dugotrajnih vjerojatnosti osmatranja izlaznih simbola. Pri tome, prvo odredite stacionarnu distribuciju stanja ( $\pi_{stac}$ ) uzastopnim množenjem 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 vasih rjesenja, upisite dugotrajnu vjerojatnost stanja 1 modela,  $p(q=1)$  kao i dugotrajnu vjerojatnost osmatranja izlaznog simbola 1,  $p(o=1)$  :

```
pi_stac=transmat0;
for i=1:T
    pi_stac=pi_stac*transmat0;
end
%dobili smo matricu koja iskazuje da je vjerojatnost svakog stanja 1/3

stac = pi_stac(1,:)*obsmat0*6; %množenjem stacionarne distribucije stanja sa
%vjerojatnostima obzervacija i veličinom uzorka koji promatramo dobivamo
%stacionirane vjerojatnosti izlaznih simbola u nekom uzorku opisanom našim modelom
```

stac						
1x6 double						
	1	2	3	4	5	6
1	1.5000	1.5000	1.3500	0.5000	0.5000	0.6500
2						
3						
4						

**IZVJEŠTAJ:** Da bi dobili dugotrajnu vjerojatnost osmatranja izlaznog simbola 1 koristimo Bayesov teorem potpune vjerojatnosti.  $P(1) = P(1|q=1)*P(q=1) \dots$  za sva stanja.

(8c) 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 :

```
%izracunavamo empirijske dugotrajne vjerojatnosti koje cemo usporedivat s
%prethodno dobivenim teorijskima
hmEmp=hist(data',[1 2 3 4 5 6]);
empStac = mean(hmEmp')/T;
%trazenje maksimalne razlike
teorStac = stac/6; %izracun teorijskih vjerojatnosti

maxDiff = 0;
for i=1:6
    if abs(teorStac(i) - empStac(i)) > maxDiff
        maxDiff = abs(teorStac(i) - empStac(i));
    end
end
```

maxDiff	0.0279
---------	--------

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

(9a) 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-izvjesnost usrednjenu preko svih nex osmotrenih nizova, te upisite dobivene rezultate u naredna tri polja (max, min i mean):

```
% izracun log-izvjesnosti svakog niza u polju "data"
llm2=zeros(nex,1); % Stupac log-izvjesnosti
for i=1:nex
    llm2(i)=dhmm_logprob(data(i,:), prior0, transmat0, obsmat0);
end

maxll = max(llm2);
minll = min(llm2);
avgll = mean(llm2);
```

maxll	-153.4338
minll	-172.3760
avgll	-163.5928

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

(10a) 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 najvise 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. Tocnost vaseg 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):