

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
Matemaatika ja statistika instituut
Matemaatilise statistika eriala

Segmenteerimine peidetud Markovi mudelite segude korral

Magistritöö (30 EAP)

Autor:

Jaak SÕNAJALG

Juhendaja:

Jüri LEMBER



Tartu, 2016

Segmenteerimine peidetud Markovi mudelite segude korral

Magistritöö

Jaak Sõnajalg

Lühikokkuvõte. Peidetud Markovi mudelite (HMM) seguga puutume kokku, kui vaatleme HMM-i parameetrite komplekti fikseerimise asemel jaotust parameetrite ruumil. Klassikaline moodus peidetud seisundite vektori hindamiseks on nn suurima tõepära meetod, mis seisneb vaatluste alusel ühe mudeli fikseerimises ning välja valitud mudelile hübriid-tõepära maksimiseeriva Viterbi algoritmi rakendamises. Siin töös tutvume alternatiivse meetodiga (nn hübriid-EM algoritm), mille puhul on eesmärk peidetud seisundite vektorit hinnata otse, HMM-i parameetreid hindamata. Hübriid-EM algoritmi väljund sõltub algjoondusest, tutvustame üht viisi algjoonduse valimiseks. Töö praktilises osas uurime kahe HMM-i segu korral, kuidas mõjutab algjoonduse valik hübriid-EM algoritmi väljundit. Lisaks võrdleme suurima tõepära meetodil ja hübriid-EM algoritmi kasutades leitud väljundjoonduste omadusi.

CERCS teaduseriala: P160 Statistika, operatsioonanalüüs, programmeerimine, finants- ja kindlustusmatemaatika.

Märksõnad: peidetud Markovi mudelite segud, HMM, segmenteerimine, hübriid-EM algoritm, Viterbi algoritm.

Segmentation in Case of Mixtures of Hidden Markov Models

Master's thesis

Jaak Sõnajalg

Abstract. The need to deal with mixtures of Hidden Markov Models (HMMs) arises when we fix a distribution on parameter space instead fixing just one set of parameters for HMM. The common way to estimate the hidden state vector is the so-called expectation maximisation method, using which we fix just one HMM and apply the Viterbi algorithm. In this paper an alternative method is described (so-called segmentation-EM algorithm), which aims to estimate the state vector directly, without estimating the parameters of the HMM. The output of this algorithm depends on the choice of initial path. We describe a method for picking initial paths. In the practical part of the thesis we will apply the algorithm on a mixture of two HMMs and see how the choice of the initial path affects the output. We will then compare the properties of the outputs obtained using the expectation maximisation method and segmentation-EM algorithm.

CERCS research specialisation: P160 Statistics, operation research, programming, actuarial mathematics.

Keywords: mixture of Hidden Markov Models, HMM, segmentation, segmentation EM algorithm, Viterbi algorithm.

Sisukord

Sissejuhatus	4
1 Peidetud Markovi mudel	5
1.1 Definiitsioon	5
1.2 Riskid ja segmenteerimine	6
1.3 <i>Forward-backward</i> valemid	9
1.4 Tinglikud üleminekutõenäosused	11
1.5 k -ploki teoreem	13
2 Peidetud Markovi mudelite segu	14
2.1 Definiitsioon	14
2.2 Omadused	15
2.2.1 Kahe HMM-i segu pole HMM	15
2.2.2 Optimaalsusprintsip HMM-ide segu korral ei kehti	16
2.2.3 Optimaalsusprintsip ei kehti HMM-ide segude korral ka siis, kui üleminekutõenäosused on iga HMM-i korral samad	18
2.2.4 k -bloki teoreem ei kehti.	19
2.3 PMAP-joondus segude korral	19
3 Algoritm hübriitõepära maksimiseeriva joonduse hindamiseks HMM-ide segu puhul	21
3.1 Viterbi-EM	21
3.2 Algoritm PMAP-joonduse hindamiseks	24
3.3 Hübriid-EM algoritm	25
3.3.1 Üks meetod optimaalse väljundjoonduse leidmiseks	27
4 Simulatsioonid	29
4.1 Hübriid-EM erinevate parameetrite korral	29
4.2 Kahe meetodi võrdlus	38
Kirjandus	44

LISAD	45
Simulatsioonides kasutatud parameetrid	45
Simulatsioonides kasutatud mudelite Viterbi joondused	46
Lihtlitsents	47

Sissejuhatus

Sageli on mõistlik mõne ajas juhuslikult edasi areneva nähtuse korral rääkida selle nähtuse seisunditest ning neist seisunditest sõltuvatest vaatlustest. Veel võib olla mõistlik eeldada, et ühelt seisundilt teisele üleminekul on mingi kindel tõenäosus ja iga seisund emiteerib teatud vaatluse mingi kindla tõenäosusega, ent realiseerunud seisundid pole meile teada. Näiteks võib selline raamistik meid aidata automaatse kõnetuvastuse juures. Siis on seisunditeks inimese poolt väljastatavad häälikud, vaatlusteks mikrofoni poolt püütud helisignaal. Antud näite juures on vaatlused kõnetuvastusprogrammidele teada, seisundid aga mitte. Pakub huvi, kuidas teada olevate vaatluste põhjal hinnata, millised seisundid vaatluste taga on.

Peidetud Markovi mudelis on seisundite ja vaatluste vektor omavahel seotud just sellisel viisil. Lisaks eeldatakse, et seisundite vektor rahuldab Markovi omadust. Tänu sellele lisaeldusele saame kasutada mitmeid võtteid, mis hõlbustavad uurimise nähtuse kirjeldamist ning vaatluste alusel seisundite hindamist. Kõigist võimalikest seisundite vektoritest ehk joondustest parima välja valimiseks on mitmeid lähenemisviise. See, millist lähenemisviisi kasutada, sõltub sellest, kuidas me joonduse headuse defineerime. Olulisemate headuse kriteeriumitega puutume käesolevas töös ka kokku.

Peidetud Markovi mudeli üleminekutõenäosused ja tõenäosused, et mingi seisund emiteerib mingi vaatluse, ei pruugi meile aga teada olla. Realistlikum on neid peidetud Markovi mudeli parameetreid kirjeldada jaotusega. Käesolevas töös vaatleme juhtu, kus võimalikke parameetrite komplekte on lõplik arv. Vaatleme sellel lihtsustatud juhul, millised on peidetud Markovi mudelite kaalutud keskmise omadused ning tutvume algoritmiga, mille abil hinnata erinevatele headuskriteeriumitele vastavaid joondusi. Töö esimeses kolmes peatükis esitletavad meetodid paneme proovile töö neljandas peatükis. Uurime simulatsioonide abil, kuidas käitub tutvustatud algoritm erinevate parameetrite korral ning võrdleme selle sooritust ühe tehisõppe-alases kirjanduses enam levinud meetodi sooritusega.

Autori panuseks on lisaks käesoleva töö koostamisele kõigi töös esitatud tulemuste iseseisev tõestamine ning simulatsioonide läbi viimiseks kasutatud algoritmide programmeerimine. Simuleerimisel kasutatud programmid on kirjutatud programmeerimiskeeles Python ning tulemuste graafiliseks esitamiseks on kasutatud tarkvara Gnuplot.

Peatükk 1

Peidetud Markovi mudel

1.1 Definiitsioon

Olgu \mathcal{X} vaatluste ruum ja $\mathcal{Y} = \{1, \dots, K\}$ lõplik seisundite hulk. Olgu Y juhuslik vektor pikkusega T , mis rahuldab Markovi omadust ja mille elemendid võtavad väärtusi seisundite hulgal \mathcal{Y} . Olgu X juhuslik vektor pikkusega T , mis võtab väärtusi hulgal \mathcal{X} . Tähistame juhuslike vektorite Y ja X realisatsioonide vastavalt y ja x . Juhuslike vektorite Y ja X ajahetkele t vastavaid elemente tähistame vastavalt Y_t ja X_t ; realisatsioonide ajahetkele t vastavaid elemente vastavalt y_t ja x_t , $t = 1, \dots, T$. Juhuslike vektorite paar (Y, X) on peidetud Markovi mudel (edaspidi HMM, ingliskeelsest nimetusest *Hidden Markov Model*), kui kehtivad järgmised tingimused:

1. juhuslike suuruste X_t tinglik sõltumatus tingimusel Y ;
2. juhusliku suuruse X_t jaotus sõltub Y kaudu ainult Y_t jaotusest.

Vektori Y algtõenäosuseid tähistame $p_{0,i}$, üleminekutõenäosuseid tähistame $p_{i,j}$, kus $i, j \in \mathcal{Y}$. Vaatluste vektori elemendi X_1 jaotust tingimusel, et sellele vastava seisundi korral kehtib $Y_1 = j$, $j \in \mathcal{Y}$, nimetame emissioonijaotuseks. Eeldame üldistust kitsendamata, et emissioonijaotustel leiduvad tihedused mingi ühise moodsu λ suhtes. Neid tihedusi nimetame emissioonitihedusteks, tähistame neid $f_j(x_t)$. Realisatsioonide y ja x ühistõepära avaldub:

$$p(y, x) := \prod_{t=1}^T p_{y_{t-1}, y_t} f_{y_t}(x_t). \quad (1.1)$$

Tähistame veel tinglikud tõenäosused

$$p(y|x) := \mathbf{P}(Y = y|X = x), \quad (1.2)$$

$$p(x|y) := \mathbf{P}(X = x|Y = y), \quad (1.3)$$

$$p_t(y_t|x) := \mathbf{P}(Y_t = y_t|X = x). \quad (1.4)$$

Kui soovime viidata juhuslike vektorite Y ja X esimesest t elemendist koosnevatele alamvektoritele (Y_1, \dots, Y_t) ja (X_1, \dots, X_t) , kasutame vastavalt tähistusi Y^t ja X^t ; vastavaid realisatsioone tähistame y^t ja x^t . Alamvektoritele $(Y_{t_1}, \dots, Y_{t_2})$ ja $(X_{t_1}, \dots, X_{t_2})$, $t_1 < t_2$, viidates kasutame tähistusi $Y_{t_1}^{t_2}$ ja $X_{t_1}^{t_2}$; vastavaid realisatsioone tähistame analoogselt vastavalt $y_{t_1}^{t_2}$ ja $x_{t_1}^{t_2}$.

1.2 Riskid ja segmenteerimine

Segmenteerimiseks nimetatakse vaatluste x_1, \dots, x_T põhjalt joonduse y_1, \dots, y_T hindamist. Kõigi võimalike joonduste seast ühe välja valimiseks kasutatakse headuse kriteeriumina riski. Fikseeritud vaatluste x korral joonduse hindamine seisneb siis sellise joonduse leidmises, mille korral on risk minimaalne. Defineerime riskid, millega käesolevas töös kokku puutume:

$$R_\infty(y|x) := 1 - p(y|x), \quad R_1(y|x) := T - \sum_{t=1}^T p_t(y_t|x) \quad (1.5)$$

ning neile vastavad logaritmilised riskid

$$\bar{R}_\infty(y|x) = -\ln p(y|x), \quad \bar{R}_1(y|x) = -\sum_{t=1}^T \ln p_t(y_t|x) \quad (1.6)$$

Paneme tähele, et riski $R_\infty(\cdot|x)$ minimiseerimine on ekvivalentne riski $\bar{R}_\infty(\cdot|x)$ minimiseerimisega. Samuti on riski $R_1(\cdot|x)$ minimiseerimine samaväärne $\bar{R}_1(\cdot|x)$ minimiseerimisega. Joondust, mis minimiseerib $\bar{R}_\infty(\cdot|x)$, nimetatakse Viterbi joonduseks. Joondust, mis minimiseerib $\bar{R}_1(\cdot|x)$, nimetatakse PMAP-joonduseks (ingl *pointwise maximum a posteriori alignment*). Riski $\bar{R}_\infty(\cdot|x)$ minimiseerides otsime joondust, mille korral on tõepära $p(y|x)$ maksimaalne. Riski $\bar{R}_1(\cdot|x)$ minimiseerimisel on eesmärgiks klassifitseerimisvigade arvu minimiseerimine. See tähendab, et me otsime fikseeritud vaatluste vektori x korral joondust $\hat{y} \in \mathcal{Y}^T$, mille korral kehtib

$$\hat{y} = \arg \min_{y \in \mathcal{Y}^T} \sum_{s \in \mathcal{Y}^T} \left(\sum_{t=1}^T l(y_t, s_t) \right) p(s|x), \quad (1.7)$$

kus $l(i, j)$ on punktiviisiline kaofunktsioon:

$$l(i, j) := \begin{cases} 0, & \text{kui } i = j \\ 1, & \text{kui } i \neq j \end{cases} \quad i, j \in \mathcal{Y}. \quad (1.8)$$

Sama vaatluste vektori korral leitud Viterbi ja PMAP-joondused võivad olla väga erinevad. Näiteks ei võeta PMAP-joonduse leidmisel arvesse üleminekutõenäosuseid, mistõttu võib PMAP-joonduse \hat{y} tinglik tõepära $p(\hat{y}|x)$ olla võrdne nulliga, st \hat{y} pole kindlasti Viterbi joondus. Viterbi joonduse korral aga võib keskmine klassifitseerimisvigade arv olla väga suur. Kahe lähenemise positiivsete omaduste kombineerimiseks defineeritakse hübriidrisk

$$\bar{R}_\alpha(y|x) = \alpha \bar{R}_\infty(y|x) + (1 - \alpha) \bar{R}_1(y|x), \quad \alpha \in [0, 1]. \quad (1.9)$$

Paneme tähele, et kuna

$$\bar{R}_\alpha(y|x) = - \left(\alpha \ln p(y|x) + (1 - \alpha) \sum_{t=1}^T \ln p_t(y_t|x) \right), \quad (1.10)$$

siis kehtib

$$\arg \min_y \bar{R}_\alpha(y|x) = \arg \max_y \left(\alpha \ln p(y|x) + (1 - \alpha) \sum_{t=1}^T \ln p_t(y_t|x) \right), \quad (1.11)$$

st hübriidrisi minimeerimine on samaväärne nn hübriidtõepära maksimeerimisega. Kuna $\ln p(y|x) = \ln p(y, x) - \ln p(x)$, $\ln p_t(y_t|x) = \ln p_t(y_t, x) - \ln p(x)$ ja kuna tõepära $\ln p(x)$ ei sõltu joondusest y , siis võime võrduses (1.11) tinglikud tõepärad soovi korral asendada ühistõepäradega:

$$\arg \min_y \bar{R}_\alpha(y|x) = \arg \max_y \left(\alpha \ln p(y, x) + (1 - \alpha) \sum_{t=1}^T \ln p_t(y_t, x) \right). \quad (1.12)$$

Nagu eespool öeldud, on üks parima joonduse välja valimise meetod sellise joonduse y leidmine, mille korral on tõenäosus $p(y|x)$, või samaväärselt tõenäosus $p(y, x)$, maksimaalne. Selle muidu eksponentsiaalse keerukusega ülesande lahendamiseks saame kasutada Viterbi algoritmi. See algoritm põhineb Bellmani optimaalsusprintsibil (vt [2], lk 450). Järgnevas veendume, et tõepära $p(y, x)$ maksimeerivat joondust y on võimalik leida dünaamilist planeerimist kasutades.

Lause 1.2.1. *Peidetud Markovi ahelate korral kehtib iga seisundi y_{t+1} korral järgnev optimaalsusprintsip:*

$$\arg \max_{y^t: y_t=i} p(y^{t+1}, x^{t+1}) = \arg \max_{y^t: y_t=i} p(y^t, x^t) \quad (1.13)$$

Tõestus. Kasutades HMM-i omadusi ja Markovi omadust:

$$p(y^{t+1}, x^{t+1}) = p(x^{t+1}|y^{t+1}) \cdot p(y^{t+1})$$

$$\begin{aligned}
&= p(x^t|y^{t+1}) \cdot p(x_{t+1}|y^{t+1}) \cdot p(y^{t+1}) \\
&= p(x^t|y^t) \cdot p(x_{t+1}|y_{t+1}) \cdot p(y^{t+1}) \\
&= p(x^t|y^t) \cdot p(x_{t+1}|y_{t+1}) \cdot p(y^t) \cdot p(y_{t+1}|y^t) \\
&= p(x^t, y^t) \cdot p(x_{t+1}|y_{t+1}) \cdot p(y_{t+1}|y^t) \\
&= p(x^t, y^t) \cdot p_{y_t, y_{t+1}} \cdot f_{y_{t+1}}(x_{t+1}).
\end{aligned}$$

Olgu $s^t = \arg \max_{y^t: y_t=i} p(y^t, x^t)$ ja olgu $y_{t+1} = j$. Siis

$$\max_{y^t: y_t=i} p(y^{t+1}, x^{t+1}) = p_{i,j} \cdot f_j(x_{t+1}) \cdot \max_{y^t: y_t=i} p(x^t, y^t).$$

Sellest järeldubki soovitud tulemus:

$$s^t = \arg \max_{y^t: y_t=i} p(y^{t+1}, x^{t+1}). \quad (1.14)$$

□

Eelnevalt oleme fikseerinud peidetud Markovi mudeli üleminekutõenäosustega $p_{i,j}$ ja emissioonitõenäosustega $f_j(x_t)$.

Sisend : Vaatlused x_1, x_2, \dots, x_T .

Arvuta iga $j \in \mathcal{Y}$ korral $\delta_1(j) = \ln p_{0,j} + \ln f_j(x_1)$.

Iga $t = 2, \dots, T$ korral:

1. Arvuta iga $j \in \mathcal{Y}$ korral $\delta_t(j) = \max_i (\delta_{t-1}(i) + \ln p_{i,j}) + \ln f_j(x_t)$.
2. Jäta meelde seisund $i_{t-1}(j) = \arg \max_i (\delta_{t-1}(i) + \ln p_{i,j})$.

Leia tagant ettepoole tulles:

$$\begin{aligned}
v_T &= \arg \max_j \delta_T(j), \\
v_t &= i_t(v_{t+1}), \quad t = T-1, T-2, \dots, 1.
\end{aligned}$$

Väljund: Viterbi joondus v .

Algoritm 1: Viterbi algoritm

Vaatleme joondusi pikkusega t , mis lõppevad seisundis j . Olgu v^t nende joonduste seas selline, mille korral tõepära $\ln p(v^t, x^t)$ on maksimaalne ja olgu $\delta_t(j)$ tõepära arvuline väärtus selle joonduse korral, st

$$\begin{aligned}
v^t &= \arg \max_{y^t: y_t=j} \ln p(y^t, x^t), \\
\delta_t(j) &= \max_{y^t: y_t=j} \ln p(y^t, x^t).
\end{aligned}$$

Lause (1.2.1) kohaselt saame suurusi $\delta_t(j)$ leida rekursiivselt. Teades iga seisundi $i \in \mathcal{Y}$ korral $\delta_{t-1}(i)$ väärtust, valime välja sellise i , mille korral suurus $\delta_{t-1}(i)p_{ij}f_j(x_t)$ on maksimaalne:

$$\begin{aligned}\delta_t(j) &= \max_{y^t: y_t=j} \ln p(y^t, x^t) \\ &= \max_{y^{t-1}: y_{t-1}=i} \left(\ln p(y^{t-1}, x^{t-1}) + \ln p_{ij} \right) + \ln f_j(x_t) \\ &= \max_i (\delta_{t-1}(i) + \ln p_{ij}) + \ln f_j(x_t).\end{aligned}$$

Kui jätame iga iteratsioonisammu korral meelde, milline on seisund i , mida eelviimasel ajahetkel läbiv joondus suuruse $\delta_t(j)$ maksimiseerib, ja leppides kokku, et mitme sellise seisundi i korral valime neist suvalise, saame Viterbi joonduse leidmiseks sobiliku algoritmi (vt algoritm (1)). Seda algoritmi nimetame Viterbi algoritmiks.

Lause 1.2.2. *Võttes algoritmis (1)*

$$\begin{aligned}\delta_1(j) &= \alpha \ln p_{0,j} + \alpha \ln f_j(x_1) + (1 - \alpha)p_1(j|x), \\ \delta_t(j) &= \max_i (\delta_{t-1}(i) + \alpha \ln p_{i,j}) + \alpha \ln f_j(x_t) + (1 - \alpha)p_t(k|x),\end{aligned}\tag{1.15}$$

saame algoritmi hübriidriski $\bar{R}_\alpha(\cdot|x)$ minimiseeriva joonduse leidmiseks.

Tõestus. Olgu D ajahetk, mille korral kehtib $2 \leq D < T$. Iga seisundi y_{D+1} korral kehtib

$$\begin{aligned}& \max_{y^D: y_D=i} \left(\alpha \ln p(y^{D+1}, x^{D+1}) + (1 - \alpha) \sum_{t=1}^{D+1} p_t(y_t|x) \right) \\ &= \max_{y^D: y_D=i} \left(\alpha \ln p(y^D, x^D) + (1 - \alpha) \sum_{t=1}^D p_t(y_t|x) \right) \\ & \quad + \alpha \ln p_{i,y_{D+1}} + \alpha \ln f_{y_{D+1}}(x_{D+1}) + (1 - \alpha)p_{D+1}(y_{D+1}|x).\end{aligned}$$

Seega saab hübriidriski $\bar{R}_\alpha(\cdot|x)$ minimiseerivat joondust leida rekursiivselt. Algoritm (1) suuruste $\delta_t(j)$ asendamisel suurustega (1.15) leiame just vajaminevad tõepärad ja algoritmi väljundiks on siis hübriidriski mõttes optimaalne joondus. \square

1.3 *Forward-backward* valemid

Mitmete meis edasises huvitavate tõepärade leidmise hõlbustamiseks defineerime iga $x^t \in \mathcal{X}^t$, $t \leq T$ ja seisundi $j \in \mathcal{Y}$ korral *forward-backward* muutujad:

$$\alpha(j, x^t) := p_t(j|x^t)p(x^t),\tag{1.16}$$

$$\beta(x_{t+1}^T | j) := \begin{cases} 1, & \text{kui } t = T \\ p(x_{t+1}^T | Y_t = j), & \text{kui } t < T. \end{cases} \quad (1.17)$$

Forward-backward muutujaid saab arvutada rekursiivselt [5]:

$$\alpha(j, x^{t+1}) = \sum_k \alpha(k, x^t) p_{kj} f_j(x_{t+1}), \quad (1.18)$$

$$\beta(x_t^T | j) = \sum_k p_{jk} f_k(x_t) \beta(x_{t+1}^T | k) \quad (1.19)$$

Joonduse hübriidriski arvutamiseks, samuti PMAP-joonduse leidmiseks vajalik t ep ara $p_t(k|x)$ avaldub *forward-backward* muutujate kaudu j rgnevalt [5]:

$$p_t(k|x) = \frac{\alpha(k, x) \cdot \beta(x_{t+1}^T | k)}{p(x)}. \quad (1.20)$$

Mida suurem on T , seda v iksemaks muutuvad t onen osused $p(x^t)$ ja seda v iksemaid v aartusi v otavad *forward-backward*-muutujad. Piisavalt suure T korral v oivad arvud muuta n i v ikeseks, et arvuti  umardab need nulliks. Eelpool defineeritud *forward-backward* muutujate asemel saab defineerida skaleeritud *forward-backward*-muutujad, mille arvutamisel pole t onen osusi $p(x^t)$ leida vaja. Sobiva viisi skaleeritud *forward*-muutuja $\tilde{\alpha}(j, x^t)$ ja skaleeritud *backward*-muutuja $\tilde{\beta}(x_{t+1}^T | j)$ defineerimiseks saame v orduses (1.20) liikmeid  umber grupeerides:

$$\tilde{\alpha}(j, x^t) := p(j|x^t), \quad (1.21)$$

$$\tilde{\beta}(x_{t+1}^T | j) := \begin{cases} 1, & \text{kui } t = T; \\ \frac{p(x_{t+1}^T | Y_t = j)}{p(x_{t+1}^T | x^t)}, & \text{kui } t < T. \end{cases} \quad (1.22)$$

T onen osus $p_t(k|x)$ avaldub siis

$$\begin{aligned} p_t(k|x) &= \frac{\alpha(k, x^t) \cdot \beta(x_{t+1}^T | k)}{p(x)} \\ &= \frac{p_t(k|x^t) p(x^t) p(x_{t+1}^T | Y_t = k)}{p(x^t, x_{t+1}^T)} \\ &= \tilde{\alpha}(k, x^t) \cdot \tilde{\beta}(x_{t+1}^T | k). \end{aligned} \quad (1.23)$$

Leidmaks eeskirja skaleeritud *forward-backward* muutujate rekursiivseks arvutamiseks, kasutame rekursioone (1.18) ja (1.19). Rekursioon skaleeritud *forward*-muutujate leidmiseks:

$$\begin{aligned}
\tilde{\alpha}(j, x_1) &= \frac{p_{0,j} f_j(x_1)}{p(x_1)}, \\
\tilde{\alpha}(j, x^t) &= \frac{\alpha_t(j)}{p(x^t)} \\
&= \frac{\sum_k \alpha(k, x^{t-1}) p_{kj} f_j(x_t)}{p(x^{t-1}) p(x_t | x^{t-1})} \\
&= \frac{\sum_k \tilde{\alpha}(k, x^{t-1}) p_{kj} f_j(x_t)}{p(x_t | x^{t-1})}.
\end{aligned} \tag{1.24}$$

Rekursiivne arvutuseeskiri skaleeritud *backward*-muutujate leidmiseks on järgmine:

$$\begin{aligned}
\tilde{\beta}(x_t^T | j) &= \frac{\beta(x_t^T | j)}{p(x_t^T | x^{t-1})} \\
&= \frac{\sum_k p_{jk} f_k(x_t) \beta(x_{t+1}^T | k)}{p(x_t^T | x^{t-1})} \\
&= \frac{p(x^{t-1}) \cdot \sum_k p_{jk} f_k(x_t) \beta(x_{t+1}^T | k)}{p(x_t^T, x^{t-1})} \\
&= \frac{p(x^{t-1}) \cdot \sum_k p_{jk} f_k(x_t) \beta(x_{t+1}^T | k)}{p(x_{t+1}^T | x^t) \cdot p(x^t)} \\
&= \frac{p(x^{t-1}) \cdot \sum_k p_{jk} f_k(x_t) \tilde{\beta}(x_{t+1}^T | k)}{p(x_t, x^{t-1})} \\
&= \frac{\sum_k p_{jk} f_k(x_t) \tilde{\beta}(x_{t+1}^T | k)}{p(x_t | x^{t-1})}.
\end{aligned} \tag{1.25}$$

Kummagi rekursiivse valemi nimetajas olevad suurused leiame paralleelselt skaleeritud *forward*-muutujate arvutamisega:

$$\begin{aligned}
p(x_1) &= \sum_j p_{0,j} f_j(x_1) \\
p(x_t | x^{t-1}) &= \sum_s f_s(x_t) \sum_j p_{t-1}(j | x^{t-1}) p_{js} \\
&= \sum_s f_s(x_t) \sum_j \tilde{\alpha}(j, x^{t-1}) p_{js}.
\end{aligned} \tag{1.26}$$

Neid suurusi kasutades saame leida vaatluste jada tõenäosuse. Et see tõenäosus võib piisavalt suure T korral olla väga väike, tasub meil leida hoopis selle tõenäosuse logaritmi:

$$\ln p(x) = \ln p(x_1) + \ln p(x_2 | x_1) + \ln p(x_3 | x^2) + \dots + \ln p(x_T | x^{T-1}). \tag{1.27}$$

1.4 Tinglikud üleminekutõenäosused

Käesoleva töö praktilises osas soovime fikseeritud vaatluste vektori x korral genereerida joondusi tinglikust jaotusest $p(y|x)$. Siin alapeatükis uurime, kuidas seda teha.

HMM-ide puhul saab näidata [4], et vaatluste vektoriga tinglikustatud seisundite vektori üleminekutõenäosuste puhul kehtib

$$\mathbf{P}(Y_{t+1} = y_{t+1} | Y_t = y_t, X = x) = \mathbf{P}(Y_{t+1} = y_{t+1} | Y_t = y_t, X_{t+1}^T = x_{t+1}^T). \quad (1.28)$$

Seda tulemust kasutades saame tuletada arvutuseeskirja tinglike üleminekutõenäosuste leidmiseks:

$$\begin{aligned} p(y_{t+1} | y_t, x_{t+1}^T) &= \frac{p(y_{t+1}, y_t, x_{t+1}^T)}{p(y_t, x_{t+1}^T)} \\ &= \frac{\sum_{y_{t+2}^T \in \mathcal{Y}^{(T-t-1)}} p(y_t^T, x_{t+1}^T)}{p(y_t, x_{t+1}^T)} \\ &= \frac{p(y_t) p(y_{t+1} | y_t) f_{y_{t+1}}(x_{t+1}) \sum_{y_{t+2}^T \in \mathcal{Y}^{(T-t-1)}} p(y_{t+2}^T, x_{t+2}^T | y_{t+1})}{p(y_t) p(x_{t+1}^T | y_t)} \\ &= \frac{p_{y_t, y_{t+1}} f_{y_{t+1}}(x_{t+1}) p(x_{t+2}^T | y_{t+1})}{p(x_{t+1}^T | y_t)} \\ &= \frac{p_{y_t, y_{t+1}} f_{y_{t+1}}(x_{t+1}) \beta(x_{t+2}^T | y_{t+1})}{\beta(x_{t+1}^T | y_t)}. \end{aligned}$$

Kui soovime kasutada skaleeritud *backward*-muutujaid, saab arvutuseeskiri tinglike üleminekutõenäosuste leidmiseks järgmise kuju:

$$\begin{aligned} p(y_{t+1} | y_t, x_{t+1}^T) &= \frac{p_{y_t, y_{t+1}} f_{y_{t+1}}(x_{t+1}) \beta(x_{t+2}^T | y_{t+1})}{\beta(x_{t+1}^T | y_t)} \\ &= \frac{p_{y_t, y_{t+1}} f_{y_{t+1}}(x_{t+1}) \tilde{\beta}(x_{t+2}^T | y_{t+1}) p(x_{t+2}^T | x^{t+1})}{\tilde{\beta}(x_{t+1}^T | y_t) p(x_{t+1}^T | x^t)} \\ &= \frac{p_{y_t, y_{t+1}} f_{y_{t+1}}(x_{t+1}) \tilde{\beta}(x_{t+2}^T | y_{t+1}) p(x^t)}{\tilde{\beta}(x_{t+1}^T | y_t) p(x^{t+1})} \\ &= \frac{p_{y_t, y_{t+1}} f_{y_{t+1}}(x_{t+1}) \tilde{\beta}(x_{t+2}^T | y_{t+1})}{\tilde{\beta}(x_{t+1}^T | y_t) p(x_{t+1} | x^t)}. \end{aligned} \quad (1.29)$$

Viimaks tuletame arvutuseeskirja ka tinglike algtõenäosuste jaoks:

$$\begin{aligned} p(y_1 | x^T) &= \frac{p(y_1, x)}{p(x)} \\ &= \frac{\sum_{y_2^T \in \mathcal{Y}^{(T-1)}} p(y, x)}{p(x)} \\ &= \frac{p_{0, y_1} f_{y_1}(x_1) \sum_{y_2^T \in \mathcal{Y}^{(T-1)}} p(y_2^T, x_2^T | y_1)}{p(x)} \\ &= \frac{p_{0, y_1} f_{y_1}(x_1) \beta(x_2^T | y_1)}{p(x)} \\ &= \frac{p_{0, y_1} f_{y_1}(x_1) \tilde{\beta}(x_2^T | y_1)}{p(x_1)}. \end{aligned} \quad (1.30)$$

1.5 k -ploki teoreem

Tutvume põgusalt artiklis [5] välja pakutud ideega leidmaks joondusi, mille korral on ühendatud Viterbi ja PMAP-joonduse positiivsed omadused. Ideeks on leida joondus y , mis maksimiseerib summa

$$p(y^k|x) + p(y_2^{k+1}|x) + p(y_3^{k+2}|x) + \cdots + p(y_{T-k+1}^T|x). \quad (1.31)$$

Valides $k = 1$ saame PMAP joonduse ja valides $k = T$ Viterbi joonduse. Vahepealsete k väärtuste korral loodame leida joondusi, mis on omadustelt Viterbi ja PMAP-joonduse vahepeal.

Selle idee edasiarendatud versiooni korral defineeritakse järgnevad suurused:

$$\bar{U}_k(s^T) := \prod_{j=1-k}^{T-1} p(s_{\max(j+1,1)}^{\min(j+k,T)}), \quad \bar{R}_k(s^T) := \ln \bar{U}_k(s^T). \quad (1.32)$$

Kehtib järgmine teoreem.

Teoreem 1.5.1. Iga $x \in \mathcal{X}^T$, $y \in \mathcal{Y}^T$ ja $k = 2, \dots, T$ korral kehtib

$$\bar{R}_k(y|x) = (k-1)\bar{R}_\infty(y|x) + \bar{R}_1(y|x) \quad (1.33)$$

Tõestus. vt [3]. □

Peatükk 2

Peidetud Markovi mudelite segu

2.1 Definiitsioon

Olgu meil m peidetud Markovi mudelit, mille vaatluste vektori ja seisundite vektori pikkus on T . Mudelite eristamiseks tähistame i -nda mudeli seisundite vektorit Y^i , vaatluste vektorit X^i , üleminekutõenäosusi $p_{k,j}(i)$ ja emissioonitihedusi $f_j(x_t|i)$. Kasutame i -nda mudeli suhtes leitud tõepärale viitamiseks tähist $p(\cdot|i)$. Igale mudelile vastab kaal π_i , nii et $\sum_i \pi_i = 1$. Nende mudelite seguks (Y, X) nimetame mudelit, mille korral

$$p(y, x) := \sum_{i=1}^m p(y, x|i)\pi_i. \quad (2.1)$$

Olgu $t_1, \dots, t_e, t'_1, \dots, t'_f$ vabalt valitud ajahetked, mille korral kehtib $1 \leq t_1 < \dots < t_e \leq T$ ja $1 \leq t'_1 < \dots < t'_f \leq T$. Järgnevas esitame ja tõestame arvutuseeskirja seisundite vektori elementide Y_{t_1}, \dots, Y_{t_e} ja vaatluste vektori elementide $X_{t'_1}, \dots, X_{t'_f}$ realisatsioonide y_{t_1}, \dots, y_{t_e} ja $x_{t'_1}, \dots, x_{t'_f}$ ühise tõepära leidmiseks. Arvutuseeskiri kehtib ka juhul, kui $e = 0$ või $f = 0$. Kui $e = 0$, leiame vaid vaatluste vektorist valitud elementidele vastavate realisatsioonide tõepära; kui $f = 0$ leiame seisundite vektorist valitud elementidele vastavate realisatsioonide tõepära.

Lause 2.1.1. *Suvaliste $U, V \subset \{1; 2; \dots; T\}$, iga seisundite komplekti $\{y_u \in \mathcal{Y}, u \in U\}$ ja iga uvate hulkade komplekti $\{H_v \subset \mathcal{X}, v \in V\}$ korral*

$$\mathbf{P}(Y_u = y_u, X_v \in H_v, u \in U, v \in V) = \sum_{i=1}^m \mathbf{P}(Y_u^i = y_u, X_v^i \in H_v, u \in U, v \in V) \cdot \pi_i. \quad (2.2)$$

Tõestus. Olgu λ mõõt, mille suhtes on defineeritud tihedused f_j hulgal \mathcal{X} . Olgu μ sellele mõõdule vastav korrutismõõt hulgal \mathcal{X}^T .

$$\mathbf{P}(Y_u = y_u, X_v \in H_v, u \in U, v \in V) =$$

$$\begin{aligned}
& \sum_{s \in \mathcal{Y}^T: (s_u = y_u, u \in U)} \int_{x \in \mathcal{X}^T: (x_v \in H_v, v \in V)} p(s, x) \, d\mu \\
&= \sum_{s \in \mathcal{Y}^T: (s_u = y_u, u \in U)} \int_{x \in \mathcal{X}^T: (x_v \in H_v, v \in V)} \left(\sum_{i=1}^m p(s, x|i) \cdot \pi_i \right) d\mu \\
&= \sum_{i=1}^m \sum_{s \in \mathcal{Y}^T: (s_u = y_u, u \in U)} \left(\int_{x \in \mathcal{X}^T: (x_v \in H_v, v \in V)} p(s, x|i) \cdot \pi_i \, d\mu \right) \\
&= \sum_{i=1}^m \mathbf{P}(Y_u^i = y_u, X_v^i \in H_v, u \in U, v \in V) \cdot \pi_i.
\end{aligned}$$

Summeerimise ja integreerimise järjekorra muutmiseks kasutasime integraali lineaarsuse omadust. □

2.2 Omadused

Siin alapeatükis uurime, kas HMM-ide segul on samasugused omadused, kui üksikul HMM-il.

2.2.1 Kahe HMM-i segu pole HMM

Vaatleme kaheseisundilisi Markovi mudeleid Y^1 ja Y^2 pikkusega T . Nende mudelite seguks nimetame mudelit Y , mille korral kehtib:

$$\mathbf{P}(Y = y) = \sum_{i=1}^2 \mathbf{P}(Y^i = y|i)\pi_i, \quad (2.3)$$

kus π_i on mudelite Y^1 ja Y^2 kaalud ning $\sum_i \pi_i = 1$.

Olgu vektorite Y^1 ja Y^2 üleminekumaatriksid vastavalt

$$\begin{pmatrix} 0.1 & 0.9 \\ 0.1 & 0.9 \end{pmatrix} \quad \text{ja} \quad \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}. \quad (2.4)$$

Seisundite 1 ja 2 algseisundiks olemise tõenäosus on Y^1 ja Y^2 puhul vastavalt $(0.2, 0.8)$ ja $(0.5, 0.5)$. Uurime, kas nende vektorite segu Y , kus Y^1 ja Y^2 kaaludeks on $\pi_1 = \pi_2 = \frac{1}{2}$, rahuldab Markovi omadust. Meeldetuletuseks märgime, et Markovi omadus on vektori Y korral rahuldatud, kui suvalise $2 < t \leq T$ korral kehtib

$$\mathbf{P}(Y_t = y_t | Y_{t-1} = y_{t-1}, \dots, Y_1 = y_1) = P(Y_t = y_t | Y_{t-1} = y_{t-1}). \quad (2.5)$$

Leiame vastavad tõenäosused $t = 3$, $y_3 = 2$, $y_2 = 2$ ja $y_1 = 2$ korral:

$$\mathbf{P}(Y_3 = 2 | Y_2 = 2) = \frac{\mathbf{P}(Y_2 = 2, Y_3 = 2)}{\mathbf{P}(Y_2 = 2)}$$

$$\begin{aligned}
&= \frac{\frac{1}{2} \cdot \mathbf{P}(Y_2^1 = 2, Y_3^1 = 2) + \frac{1}{2} \cdot \mathbf{P}(Y_2^2 = 2, Y_3^2 = 2)}{\frac{1}{2} \cdot \mathbf{P}(Y_2^1 = 2) + \frac{1}{2} \cdot \mathbf{P}(Y_2^2 = 2)} \\
&= \frac{(0.2 \cdot 0.9 + 0.8 \cdot 0.9) \cdot 0.9 + (0.5 \cdot 0.5 + 0.5 \cdot 0.5) \cdot 0.5}{(0.2 \cdot 0.9 + 0.8 \cdot 0.9) + (0.5 \cdot 0.5 + 0.5 \cdot 0.5)} \\
&= \frac{53}{70},
\end{aligned}$$

$$\begin{aligned}
\mathbf{P}(Y_3 = 2 | Y_2 = 2, Y_1 = 2) &= \\
&= \frac{\mathbf{P}(Y_1 = 2, Y_2 = 2, Y_3 = 2)}{\mathbf{P}(Y_1 = 2, Y_2 = 2)} \\
&= \frac{\frac{1}{2} \cdot \mathbf{P}(Y_1^1 = 2, Y_2^1 = 2, Y_3^1 = 2) + \frac{1}{2} \cdot \mathbf{P}(Y_1^2 = 2, Y_2^2 = 2, Y_3^2 = 2)}{\frac{1}{2} \cdot \mathbf{P}(Y_1^1 = 2, Y_2^1 = 2) + \frac{1}{2} \cdot \mathbf{P}(Y_1^2 = 2, Y_2^2 = 2)} \\
&= \frac{0.8 \cdot 0.9 \cdot 0.9 + 0.5 \cdot 0.5 \cdot 0.5}{0.8 \cdot 0.9 + 0.5 \cdot 0.5} \\
&= \frac{773}{970}
\end{aligned}$$

Leitud tõenäosused ei ole võrdsed, seega ei rahulda segu Y Markovi omadust. Kuna Y ei rahulda Markovi omadust, ei ole (Y, X) ühegi juhusliku vektori X korral HMM.

2.2.2 Optimaalsusprintsiiip HMM-ide segu korral ei kehti

Tahame teada, kas HMM-ide segu (2.1) korral saab suurima tõepäraga joonduse leidmiseks kasutada Viterbi algoritmi. Selleks uurime, kas segu korral kehtib optimaalsusprintsiiip (1.13).

Olgu v^t joondus, mille korral kehtib $p(v^t, x^t) = \max_{y^t} p(y^t, x^t)$. Vaatleme kõiki joondusi pikkusega $t + 1$, mis on ajahetkel t seisundis v_t . Kui optimaalsusprintsiiip kehtiks, peaks neist suurima tõepäraga joondus olema ajahetkeni t identne joondusega v^t , st peaks kehtima

$$v^t = \arg \max_{y^t: y_t = v_t} p(y^{t+1}, x^{t+1}).$$

Olgu $t = 5$, $\mathcal{Y} = \{1, 2\}$ ja $\mathcal{X} = \{A, B\}$. Olgu meil kaks mudelit kaaludega $\pi_1 = 0,25$ ja $\pi_2 = 0,75$. Olgu mudelite algtõenäosused ja üleminekumaatriksid vastavalt

$$P_1^0 = \begin{pmatrix} 0,3 \\ 0,7 \end{pmatrix}, \quad P_1 = \begin{pmatrix} 0,4 & 0,6 \\ 0,8 & 0,2 \end{pmatrix} \quad \text{ja} \quad P_2^0 = \begin{pmatrix} 0,8 \\ 0,2 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 0,3 & 0,7 \\ 0,6 & 0,4 \end{pmatrix}$$

ning olgu emissioonitõenäosused¹

¹Emissioonitõenäosused on siin ja edaspidi esitatud maatrikskujul. Maatriksi i -ndale reale vastab seisundite ruumi $\mathcal{Y} = \{\tilde{y}_1; \tilde{y}_2; \dots; \tilde{y}_a\}$ element \tilde{y}_i , j -ndale veerule vaatluste ruumi $\mathcal{X} = \{\tilde{x}_1; \tilde{x}_2; \dots; \tilde{x}_b\}$ element \tilde{x}_j ja i -ndas reas ning j -ndas veerus asuvale elemendile emissioonitõenäosus $f_{\tilde{y}_i}(\tilde{x}_j)$.

$$f_1 = \begin{pmatrix} 0,2 & 0,8 \\ 0,9 & 0,1 \end{pmatrix} \quad \text{ja} \quad f_2 = \begin{pmatrix} 0,7 & 0,3 \\ 0,2 & 0,8 \end{pmatrix}$$

Olgu realiseerunud vaatlusteks $x^6 = (A, B, B, A, A, B)$. Siis tee pikkusega 5, mis maksimiseerib $p(y^5, x^5)$, on $v^5 = (1, 2, 2, 1, 2)$. Samas joondus w^6 , mis maksimiseerib $p(y^6, x^6)$ selliste joonduste y^6 seas, mis läbivad ajahetkel 5 seisundit 2, on $(2, 1, 1, 2, 2, 1)$. Näeme, et joonduse v^5 seisundid erinevad joonduse w^6 esimesele viiele ajahetkele vastavatest seisundidest moodustatud vektorist. Seega ei kehti segude (2.1) korral optimaalsusprintsiiip (1.13).

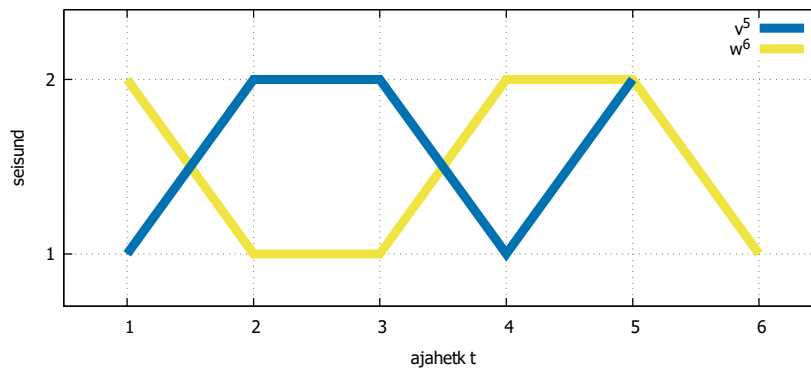
y^5	$p(y^5 x^5)$
(1, 2, 2, 1, 2)	0,124
(2, 1, 1, 2, 2)	0,088

Tabel 2.1: Mõni joondus pikkusega 5 ja neile vastavad tõepärad.

Tabelis (2.1) on esitatud joondusele v^5 vastav tõepära koos joonduse w^6 algusosale vastava tõepäraga. Tabelis (2.2) on esitatud joondusele w^6 vastav tõepära koos joonduse v^5 ühe seisundi võrra pikendamise teel saadud joonduste tõepäradega. Joonduse v^5 jätkamisel saadavate joonduste tõepärad on mõlemad väiksemad, kui joondusele w^6 vastav tõepära. Joonduseid v^5 ja v^6 on kujutatud joonisel (2.1).

y^6	$p(y^6 x^6)$
(2, 1, 1, 2, 2, 1)	0,096
(1, 2, 2, 1, 2, 1)	0,038
(1, 2, 2, 1, 2, 2)	0,068

Tabel 2.2: Mõni joondus pikkusega 6 ja neile vastavad tõepärad.



Joonis 2.1: Optimaalsusprintsiiip ei kehti. Joondus v^5 ei ühti joonduse w^6 algusosaga.

2.2.3 Optimaalsusprintsiiip ei kehti HMM-ide segude korral ka siis, kui üleminekutõenäosused on iga HMM-i korral samad

Olgu $\mathcal{Y} = \{1, 2\}$ ja $\mathcal{X} = \{A, B\}$. Olgu meil kaks mudelit kaaludega $\pi_1 = 0,25$ ja $\pi_2 = 0,75$. Olgu mõlema mudeli puhul algõtõenäosused ja üleminekumaatriksid

$$P_0 = \begin{pmatrix} 0,8 \\ 0,2 \end{pmatrix}, \quad P = \begin{pmatrix} 0,3 & 0,7 \\ 0,6 & 0,4 \end{pmatrix}$$

ning olgu emissioonitõenäosused

$$f_1 = \begin{pmatrix} 0,2 & 0,8 \\ 0,9 & 0,1 \end{pmatrix} \quad \text{ja} \quad f_2 = \begin{pmatrix} 0,07 & 0,93 \\ 0,02 & 0,98 \end{pmatrix}.$$

Olgu realiseerunud vaatlusteks $x^3 = (A, B, B)$. Siis joondus pikkusega 2, mis maksimeerib $p(y^2, x^2)$, on $v^2 = (2, 1)$. Joondus w^3 , mille korral $p(w^3, x^3) = \max_{y^3: y_2=1} p(y^3, x^3)$, on $(1, 1, 2)$. Joondus v^2 ei ühti joonduse w^3 algusosaga, seega ei kehti segude (2.1) korral optimaalsusprintsiiip (1.13) isegi siis, kui üleminekutõenäosused on iga mudeli korral samad.

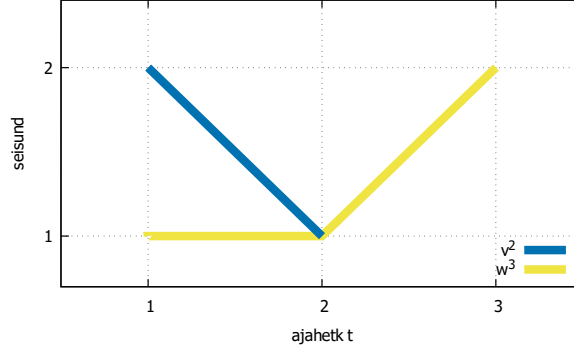
y^2	$p(y^2 x^2)$
(2, 1)	0,294
(1, 1)	0,269

Tabel 2.3: Mõni joondus pikkusega 2 ja neile vastavad tõepärad.

y^3	$p(y^2 x^3)$
(1, 1, 2)	0,163
(2, 1, 1)	0,106
(2, 1, 2)	0,050

Tabel 2.4: Mõni joondus pikkusega 3 ja neile vastavad tõepärad.

Tabelis (2.3) on esitatud joondusele v^2 vastav tõepära koos joonduse w^3 algusosale vastava tõepäraga. Tabelis (2.4) on esitatud joondusele w^3 vastav tõepära koos joonduse v^2 ühe seisundi võrra pikendamise teel saadud joonduste tõepäradega. Joonduse v^2 jätkamisel saadavate joonduste tõepärad on mõlemad väiksemad, kui joondusele w^3 vastav tõepära. Joonduseid v^2 ja w^3 on kujutatud joonisel (2.1).



Joonis 2.2: Optimaalsusprintsiiip ei kehti. Joondus v^2 ei ühti joonduse w^3 algusosaga.

2.2.4 k -bloki teoreem ei kehti.

Näitame kontranäite abil, et teoreem (1.5.1) segude korral ei kehti. Olgu kahe peidetud Markovi mudeli segu samasugune nagu ala-alapeatükis (2.2.2). Olgu joondus $s^6 = (1, 2, 1, 1, 2, 1)$ ja olgu realiseerunud vaatlusteks $x^6 = (A, B, B, A, B, A)$. Valime ploki suuruseks $k = 3$. Siis

$$\begin{aligned} (k-1)\bar{R}_\infty(s^n|x^n) + \bar{R}_1(s^n|x^n) &\approx 5,950, \\ \bar{R}_3(s^n|x^n) &\approx 6,758. \end{aligned}$$

2.3 PMAP-joondus segude korral

PMAP-joondus on joondus, mis minimiseerib $R_1(\cdot|x)$ ja $\bar{R}_1(\cdot|x)$. Teisisõnu, tegu on joondusega u , mis rahuldab iga ajahetke t korral võrdust

$$u_t = \arg \max_k p_t(k, x^T).$$

HMM-ide segu (2.1) korral tähendab see, et PMAP-joondus u rahuldab järgmist võrdust.

$$u_t = \arg \max_k \sum_{i=1}^m p_t(k, x|i)\pi_i. \quad (2.6)$$

Skaleeritud *forward-backward* valemite abil oskame me iga mudeli i , ajahetke t ja klassi k korral arvutada tõenäosusi $p_t(k|x, i)$ (vt (1.23)). Kui T pole liiga suur, saame iga mudeli korral arvutada ka tõenäosusi $p(x|i)$. Nii saame arvutada PMAP-joonduse leidmiseks vajalikud tõenäosused:

$$p_t(k, x|i) = p_t(k|x, i) \cdot p(x|i).$$

Kui T on suur ja eelistame vaatluste jada tõenäosuse asemel kasutada selle logaritmi $\ln p(x|i)$, saame võrduses (3.2) maksimiseeritava tõenäosuse logaritmi leida kasutades järgnevat võtet, millele viitame edaspidi nimetusega „logaritnivõte“.

Olgu a_1, \dots, a_m positiivsed arvud. Siis

$$\begin{aligned} \ln \left(\sum_i a_i \right) &= \ln a_1 + \ln \left(1 + \frac{a_2}{a_1} + \dots + \frac{a_m}{a_1} \right) \\ &= \ln a_1 + \ln (1 + \exp [\ln a_2 - \ln a_1] + \dots + \exp [\ln a_m - \ln a_1]) . \end{aligned} \tag{2.7}$$

Asendades tehtes (2.7) suurused a_i suurustega $p_t(k|x, i)p(x|i)\pi_i$, saame eeskirja tõenäosuste $p_t(k, x)$ leidmiseks. Nimelt siis $\ln a_i = \ln p_t(k|x, i) + \ln p(x|i) + \ln \pi_i$, mida oskame leida: tõenäosused $\ln p_t(k|x, i)$ saame võrdusest (1.23) ja suurused $\ln p(x|i)$ võrdusest (1.27).

Peatükk 3

Algoritm hübriitõepära maksimiseeriva joonduse hindamiseks HMM-ide segu puhul

Eelmises peatükis nägime, et HMM-ide segu korral optimaalsusprintsip (1.13) ei kehti. Seega ei saa HMM-ide segu korral tõepära maksimiseeriva joonduse leidmiseks kasutada Viterbi algoritmi.

Käesolevas peatükis jõuame samm-sammult iteratiivse algoritmini, mille abil saame HMM-ide segu korral hübriidtõepära maksimiseerivat joondust hinnata. Joonduse hindamise all peame silmas võimalikult kõrge hübriidtõepäraga joonduse leidmist. Muuhulgas saab seda algoritmi kasutada ka Viterbi joonduse hindamiseks. Esmalt tutvustame iteratiivseid algoritme, mida saab kasutada hübriidtõepära maksimiseerimiseks $\alpha = 0$ ja $\alpha = 1$ korral. Nende algoritmide juures kasutatavaid ideid kombineerides jõuame algoritmini, mille abil saab hübriidtõepära maksimiseerida suvalise α väärtuse korral.

Iteratiivne algoritm, milleni jõuame, kasutab hübriidriski minimiseeriva joonduse hindamiseks algjoondust. Käesoleva peatüki lõpus tutvume ühe ideega, mille alusel algjoondust valida.

3.1 Viterbi-EM

Tähistame

$$p(i|y, x) := \frac{p(y, x|i)\pi_i}{\sum_i p(y, x|i)\pi_i} = \frac{p(y, x|i)\pi_i}{p(y, x)}. \quad (3.1)$$

Viterbi-EM on algoritm, mis iga algjoonduse $y^{(0)}$ korral leiab iteratiivselt joondusi $y^{(1)}, y^{(2)}, \dots$, kusjuures

$$y^{(l+1)} = \arg \max_y \sum_i \ln (p(y, x|i)) p(i|y^{(l)}, x). \quad (3.2)$$

Viterbi-EM algoritmi eesmärk on maksimiseerida hübriidtõepära $\alpha = 1$ korral. Näeme järgnevas, et selle algoritmi iga iteratsioonisamm kasvatab sellele α väärtusele vastavat hübriidtõepära.

Lause 3.1.1. Iga Viterbi-EM algoritmi iteratsioonisamm kasvatab tõepära, st

$$\ln p(y^{(l+1)}|x) \geq \ln p(y^{(l)}|x). \quad (3.3)$$

Tõestus. Tähistame

$$p(y, i|x) := \frac{p(y, x|i) \cdot \pi_i}{p(x)}. \quad (3.4)$$

Siis kehtib:

$$p(y, i|x) = \frac{p(y, x, i)}{p(x)} = \frac{p(i|y, x) \cdot p(x, y)}{p(x)} = p(i|x, y) \cdot p(y|x), \quad (3.5)$$

kus $p(y, x, i)$ on i -nda mudeli ja realisatsioonide y ja x ühistõepära. Eelnevast võrduste-reast järeldub, et iga mudeli i korral kehtib

$$\ln p(y|x) = \ln p(y, i|x) - \ln p(i|y, x).$$

Nii saame üle kõigi mudelite summeerides anda meid huvitava tõepära logaritmile järgneva kuju.

$$\ln p(y|x) = \sum_i \ln (p(y, i|x)) p(i|y^{(l)}, x) - \sum_i \ln (p(i|y, x)) p(i|y^{(l)}, x).$$

Näitame nüüd, et $\ln \frac{p(y^{(l+1)}|x)}{p(y^{(l)}|x)} \geq 0$ ehk kehtib $p(y^{(l+1)}|x) \geq p(y^{(l)}|x)$:

$$\begin{aligned} \ln p(y^{(l+1)}|x) - \ln p(y^{(l)}|x) &= \\ &= \sum_i \ln (p(y^{(l+1)}, i|x)) p(i|y^{(l)}, x) - \sum_i \ln (p(i|y^{(l+1)}, x)) p(i|y^{(l)}, x) \\ &\quad - \left(\sum_i \ln (p(y^{(l)}, i|x)) p(i|y^{(l)}, x) - \sum_i \ln (p(i|y^{(l)}, x)) p(i|y^{(l)}, x) \right) \\ &= \sum_i \ln (p(y^{(l+1)}, i|x)) p(i|y^{(l)}, x) - \sum_i \ln (p(y^{(l)}, i|x)) p(i|y^{(l)}, x) \\ &\quad - \sum_i \ln (p(i|y^{(l+1)}, x)) p(i|y^{(l)}, x) + \sum_i \ln (p(i|y^{(l)}, x)) p(i|y^{(l)}, x) \end{aligned}$$

$$\begin{aligned}
&\geq \sum_i -\ln \left(\frac{p(i|y^{(l+1)}, x)}{p(i|y^{(l)}, x)} \right) p(i|y^{(l)}, x) \\
&\geq -\ln \sum_i \frac{p(i|y^{(l+1)}, x)}{p(i|y^{(l)}, x)} p(i|y^{(l)}, x) \\
&= -\ln \sum_i p(i|y^{(l+1)}, x) = 0.
\end{aligned}$$

Siin tuleneb esimene võrratus võrdusest (3.2). Nimelt kehtib $\ln p(y, x|i) = \ln p(y, i|x) + \ln p(x) - \ln \pi_i$. Võrduse parema poole liidetavatest sõltub y -st ainult $\ln p(y, i|x)$, mistõttu on (3.2) ekvivalentne järgmise võrdusega:

$$y^{(l+1)} = \arg \max_y \sum_i \ln (p(y, i|x)) p(i|y^{(l)}, x).$$

Teine võrratus tuleneb Jensen'i võrratusest. Nimelt on $-\ln(\cdot)$ kumer funktsioon. □

Järeldus 3.1.1. *Viterbi-EM algoritmi puhul kehtib*

$$p(y^{(l+1)}|x) \geq p(y^{(l)}|x). \quad (3.6)$$

Tõestus. Tuleneb logaritmi omadustest. □

Näitame järgnevalt, et kui asendada Viterbi algoritmis (algoritm (1)) üleminekutõenäosused ja emissioonitihedused teiste, sobivalt defineeritud suurustega, saame algoritmi joonduste (3.2) leidmiseks.

Teoreem 3.1.1. *Joonduse (3.2) võib leida Viterbi algoritmi abil, kus üleminekutõenäosuste (ka algõenäosuste) ja emissioonitõenäosuste (või emissioonitiheduste) asemel kasutatakse vastavalt suurusi*

$$u_{r,s} := \exp \left[\sum_i \ln (p_{r,s}(i)) p(i|y^{(l)}, x) \right], \quad r = 0, \dots, K, \quad s = 1, \dots, K. \quad (3.7)$$

ja

$$h_k(x_t) := \exp \left[\sum_i \ln (f_k(x_t|i)) p(i|y^{(l)}, x) \right], \quad k = 1, \dots, K. \quad (3.8)$$

Tõestus. Võrduses (3.2) maksimiseeritav suurus avaldub suuruste $u_{r,s}$ ja $h_k(x_t)$ kaudu järgnevalt:

$$\begin{aligned}
&\sum_i \ln (p(y, x|i)) p(i|y^{(l)}, x) = \\
&\sum_i \sum_{t=0}^{T-1} \ln p_{t,t+1}(i) p(i|y^{(l)}, x) + \sum_i \sum_{t=1}^T \ln f_{y_t}(x_t|i) p(i|y^{(l)}, x)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{t=0}^{T-1} \ln \exp \left[\sum_i \ln p_{t,t+1}(i) p(i|y^{(t)}, x) \right] + \sum_{t=1}^T \ln \exp \left[\sum_i \ln f_{y_t}(x_t|i) p(i|y^{(t)}, x) \right] \\
&= \sum_{t=0}^{T-1} \ln u_{y_t, y_{t+1}} + \sum_{t=1}^T \ln h_{y_t}(x_t).
\end{aligned}$$

Olgu s joondus pikkusega T ja olgu D mingi ajahetk, mille korral kehtib $1 \leq D < T$. Näitame nüüd, et iga seisundi y_{D+1} korral kehtib

$$\arg \max_{y^D: y_D=k} \sum_i \ln \left(p(y^{D+1}, x|i) \right) p(i|s, x) = \arg \max_{y^D: y_D=k} \sum_i \ln \left(p(y^D, x|i) \right) p(i|s, x).$$

Olgu $y_{D+1} = j$. Siis

$$\begin{aligned}
&\max_{y^D: y_D=k} \sum_i \ln \left(p(y^{D+1}, x|i) \right) p(i|s, x) \\
&= \max_{y^D: y_D=k} \left[\sum_{t=0}^D \ln u_{t,t+1} + \sum_{t=1}^{D+1} \ln h_{y_t}(x_t) \right] \\
&= \ln u_{kj} + \ln h_j(x_{D+1}) + \max_{y^D: y_D=k} \left[\sum_{t=0}^{D-1} \ln u_{y_t, t+1} + \sum_{t=1}^D \ln h_{y_t}(x_t) \right].
\end{aligned}$$

Seega saab joondust (3.2) leida Viterbi algoritmi abil, kus iga seisundi $j \in \mathcal{Y}$ korral

$$\begin{aligned}
\delta_1(j) &= \ln u_{0, y_1} + \ln h_j(x_1), \\
\delta_t(j) &= \max_{r \in \mathcal{Y}} (\delta_{t-1}(r) + \ln u_{r,j}) + \ln h_j(x_t).
\end{aligned}$$

□

Paneme tähele, et fikseeritud seisundi $r \in \mathcal{Y}$ korral ei pruugi summad $\sum_{s \in \mathcal{Y}} u_{r,s}$ ja $\sum_{x \in \mathcal{X}} h_k(x)$ võrduda ühega. See meid aga ei sega. Kuna suurused $u_{r,s}$ ja $h_k(x_t)$ on alati mittenegatiivsed, on tuletatud algoritmi kasutamine õigustatud.

3.2 Algoritm PMAP-joonduse hindamiseks

Eelmises peatükis nägime, et PMAP-joondust saab HMM-ide segude korral leida. Seetõttu pole PMAP-joondust hindavat algoritmi meile vaja. Siiski aitab sellise algoritmi uurimine meil hiljem leida algoritmi, mida kasutades saame hinnata hübriidriski minimeerivat joondust.

Tähistame esmalt

$$p(i|Y_t = k, x) := \frac{p_t(k, x|i)\pi_i}{\sum_i p_t(k, x|i)\pi_i} = \frac{p_t(k, x|i)\pi_i}{p_t(k, x)}. \quad (3.9)$$

Olgu $y^{(0)}, y^{(1)}, y^{(2)}, \dots$ joonduste jada, kusjuures

$$y^{(l+1)} = \arg \max_y \sum_t \sum_i (\ln p_t(y_t|x, i)) p(i|Y_t = y_t^{(l)}, x). \quad (3.10)$$

Lause 3.2.1. *Tingimuse (3.10) järgi leitud joonduste korral*

$$p_t(y_t^{(l+1)}|x) \geq p_t(y_t^{(l)}|x). \quad (3.11)$$

Tõestus. Olgu

$$p_t(k, i|x) := \frac{p_t(k, x|i) \cdot \pi_i}{p(x)}. \quad (3.12)$$

Analoogselt Teoreemi (3.1.1) tõestusega saame, et

$$\begin{aligned} \ln p_t(y_t^{(l+1)}|x) - \ln p_t(y_t^{(l)}|x) &= \\ &= \sum_i \ln (p_t(y_t^{(l+1)}, i|x)) p(i|y_t^{(l)}, x) - \sum_i \ln (p_t(y_t^{(l)}, i|x)) p(i|y_t^{(l)}, x) \\ &\quad - \sum_i \ln (p_t(i|y_t^{(l+1)}, x)) p(i|y_t^{(l)}, x) + \sum_i \ln (p_t(i|y_t^{(l)}, x)) p(i|y_t^{(l)}, x) \\ &\geq \sum_i -\ln \left(\frac{p_t(i|y_t^{(l+1)}, x)}{p_t(i|y_t^{(l)}, x)} \right) p(i|y_t^{(l)}, x) \\ &\geq -\ln \sum_i \frac{p_t(i|y_t^{(l+1)}, x)}{p_t(i|y_t^{(l)}, x)} p(i|y_t^{(l)}, x) = 0. \end{aligned}$$

□

3.3 Hübrid-EM algoritm

Kasutame eelmistes alapeatükkides esitatud ideid meile huvi pakuva algoritmi leidmiseks. Viidates meie eesmärgile hinnata hübridriski minimiseerivat joondust, viitame tuletatavale algoritmile nimetusega „hübrid-EM algoritm“.

Olgu $\alpha \in [0, 1]$, olgu $y^{(0)}, y^{(1)}, y^{(2)}, \dots$ joonduste jada, kusjuures

$$y^{(l+1)} = \arg \max_y \left[\alpha \sum_i \ln (p(y, x|i)) p(i|y^{(l)}, x) + (1 - \alpha) \sum_t \sum_i (\ln p_t(y_t|x, i)) p(i|Y_t = y_t^{(l)}, x) \right] \quad (3.13)$$

Lause 3.3.1. *Iga hübridalgoritmi iteratsioonisamm kasvatab tõepära, st (3.13) järgi leitud joonduste korral kehtib*

$$\alpha \ln p(y^{(l+1)}|x) + (1 - \alpha) \sum_t \ln p_t(y_t^{(l+1)}|x) \geq \alpha \ln p(y^{(l)}|x) + (1 - \alpha) \sum_t \ln p_t(y_t^{(l)}|x). \quad (3.14)$$

Tõestus. Võrratuse tõestamiseks kasutame eelmises kahes alapeatükis kasutatud võtteid.

$$\begin{aligned}
& \alpha \ln p(y^{(l+1)}|x) + (1 - \alpha) \sum_t \ln p_t(y_t^{(l+1)}|x) - \left[\alpha \ln p(y^{(l)}|x) + (1 - \alpha) \sum_t \ln p_t(y_t^{(l)}|x) \right] \\
&= \alpha \sum_i \ln \left(p(y^{(l+1)}, i|x) \right) p(i|y^{(l)}, x) - \alpha \sum_i \ln \left(p(i|y^{(l+1)}, x) \right) p(i|y^{(l)}, x) \\
&\quad - \left[\alpha \sum_i \ln \left(p(y^{(l)}, i|x) \right) p(i|y^{(l)}, x) - \alpha \sum_i \ln \left(p(i|y^{(l)}, x) \right) p(i|y^{(l)}, x) \right] \\
&\quad + (1 - \alpha) \sum_t \sum_i \ln \left(p_t(y_t^{(l+1)}, i|x) \right) p(i|y_t^{(l)}, x) - (1 - \alpha) \sum_t \sum_i \ln \left(p_t(i|y_t^{(l+1)}, x) \right) p(i|y_t^{(l)}, x) \\
&\quad - \left[(1 - \alpha) \sum_t \sum_i \ln \left(p_t(y_t^{(l)}, i|x) \right) p(i|y_t^{(l)}, x) - (1 - \alpha) \sum_t \sum_i \ln \left(p_t(i|y_t^{(l)}, x) \right) p(i|y_t^{(l)}, x) \right] \\
&= \alpha \sum_i \ln \left(p(y^{(l+1)}, i|x) \right) p(i|y^{(l)}, x) + (1 - \alpha) \sum_t \sum_i \ln \left(p_t(y_t^{(l+1)}, i|x) \right) p(i|y_t^{(l)}, x) \\
&\quad - \left[\alpha \sum_i \ln \left(p(y^{(l)}, i|x) \right) p(i|y^{(l)}, x) + (1 - \alpha) \sum_t \sum_i \ln \left(p_t(y_t^{(l)}, i|x) \right) p(i|y_t^{(l)}, x) \right] \\
&\quad + \alpha \sum_i - \ln \left(\frac{p(i|y^{(l+1)}, x)}{p(i|y^{(l)}, x)} \right) p(i|y^{(l)}, x) + (1 - \alpha) \sum_t \sum_i - \ln \left(\frac{p_t(i|y^{(l+1)}, x)}{p_t(i|y^{(l)}, x)} \right) p_t(i|y^{(l)}, x) \\
&\geq - \alpha \ln \sum_i \left(\frac{p(i|y^{(l+1)}, x)}{p(i|y^{(l)}, x)} \right) p(i|y^{(l)}, x) - (1 - \alpha) \sum_t \ln \sum_i \left(\frac{p_t(i|y^{(l+1)}, x)}{p_t(i|y^{(l)}, x)} \right) p_t(i|y^{(l)}, x) \\
&= - \alpha \ln \sum_i p(i|y^{(l+1)}, x) - (1 - \alpha) \sum_t \ln \sum_i p_t(i|y^{(l+1)}, x) \\
&= 0.
\end{aligned}$$

Võrratus tuleneb joonduste leidmise eeskirjast (3.13) ja Jenseni võrratusest. \square

Järelduse (3.1.1) ekvivalenti siin tõestada ei saa, st järjestikused joondused ei pruugi kasvatada tõepära

$$\alpha p(y^{(l)}|x) + (1 - \alpha) \sum_t p_t(y_t^{(l)}|x).$$

Järgnevalt esitame algoritmi hübriidriski minimeeriva joonduse hindamiseks ning tõestame, et selle kasutamine on õigustatud.

Teoreem 3.3.1. *Joondusi (3.13) saab leida Viterbi tüüpi algoritmiga.*

Tõestus. Avaldame eeskirjas (3.13) maksimiseeritava suuruse Teoreemis (3.1.1) defineeritud suuruste $u_{r,s}$ ning $h_k(x_t)$ ja suuruste

$$q_t(k) := \sum_i (\ln p_t(k|x, i)) p(i|Y_t = y_t^{(l)}, x) \quad (3.15)$$

kaudu:

$$\begin{aligned} & \alpha \sum_i \ln(p(y, x|i)) p(i|y^{(l)}, x) + (1 - \alpha) \sum_{t=1}^T \sum_i (\ln p_t(y_t|x, i)) p(i|Y_t = y_t^{(l)}, x) \\ & = \alpha \sum_{t=0}^{T-1} \ln u_{y_t, y_{t+1}} + \alpha \sum_{t=1}^T \ln h_{y_t}(x_t) + (1 - \alpha) \sum_{t=1}^T q_t(y_t). \end{aligned}$$

Olgu s mingi joondus pikkusega T ning olgu D mingi ajahetk, mille korral kehtib $1 \leq D < T$. Näitame nüüd, et iga seisundi y_{D+1} korral kehtib

$$\begin{aligned} \arg \max_{y^D: Y_D=k} & \left[\alpha \sum_i \ln(p(y^{(D+1)}, x|i)) p(i|s, x) + (1 - \alpha) \sum_{t=1}^{D+1} \sum_i (\ln p_t(y_t|x, i)) p(i|Y_t = s_t, x) \right] = \\ & \arg \max_{y^D: Y_D=k} \left[\alpha \sum_i \ln(p(y^D, x|i)) p(i|s, x) + (1 - \alpha) \sum_{t=1}^D \sum_i (\ln p_t(y_t|x, i)) p(i|Y_t = s_t, x) \right]. \end{aligned}$$

Olgu $y_{D+1} = j$. Siis

$$\begin{aligned} & \max_{y^D: Y_D=k} \left[\alpha \sum_i \ln(p(y^{(D+1)}, x|i)) p(i|s, x) + (1 - \alpha) \sum_{t=1}^{D+1} \sum_i (\ln p_t(y_t|x, i)) p(i|Y_t = s_t, x) \right] \\ & = \max_{y^D: Y_D=k} \left[\alpha \sum_{t=0}^D \ln u_{y_t, y_{t+1}} + \alpha \sum_{t=1}^{D+1} \ln h_{y_t}(x_t) + (1 - \alpha) \sum_{t=1}^{D+1} q_t(y_t) \right] \\ & = \alpha \ln u_{kj} + \alpha \ln h_j(x_{D+1}) + (1 - \alpha) q_{D+1}(j) \\ & \quad + \max_{y^D: Y_D=k} \left[\alpha \sum_{t=0}^{D-1} \ln u_{y_t, y_{t+1}} + \alpha \sum_{t=1}^D \ln h_{y_t}(x_t) + (1 - \alpha) \sum_{t=1}^D q_t(y_t) \right]. \end{aligned}$$

Seega saab joondusi (3.13) leida Viterbi algoritmiga, kus

$$\delta_{t+1}(k) = \max_j (\delta_t(j) + \alpha \ln u_{jk}) + \alpha \ln h_k(x_{t+1}) + (1 - \alpha) q_{t+1}(k). \quad (3.16)$$

□

3.3.1 Üks meetod optimaalse väljundjoonduse leidmiseks

Nagu käesoleva töö praktilises osas näeme, võib hübriid-EM algoritmi väljundjoonduste hübriidtoepära erinevate algjoonduste korral olla väga erinev. Loomulikult on meie eesmärgiks alati sellise algjoonduse kasutamine, mille korral on väljundjoonduse hübriidtoepära võimalikult suur. Tutvustame siin üht ideed hübriid-EM algoritmile sobiva algjoonduse leidmiseks.

Jagame lõigu $[0, 1]$ r -ks võrdseks osaks pikkusega Δ . Olgu $\alpha_0, \dots, \alpha_{r+1}$ arvud lõigul $[0, 1]$ sellised, et $\alpha_0 = 0$, $\alpha_{r+1} = 1$ ja kehtigu iga $i \in 0, \dots, r$ korral $\alpha_{i+1} = \alpha_i + \Delta$.

Alapeatükis (2.3) nägime, kuidas leida $\alpha = \alpha_0$ korral hübriidriski minimiseerivat joondust ehk PMAP-joondust y_{α_0} . Kasutame joondust y_{α_0} algjoondusena parameetri α väärtuse α_1 korral. Kui Δ on piisavalt väike ja kui kehtib $\ln p(y_{\alpha_0}, x) \neq 0$, on y_{α_0} ilmselt ka

α_1 korral hübriidtõepära poolest optimaalne joondus. Kui ta seda pole, loodame joondust y_{α_0} algjoondusena kasutades leida α_1 korral hübriidtõepära poolest optimaalse joonduse y_{α_1} . Joondust y_{α_1} parameetri α väärtuse α_2 korral algjoondusena kasutades loodame leida optimaalset joondust y_{α_2} ja nii edasi.

Selliste väikeste sammude kaupa parameetri α väärtust kasvatades loodame hübriid-tõepära maksimiseerivat joondust leida ka $\alpha = \alpha_{r+1}$ korral.

Peatükk 4

Simulatsioonid

Käesoleva töö praktiline osa koosneb kahest poolest. Esmalt uurime fikseeritud vaatluste vektori x korral, kuidas sõltub väljundjoondus parameetri α ja algjoonduse $y^{(0)}$ valikust. Teeme seda kahe segu korral.

Sageli hinnatakse HMM-ide segu segmenteerimisülesande lahendamiseks esmalt kõige tõenäolisem mudel ning seejärel kasutatakse seda mudelit huvipakkuvate joonduste leidmiseks (vt nt [1]). Praktilise osa teises pooles võrdleme seda meetodit kasutades leitud joonduste omadusi hübriid-EM algoritmi väljundjoonduste omadustega.

4.1 Hübriid-EM erinevate parameetrite korral

Eelmises peatükis jõudsimme algoritmini, mille abil loodame leida hübriidriski $\bar{R}_\alpha(\cdot|x)$ minimeerivat ehk hübriidtõepära

$$\alpha \ln p(s|x) + (1 - \alpha) \sum_{t=1}^T \ln p_t(s_t|x) \quad (4.1)$$

maksimiseerivat joondust. Selle algoritmi pseudokood on algoritm (2). Uurime kahe lihtsa näite korral, kuidas algoritm käitub erinevate parameetri α väärtuste ja erinevate algjoonduste $y^{(0)}$ korral. Vaatleme üksteise järel kaht kahe HMM-i segu. Esimese segu (segu A) korral on kõik üleminekutõenäosused nullist erinevad, teise segu (segu B) korral mitte. Edasises kasutame mingi α väärtuse korral ja algjoondust $y^{(0)}$ kasutades leitud algoritmi (2) väljundjoondusele viitamiseks tähist $v_\alpha(y^{(0)})$.

Olgu seisundite ruum ja vaatluste ruum vastavalt

$$\mathcal{Y} = \{1, 2, 3, 4, 5\},$$
$$\mathcal{X} = \{\text{„A“}, \text{„B“}, \text{„C“}, \text{„D“}, \text{„E“}, \text{„F“}, \text{„G“}\}.$$

Eelnevalt oleme fikseerinud m peidetud Markovi mudelit üleminekutõenäosustega $p_{k,l}(i)$ ja emissioonitõenäosustega f_l ning mudelite kaalud π_i .

Sisend : Vaatlused x_1, x_2, \dots, x_T , algjoondus $y^{(0)}$, parameeter α ja peatumiskriteeriumi määrav ε .

Iga $t = 1, \dots, T$ korral:

1. Arvuta valemi (1.26) järgi $p(x_t|x^{t-1})$.
2. Arvuta iga $j \in \mathcal{Y}$ korral valemit (1.24) kasutades $\tilde{\alpha}(j, x^t)$.

Iga $t = 1, \dots, T$ korral:

1. Arvuta iga $j \in \mathcal{Y}$ korral valemit (1.25) kasutades $\tilde{\beta}(x_{t+1}^T|j)$.

Leia suuruste $p(x_t|x^{t-1})$ abil $\ln p(x)$ (vt valem (1.27)) ja suuruste $\tilde{\alpha}(j, x^t)$ ja $\tilde{\beta}(x_{t+1}^T|j)$ abil suurused $p_t(j|x)$ (vt valem (1.23)).

Olgu $l = 0$ ja $\bar{R}_\alpha(y^{(-1)}|x) = \infty$.

Kuni $\bar{R}_\alpha(y^{(l-1)}|x) - \bar{R}_\alpha(y^{(l)}|x) > \varepsilon$:

1. **Leia** iga mudeli $i = 1, \dots, m$ korral nn logaritnivõtet kasutades $\ln p(i|y^{(l)}, x) = \ln(p(y^{(l)}, x|i)\pi_i) - \ln(\sum_{j=1}^m p(y^{(l)}, x|j)\pi_j)$
2. **Leia** iga ajahetke t , iga mudeli $i = 1, \dots, m$ ja $j \in \mathcal{Y}$ korral nn logaritnivõtet kasutades $\ln p_t(i|y_t^{(l)}, x) = \ln(p_t(y_t^{(l)}, x|i)\pi_i) - \ln(\sum_{j=1}^m p_t(y_t^{(l)}, x|j)\pi_j)$.
3. **Arvuta** iga $r = 0, \dots, K$, iga $s = 1, \dots, K$ korral nn üleminekutõenäosused $u_{r,s}$ (vt (3.7)).
4. **Arvuta** iga $k \in \mathcal{Y}$ ja iga üksiku vaatluse x_t korral nn emissioonitõenäosused $h_k(x_t)$ (vt (3.8)),
5. **Arvuta** iga ajahetke t ja iga $k \in \mathcal{Y}$ korral suurused $q_t(k)$ (vt (3.15))
6. **Arvuta** iga $j \in \mathcal{Y}$ korral $\delta_1(j) = \alpha(\ln u_{0,j} + \ln h_j(x_1)) + (1 - \alpha)q_1(j)$.

Iga $t = 2, \dots, T$ korral:

1. Arvuta iga $j \in \mathcal{Y}$ korral $\delta_t(j) = \max_i(\delta_{t-1}(i) + \alpha \ln u_{ij}) + \alpha \ln h_j(x_t) + (1 - \alpha)q_t(j)$.
2. Jäta meelde seisund $i_{t-1}(j) = \arg \max_i(\delta_{t-1}(i) + \alpha \ln u_{ij})$.

Leia tagant ettepoole tülles:

$$y_T^{(l+1)} = \arg \max_j \delta_T(j),$$

$$y_t^{(l+1)} = i_t(y_{t+1}^{(l+1)}), \quad t = T - 1, T - 2, \dots, 1.$$

Väljund: Joondus $y^{(l)}$.

Algoritm 2: Hübriid-EM algoritm

Fikseerime $T = 25$. Vaatleme kahe HMM-i segu, kus mudelite kaaludeks on $\pi_1 = 0,25$ ja $\pi_2 = 0,75$. Mudelite üleminekumaatriksid olgu vastavalt P_1^A ja P_2^A , algjaotused vastavalt P_1^0 ja P_2^0 ning emissioonitõenäosuste maatriksid vastavalt f_1 ja f_2 . Üleminekumaatriksid, algjaotused ja emissioonitõenäosuste maatriksid on esitatud käesoleva töö lisas. Edaspidi viitame sellele segule nimega „segu A “.

Simuleerime segust A vaatluste vektori realisatsiooni x . Selleks valime juhuslikult ühe mudeli kahest, kusjuures esimese mudeli valimise tõenäosus on π_1 ja teise mudeli valimise tõenäosus π_2 . Seejärel genereerime valituks osutunud mudeli i algjaotust P_i^0 ja üleminekumaatriksit P_i^A kasutades seisundite vektori y . See tähendab, et seisundi y_1 genereerime juhuslikult algjaotuse P_i^0 kohaselt. Iga edasise seisundi y_t realisatsiooni k genereerime eelmisele seisundile y_{t-1} üleminekumaatriksis P_i^A vastavate üleminekutõenäosuste $p_{y_{t-1},k}$ kohaselt.

Vaatluste vektori x genereerimiseks kasutame nüüd emissioonitõenäosuste maatriksit f_i . Vaatluse x_t genereerimiseks valime juhuslikult ühe vaatluse hulgast \mathcal{X} , kusjuures mingi hulga \mathcal{X} elemendi valituks osutumise tõenäosusena kasutame tõenäosusi emissioonitõenäosuste maatriksi f_i sellelt realt, mis vastab eespool genereeritud seisundile y_t .

Kirjeldatud viisil toimides osutus meil valituks mudel 2. Realiseerunud seisundite vektor ja vaatluste vektor on esitatud käesoleva töö lisas.

Parameetri α väärtusena kasutame kõiki arve hulgast $\{0,000; 0,001; 0,002, \dots, 1,000\}$.

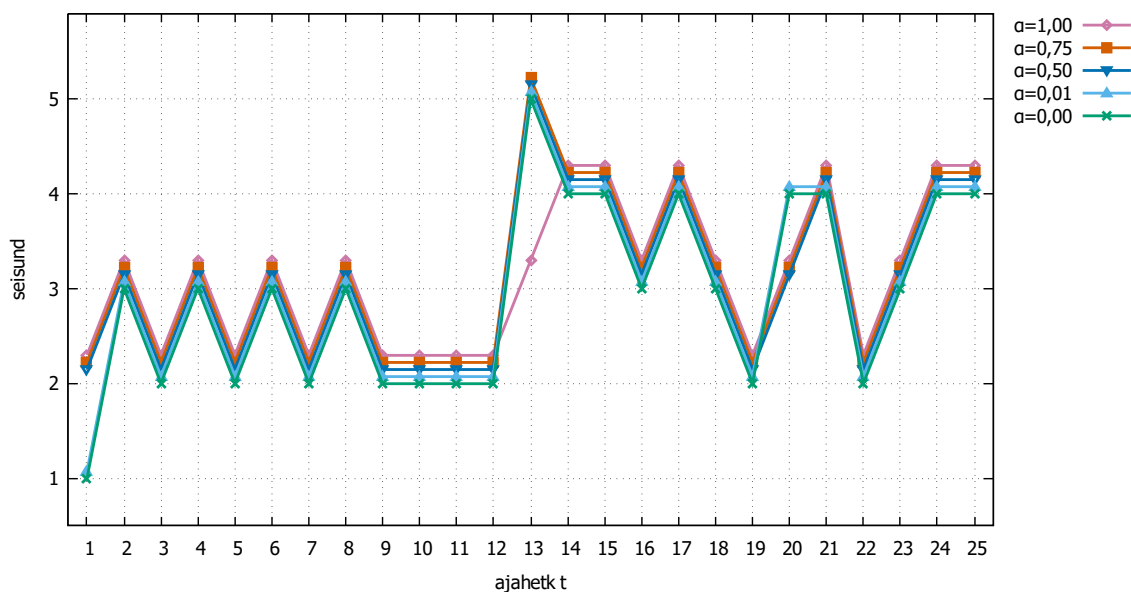
Kasutame edasises algjoonduse $y^{(0)}$ rollis üheksat erinevat joondust ja üht joonduste peret:

- vastavalt ainult ühtedest, kahtedest, kolmedest, neljadest ja viitest koosnevad joondused $y_1^{(0)}$, $y_2^{(0)}$, $y_3^{(0)}$, $y_4^{(0)}$ ja $y_5^{(0)}$;
- joondus $y_6^{(0)} = (3, 2, 0, \dots, 0)$;
- joondus $y_7^{(0)} = (3, 3, 0, \dots, 0)$;
- kummalegi mudelile vastavad Viterbi joondused $y_{v1,A}^{(0)}$ ja $y_{v2,A}^{(0)}$;
- joonduste pere $\{y_\alpha^{(0)}\}$, kus $y_{\alpha_i}^{(0)}$ on hübriid-EM algoritmi väljund algjoonduse $y_{\alpha_{i-1}}^{(0)}$ korral ja kus $y_{\alpha_1}^{(0)}$ leidmiseks on algjoondusena kasutatud PMAP-joondust (loe lähemalt ala-alapeatükis (3.3.1)).

Algjoonduse $y_6^{(0)}$ valimise motiiviks on meie kavatsus segu B esimese mudeli korral seada üleminekutõenäosus $p_{3,2}$ võrdseks nulliga. Segu B teise mudeli üleminekumaatriksis võrdsustame nulliga tõenäosuse $p_{3,3}$, seetõttu on huvitav vaadelda, kuidas käitub algoritm algjoonduse $y_7^{(0)}$ korral. Viterbi joondused saab leida rakendades algoritmi (2) ainult ühest mudelist koosnevale „segule“ ja võttes $\alpha = 1$. Kummalegi mudelile vastavad Viterbi joondused on esitatud käesoleva töö lisas.

Joonisel (4.1) on kujutatud joondusi $v_\alpha(y_1^{(0)})$ viie erineva α väärtuse korral. Samasugused joondused saame ka algjoondusi $y_4^{(0)}, y_5^{(0)}, y_6^{(0)}, y_7^{(0)}, y_{v2,A}^{(0)}$ ja algjoonduste peret $\{y_\alpha^{(0)}\}$ kasutades. Näeme, et väljundjoondused erinevad üksteisest õige vähe. Näeme kaht identsete joonduste paari: $v_{0,5}(y_1^{(0)})$ ja $v_{0,75}(y_1^{(0)})$ ning $v_0(y_1^{(0)})$ ja $v_{0,01}(y_1^{(0)})$.

Joonisel (4.2) on kujutatud joondusi $v_\alpha(y_2^{(0)})$ nende samade viie α väärtuse korral. Samasugused väljundjoondused saame ka algjoondusi $y_3^{(0)}$ ja $y_{v1,A}^{(0)}$ kasutades. Ka siin on parameetri α väärtustele 0 ja 0,01 vastavad väljundjoonduse identsed. Parameetri α väärtustele 0,5 ja 0,75 vastavad joondused on siin aga üksteisest erinevad.

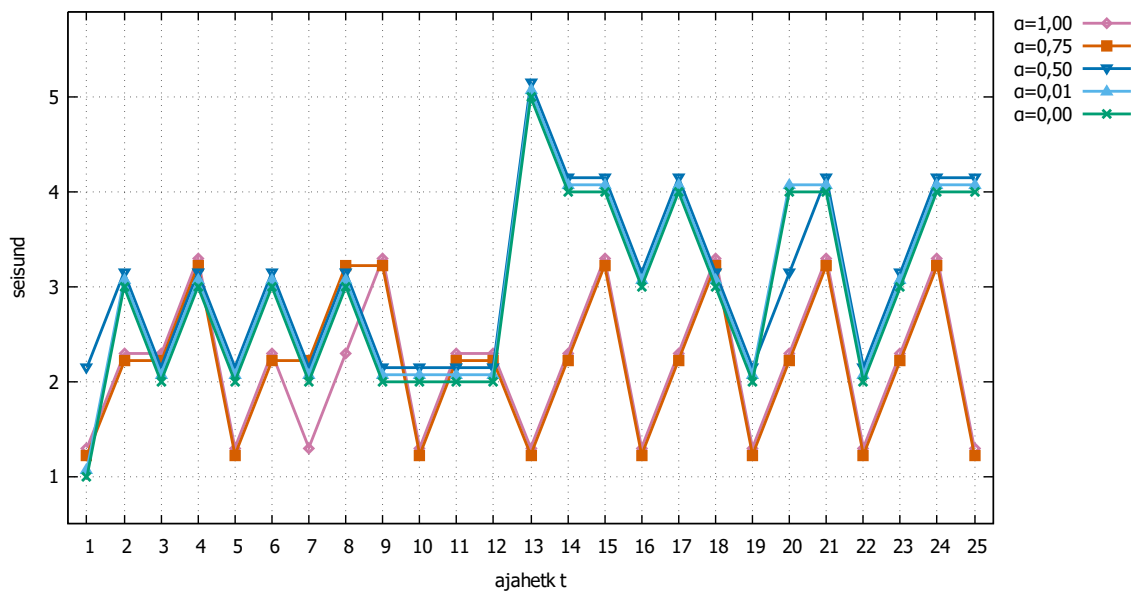


Joonis 4.1: Hübrid-EM algoritmi väljundid segu A korral, kui kasutada algjoondusena joondust $y_1^{(0)}, y_4^{(0)}, y_5^{(0)}, y_6^{(0)}, y_7^{(0)}, y_{v2,A}^{(0)}$ või algjoonduste peret $\{y_\alpha^{(0)}\}$. Väljundjoonduste paremaks eristamiseks on joonduste graafikuid siin ja edaspidi pisut y -telje suunal nihutatud.

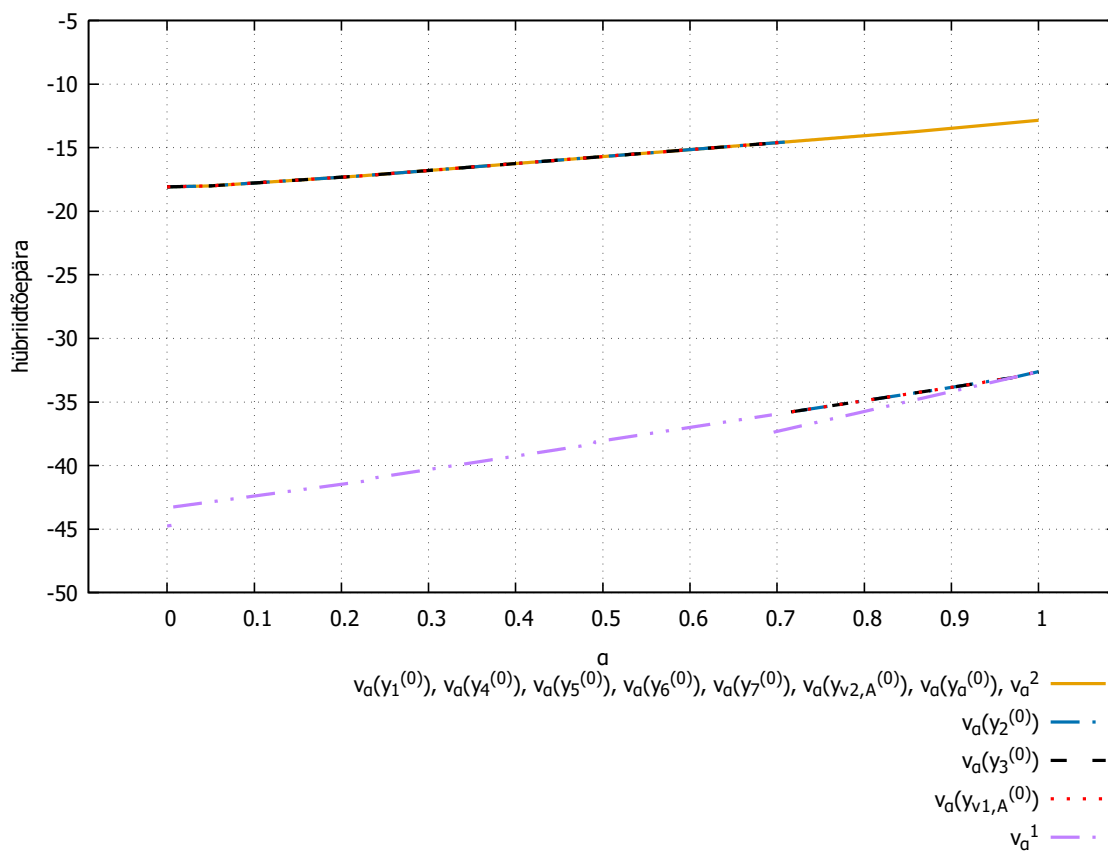
Jooniseid (4.1) ja (4.2) omavahel võrreldes näeme, et joondused $v_\alpha(y_1^{(0)})$ ja $v_\alpha(y_2^{(0)})$ on parameetri α väärtuste 1,00 ja 0,75 korral erinevad. Kuna meie eesmärk on maksimiseerida tõepära (4.1), pakub meile huvi, millist algjoondust või algjoonduste peret kasutades saadud väljundjoondus on mingi α väärtuse korral suurima hübriidtõepäraga. Joonisel (4.3) on kujutatud algoritmi väljundjoonduste hübriidtõepärad erinevate α väärtuste korral. Võrdselt head väljundjoondused saame iga α korral kasutades algjoondusi $y_1^{(0)}, y_4^{(0)}, y_5^{(0)}, y_6^{(0)}, y_7^{(0)}, y_{v2,A}^{(0)}$ ja algjoonduste peret $\{y_\alpha^{(0)}\}$.

Kehvema tulemuse annavad algjoondused $y_2^{(0)}, y_3^{(0)}$ ja $y_{v1,A}^{(0)}$. Nende kolme algjoonduse puhul langeb hübriidtõepära graafik kusagil $\alpha = 0,7$ läheduses märgatavalt madalamale. Parameetri α väärtused, mille korral hübriidtõepära graafik madalamale nivoole kukub, on algjoonduseti pisut erinevad.

Uurime nüüd algjoonduste pere $\{y_\alpha^{(0)}\}$ ja algjoonduse $y_{v1,A}^{(0)}$ näitel, kuidas käituvad väl-



Joonis 4.2: Hübrid-EM algoritmi väljundid segu A korral, kui kasutada algoondusena joondust $y_2^{(0)}$, $y_3^{(0)}$ või $y_{v1,A}^{(0)}$.



Joonis 4.3: Erinevate algoonduste korral leitud väljundjoonduste hübriditõepära. Mitmele väljundjoondusele vastab üks joon, kui need on võrdsed iga arvutamisel kasutatud α korral.

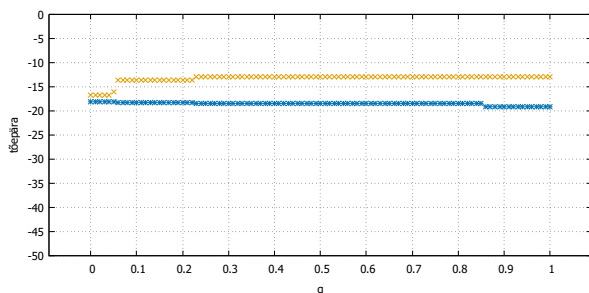
jundjoonduse hübriidtõepära komponendid erinevate α väärtuste korral. Teisisõnu, uurime hübriidtõepära nn Viterbi komponendi

$$\ln p(v|x) \quad (4.2)$$

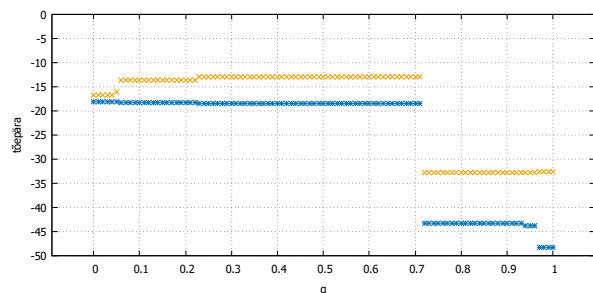
ja nn PMAP komponendi

$$\sum_t \ln p_t(v|x) \quad (4.3)$$

muutumist, kui muutub α . Tulemused on esitatud joonisel (4.4). Väljundjoonduste $v_\alpha(y_\alpha^{(0)})$ hübriidtõepära komponendid käituvad ootuspäraselt: mida suurem on α , seda suurem on Viterbi komponendi väärtus; $(1 - \alpha)$ kasvades kasvab ka PMAP-komponent. Väljundjoonduse $v_\alpha(y_{v1,A}^{(0)})$ hübriidtõepära komponentide väärtused langevad alates α väärtusest 0,716 märgatavalt.



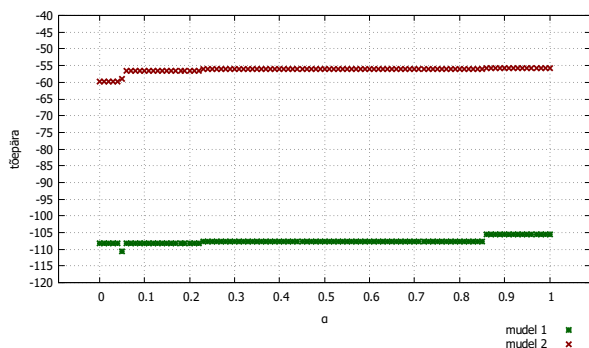
(a) Väljundjoondused $v_\alpha(y_\alpha^{(0)})$.



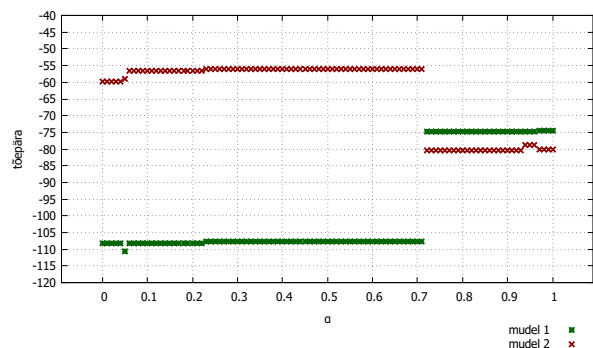
(b) Väljundjoondused $v_\alpha(y_{v1,A}^{(0)})$.

Joonis 4.4: Kahe erineva algjoonduse põhjal leitud väljundjoonduste hübriidtõepära komponendid.

Leiame nende samade algjoonduste puhul, milline on väljundjoonduste tõepära kummagi mudeli suhtes, st leiame erinevate α väärtuste korral

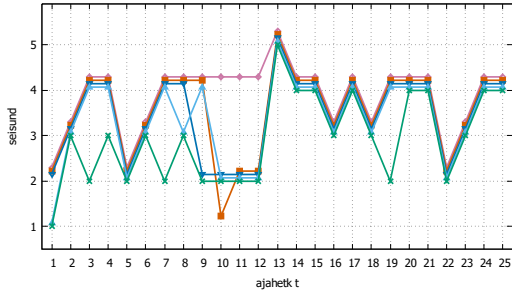


(a) Väljundjoondused $v_\alpha(y_\alpha^{(0)})$.

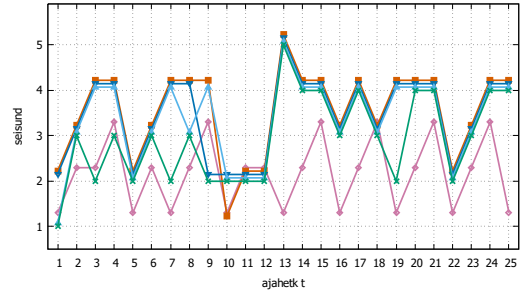


(b) Väljundjoondused $v_\alpha(y_{v1,A}^{(0)})$.

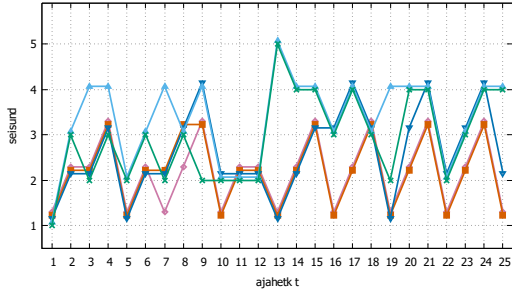
Joonis 4.5: Kahe erineva algjoonduse põhjal leitud väljundjoonduste tõepärad kummagi mudeli suhtes.



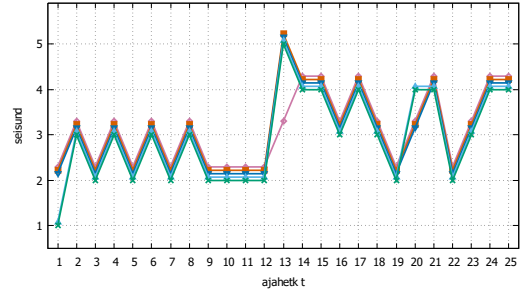
(a) Joondused $v_\alpha(y_1^{(0)})$, $v_\alpha(y_4^{(0)})$ ja $v_\alpha(y_5^{(0)})$.



(b) Joondused $v_\alpha(y_2^{(0)})$ ja $v_\alpha(y_{v1,B}^{(0)})$.



(c) Joondused $v_\alpha(y_3^{(0)})$ ja $v_\alpha(y_7^{(0)})$.



(d) Joondused $v_\alpha(y_6^{(0)})$, $v_\alpha(y_{v2,B}^{(0)})$ ja $v_\alpha(y_\alpha^{(0)})$.

Joonis 4.6: Hübriid-EM algoritmi väljundjoondused erinevate algjoonduste ja α väärtuste korral (segu B).

$$\ln p(v, x|i), \quad i = 1, 2. \quad (4.4)$$

Tulemusi on kujutatud joonisel (4.5). Joonisel (4.5a) kajastub, et joonduste $v_\alpha(y_\alpha^{(0)})$ leidmisel on algoritm (2) arvestanud rohkem mudeliga 2 - väljundjoonduste tõepära on mudeli 2 suhtes märgatavalt suurem, kui mudeli 1 suhtes. See on mudelite kaalusid silmas pidades igati loogiline. Jooniselt (4.5b) näeme, et kui α ületab teatud piiri, pääseb algjoonduse $y_{v1,A}^{(0)}$ puhul enam mõjule mudel 1. Väljundjoonduse tõepära on väiksema kaaluga mudeli suhtes suurem, mis seletab ka, miks joonduste $v_\alpha(y_{v1,A}^{(0)})$ hübriidtõepära suuremate α väärtuste korral mitmete teiste algjoonduste põhjal leitud väljundjoonduste omast madalam on.

Märgime, et algoritm (2) koondub juba mõne iteratsioonisammu järel. Keskmine ja maksimaalne iteratsioonisammude arv erinevate algjoonduste korral on esitatud tabelis (4.1). Maksimaalne iteratsioonisammude arv on 5. See tähendab, et iga algjoonduse ja α väärtuse korral on viiendal iteratsioonisammul leitud joondus samasugune, kui neljandal sammul leitud joondus. Kuna koondumine on nii kiire, on meil algoritmi (2) peatumiskriteeriumi määrava muutuja ε väärtuseks 0. Kahest madalam keskmine iteratsioonisammude arv näitab, et leidub α väärtusi, mille korral algoritm koondub vaid ühe sammuga ehk algoritm väljastab algjoonduse või sellega võrdse hübriidtõepäraga joonduse. Algjoonduste pere $\{y_\alpha^{(0)}\}$ korral koondub algoritm pea alati ühe sammuga.

Muudame nüüd mudelite üleminekutõenäosusi. Kui varem oli esimese mudeli puhul

Algjoendus	$y_1^{(0)}$	$y_2^{(0)}$	$y_3^{(0)}$	$y_4^{(0)}$	$y_5^{(0)}$	$y_6^{(0)}$	$y_7^{(0)}$	$y_{v1,A}^{(0)}$	$y_{v2,A}^{(0)}$	$y_\alpha^{(0)}$
Keskmine	2,96	2,73	2,80	2,00	2,00	2,02	2,96	2,73	1,86	1,00
Maksimum	3	4	5	2	2	3	3	5	2	2

Tabel 4.1: Algoritmi (2) iteratsioonisammude arv erinevate algjoonduste korral (segu A).

üleminekutõenäosus $p_{3,2}$ võrdne 0,01-ga, siis võrdsustame selle nüüd nulliga. Sarnaselt käitume teise mudeli üleminekutõenäosusega $p_{3,3}$. Nii saame üleminekumaatriksid P_1^B ja P_2^B (vt lisa). Võtame need üleminekumaatriksid kasutusele varem kasutatud ülemineku-maatriksite P_1^A ja P_2^A asemel ja jätame ülejäänud segu A juures kasutatud parameetrid samaks. Nii saadud segule viitame edaspidi nimetusega „segu B “.

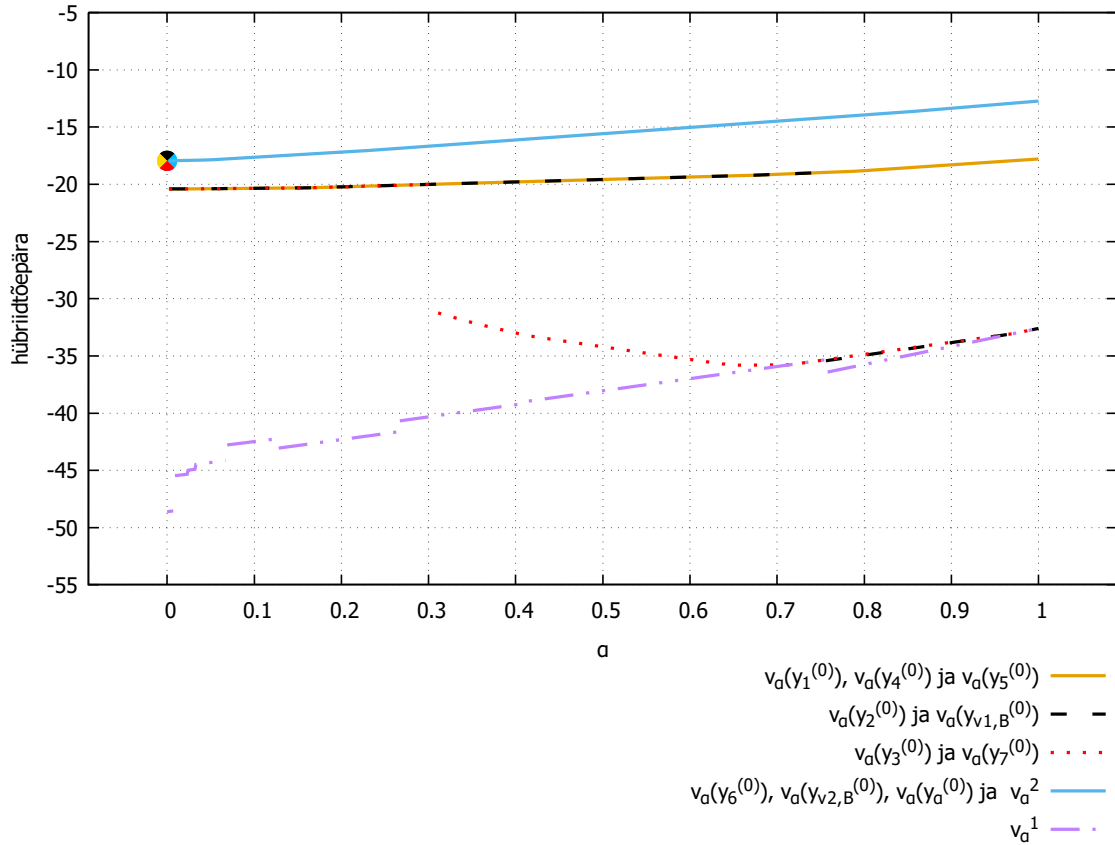
Kasutame sama vaatlustevektorit x , mis segu A korral, et näha, millist efekti omab kahe üleminekutõenäosuse nulliga võrdsustamine. Realisatsiooni x tõepära $\ln p(x)$ on segu B korral pea identne selle tõepäraga segu A korral.

Ka selle segu korral soovime algjoondustena kasutada mudelite Viterbi joondusi. Tähistame neid vastavalt $y_{v1,B}^{(0)}$ ja $y_{v2,B}^{(0)}$. Osutub, et need on identsed segu A mudelite Viterbi joondustega $y_{v1,A}^{(0)}$ ja $y_{v2,A}^{(0)}$.

Joonisel (4.6) on kujutatud erinevate algjoonduste ja α väärtuste korral leitud hübriid-EM algoritmi väljundjoondusi. Meie poolt välja valitud üheksat erinevat algjoondust, üht algjoonduste peret ja viit erinevat α väärtust kasutades saame neli väljundjoonduste komplekti. Algjoondused $y_6^{(0)}$, $y_{v2,B}^{(0)}$ ja algjoonduste pere $\{y_\alpha^{(0)}\}$ annavad nende α väärtuste korral täpselt samasugused väljundjoondused kui hübriidtõepära suhtes kõige paremaid tulemusi andnud algjoondused segu A korral, st joonised (4.6d) ja (4.1) on identsed. Joonisel (4.6a) kujutatud joondused $v_\alpha(y_1^{(0)})$ on iga α väärtuse korral erinevad. Ka joonistel (4.6b) ja (4.6c) ei leidu kaht ühesugust joondust.

Nägime, et joondused võivad erinevate algjoonduste korral olla väga erinevad. Taas huvitab meid, millised algjoondused annavad mingi α väärtuse korral suurima hübriid-tõepäraga väljundjoondusi. Erinevate algjoonduste põhjal leitud väljundjoonduste hübriidtõepärade võrdlus on joonis (4.7). Kui $\alpha = 0$, on kõik väljundjoondused optimaalse hübriidtõepäraga. Nimelt on kõikidele algjoondustele vastavad väljundjoondused $\alpha = 0$ korral võrdse hübriidtõepäraga ja teame, et üks neist - $v_0(y_0^{(0)})$ - on kindlasti PMAP-joondus. Juba järgmise α väärtuse $\alpha = 0,001$ korral jagunevad väljundjoondused kaheks. Kõrgem hübriidtõepära on siis algjoonduste $y_6^{(0)}$ ja $y_{v2,B}^{(0)}$ ning algjoonduste pere $\{y_\alpha^{(0)}\}$ põhjal leitud väljundjoondustel; kõigi teiste väljundjoonduste hübriidtõepära on madalam. Parameetri α kasvades esineb neljale algjoondusele vastavatel väljundjoondusel veel üks kord sellist madalamale nivoole langemist. Väljundjoondused $v_\alpha(y_6^{(0)})$, $v_\alpha(y_{v2,B}^{(0)})$ ja $v_\alpha(y_\alpha^{(0)})$ annavad iga α väärtuse korral parima tulemuse.

Väljundjoonduste hübriidtõepära Viterbi komponendi (4.2) ning PMAP komponendi (4.3) väärtus ja tõepära kummagi mudeli suhtes erinevate α väärtuste korral on kujutatud

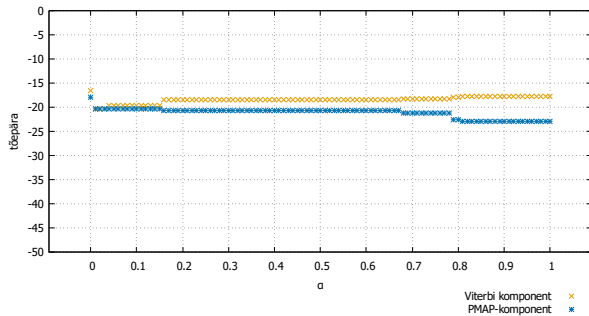


Joonis 4.7: Erinevate algoonduste korral leitud väljundjoonduste hübriiditõepära.

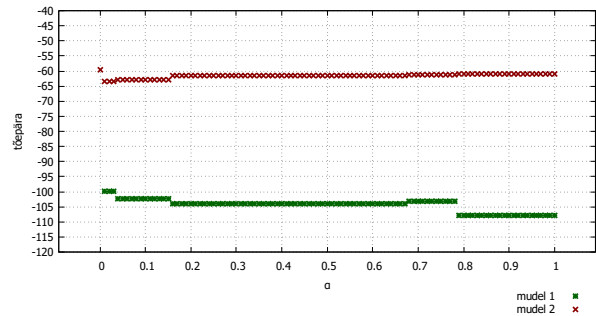
joonisel (4.8). Algoonduste $\{y_\alpha^{(0)}\}$ puhul kehtib iga α korral $\ln p(v_\alpha(y_\alpha^{(0)}), x|i=1) = -\infty$. See tähendab, et iga α korral sisaldab algoondusele $y_\alpha^{(0)}$ vastav väljundjoendus üleminekut seisundist 3 seisundisse 2, mis on mudeli 1 korral keelatud. Et mudeli 2 kaal on mudeli 1 kaalust märksa suurem, on need väljundjoondused hübriiditõepära silmas pidades siiski teistest paremad. Nagu oli näha ka jooniselt (4.7), on algoonduste $y_3^{(0)}$ ja $y_{v1,B}^{(0)}$ põhjal leitud väljundjoondused alates mingist α väärtusest väga kehvade omadustega. Mõlemal juhul arvestatakse ehk liigselt mudeliga 1 (vaata jooniseid (4.8d) ja (4.8f)).

Joonistel (4.8a), (4.8c) kui (4.8e) näeme, et Viterbi komponent on maksimaalne väljundjoonduste puhul, mis on leitud $\alpha = 0$ korral. On selge, et selliselt käituvat algoondust meil kasutada ei maksa. Joonisel (4.8g) näeme heale algoondusele omast pilti. Parameetri α väärtuste kasvades kasvab väljundjoondustele vastav Viterbi komponent monotoonselt, samal ajal, kui PMAP-komponent kahaneb monotoonselt.

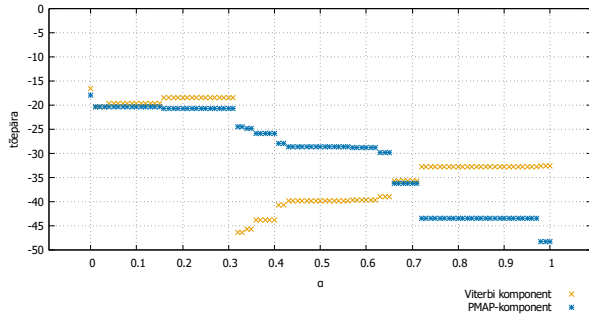
Algoritmi koondumiskiirused segu B korral on esitatud tabelis (4.2). Koondumiskiirused on sarnased segu A korral leitud koondumiskiirustega. Seguga A võrreldes on enim muutunud algoondustele $y_3^{(0)}$ ja $y_7^{(0)}$ vastavad keskmised iteratsioonisammude arvud.



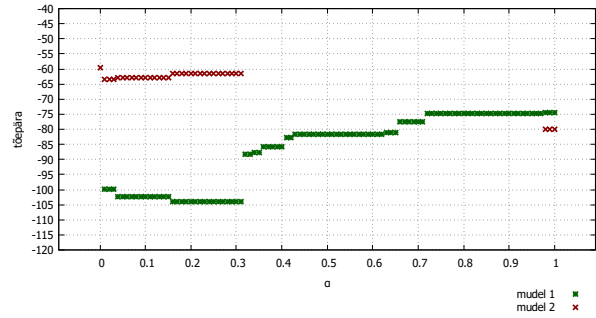
(a) Joondus $v_\alpha(y_1^{(0)})$.



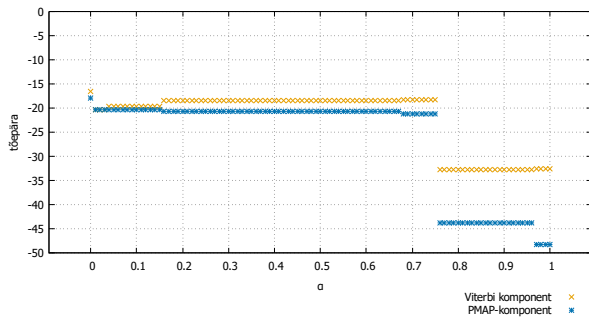
(b) Joondus $v_\alpha(y_1^{(0)})$.



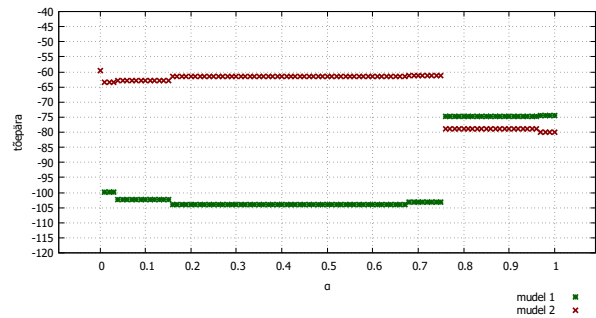
(c) Joondus $v_\alpha(y_3^{(0)})$.



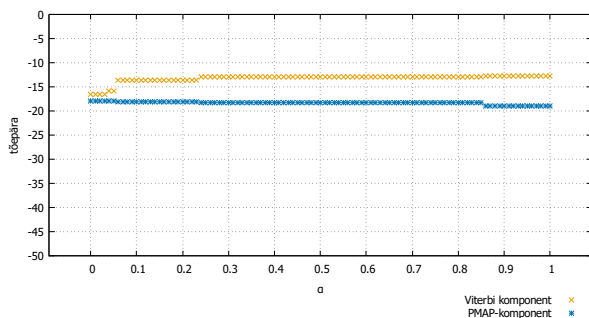
(d) Joondus $v_\alpha(y_3^{(0)})$.



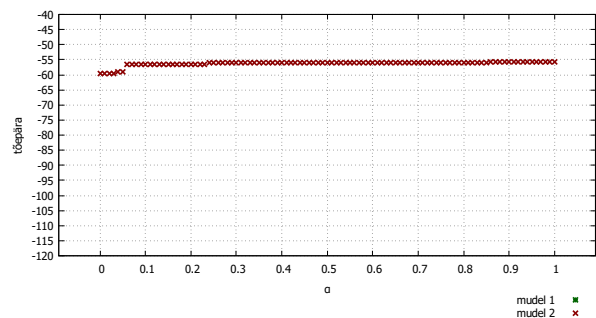
(e) Joondus $v_\alpha(y_{v1,B}^{(0)})$.



(f) Joondus $v_\alpha(y_{v1,B}^{(0)})$.



(g) Joondus $v_\alpha(y_\alpha^{(0)})$.



(h) Joondus $v_\alpha(y_\alpha^{(0)})$.

Joonis 4.8: Erinevate algjoonduste põhjal leitud väljundjoonduste hübriidtoepära komponendid (vasakpoolne tulp) ja toepärad kummagi mudeli suhtes (parempoolne tulp).

4.2 Kahe meetodi võrdlus

HMM-ide segu korral on tüüpiline lahendus segmenteerimisülesandele järgmine. Esmalt leitakse mudel, mille korral vaatlusvektori toepära on maksimaalne. Seejärel rakendatakse

Algjoendus	$y_1^{(0)}$	$y_2^{(0)}$	$y_3^{(0)}$	$y_4^{(0)}$	$y_5^{(0)}$	$y_6^{(0)}$	$y_7^{(0)}$	$y_{v1,A}^{(0)}$	$y_{v2,A}^{(0)}$	$y_\alpha^{(0)}$
Keskmine	2,97	2,78	2,28	2,00	2,00	2,00	2,28	2,79	1,86	1,00
Maksimum	3	5	3	2	2	2	3	5	2	2

Tabel 4.2: Algoritmi (2) iteratsioonisammude arv erinevate algjoonduste korral (segu B).

välja valitud mudelile hübriid-EM algoritmi. Tulemuseks saadakse joendus, mis minimizeerib antud mudeli korral hübriidrisiki $\bar{R}_\alpha(\cdot|x)$. Edaspidi viitame sellele toimumisviisile nimetusega "STP-meetod". Soovime näha, kui heade omadustega on sellisel viisil saadud väljundjoendus võrreldes väljundjoondusega, mille saame hübriid-EM algoritmi otse segule rakendades. Tähistame hübriid-EM algoritmi kasutades saadud väljundjoondust \check{y} ja STP-meetodit kasutades leitud joondust \hat{y} . Et kaht meetodit võrdsetel alustel kõrvutada, arvestame ka STP-meetodi korral mudelite kaalusid. Sellest tulenevalt on STP-meetodil välja valitav mudel fikseeritud vaatluste vektori x korral selline, mis rahuldab võrdust

$$\hat{i} = \arg \max_{i=1,\dots,m} p(x|i)\pi_i. \quad (4.5)$$

Joonistel (4.3) ja (4.7) on muuhulgas kujutatud ka mudeli 1 suhtes optimaalsetele väljundjoondustele v_α^1 ja mudeli 2 suhtes optimaalsetele väljundjoondustele v_α^2 vastav hübriidtoepära. Kuna seal fikseeritud vaatluste vektori x korral otsustades olnuks STP-meetodil osutunud valituks mudel 2, illustreerivad mõlemad joonised olukordi, kus \check{y} ja \hat{y} on võrdselt head. Teisisõnu, ainult mudelit 2 kasutades saadud väljundjoendus v_α^2 on sama suure hübriidtoepäraga, kui parim hübriid-EM algoritmi väljundjoendus.

Võrdleme joondusi \check{y} ja \hat{y} kolme kriteeriumi alusel. Esimeseks kriteeriumiks on joondusele vastav keskmine klassifitseerimisvigade arv. Selleks defineerime punktiviisilist kaofunktsiooni (1.8) kasutades kaofunktsiooni, mis loeb kokku, mitme ajahetke t korral kahe vektori y ja s elemendid üksteisest erinevad.

$$L(y, s) := \sum_{t=1}^T l(y_t, s_t). \quad (4.6)$$

Joondusele \check{y} ja joondusele \hat{y} vastav keskmine vigade arv on fikseeritud vaatluste vektori x korral

$$\sum_{y \in \mathcal{Y}} L(y, v) p(y|x), \quad v \in \{\check{y}, \hat{y}\}. \quad (4.7)$$

Nende suuruste hindamiseks genereerime n_y joondust jaotusest $p(y|x)$. Tähistame neid joondusi y_1, \dots, y_{n_y} . Genereeritud joondusi kasutades leiame hinnangud suurustele (4.7):

$$\frac{1}{n_y} \sum_{j=1}^{n_y} L(y_j, v), \quad v \in \{\check{y}, \hat{y}\}. \quad (4.8)$$

Andmaks lõplikku hinnangut keskmisele vigade arvule, genereerime jaotusest $p(x)$ n_x vektorit, mida tähistame x_1, \dots, x_{n_x} , leiame iga vaatluste vektori korral hinnangud (4.8) ja arvutame hinnangute aritmeetilise keskmise.

Teiseks kriteeriumiks on joondusele vastav tõepära. Selle hindamiseks jätame fikseeritud vaatluste vektori x korral tinglikust jaotusest joondusi genereerides meelde, mitu joondust mingist mudelist genereeriti. Tähistame vastavad arvud n_1, \dots, n_m . Joonduste \check{y} ja \hat{y} tõepära alusel võrdlemisel kasutame fikseeritud vaatluste x korral hinnanguid

$$\sum_{i=1}^m \frac{n_i}{n_y} \ln p(y|x, i), \quad v \in \{\check{y}, \hat{y}\}. \quad (4.9)$$

Lõplike hinnangute leidmiseks arvutame vektoritele x_1, \dots, x_{n_x} vastavate hinnangute (4.9) aritmeetilise keskmise.

Kolmandaks kriteeriumiks on hübriidtõepära. Arvutame vaatluste vektorite x_1, \dots, x_{n_x} korral hübriidtõepärad

$$\alpha \ln p(v|x) + (1 - \alpha) \sum_{t=1}^T \ln p_t(v_t|x), \quad v \in \{\check{y}, \hat{y}\}. \quad (4.10)$$

Lõpliku hinnanguna kasutame vektoritele x_1, \dots, x_{n_x} vastavate hinnangute (4.9) aritmeetilist keskmist.

Tinglikust jaotusest $p(y|x)$ joonduste genereerimiseks paneme tähele, et iga mudeli i , $i = 1, \dots, m$, korral kehtib

$$p(y, i|x) = p(y|x, i)p(i|x), \quad (4.11)$$

kus

$$p(i|x) := \frac{p(x|i)\pi_i}{p(x)}. \quad (4.12)$$

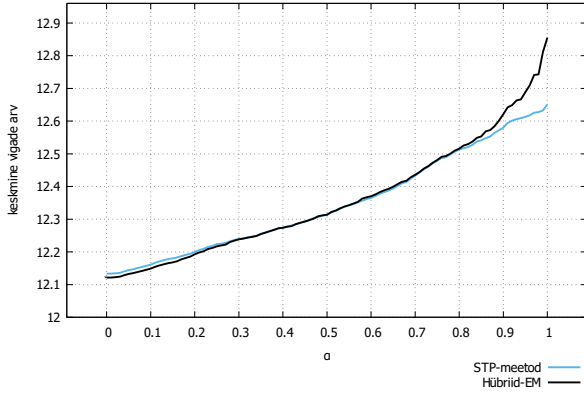
Summeerides üle mudelite, saame

$$p(y|x) = \sum_{i=1}^m p(y|x, i)p(i|x). \quad (4.13)$$

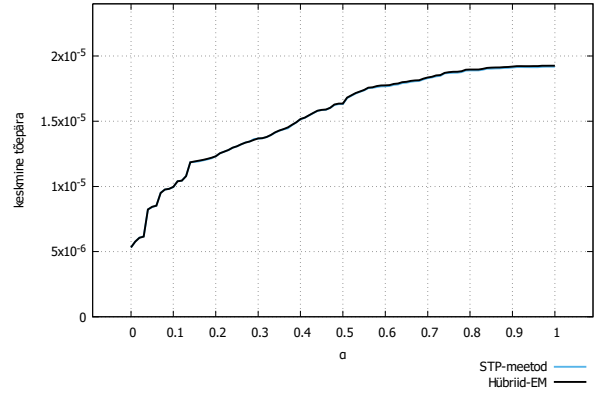
Seega tuleb tinglikust jaotusest $p(y|x)$ joonduse genereerimiseks valida vastavalt tõenäosustele $p(i|x)$ välja üks mudel ning seejärel genereerida joondus sellele mudelile vastavast jaotusest $p(y|x, i)$. Alapeatükis (1.4) nägime, et tinglikud üleminekutõenäosused ja algtõenäosused saab leida järgnevaid arvutuseeskirju kasutades:

$$p(y_{t+1}|y_t, x_{t+1}^T) = \frac{p_{y_t, y_{t+1}} f_{y_{t+1}}(x_{t+1}) \tilde{\beta}(x_{t+2}^T | y_{t+1})}{\tilde{\beta}(x_{t+1}^T | y_t) p(x_{t+1} | x^t)}, \quad (4.14)$$

$$p(y_1 | x^T) = \frac{p_{0, y_1} f_{y_1}(x_1) \tilde{\beta}(x_2^T | y_1)}{p(x_1)}. \quad (4.15)$$



(a) Keskmine vigade arv.



(b) Keskmine tõepära.

Joonis 4.9: Kahe meetodi väljundjoonduste omadused segu A korral.

Eespool nägime, et hübriid-EM algoritm võib erinevate algjoonduste korral anda väga erinevate omadustega väljundjoondusi. Heade väljundjoonduste leidmiseks kasutame alalapeatükis (3.3.1) tutvustatud meetodi pisut täiustatud versiooni.

Parameetri α väärtuse 0 korral kasutame algjoondusena segu PMAP-joondust. Et selle joonduse korral on hübriid-EM algoritmi kindlasti maksimaalne, väljastab hübriid-EM algoritm selle joonduse või mõne selle joondusega võrdse tõepäraga joonduse. Leitud joondust $y_{\alpha_0}^{(0)}$ kasutame ühe algjoondusena järgmise α väärtuse α_1 korral. Teisteks kasutatavateks algjoondusteks on mudelite Viterbi joondused $y_{v_1}^{(0)}, \dots, y_{v_m}^{(0)}$. Väljundjoondustest $v_{\alpha_1}(y_{\alpha_0}^{(0)}), v_{\alpha_1}(y_{v_1}^{(0)}), \dots, v_{\alpha_1}(y_{v_m}^{(0)})$ valime joonduseks \check{y} suurima hübriid-EM tõepäraga joonduse. Selle joonduse võtame järgmise α väärtuse α_2 korral algjoonduse $y_{\alpha_1}^{(0)}$ rolli. Leiame väljundjoondused $v_{\alpha_2}(y_{\alpha_1}^{(0)}), v_{\alpha_2}(y_{v_1}^{(0)}), \dots, v_{\alpha_2}(y_{v_m}^{(0)})$ ning võtame neist parima joonduseks \check{y} . Protsess jätkub analoogselt kuni parameetri α väärtuseni $\alpha = 1$.

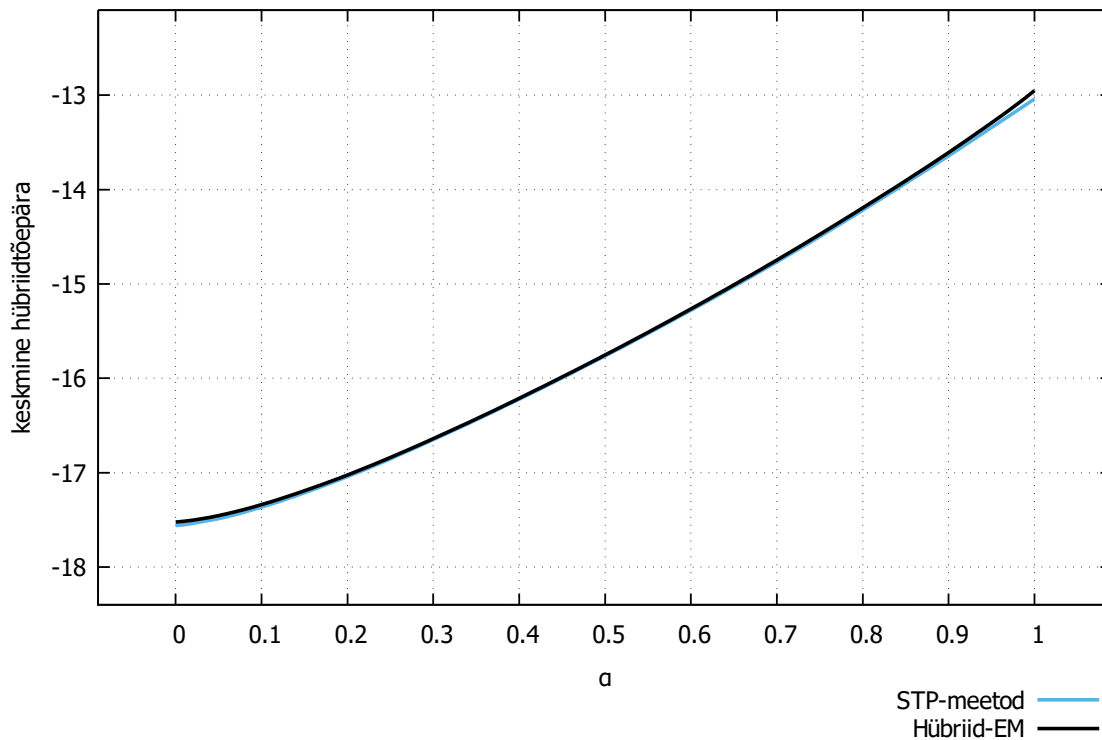
Võtame $n_x = 400$ ja $n_y = 1000$. Kasutame parameetri α väärtusi hulgast

$$\{0,00; 0,01; 0,02; \dots; 1,00\}.$$

Hinnangud keskmisele vigade arvule ja keskmisele tõepärale segu A korral on kujutatud joonisel (4.9). Olgugi, et väikeste α väärtuste korral on hübriid-EM algoritmi väljundjoondustele vastav keskmine vigade arv pisut väiksem ja suurte α väärtuste korral neile vastav keskmine tõepära pisut suurem kui STP-meetodi väljundjoondustel, on kahel meetodil saadud joondused keskmiselt küllaltki sarnaste omadustega. Mudelite üleminekutõenäosused ja emissioonitõenäosused on selle segu korral küllaltki erinevad, mistõttu erinevad tinglikud tõenäosused $p(i|x)$ sageli üksteisest mitme suurusjärgu võrra. Seetõttu genereeritakse fikseeritud vaatluste vektori korral joondused y_1, \dots, y_{n_y} sageli ainult üht mudelit kasutades. STP-meetodi puhul toimub joonduse \hat{y} hindamine sedasama mudelit kasutades. Ka hübriid-EM algoritm teeb valdavalt otsuseid vaid sama mudeliga arvestades. Seetõttu on joondused \check{y} ja \hat{y} paljude vaatlusvektorite x_i , $i = 1, \dots, n_x$, korral

identsed.

Joonisel (4.10) on kujutatud hinnanguid hübriid-EM algoritmi ja STP-meetodi väljundjoenduste keskmisele hübriidtõepärale. Näeme, et väljundjoendused on nende meetodite puhul keskmiselt pea võrdse keskmise hübriidtõepäraga.

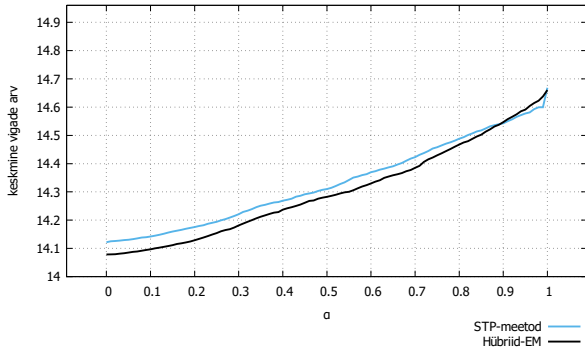


Joonis 4.10: Hübriid-EM algoritmi ja STP-meetodit kasutades leitud joenduste keskmine hübriidtõepära (segu A).

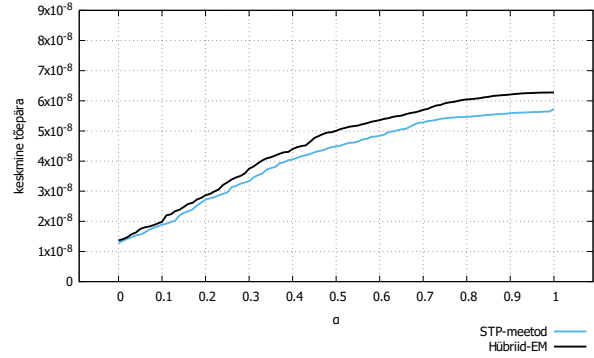
Soovime näha, kuidas käitub kumbki meetod juhul, kui mudel vaatluste vektori põhjal nii lihtsasti tuvastatav pole. Selleks asendame segus A teise mudeli üleminekumaatriksid ja emissioonitõenäosuste maatriksid vastavalt maatriksitega P_3 ja f_3 (vt lisa). Nii saame segus C . Hinnangud keskmisele vigade arvule ja keskmisele tõepärale selle segus korral on esitatud joonisel (4.11). Näeme, et mõlema kriteeriumi järgi otsustades on hübriid-EM algoritm STP-meetodist selgelt parem. Väikeste α väärtuste korral on keskmine väljundjoendustele vastav vigade arv hübriid-EM algoritmi puhul märgatavalt väiksem. Enamgi veel, ka α kasvades on hübriid-EM väljundjoendustele vastav keskmine vigade arv väiksem, kui STP-meetodil leitud väljundjoendustel.

Hübriid-EM algoritmi kasutades õnnestub paremini ka Viterbi joenduse hindamine. Ka suhteliselt väikeste α väärtuste korral leitud väljundjoenduste puhul on hübriid-EM algoritmiga leitud väljundjoenduste tõepära keskmiselt suurem, kui STP-meetodi puhul.

Hinnangud meetoditele vastavatele hübriidtõepäraledele on esitatud joonisel (4.12). Näeme, et iga parameetri α väärtuse korral kehtib



(a) Keskmine vigade arv.

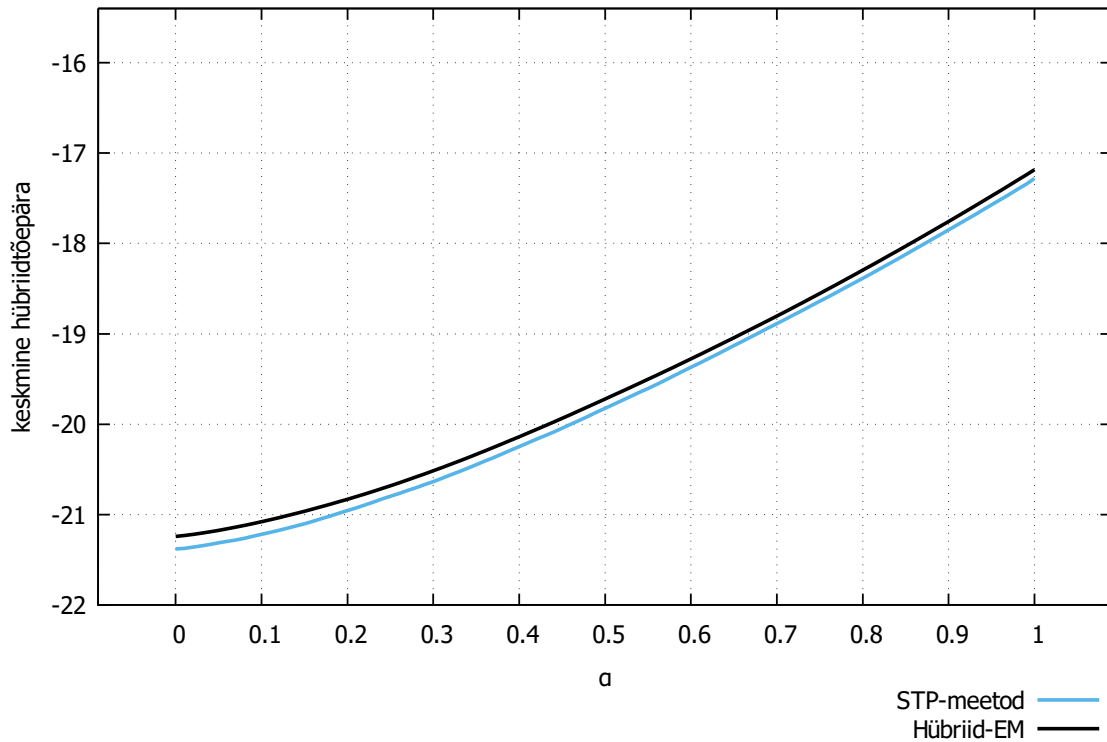


(b) Keskmine tõepära.

Joonis 4.11: Kahe meetodi väljundjoonduste omadused segu C korral.

$$\frac{1}{n_x} \sum_{j=1}^{n_x} \bar{R}_\alpha(\check{y}|x_j) > \frac{1}{n_x} \sum_{j=1}^{n_x} \bar{R}_\alpha(\hat{y}|x_j), \quad (4.16)$$

st hübriid-EM algoritm saavutab kõigi kolme eespool defineeritud kriteeriumi järgi paremad tulemused, kui STP-meetod.



Joonis 4.12: Hübriid-EM algoritmi ja STP-meetodit kasutades leitud joonduste keskmine hübriidtõepära (segu C).

Kirjandus

- [1] Bilmes, J. A. (1998). A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models.
- [2] Kaasik, Ü., Kivistik, L. (1982). *Operatsioonianalüüs*. Kirjastus „Valgus“, Tallinn.
- [3] Lember J., Koloydenko A. A., (2014). Bridging Viterbi and Posterior Decoding: A Risk Approach to Hidden Path Inference Based on Hidden Markov Models. *Journal of Machine Learning Research* 15, lk-d 1-58.
- [4] Lember, J., (2013). Tehisõpe I (loengukonspekt), *MTMS.02.046 Tehisõpe I*. Tartu Ülikool.
- [5] Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, vol 77, nr 2, lk-d 257-286.

LISAD

Simulatsioonides kasutatud parameetrid

Aljaotused ja üleminekumaatriksid:

$$\begin{array}{l}
 P_1^0 = \begin{pmatrix} 0,30 \\ 0,29 \\ 0,01 \\ 0,30 \\ 0,10 \end{pmatrix} \quad P_1^A = \begin{pmatrix} 0,25 & 0,30 & 0,20 & 0,10 & 0,15 \\ 0,30 & 0,20 & 0,40 & 0,05 & 0,05 \\ 0,39 & \mathbf{0,01} & \mathbf{0,20} & 0,30 & 0,10 \\ 0,10 & 0,10 & 0,30 & 0,30 & 0,20 \\ 0,05 & 0,15 & 0,10 & 0,40 & 0,30 \end{pmatrix} \quad P_1^B = \begin{pmatrix} 0,25 & 0,30 & 0,20 & 0,10 & 0,15 \\ 0,30 & 0,20 & 0,40 & 0,05 & 0,05 \\ 0,39 & \mathbf{0,00} & \mathbf{0,21} & 0,30 & 0,10 \\ 0,10 & 0,10 & 0,30 & 0,30 & 0,20 \\ 0,05 & 0,15 & 0,10 & 0,40 & 0,30 \end{pmatrix} \\
 P_2^0 = \begin{pmatrix} 0,30 \\ 0,19 \\ 0,21 \\ 0,25 \\ 0,05 \end{pmatrix} \quad P_2^A = \begin{pmatrix} 0,20 & 0,40 & 0,10 & 0,20 & 0,10 \\ 0,10 & 0,19 & 0,60 & 0,01 & 0,10 \\ 0,10 & \mathbf{0,49} & \mathbf{0,01} & 0,30 & 0,10 \\ 0,10 & 0,10 & 0,30 & 0,34 & 0,16 \\ 0,10 & 0,13 & 0,13 & 0,32 & 0,32 \end{pmatrix} \quad P_2^B = \begin{pmatrix} 0,20 & 0,40 & 0,10 & 0,20 & 0,10 \\ 0,10 & 0,19 & 0,60 & 0,01 & 0,10 \\ 0,10 & \mathbf{0,50} & \mathbf{0,00} & 0,30 & 0,10 \\ 0,10 & 0,10 & 0,30 & 0,34 & 0,16 \\ 0,10 & 0,13 & 0,13 & 0,32 & 0,32 \end{pmatrix} \\
 P_3^0 = P_2^0, \quad P_3 = \begin{pmatrix} 0,25 & 0,30 & 0,20 & 0,10 & 0,15 \\ 0,30 & 0,20 & 0,40 & 0,10 & 0,00 \\ 0,40 & 0,00 & 0,20 & 0,30 & 0,10 \\ 0,10 & 0,10 & 0,30 & 0,30 & 0,20 \\ 0,00 & 0,15 & 0,15 & 0,40 & 0,30 \end{pmatrix}
 \end{array}$$

Emissioonitõenäosused:

$$f_1 = \begin{pmatrix} 0,40 & 0,10 & 0,08 & 0,02 & 0,10 & 0,10 & 0,20 \\ 0,33 & 0,20 & 0,15 & 0,10 & 0,20 & 0,01 & 0,01 \\ 0,01 & 0,20 & 0,01 & 0,20 & 0,08 & 0,30 & 0,20 \\ 0,30 & 0,10 & 0,01 & 0,20 & 0,10 & 0,09 & 0,20 \\ 0,01 & 0,10 & 0,01 & 0,70 & 0,02 & 0,05 & 0,11 \end{pmatrix}$$

$$f_2 = \begin{pmatrix} 0,21 & 0,03 & 0,40 & 0,09 & 0,07 & 0,10 & 0,10 \\ 0,01 & 0,02 & 0,60 & 0,19 & 0,08 & 0,05 & 0,05 \\ 0,19 & 0,20 & 0,01 & 0,06 & 0,30 & 0,20 & 0,04 \\ 0,01 & 0,50 & 0,30 & 0,10 & 0,02 & 0,02 & 0,05 \\ 0,10 & 0,20 & 0,30 & 0,04 & 0,10 & 0,06 & 0,20 \end{pmatrix}$$

$$f_3 = \begin{pmatrix} 0,30 & 0,20 & 0,08 & 0,02 & 0,10 & 0,10 & 0,20 \\ 0,33 & 0,15 & 0,20 & 0,10 & 0,20 & 0,01 & 0,01 \\ 0,01 & 0,20 & 0,01 & 0,20 & 0,08 & 0,30 & 0,20 \\ 0,30 & 0,10 & 0,01 & 0,20 & 0,20 & 0,09 & 0,10 \\ 0,11 & 0,10 & 0,01 & 0,60 & 0,02 & 0,15 & 0,01 \end{pmatrix}$$

$$p_3^0 = P_2^0, \quad P_3 = \begin{pmatrix} 0,25 & 0,30 & 0,20 & 0,10 & 0,15 \\ 0,30 & 0,20 & 0,40 & 0,10 & 0,00 \\ 0,40 & 0,00 & 0,20 & 0,30 & 0,10 \\ 0,10 & 0,10 & 0,30 & 0,30 & 0,20 \\ 0,00 & 0,15 & 0,15 & 0,40 & 0,30 \end{pmatrix}$$

Tegelik seisundite vektor:

[2, 3, 5, 5, 5, 5, 1, 4, 3, 2, 1, 2, 3, 4, 4, 5, 4, 3, 5, 5, 4, 2, 3, 2, 2].

Tegelik vaatluste vektor:

['G', 'E', 'C', 'B', 'E', 'E', 'C', 'B', 'D', 'C', 'C', 'C', 'G', 'C', 'B', 'E', 'B', 'F', 'C', 'B', 'B', 'C', 'E', 'B', 'C'].

Simulatsioonides kasutatud mudelite Viterbi joondused

Olgu vaatluste vektor x^T järgnev:

['G', 'E', 'C', 'B', 'E', 'E', 'C', 'B', 'D', 'C', 'C', 'C', 'G', 'C', 'B', 'E', 'B', 'F', 'C', 'B', 'B', 'C', 'E', 'B', 'C'].

Siis on mudelite Viterbi joondused segu A ja segu B korral samad. Esimese mudeli Viterbi joondus:

[0, 1, 1, 2, 0, 1, 0, 1, 2, 0, 1, 1, 0, 1, 2, 0, 1, 2, 0, 1, 2, 0, 1, 2, 0].

Teise mudeli Viterbi joondus:

[1, 2, 1, 2, 1, 2, 1, 2, 1, 1, 1, 1, 2, 3, 3, 2, 3, 2, 1, 2, 3, 1, 2, 3, 3].

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Jaak Sõnajalg,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Segmenteerimine peidetud Markovi mudelite segude korral“, mille juhendaja on Jüri Lember,
 - 1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace'is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 12.05.2016