



پرديس علوم
دانشكده رياضي، آمار و علوم كامپيوتر

مدل‌های مارکف پنهان و نیمه مارکف پنهان در نرم افزار R و کاربردهای آن

نگارنده

زهرآ دهقانی قبادی

استاد راهنما: دکتر سید مرتضی امینی

پروژه کارشناسی

رشته آمار و کاربردها

دی ماه ۱۳۹۸

چکیده

مدل‌های مارکف پنهان ابزار قدرتمندی برای مدل‌بندی دنباله‌هایی است که توسط یک فرایند گسسته غیر قابل مشاهده تولید شده‌اند. در مدل مارکف پنهان، توزیع زمان انتظار در هر حالت فرایند پنهان توزیع هندسی است که توزیعی فاقد حافظه بوده و در نتیجه باعث کاهش قدرت پیش‌بینی مدل مارکف پنهان می‌شود. برای رفع این مشکل، مدل‌های نیمه مارکف پنهان به عنوان تعمیمی از مدل‌های مارکف پنهان می‌توانند توزیع‌های زمان انتظار متفاوتی داشته باشند. این مدل‌ها در مدل‌سازی سری‌های زمانی، تشخیص صوت، مدل‌بندی دنباله‌های ژنتیکی، مسائل مالی، تحلیل بقا و مانند آن کاربرد دارند. در این پروژه پس از شرح نظری مدل‌های مارکف و نیمه مارکف پنهان و الگوریتم‌های آن به بررسی بسته‌های نرم‌افزاری موجود در نرم افزار R و تحلیل داده‌های واقعی براساس این مدل‌ها می‌پردازیم.

پیشگفتار

تحلیل داده‌های دنباله‌ای امروزه از اهمیت بالایی برخوردار است. سری‌های زمانی، داده‌های زیست‌شناسی و ژنتیکی، صوت و ویدئو و بسیاری دیگر از داده‌ها که امروزه در تحلیل‌ها مورد استفاده قرار می‌گیرند از نوع داده‌های دنباله‌ای هستند. برای تحلیل داده‌های دنباله‌ای دو دسته مدل‌های احتمالی و مدل‌های شبکه عصبی مورد استفاده قرار می‌گیرند. مدل‌های سری زمانی کلاسیک، مدل‌های مارکوف پنهان و نیمه پنهان و شبکه‌های بیزی پویا از جمله پرکاربردترین مدل‌های احتمالی در تحلیل داده‌های دنباله‌ای هستند. از طرف دیگر شبکه‌های عصبی بازگشتی ابزاری برای تحلیل داده‌های دنباله‌ای است. با این حال شبکه‌های عصبی برای برازش به حجم نمونه بالایی نیاز دارند و مدل‌های کلاسیک سری زمانی نیز فرضیات قوی‌ای دارند که ممکن است همواره برقرار نباشد. از این جهت مدل‌های مارکوف پنهان، نیمه پنهان و شبکه‌های بیزی پویا در زمینه‌های مختلف مانند تشخیص صوت، قابلیت اعتماد و مالی کاربرد گسترده‌ای پیدا کرده‌اند.

مدل‌های پنهان مارکوف برای اولین بار در مجموعه مقالات آماری لئونارد باوم و سایر نویسندگان [۱] در نیمه دوم دهه ۱۹۶۰ توضیح داده شد. یکی از اولین کاربردهای HMM تشخیص گفتار بوده که در اواسط دهه ۱۹۷۰ شروع شد. HMM در نیمه دوم ۱۹۸۰ وارد حوزه آنالیز دنباله‌های بیولوژیکی، به‌طور خاص DNA شد [۲]. از آن پس، کاربرد آن در بیوانفورماتیک گسترش یافت، همچنین این مدل‌ها در تشخیص چهره، پیش‌بینی نرخ ارز در روزهای مختلف، پیش‌بینی میزان بارش باران در روزهای متوالی یک سال و مانند آن کاربرد دارد. مدل‌های نیمه مارکوف پنهان نیز برای اولین بار در مقاله‌ای توسط لئونارد باوم و تد پتری در سال ۱۹۹۶ مطرح شد [۳]، این مدل‌ها در احتمال و مدل‌سازی آماری اهمیت فزاینده‌ای دارند و در برنامه‌های کاربردی

مربوط به نظریه صف‌بندی، قابلیت اطمینان و نگهداری، تجزیه و تحلیل بقا، ارزیابی عملکرد، زیست‌شناسی، تجزیه و تحلیل DNA فرآیندهای خطر، بیمه و امور مالی، مدل‌سازی زلزله و مانند آن به کار می‌رود، در فصل اول این پروژه به بررسی مدل‌های مارکف پنهان و در فصل دوم به بررسی مدل نیمه مارکف پنهان و در فصل آخر به تحلیل یک مجموعه داده واقعی بر اساس هر دو مدل پرداخته و به مقایسه نتایج می‌پردازیم.

زمستان ۱۳۹۸

زهره دهقانی قبادی

فهرست مطالب

۱	۱	مدل‌های مارکف پنهان
۱	۱.۱	مروری بر تعاریف اساسی فرایندهای تصادفی
۲	۲.۱	تعریف مدل‌های مارکف پنهان
۴	۳.۱	تابع درستنمایی مدل مارکف پنهان
۵	۴.۱	الگوریتم EM
۷	۱.۴.۱	الگوریتم EM برای مدل مارکف پنهان
۸	۲.۴.۱	الگوریتم پیش‌رو - پس‌رو
۹	۳.۴.۱	الگوریتم پیش‌رو - پس‌رو برای مدل مارکف پنهان
۱۱	۴.۴.۱	مرحله M (ماکزیمم‌سازی) برای توزیع گسسته مشاهده‌ها
۱۲	۵.۴.۱	مرحله M (ماکزیمم‌سازی) برای توزیع‌های پواسون مشاهده‌ها
۱۲	۶.۴.۱	مرحله M برای توزیع مشاهده‌های نرمال چند متغیره
۱۴	۵.۱	الگوریتم ویتربی
۱۵	۱.۵.۱	الگوریتم ویتربی برای مدل‌های مارکف پنهان
۱۹	۲	مدل‌های نیمه مارکف پنهان
۲۰	۱.۲	تعریف مدل‌های نیمه پنهان مارکف
۲۱	۲.۲	تابع درستنمایی مدل نیمه مارکف پنهان

۲۲	الگوریتم EM برای مدل نیمه مارکف پنهان	۳.۲
۲۲	مرحله اول الگوریتم (مرحله E)	۱.۳.۲
۲۳	مرحله دوم الگوریتم (ماکزیمم سازی)	۲.۳.۲
۳۰	مدل های مختلف توزیع زمان انتظار	۴.۲
۳۱	الگوریتم ویتربی برای مدل نیمه مارکف پنهان	۵.۲
۳۵	تحلیل و پیش بینی داده های بورس آمریکا	۳
۳۵	معرفی داده	۱.۳

لیست تصاویر

۳۶	نمودار مربوط به متغیرهای، close open و volume	۱.۳
۳۷	نمودارهای سری‌های ایستاشده	۲.۳
۴۶	نمودار مشاهدات چندگانه مدل مارکف پنهان	۳.۳
۴۷	پیش‌بینی با مدل مارکوف پنهان	۴.۳
۴۹	پیش‌بینی با مدل مارکوف پنهان	۵.۳
۵۳	نمودار دنباله لگاریتم درست‌نمایی مدل نیمه مارکف پنهان با زمان انتظار پواسون	۶.۳
۵۵	پیش‌بینی با مدل نیمه مارکوف پنهان	۷.۳
۵۶	توزیع زمان انتظار ناپارامتری	۸.۳
۵۹	نمودار دنباله لگاریتم درست‌نمایی مدل نیمه مارکف پنهان با زمان انتظار ناپارامتری	۹.۳
۶۱	پیش‌بینی با مدل نیمه مارکوف پنهان	۱۰.۳

فصل ۱

مدل‌های مارکف پنهان

۱.۱ مروری بر تعاریف اساسی فرایندهای تصادفی

ابتدا به تعریف فرایندهای تصادفی و سپس زنجیره‌های مارکف می‌پردازیم.

تعریف ۱.۱.۱ (فرایندهای تصادفی) $\{S_t: t \in T\}$ را یک فرایند تصادفی با مجموعه اندیس گذار T و فضای حالت E می‌گوییم، اگر $A \in E$ یا $S_t \in T$ و $S_t \in A$ باشد می‌گوییم فرایند در زمان t (یا مرحله) t در مجموعه A یا در حالت S قرار دارد، اگر T مجموعه‌ای شماره‌ای باشد فرایند را زمان گسسته می‌نامیم و اگر مجموعه‌هایی به صورت $[0, \infty]$ یا $(-\infty, \infty)$ باشد، آن را زمان پیوسته می‌نامیم، برای هر ω از فضای نمونه مجموعه $\{S_t: t \in T\}$ را که یک زیر مجموعه‌ای از E است، تحقق یا مسیر نمونه‌ای می‌نامیم.

تعریف ۲.۱.۱ (زنجیر) به فرایند تصادفی که فضای حالت آن متناهی یا نامتناهی شماره‌ای باشد، یک زنجیر گفته می‌شود.

تعریف ۳.۱.۱ (زنجیره‌های مارکف و احتمال‌های انتقال یک مرحله‌ای) فرض کنید فرایند تصادفی $\{S_n: n = 0, 1, 2, \dots\}$ با زمان گسسته T و فضای حالت شماره‌ای E باشد. گوییم این فرایند یک زنجیر مارکف است،

اگر به ازای هر $n \geq 1$ و هر S_1, S_2, \dots, S_n از حالت‌ها برابری زیر برقرار باشد

$$P(S_{n+1} = j | S_1 = i_1, \dots, S_{n-1} = i_{n-1}, S_n = i_n) = P(S_{n+1} = j | S_n = i_n)$$

یعنی فقط اطلاع از حالت فرایند در مرحله n ، برای تعیین توزیع فرایند در مرحله $n+1$ کفایت می‌کند و اطلاعات قبل از آن موثر نخواهد بود، احتمال شرطی $P(S_{n+1} = j | S_n = i)$ را احتمال انتقال یک مرحله‌های از حالت i در مرحله n به j در مرحله $n+1$ می‌نامیم. این احتمال فقط به i ، j و n وابسته است، اگر به n وابسته نباشد چنین زنجیر مارکفی را زنجیر مارکف همگن می‌نامیم. در زنجیر مارکف همگن، احتمال‌های انتقال را با P_{ij} نشان می‌دهیم.

قضیه ۴.۱.۱ احتمال‌های انتقال دارای دو ویژگی زیر هستند

$$(الف) \quad P_{ij} \geq 0, \quad \forall i, j$$

$$(ب) \quad \sum_{j \in E} P_{ij} = 1, \quad \forall i$$

تعریف ۵.۱.۱ (توزیع آغازین) توزیع S_1 را توزیع آغازین فرایند می‌نامیم، توزیع آغازین را اغلب به صورت بردار سطری به نام π و مولفه i ام آن $P(S_1 = i)$ است، و با نماد π_i نشان می‌دهیم.

قضیه ۶.۱.۱ فرض کنیم π توزیع آغازین و (P_{xy}) ماتریس احتمال‌های انتقال زنجیر مارکف باشند، در این صورت به ازای هر n و i_1, i_2, \dots, i_n از فضای حالت داریم

$$P(S_n = i_n, S_{n-1} = i_{n-1}, \dots, S_1 = i_1) = \pi_{i_1} P_{i_1 i_2} \cdots P_{i_{n-1} i_n}$$

۲.۱ تعریف مدل‌های مارکف پنهان

در مدل مارکف پنهان، هر حالت متناظر با یک رویداد قابل مشاهده است، به مدل مارکفی که در آن مشاهدات توابع احتمالاتی از حالت‌ها باشند، یعنی یک مدل تصادفی با یک فرایند زیرین (غیر قابل مشاهده) تصادفی

که تنها توسط مجموعه‌ای از فرایندهای تصادفی که دنباله مشاهدات را تولید می‌کنند قابل مشاهده است، مدل مارکف پنهان می‌گویند.

در مدل مارکف پنهان، دنباله حالات سیستم قابل مشاهده نیست، این ویژگی سبب شده است که چنین سیستمی پنهان نامیده شود، ولی با توجه به دنباله مشاهدات، می‌توان آن را حدس زد یا به بیان بهتر احتمال آن را محاسبه کرد.

باید توجه داشت که به ازای هر دنباله مشاهده، تعداد زیادی دنباله حالت موجود است، که می‌تواند همان دنباله مشاهده را با احتمال‌های متفاوت تولید کند.

- در مدل مارکف پنهان علاوه بر حرکت تصادفی بین حالت‌ها، خروجی مشاهده شده هم تصادفی است.
- مدل مارکف پنهان در واقع نوعی مدل مارکف تودرتو است، بدین ترتیب که مدل مارکف اصلی انتقال بین حالت‌ها را نشان می‌دهد و در هر حالت، مشاهده با توجه به یک مدل مارکف وابسته به آن حالت انجام می‌شود.

پارامترهای یک مدل مارکف پنهان، با توزیع گسسته مشاهده‌ها عبارت‌اند از:

۱- ماتریس احتمال انتقال $P = ((P_{ij}))$ مجموعه‌ای از احتمال‌های انتقال بین حالت‌ها است و به صورت

$$P_{ij} = P(S_{t+1} = j | S_t = i), \quad j \leq N, i \geq 1$$

حالت فعلی است.

۲- توزیع احتمال مشاهده‌ها به صورت زیر تعریف می‌شود

$$f_j(O_t) = P(O_t | S_t = j)$$

در واقع $f_j(O_t)$ تابع جرم یا تابع چگالی شرطی O_t به شرط $S_t = j$ است.

۳- توزیع احتمال آغازین به صورت زیر تعریف می‌شود

$$\pi = (\pi_i), \quad \pi_i = P(S_1 = i)$$

برای هر مدل مارکف پنهان فرض های زیر در نظر گرفته می‌شود:

۱- فرض مارکف: خاصیت مارکف در فرایندهای زمان گسسته به صورت

$$P(S_{n+1} = j | S_0 = i_0, \dots, S_{n-1} = i_{n-1}, S_n = i_n) = P(S_{n+1} = j | S_n = i_n)$$

و در فرایندهای زمان پیوسته به صورت

$$P(S_{t+h} = j | S_k = i \quad \forall k \leq t) = P(S_{t+h} = j | S_t = i) \quad \forall h \geq 0$$

تعریف می‌شود. به بیان دیگر فرض می‌شود که حالت بعدی تنها به حالت فعلی وابسته است، فرآیند تصادفی خاصیت مارکف را دارد اگر توزیع احتمال شرطی حالت‌های آینده این روند (مشروط به شرایط گذشته و حال) فقط به وضعیت فعلی بستگی دارد، نه به توالی وقایعی که پیش از آن وجود داشته است. فرآیندی با این خاصیت یک فرآیند مارکف نامیده می‌شود.

۲- (فرض ایستایی) احتمال انتقال در بین حالت‌ها از زمان رخداد انتقال مستقل است، به عبارت دیگر برای

هر $t_1 \neq t_2$ داریم

$$P(S_{t_1+1} = j | S_{t_1} = i) = P(S_{t_2+1} = j | S_{t_2} = i)$$

۳- (فرض استقلال خروجی) دنباله مشاهده‌ها به شرط حالت‌ها از یکدیگر و از حالت‌های زمان‌های دیگر

مستقل هستند، یعنی با فرض اینکه مدل HMM به نام θ داشته باشیم، که به صورت $\theta = (A, B, \Pi)$

باشد، خواهیم داشت

$$f(O | S_1, \dots, S_T) = \prod_{t=1}^T f(O_t | S_t)$$

۳.۱ تابع درستنمایی مدل مارکف پنهان

با توجه به آنچه گفته شد، تابع درستنمایی مدل مارکف پنهان به صورت زیر است

$$P(O_1, \dots, O_N) = \sum_{i_1, \dots, i_T} P(O_1, \dots, O_T, S_1, \dots, S_T) \quad (1.1)$$

که در آن

$$\begin{aligned} P(O_1, \dots, O_T; S_1, \dots, S_T) &= P(S_1) \prod_{t=1}^T P(S_t | S_{t-1}) \prod_{t=0}^T P(O_t | S_t) \\ &= \pi \cdot \prod_{t=1}^T P_{S_t S_{t-1}} \prod_{t=0}^T f_j(O_t) \end{aligned}$$

همان‌طور که از رابطه (۱.۱) می‌توان دید، تابع درستمایی مدل مارکف پنهان شکل پیچیده‌ای داشته و معادلات درستمایی به جواب‌های جبری معینی نمی‌انجامد، پس برای به دست آوردن یک جواب عددی مناسب برای برآورد ماکزیمم درستمایی پارامترها از الگوریتم EM استفاده می‌شود.

۴.۱ الگوریتم EM

در بسیاری از مسائل مربوط به برآورد پارامترهای یک مدل آمیخته، با شکلی از تابع درستمایی مواجه می‌شویم که امکان بیشینه‌سازی آن به روش جبری وجود ندارد. در واقع شکل جمعی تابع درستمایی باعث می‌شود معادله‌های درستمایی جواب جبری نداشته باشند. در چنین مواقعی با در نظر گرفتن متغیر پنهان، مدل تابع درستمایی ساده‌تر شده، در واقع این روش بر اساس وجود متغیر پنهان، امکان برآورد پارامترهای مدل آمیخته را میسر می‌سازد.

الگوریتم EM یک روش بهینه‌سازی برای حل معادلات درستمایی در شرایطی است که تعدادی متغیر پنهان و یا نمونه‌های مشاهده نشده در مدل وجود داشته باشد. اساس کار این الگوریتم عمده‌سازی است، به این معنی که تابع بزرگ‌تر از تابع هدف می‌یابیم و مینیمم کننده آن به صورت تکراری به جواب مینیمم کننده تابع هدف همگرا می‌شود.

فرض کنید $O = (O_1, \dots, O_n)$ متغیرهای مشاهده شده و $S = (S_1, \dots, S_m)$ متغیرهای پنهان باشند. در

این صورت لگاریتم تابع درستمایی داده‌ها به صورت زیر است

$$l(\theta) = \log P(O; \theta) = \log \sum_S P(O, S; \theta) + \log P(S|O; \theta)$$

با امید گرفتن نسبت به $P(S|O; \theta^{(t)})$ از طرفین داریم

$$\begin{aligned} l(\theta) &= \sum_S P(S|O; \theta^{(t)}) \log P(O, S; \theta) - \sum_S P(S|O; \theta) \log P(S|O; \theta) \\ &= Q(\theta|\theta^{(t)}) + H(\theta|\theta^{(t)}) \end{aligned}$$

که در آن $\theta^{(t)}$ یک مقدار θ در تکرار t ام الگوریتم است.

حال فرض کنید $\theta^{(t+1)}$ مقداری باشد که $Q(\theta|\theta^{(t)})$ را ماکزیمم کند، از طرف دیگر داریم

$$\begin{aligned} H(\theta^{(t+1)}|\theta^{(t)}) &= \sum_S P(S|O; \theta^{(t)}) \log \frac{P(S|O; \theta^{(t+1)})}{P(S|O; \theta^{(t)})} \\ &\leq \sum_S P(S|O; \theta^{(t)}) \left(\frac{P(S|O; \theta^{(t+1)})}{P(S|O; \theta^{(t)})} - 1 \right) \\ &= 0 \end{aligned}$$

بنابراین مقدار $\theta^{(t)}$ تابع $l(\theta)$ را بهبود می‌بخشد.

بدین ترتیب الگوریتم EM دارای دو مرحله به صورت زیر است:

۱- (مرحله E) تابع زیر را به دست می‌آوریم

$$Q(\theta|\theta^{(t)}) = \sum_S P(S|O; \theta^{(t)}) \log P(O, S; \theta)$$

۲- (مرحله M) تابع $Q(\theta|\theta^{(t)})$ را نسبت به θ ماکزیمم کرده و $\theta^{(t+1)}$ را به دست می‌آوریم.

این فرایند تا همگرایی دنباله $\theta^{(t+1)}$ ادامه می‌یابد.

۱.۴.۱ الگوریتم EM برای مدل مارکف پنهان

(مرحله E) احتمال توام هر دو حالت متوالی به شرط مشاهده‌ها به صورت زیر به دست می‌آید

$$\begin{aligned}
 \xi_t(i, j) &= P(S_t = i, S_{t+1} = j | O, \theta) \\
 &= \frac{P(O | S_t = i, S_{t+1} = j, \theta) P(S_t = i, S_{t+1} = j | \theta)}{P(O | \theta)} \\
 &= \frac{P(O | S_t = i, S_{t+1} = j, \theta) P(S_{t+1} = j | S_t = i, \theta) P(S_t = i | \theta)}{P(O | \theta)} \\
 &= \frac{1}{P(O | \theta)} P(O_1, \dots, O_t | S_t = i, \theta) P(O_{t+1} | S_{t+1} = j, \theta) \\
 &\quad P(O_{t+2}, \dots, O_T | S_{t+1} = j, \theta) P_{ij} P(S_t = i | \theta) \\
 &= \frac{1}{P(O | \theta)} P(O_1, \dots, O_t, S_t = i, \theta) P(O_{t+1} | S_{t+1} = j, \theta) \\
 &\quad P(O_{t+2}, \dots, O_T | S_{t+1} = j, \theta) P_{ij}
 \end{aligned}$$

بنابراین

$$\xi(i, j) = \frac{\alpha_t(i) f_j(O_{t+1}) \beta_{t+1}(j) P_{ij}}{\sum_k \sum_l \alpha_t(k) P_{kl} f_l(O_{t+1}) \beta_{t+1}(l)}$$

همچنین احتمال حالت i ام به شرط مشاهده‌ها به صورت زیر محاسبه می‌شود

$$\gamma_t(i) = P(S_t = i | O, \theta) = \sum_{j=1}^N \xi_t(i, j)$$

در نهایت داریم

$$\begin{aligned}
\gamma_t(i) &= P(S_t = i | O, \theta) \\
&= \frac{P(O | S_t = i, \theta) P(S_t = i | \theta)}{P(O | \theta)} \\
&= \frac{P(O_1, \dots, O_t | S_t = i, \theta) P(O_{t+1}, \dots, O_T | S_t = i, \theta) P(S_t = i | \theta)}{\sum_{j=1}^N P(O, S_t = i | \theta)} \\
&= \frac{P(O_1, \dots, O_t, S_t = i | \theta) P(O_{t+1}, \dots, O_T | S_t = i, \theta)}{\sum_{j=1}^N P(O | S_t = i, \theta) P(S_t = j)} \\
&= \frac{\alpha_t(i) \beta_t(i)}{\sum_{j=1}^N \alpha_t(j) \beta_t(j)}
\end{aligned}$$

۲.۴.۱ الگوریتم پیش‌رو - پس‌رو

همانطور که دیدیم برای محاسبه احتمال‌های $\xi_t(i, j)$ و $\gamma_t(i)$ به احتمال‌های $\alpha_t(i)$ و $\beta_t(i)$ نیاز داریم. این احتمال‌ها توسط یک سری فرمول‌های بازگشتی و یا پیش‌رونده در الگوریتم تحت عنوان الگوریتم پیش‌رو-پس‌رو محاسبه می‌شوند.

الگوریتم پیش‌رو-پس‌رو یک الگوریتم در حوزه استنباط آماری است که برای مدل پنهان و نیمه پنهان مارکف به کار می‌رود، در این الگوریتم احتمال پسین توزیع حاشیه ای تمام متغیرهای حالت پنهان با توالی مشاهدات $\{O_{1:t} := O_1, \dots, O_t\}$ را محاسبه می‌کند، یعنی برای تمام متغیرهای حالت پنهان $X_k \in \{X_1, \dots, X_t\}$ را محاسبه می‌کند، به این عمل استنباطی، توزیع $P(X_K | O_{1:t})$ «صاف کردن» می‌گویند، این الگوریتم از اصل «برنامه نویسی پویا» برای محاسبه کارآمد مقادیر مورد نیاز برای به دست آوردن احتمال پسین توزیع حاشیه ای در دو مسیر استفاده می‌کند.

اولین مسیر به جلو (پیش) حرکت می‌کند در حالی که مسیر دوم در همان زمان به عقب (پس) می‌رود، از این رو نام پیش‌رو پس‌رو را برای این الگوریتم انتخاب کرده‌اند.

در مسیر اول، الگوریتم مجموعه احتمالاتی را محاسبه می‌کند که برای تمام k ها، احتمال پایان یافتن در یکی از k حالت مشاهده شده اول در توالی است، یعنی $P(X_k|O_{1:t})$.

در مسیر دوم، الگوریتم مجموعه احتمالاتی را محاسبه می‌کند که احتمال مشاهده‌های باقی‌مانده با شروع از نقطه k را به ما می‌دهد، یعنی $P(O_{k+1:t}|X_k)$. این دو مجموعه توزیع‌های احتمالی با ترکیب شدن با هم می‌توانند توزیع هر حالتی در هر زمانی را با داشتن توالی آن به دست آورند

$$P(X_k|O_{1:t}) = P(X_k|O_{1:k}, O_{k+1:t}) \\ \propto P(O_{k+1:t}|X_k)P(O_{1:k}, X_k)$$

در قدم آخر با استفاده از قضیه بیز و استقلال مشروط $O_{1:k}$ و $O_{k+1:t}$ مقادیر X_k . این الگوریتم شامل سه مرحله است :

(۱) محاسبه کردن احتمالات رو به جلو (پسین)

(۲) محاسبه کردن احتمالات رو به عقب (پیشین)

(۳) محاسبه کردن مقادیر «صاف شده»

۳.۴.۱ الگوریتم پیش رو - پس رو برای مدل مارکف پنهان

مراحل الگوریتم پیش رو به شرح زیر است:

گام اول: مقدار دهی اولیه در زمان $t = 0$ به صورت زیر

$$\alpha_0(i) = \pi_i$$

گام دوم: محاسبه مقدار $\alpha_{t+1}(j)$ از روی مقادیر $\alpha_t(i)$ برای تمامی حالات و مشاهدات

$$\alpha_{t+1}(j) = \sum_{i=1}^N \alpha_t(i) P_{ij} f_j(O_{t+1}) \quad (2.1)$$

مراحل الگوریتم پس‌رو:

گام اول: مقداردهی اولیه $\beta_t(i)$ در زمان $t = T$ به صورت زیر

$$\beta_T(i) = 1 \quad i = 1, \dots, N$$

گام دوم: محاسبه مقدار $\beta_t(i)$ از روی مقادیر $\beta_{t+1}(j)$ به صورت زیر

$$\beta_t(i) = \sum_j P_{ij} f_j(O_{t+1}) \beta_{t+1}(j) \quad (3.1)$$

برای اثبات روابط (۲.۱) و (۳.۱) توجه کنید که

$$\begin{aligned} \beta_t(i) &= P(O_{t+1}, \dots, O_T | S_t = i, \theta) \\ &= \sum_j P(O_{t+1}, \dots, O_T, S_{t+1} = j, S_t = i, \theta) \\ &= \sum_j P(O_{t+1}, \dots, O_T | S_{t+1} = j, S_t = i, \theta) P(S_{t+1} = j | S_t = i, \theta) \\ &= \sum_j P(O_{t+1} | S_{t+1} = j, S_t = i, \theta) P(O_{t+2}, \dots, O_T | S_{t+1} = j, S_t = i, \theta) \\ &\quad \times P(S_{t+1} = j | S_t = i, \theta) \\ &= \sum_j P(O_{t+1} | S_{t+1} = j, \theta) P(O_{t+2}, \dots, O_T | S_{t+1} = j, \theta) P(S_{t+1} = j | S_t = i, \theta) \\ &= \sum_{j=1}^N P_{ij} f_j(O_{t+1}) \beta_{t+1}(j) \end{aligned}$$

و

$$\begin{aligned}
\alpha_{t+1}(j) &= P(O_1, \dots, O_{t+1}, S_{t+1} = j | \theta) \\
&= P(O_1, \dots, O_{t+1} | S_{t+1} = j, \theta) P(S_{t+1} = j | \theta) \\
&= P(O_1, \dots, O_t | S_{t+1} = j, \theta) P(O_{t+1} | S_{t+1} = j, \theta) P(S_{t+1} = j | \theta) \\
&= P(O_1, \dots, O_t, S_{t+1} = j | \theta) P(O_{t+1} | S_{t+1} = j, \theta) \\
&= P(O_{t+1} | S_{t+1} = j, \theta) \sum_i P(O_1, \dots, O_t, S_t = i, S_{t+1} = j | \theta) \\
&= P(O_{t+1} | S_{t+1} = j, \theta) \sum_i P(O_1, \dots, O_t, S_{t+1} = j | S_t = i, \theta) P(S_t = i, \theta) \\
&= P(O_{t+1} | S_{t+1} = j, \theta) \sum_i P(O_1, \dots, O_t, S_t = i | \theta) \\
&\quad \times P(S_{t+1} = j, S_t = i, \theta) P(S_t = i | \theta) \\
&= P(O_{t+1} | S_{t+1} = j, \theta) \sum_i P(O_1, \dots, O_t, S_t = i | \theta) P(S_{t+1} = j | S_t = i, \theta) \\
&= P(O_{t+1} | S_{t+1} = j, \theta) \sum_i \alpha_t(i) P(S_{t+1} = j | S_t = i, \theta)
\end{aligned}$$

۴.۴.۱ مرحله M (ماکزیم سازی) برای توزیع گسسته مشاهدهها

اگر مدل مشاهدهها مدل گسسته باشد پارامترهای مدل به صورت زیر به روزسانی می شوند

$$\pi_i^{(k+1)} = \frac{\sum_{k=1}^K \gamma_t^{(k)}(i)}{K}$$

$$P_{ij}^{(k+1)} = \frac{\sum_{k=1}^K \sum_{t=1}^{T_k-1} \xi_t^{(k)}(i, j)}{\sum_{k=1}^K \sum_{t=1}^{T_k-1} \gamma_t^{(k)}(i)}$$

$$f_j(m)^{(k+1)} = \frac{\sum_{k=1}^K \sum_{t=1}^{T_k-1} \gamma_t^{(k)}(i) I(O_t^{(k)} = m)}{\sum_{k=1}^K \sum_{t=1}^{T_k-1} \gamma_t^{(k)}(i)}$$

۵.۴.۱ مرحله M (ماکزیمم‌سازی) برای توزیع‌های پواسون مشاهده‌ها

می‌دانیم فرم کلی تابع Q برای مشاهدات به صورت زیر است

$$\begin{aligned} Q_b(\{f_j(O_t^{T-1})\}|\theta^{(k)}) &= \sum_{t=0}^{T-1} P(S_t = j | O_t^{T-1} = o_t^{T-1}; \theta^{(k)}) \log f_j(O_t) \\ &= \sum_{t=0}^{T-1} \gamma_t^{(k)}(j) \log f_j(O_t) \\ &= \sum_{t=0}^{T-1} \gamma_j^{(k)}(t) [O_t \log \lambda_j - \lambda_j - \log O_t!] \end{aligned}$$

با مشتق گرفتن از Q_b نسبت به λ_j و مساوی صفر قرار دادن داریم

$$\frac{\partial}{\partial \lambda_j} Q_b(\{f_j(O_t^{T-1})\}|\theta^{(k)}) = \sum_{t=0}^{T-1} \gamma_t^{(k)}(j) \left[\frac{O_t}{\lambda_j} - 1 \right] = 0$$

در نتیجه

$$\lambda_j^{(k+1)} = \frac{\sum_{t=0}^{T-1} \gamma_t^{(k)}(j) O_t}{\sum_{t=0}^{T-1} \gamma_t^{(k)}(j)}$$

۶.۴.۱ مرحله M برای توزیع مشاهده‌های نرمال چند متغیره

در این حالت داریم

$$f_j(O_t) = \frac{1}{\sqrt{2\pi}^{\frac{p}{2}} |\Sigma_j|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (O_t - \mu_j)^T \Sigma_j^{-1} (O_t - \mu_j)\right)$$

بنابراین

$$Q_b(\{f_j(O_t^{T-1})\}|\theta^{(k)}) = \sum_{t=0}^{T-1} \gamma_t^{(k)}(j) \left[c - \frac{1}{2} \log(|\Sigma_j|) - \frac{1}{2} (O_t - \mu_j)^T \Sigma_j^{-1} (O_t - \mu_j) \right]$$

که به معادله درستمایی زیر برای μ_j منجر می شود

$$\sum_{t=\cdot}^{T-1} \gamma_t^{(k)}(j) \Sigma_j^{-1} (O_t - \mu_j) = \cdot$$

در نتیجه

$$\mu_j = \frac{\sum_{t=\cdot}^{T-1} \gamma_t^{(k)}(j) O_t}{\sum_{t=\cdot}^{T-1} \gamma_t^{(k)}(j)}$$

از طرفی رابطه زیر نیز برقرار است

$$\begin{aligned} \frac{1}{\Psi} (|\Sigma_j|^{-1}) \sum_{t=\cdot}^{T-1} \gamma_t^{(k)}(j) - \frac{1}{\Psi} \sum_{t=\cdot}^{T-1} \gamma_t^{(k)}(j) \text{tr}(\Sigma_j^{-1} (O_t - \mu_j)(O_t - \mu_j)^T) \\ = \frac{1}{\Psi} \log(|\Sigma_j^{-1}|) \sum_{t=\cdot}^{T-1} \gamma_t^{(k)}(j) - \frac{1}{\Psi} \sum_{t=\cdot}^{T-1} \gamma_t^{(k)}(j) \text{tr}(\Sigma_j^{-1} N_{jt}) \end{aligned}$$

که در آن $N_{jt} := (O_t - \mu_j)(O_t - \mu_j)^T$ داریم

$$\begin{aligned} \frac{1}{\Psi} \sum_{t=\cdot}^{T-1} \gamma_t^{(k)}(j) (\Psi \Sigma_j - \text{diag}(\Sigma_j)) - \frac{1}{\Psi} \sum_{t=\cdot}^{T-1} \gamma_t^{(k)}(j) (\Psi N_{jt} - \text{diag}(N_{jt})) \\ = \frac{1}{\Psi} \sum_{t=\cdot}^{T-1} \gamma_t^{(k)}(j) (\Psi M_{jt} - \text{diag}(M_{jt})) = \Psi S - \text{diag}(S) \end{aligned}$$

که در آن $S := \sum_{t=\cdot}^{T-1} \gamma_t^{(k)}(j) M_{jt}$ و $M_{jt} := \Sigma_j - N_{jt}$ داریم

$$\Psi S - \text{diag}(S) = \cdot \implies S = \cdot \implies \sum_{t=\cdot}^{T-1} \gamma_t^{(k)}(j) (\Sigma_j - N_{jt}) = \cdot$$

در نتیجه

$$\Sigma_j^{(k+1)} = \frac{\sum_{t=\cdot}^{T-1} \gamma_t^{(k)}(j) (O_t - \mu_j^{(k+1)})(O_t - \mu_j^{(k+1)})^T}{\sum_{t=\cdot}^{T-1} \gamma_t^{(k)}(j)}$$

بنابراین برآورد پارامترهای به صورت زیر است

$$\mu_j^{(k+1)} = \frac{\sum_{t=0}^{T-1} \gamma_t^{(k)}(j) O_t}{\sum_{t=0}^{T-1} \gamma_t^{(k)}(j)}$$

$$\Sigma_j^{(k+1)} = \frac{\sum_{t=0}^{T-1} \gamma_t^{(k)}(j) (O_t - \mu_j^{(k+1)}) (O_t - \mu_j^{(k+1)})^T}{\sum_{t=0}^{T-1} \gamma_t^{(k)}(j)}$$

۵.۱ الگوریتم ویتربی

الگوریتم ویتربی، الگوریتمی پویا برای پیدا کردن محتمل‌ترین مسیر از حالت‌های پنهان، با داشتن یک توالی از مشاهدات است. این الگوریتم اغلب در مواردی به کار می‌رود که با داشتن یک مدل پنهان مارکف و توالی‌ای از مشاهدات، می‌خواهیم بدانیم چه توالی‌ای از حالت‌ها این مشاهدات را تولید کرده‌اند. در بسیاری از کاربردهای مدل‌های پنهان مارکف، متغیرهای پنهان تفسیر معناداری دارند و در نتیجه یکی از مهم‌ترین مسائل در این حیطه، پیدا کردن محتمل‌ترین توالی از متغیرهای پنهان با داشتن یک توالی از مشاهدات است. به‌عنوان مثال، در حوزه بازشناسی گفتار، می‌خواهیم یک توالی از واج‌ها با استفاده از توالی‌ای از آواها داشته باشیم.

این مسئله نباید با مسئله پیدا کردن محتمل‌ترین مجموعه از حالت‌های پنهان اشتباه گرفته شود. مسئله دوم می‌تواند با استفاده از الگوریتم پس‌رو-پیش‌رو حل شود. بدین صورت که ابتدا توزیع حاشیه‌ای برای هر متغیر پنهان را به‌دست آورده و سپس جداگانه آن‌ها را بیشینه می‌کنیم. اما در حالت کلی، مسئله پیدا کردن محتمل‌ترین توالی پراهمیت‌تر بوده و الگوریتم بهینه‌ارائه‌شده برای آن، الگوریتم جمع-بیشینه که در حیطه مدل‌های پنهان مارکف، به آن الگوریتم ویتربی گفته می‌شود استفاده کرد. در مسائل طبیعی همیشه واقعیت منطبق بر محتمل‌ترین مسیر نیست. اما در بسیاری از مواقع محتمل‌ترین مسیر نیز اطلاعات خوبی در اختیار می‌گذارد.

در واقع ما با داشتن دنباله مشاهد $O = O_1, \dots, O_T$ و مدل θ می‌خواهیم بهترین دنباله حالات

مانند $Q = S_1 S_2 \dots S_T$ را متناسب با مشاهدات بیابیم.

چند راه حل برای مسئله بالا وجود دارد:

۱- راه حل کامل: بررسی تمام دنباله حالات ممکن و انتخاب بهترین آنها.

۲- برای هر مشاهده، یک حالت که به آن شبیه تر است را انتخاب کنیم.

۳- در هر مرحله حالتی را انتخاب کنیم که احتمال $\alpha_t(i)$ آن بیشتر باشد.

۴- الگوریتم ویتربی: بیشینه کردن $P(O, Q|\theta)$.

راه اول راه حلی بسیار زمان بر است و محاسبات آن از مرتبه $O(T * N^T)$ است، که N تعداد حالات و T طول دنباله مشاهدات است. راه حل دوم و سوم نیز مشکلاتی دارند، به عنوان مثال ممکن است سبب تولید یک دنباله غیر ممکن شوند، دنباله ای که در آن انتقال از یک حالت به حالت دیگر ممکن نیست، یا سبب تولید دنباله بهینه محلی شوند. در ادامه به راه حل ۴ یعنی الگوریتم ویتربی می پردازیم.

۱.۵.۱ الگوریتم ویتربی برای مدل های مارکف پنهان

محتمل ترین دنباله حالتی که از مشاهده اول تا مشاهده t ام را شامل شده و در حالت i ام پایان یافته است را به صورت زیر تعریف می کنیم

$$\delta_t(i) = \max_{S_1, \dots, S_{t-1}} P(S_1 S_2 \dots S_{t-1}, S_t = i, O_1, \dots, O_t | \theta)$$

$$\delta_{t+1}(j) = \max_i \delta_t(i) P_{ij} f_j(O_{t+1})$$

پس به کمک فرمول های بازگشتی بالا $\delta_{t+1}(j)$ را می توان از $\delta_t(i)$ محاسبه کرد.

مراحل الگوریتم ویتربی:

گام (۱) مقدار دهی اولیه $\delta_t(i)$ ها به صورت زیر

$$\delta_1(i) = \pi_i f_i(O_1)$$

$$\Psi_1(i) = \cdot \quad i = 1, \dots, N$$

گام (۲) مرحله بازگشتی

$$\delta_t(j) = \max_{1 \leq i \leq N} \delta_{t-1}(i) P_{ij} f_j(O_t)$$

$$\Psi_t(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) P_{ij}]$$

که عبارت داخل [] حاوی اندیس بهترین حالات است.

گام (۳) محاسبه

$$q_t = \Psi_{t+1}(S_{t+1})$$

استفاده از الگوریتم در عمل: به دلیل ضرب مقادیر احتمالها (که از ۱ کمتر هستند)، با افزایش طول دنباله مشاهده، مقادیر $\delta_t(i)$ به صفر می‌رسند. استفاده از الگوریتم احتمال، محاسبات را ساده تر می‌کند.

قرار می‌دهیم

$$\pi_i^* = \log(\pi_i)$$

$$f_j^*(O_t) = \log(f_j(O_t))$$

$$P_{ij}^* = \log(P_{ij})$$

مراحل الگوریتم ویتربی در دامنه الگوریتم:

گام (۱): مقداردهی اولیه به صورت زیر

$$\delta_1^*(i) = \pi_i^* + f_i^*(O_1) \quad \Psi_1(i) = \cdot$$

گام (۲): مرحله بازگشتی

$$\delta_1^*(i) = \pi_i^*(O_1) \quad \Psi_1(i) = \cdot$$

$$\delta_t^*(j) = \max_{1 \leq i \leq N} [\delta_{t-1}^*(i) + P_{ij}^*] + f_j^*(O_t)$$

$$\Psi_t(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}^*(i) + P_{ij}^*]$$

گام (۳): محاسبه دنباله محتمل ترین حالت ها به صورت زیر

$$q_T^* = \arg \max_{1 \leq i \leq N} [\delta_T^*(i)]$$

فصل ۲

مدل های نیمه مارکف پنهان

همانطور که قبلا گفتیم یکی از مشکلات مدل مارکف پنهان توزیع زمان انتظار آن است که از توزیع هندسی پیروی می کند، در واقع داریم

$$\begin{aligned} & P(S_{t+1} = i, \dots, S_{t+m-1} = i, S_{t+m} \neq i | S_t = i) \\ &= P(S_{t+m} \neq i | S_{t+m-1} = i, \dots, S_{t+1} = i) P(S_{t+m-1} = i | S_{t+m-2} = i, \dots, S_{t+1} = i) \\ & P(S_{t+m-2} = i | S_{t+m-3} = i, \dots, S_{t+1} = i) \dots P(S_{t+1} = i | S_t = i) \\ &= P(S_{t+m} \neq i | S_{t+m-1} = i) P(S_{t+m-1} = i | S_{t+m-2} = i) \dots P(S_{t+1} = i | S_t = i) \\ &= (1 - P_{ii}) P_{ii}^{m-1} \end{aligned}$$

با توجه به فقدان حافظه توزیع هندسی، مدل مارکف پنهان قدرت پیش بینی بسیار پایینی دارد. مدل نیمه مارکف پنهان با جایگزینی توزیع زمان انتظار هر حالت با توزیع های مناسب دیگر این مشکل را برطرف می سازد. مدل نیمه پنهان مارکف یک مدل آماری با همان ساختار مدل پنهان مارکف است، به جز اینکه فرایند نیمه مارکف است نه مارکف، به این معنا که احتمال تغییر وضعیت پنهان، بستگی به مدت زمانی دارد که از زمان ورود به وضعیت فعلی سپری شده است، که این برخلاف مدل های پنهان مارکف است که در آن احتمال تغییر وضعیت با توجه به ماندن در آن حالت تا آن زمان تعیین می شود.

در واقع مدل های نیمه مارکف پنهان تعمیمی از مدل های مارکف پنهان هستند که در آنها توزیع زمان انتظار هندسی نیست و می توانند توزیع های زمان انتظار متفاوتی داشته باشند. یعنی در مدل های نیمه مارکف پنهان توزیع مدت زمان ماندن در هر حالت متفاوت است.

۱.۲ تعریف مدل های نیمه پنهان مارکف

پارامترهای مدل نیمه مارکف پنهان عبارتند از

$$\pi_i := P(S_i = i), \quad \sum_i \pi_i = 1$$

$$P_{ij} := P(S_{t+1} = j | S_{t+1} \neq i, S_t = i), \quad \sum_{j \neq i} P_{ij} = 1, \quad P_{ii} = 0, \quad 1 \leq i, j \leq N$$

و برای $u \in \{1, \dots, M_j\}$

$$d_j(u) := P(S_{t+u+1} \neq j, S_{t+u-\nu} = j, \nu = 0, \dots, u-1 | S_{t+1} = j, S_t \neq j)$$

در عبارت بالا M_j حد بالای زمان گذرانده شده در حالت j است که می تواند بی نهایت باشد. همچنین تعریف می کنیم

$$D_j(u) := \sum_{\nu \geq u} d_j(\nu)$$

اگر فرایند در حالت j در زمان صفر آغاز شده باشد رابطه زیر برقرار است

$$P(S_t \neq j, S_{t-\nu} = j, \nu = 1, \dots, t) = d_j(t) \Pi_j$$

همچنین مانند مدل مارکف پنهان

$$P(O_t = o_t | O_t^{T-1} = o_t^{T-1}, S_t^{T-1} = s_t^{T-1}) = P(O_t = o_t | S_t = s_t)$$

علاوه بر آن فرض می‌کنیم که $u_0 + u_1 + \dots + u_R = T$. همچنین داریم

$$f_j(O_t) = P(O_t = o_t | S_t = j)$$

در صورت گسسته بودن توزیع مشاهده‌ها $\sum_k f_j(k) = 1$ و در صورت پیوسته بودن

$$\int f_j(z) dz = 1 \text{ است.}$$

۲.۲ تابع درستنمایی مدل نیمه مارکف پنهان

ابتدا به تعریف نمادهای زیر می‌پردازیم

$$\{O_{t_1}^{t_1} = o_{t_1}^{t_1}\} := \{O_{t_1} = o_{t_1}, \dots, O_{t_1} = o_{t_1}\}$$

$$\{S_{t_1}^{t_1} = s_{t_1}^{t_1}\} := \{S_{t_1} = s_{t_1}, \dots, S_{t_1} = s_{t_1}\}$$

تابع احتمال توام مشاهده‌ها و حالت‌های پنهان به صورت زیر است

$$\begin{aligned} L(S_{1:T-1}^*, O_{1:T-1}^* | \theta) &= P(S_{1:T-1}^* = s_{1:T-1}^*, O_{1:T-1}^* = o_{1:T-1}^* | \theta) \\ &= P(O_{1:T-1}^* = o_{1:T-1}^* | S_{1:T-1}^* = s_{1:T-1}^*, \theta) P(S_{1:T-1}^* = s_{1:T-1}^* | \theta) \\ &= \prod_{t=1}^{T-1} f_{S_t}(O_t) \pi_{S_t^*} d_{S_t^*}(u_r) \prod_{r=1}^R P_{S_{r-1}^*, S_r^*} d_{S_r^*}(u_r) \end{aligned}$$

که در آن S_r^* برابر با $r + 1$ امین حالت ملاقات، و u_r زمان سپری شده در حالت S_r^* است.

بنابراین تابع درستنمایی بدون در نظر گرفتن حالت‌های پنهان به صورت زیر است

$$L(\theta) = \sum_{S_1, \dots, S_{T-1}} L(S_t, O_t | \theta)$$

۳.۲ الگوریتم EM برای مدل نیمه مارکف پنهان

۱.۳.۲ مرحله اول الگوریتم (مرحله E)

تابع شبه درستنمایی برابر است با

$$Q(\theta|\theta^{(k)}) = E[\log L(S_{:}^{T-1+u}, O_{:}^{T-1}|\theta)|O_{:}^{T-1} = o_{:}^{T-1}, \theta^{(k)}]$$

بنابراین

$$Q(\theta|\theta^{(k)}) = \sum_{S_1, \dots, S_{T-1}} \sum_{u_T^*} [\log L(S_{:}^{T-1+u}, O_{:}^{T-1}|\theta)P(S_{:}^{T-1+u} = s_{:}^{T-1+u}|O_{:}^{T-1} = o_{:}^{T-1}; \theta^{(k)})]$$

$$\begin{aligned} Q(\theta|\theta^{(k)}) &= \sum_{S_1, \dots, S_{T-1}} \sum_{u_T^*} \log \pi_{S^*} P(S_{:}^{T-1+u} = s_{:}^{T-1+u}|O_{:}^{T-1} = o_{:}^{T-1}, \theta^{(k)}) \\ &+ \sum_{S_1, \dots, S_{T-1}} \sum_{u_T^*} \sum_{r=1}^R \log P_{S_{r-1}^*, S_r^*} P(S_{:}^{T-1+u} = s_{:}^{T-1+u}|O_{:}^{T-1} = o_{:}^{T-1}, \theta^{(k)}) \\ &+ \sum_{S_1, \dots, S_{T-1}} \sum_{u_T^*} \sum_{r=1}^R \log d_{S_r^*}(u_r) P(S_{:}^{T-1+u} = s_{:}^{T-1+u}|O_{:}^{T-1} = o_{:}^{T-1}, \theta^{(k)}) \\ &+ \sum_{S_1, \dots, S_{T-1}} \sum_{u_T^*} \sum_{t=\cdot}^{T-1} \log f_{S_t}(O_t) P(S_{:}^{T-1+u} = s_{:}^{T-1+u}|O_{:}^{T-1} = o_{:}^{T-1}, \theta^{(k)}) \\ &= Q(\pi|\theta^{(k)}) + Q(P|\theta^{(k)}) + Q(d|\theta^{(k)}) + Q(f|\theta^{(k)}) \end{aligned}$$

که در آن

$$\begin{aligned} Q(\pi|\theta^{(k)}) &= \sum_{S_1, \dots, S_{T-1}} \sum_{u_T^*} [\log L(S_{:}^{T-1+u}, O_{:}^{T-1}|\theta)P(S_{:}^{T-1+u} = s_{:}^{T-1+u}|O_{:}^{T-1} = o_{:}^{T-1}|\theta^{(k)})] \\ &= \sum_{j=\cdot}^{J-1} P(S_{:} = j|O_{:}^{T-1} = o_{:}^{T-1}, \theta^{(k)}) \log \Pi_j \end{aligned}$$

$$\begin{aligned}
Q(P|\theta^{(k)}) &= \sum_{S_1, \dots, S_{T-1}} \sum_{u_T^*} \sum_{r=1}^R \log P_{S_{r-1}^*, S_r^*} P(S_{T-1+u}^{T-1+u} = s_{T-1+u}^{T-1+u} | O_{T-1}^{T-1} = o_{T-1}^{T-1}, \theta^{(k)}) \\
&= \sum_{i=0}^{J-1} \sum_{j \neq i} \sum_{t=2}^{T-1} P(S_{t+1} = j, S_t = i | O_{T-1}^{T-1} = o_{T-1}^{T-1}, \theta^{(k)}) \log P_{ij} \\
& \quad t \in \{0, \dots, T-2\}, j = 0, \dots, J-1
\end{aligned}$$

$$\begin{aligned}
Q(d|\theta^{(k)}) &= \sum_{S_1, \dots, S_{T-1}} \sum_{u_T^*} \sum_{r=1}^R \log d_{S_r^*}(u_r) P(S_{T-1+u}^{T-1+u} = s_{T-1+u}^{T-1+u} | O_{T-1}^{T-1} = o_{T-1}^{T-1}, \theta^{(k)}) \\
&= \sum_u \sum_{t=2}^{T-2} P(S_{t+u+1} \neq j, S_{t+u-v} = j, v = \{0, \dots, u-1\}, S_t \neq j | O_{T-1}^{T-1} = o_{T-1}^{T-1}, \theta^{(k)}) \\
& \quad + P(S_u \neq j, S_{u-v} = j, v = \{1, \dots, u\} | O_{T-1}^{T-1} = o_{T-1}^{T-1}, \theta^{(k)}) \log d_j(u)
\end{aligned}$$

و

$$\begin{aligned}
Q(f|\theta^{(k)}) &= \sum_{S_1, \dots, S_{T-1}} \sum_{u_T^*} \sum_{t=0}^{T-1} \log f_{S_t}(O_t) P(S_{T-1+u}^{T-1+u} = s_{T-1+u}^{T-1+u} | O_{T-1}^{T-1} = o_{T-1}^{T-1}, \theta^{(k)}) \\
&= \sum_{j=0}^{J-1} \sum_{t=0}^{T-1} P(S_t = j | O_{T-1}^{T-1} = o_{T-1}^{T-1}, \theta^{(k)}) \log f_j(O_t)
\end{aligned}$$

۲.۳.۲ مرحله دوم الگوریتم (ماکزیمم سازی)

این الگوریتم برای محاسبه احتمال زیر به ازای هر حالت j و $t = 0, \dots, T-1$ به کار می‌رود

$$F_j(t) = P(S_{t+1} \neq j, S_t = j | O^t = o^t)$$

همچنین

$$\begin{aligned}
N_t &:= P(O_t = o_t | O^{t-1} = o^{t-1}) \\
&= \sum_j P(S_t = j, O_t = o_t | O^{T-1} = o^{T-1}) \\
&= \sum_j f_j(O_t) \sum_{u=1}^t \prod_{\nu=1}^{u-1} \frac{f_j(O_{t-\nu})}{N_{t-\nu}} D_j(u) \sum_{i \neq j} P_{ij} F_i(t-u) \\
&\quad + \prod_{\nu=1}^t \frac{f_j(O_{t-\nu})}{N_{t-\nu}} D_j(t+1) \pi_j
\end{aligned}$$

داریم

$$\begin{aligned}
F_j(\cdot) &= P(S_1 \neq j, S_{\cdot} = j | O_{\cdot} = o_{\cdot}) \\
&= \pi_j d_j(1)
\end{aligned}$$

$$\begin{aligned}
F_j(t) &= P(S_{t+1} \neq j, S_t = j | O^t = o^t) \\
&= \sum_{u=1}^t \sum_{i \neq j} P(S_{t+1} \neq j, S_{t-u} = j, \nu = \{\cdot, \dots, u-1\}, S_{t-u} = i | O^t = o^t) \\
&\quad + P(S_{t+1} \neq j, S_{t-\nu} = j, \nu = \{\cdot, \dots, t\} | O^t = o^t) \\
&= \sum_{u=1}^t \frac{P(O_{t-u+1}^t = o_{t-u+1}^t | S_{t-\nu} = j, \nu = \{\cdot, \dots, u-1\})}{P(O_{t-u+1}^t = o_{t-u+1}^t | O^{t-u} = o^{t-u})} \\
&\quad \times P(S_{t+1} \neq j, S_{t-\nu} = j, \nu = \{\cdot, \dots, u-2\} | S_{t-u+1} = j, S_{t-u} \neq j) \\
&\quad \times \sum_{i \neq j} P(S_{t-u+1} = j | S_{t-u+1} \neq i, S_{t-u} = i) \\
&\quad \times P(S_{t-u+1} \neq i, S_{t-u} = i | O^{t-u} = o^{t-u}) \\
&\quad + \frac{P(O^t = o^t | S_{t-\nu} = j, \nu = \{\cdot, \dots, t\}) P(S_{t+1} \neq j, S_{t-\nu} = j, \nu = \{\cdot, \dots, t\})}{P(O^t = o^t)}
\end{aligned}$$

$$\begin{aligned}
&= \frac{f_j(O_t)}{N_t} \sum_{u=1}^t \prod_{\nu=1}^{u-1} \frac{f_j(O_{t-\nu})}{N_{t-\nu}} d_j(\nu) \sum_{i \neq j} P_{ij} F_i(t-\nu) \\
&+ \prod_{\nu=1}^t \frac{f_j(O_{t-\nu})}{N_{t-\nu}} d_j(t+1) \pi_j
\end{aligned}$$

رابطه بالا برای $t \in \{0, \dots, T-2\}$ و $j \in \{0, \dots, T-2\}$ برقرار است، حال $t = T-1$ داریم

$$\begin{aligned}
F_j(T-1) &= P(S_{T-1} = j | O_{1:T-1} = o_{1:T-1}) \\
&= \frac{f_j(O_{T-1})}{N_{T-1}} \sum_{u=1}^{T-1} \prod_{\nu=1}^{u-1} \frac{f_j(O_{T-1-\nu})}{N_{T-1-\nu}} D_j(u) \sum_{i \neq j} P_{ij} F_i(T-1-u) \\
&+ \prod_{\nu=1}^{T-1} \frac{f_j(O_{T-1-\nu})}{N_{T-1-\nu}} D_j(T) \pi_j
\end{aligned}$$

الگوریتم پس رو برای محاسبه احتمال زیر به ازای هر حالت j و $t = 0, \dots, T-1$ به کار می رود

$$\gamma_t^{(k)}(j) = P(S_t = j | O_{1:T-1} = o_{1:T-1})$$

با مقدار اولیه $t = T-1$ شروع می کنیم

$$\gamma_{T-1}^{(k)}(j) = P(S_{T-1} = j | O_{1:T-1} = o_{1:T-1}) = F_j(T-1)$$

به ازای هر حالت j و $t = 0, \dots, T-2$ داریم

$$\begin{aligned}
\gamma_t^{(k)}(j) &= P(S_t = j | O_{1:T-1} = o_{1:T-1}) \\
&= P(S_{t+1} \neq j, S_t = j | O_{1:T-1} = o_{1:T-1}) + P(S_{t+1} = j | O_{1:T-1} = o_{1:T-1}) \\
&- P(S_{t+1} = j, S_t \neq j | O_{1:T-1} = o_{1:T-1}) \\
&= L_{1j}(t) + \gamma_{t+1}^{(k)}(j) - P(S_{t+1} = j, S_t \neq j | O_{1:T-1} = o_{1:T-1})
\end{aligned}$$

که $L_{1j}(t)$ به صورت زیر محاسبه می گردد

$$\begin{aligned}
L_{\setminus j}(t) &= P(S_{t+\setminus} \neq j, S_t = j | O_{\cdot}^{T-\setminus} = o_{\cdot}^{T-\setminus}) \\
&= \sum_{k \neq j} \sum_{u=\setminus}^{T-\setminus-t} P(S_{t+u+\setminus} \neq k, S_{t+u-\nu} = k, \nu = \{\cdot, \dots, u - \setminus\}, S_t = j | O_{\cdot}^{T-\setminus} = o_{\cdot}^{T-\setminus}) \\
&+ P(S_{T-\setminus-\nu} = k, \nu = \{\cdot, \dots, T - \setminus - t\}, S_t = j | O_{\cdot}^{T-\setminus} = o_{\cdot}^{T-\setminus})
\end{aligned}$$

همچنین بخشی از رابطه بالا به صورت زیر محاسبه می شود

$$\begin{aligned}
&P(S_{t+u+\setminus} \neq k, S_{t+u-\nu} = k, \nu = \{\cdot, \dots, u - \setminus\}, S_t = j | O_{\cdot}^{T-\setminus} = o_{\cdot}^{T-\setminus}) \\
&= \frac{P(S_{t+u+\setminus} \neq k, S_{t+u-\nu} = k, \nu = \{\cdot, \dots, u - \setminus\}, S_t = j, O_{\cdot}^{T-\setminus} = o_{\cdot}^{T-\setminus})}{P(S_{t+u+\setminus} \neq k, S_{t+u} = k, O_{\cdot}^{T-\setminus} = o_{\cdot}^{T-\setminus})} \\
&\times P(S_{t+u+\setminus} \neq k, S_{t+u} = k, O_{\cdot}^{T-\setminus} = o_{\cdot}^{T-\setminus}) \\
&= \frac{P(O_{t+u+\setminus}^{T-\setminus} = o_{t+u+\setminus}^{T-\setminus} | S_{t+u+\setminus} \neq k, S_{t+u} = k)}{P(O_{t+u+\setminus}^{T-\setminus} = o_{t+u+\setminus}^{T-\setminus} | S_{t+u+\setminus} \neq k, S_{t+u} = k)} \\
&\times \frac{P(S_{t+u+\setminus} \neq k, S_{t+u} = k | O_{\cdot}^{T-\setminus} = o_{\cdot}^{T-\setminus})}{P(S_{t+u+\setminus} \neq k, S_{t+u} = k | O_{\cdot}^{t+u} = o_{\cdot}^{t+u})} \\
&\times \frac{P(O_{t+\setminus}^{t+u} = o_{t+\setminus}^{t+u} | S_{t+u-\nu} = k, \nu = \{\cdot, \dots, u - \setminus\})}{P(O_{t+\setminus}^{t+u} = o_{t+\setminus}^{t+u} | O_{\cdot}^t = o_{\cdot}^t)} \\
&\times P(S_{t+u+\setminus} \neq k, S_{t+u-\nu} = k, \nu = \{\cdot, \dots, u - \setminus\} | S_{t+\setminus} = k, S_t \neq k) \\
&\times P(S_{t+\setminus} = k | S_{t+\setminus} \neq j, S_t = j) P(S_{t+\setminus} \neq j, S_t = j | O_{\cdot}^t = o_{\cdot}^t) \\
&= \frac{L_{\setminus k}(t+u)}{F_k(t+u)} \prod_{\nu=\cdot}^{u-\setminus} \frac{f_k(O_{t+u-\nu})}{N_{t+u-\nu}} d_k(u) P_{jk} F_j(t)
\end{aligned}$$

از آنجا که

$$D_j(u) = \sum_{v \geq u} d_j(v)$$

رابطه زیر به دست می‌آید

$$L_{\backslash j}(t) = \sum_{k \neq j} \sum_{u=1}^{T-\backslash-t} \frac{L_{\backslash k}(t+u)}{F_k(t+u)} \prod_{\nu=\cdot}^{u-1} \frac{b_k(O_{t+u-\nu})}{N_{t+u-\nu}} d_k(u) \\ + \prod_{u=\cdot}^{T-\backslash-t} \frac{f_k(O_{T-\backslash-\nu})}{N_{T-\backslash-\nu}} D_k(T-\backslash-t) P_{jk} F_j(t)$$

بنابراین احتمال مورد نظر برابر با عبارت زیر است

$$L_{\backslash j}(t) = P(S_{t+\backslash} = j, S_t \neq j | O_{\cdot}^{T-\backslash} = o_{\cdot}^{T-\backslash}) \\ = \sum_{u=1}^{T-\backslash-t} \sum_{i \neq j} P(S_{t+u+\backslash} \neq j, S_{t+u-\nu} = j, v = \{\cdot, \dots, u-1\}, S_t = i | O_{\cdot}^{T-\backslash}) \\ + \sum_{i \neq j} P(S_{T-\backslash-\nu} = j, v = \{\cdot, \dots, T-\backslash\}, S_t = i | O_{\cdot}^{T-\backslash} = o_{\cdot}^{T-\backslash}) \\ = \sum_{u=\backslash}^{T-\backslash-t} \frac{L_{\backslash j}(t+u)}{F_j(t+u)} \prod_{\nu=\cdot}^{u-1} \frac{f_j(O_{t+u-\nu})}{N_{t+u-\nu}} d_j(u) \\ + \prod_{\nu=\cdot}^{T-\backslash-t} \frac{f_j(O_{T-\backslash-\nu})}{N_{T-\backslash-\nu}} D_j(T-\backslash-t) \sum_{i \neq j} P_{ij} F_i(t)$$

در نتیجه

$$P(S_{t+\backslash} = j, S_t \neq j | O_{\cdot}^{T-\backslash} = o_{\cdot}^{T-\backslash}) \\ = \sum_{u=1}^{T-\backslash-t} \frac{L_{\backslash j}(t+u)}{F_j(t+u)} \sum_{v=\cdot}^{u-1} \frac{f_j(O_{t+u-\nu})}{N_{t+u-\nu}} d_j(u) \\ + \prod_{\nu=\cdot}^{T-\backslash-t} \frac{f_j(O_{T-\backslash-\nu})}{N_{T-\backslash-\nu}} D_j(T-\backslash-t) \sum_{i \neq j} P_{ij} F_i(t)$$

$$\pi_j^{(k+1)} = P(S_{\cdot} = j | O_{\cdot}^{T-\backslash} = o_{\cdot}^{T-\backslash}; \theta^{(k)})$$

$$P_{ij}^k = \frac{\sum_{t=\cdot}^{T-\backslash} P(S_{t+\backslash} = j, S_t = i | O_{\cdot}^{T-\backslash} = o_{\cdot}^{T-\backslash}; \theta^{(k)})}{\sum_{t=\cdot}^{T-\backslash} P(S_{t+\backslash} \neq i, S_t = i | O_{\cdot}^{T-\backslash} = o_{\cdot}^{T-\backslash}; \theta^{(k)})}$$

از طرفی برای هر حالت غیرجاذب j داریم

$$\begin{aligned} Q_d(\{d_j\}|\theta^{(k)}) &= \sum_u \sum_{t=\cdot}^{T-2} P(S_{t+u-1} \neq j, S_{t+u-2} = j, \nu = \{\cdot, \dots, u-1\} | S_t \neq j | O_{\cdot}^{T-1}; \theta^{(k)}) \\ &+ P(S_u \neq j, S_{u-2} = j, \nu = \{1, \dots, u\} | O_{\cdot}^{T-1} = o_{\cdot}^{T-1}; \theta^{(k)}) \log d_j(u) \\ &= \sum_u \eta_{j,u}^k \log d_j(u) \end{aligned}$$

برای $u \leq T - 2 - t$ در رابطه بالا داریم

$$\begin{aligned} &P(S_{t+u+1} \neq j, S_{t+u-2} = j, \nu = \{\cdot, \dots, u-1\}, S_t \neq j | O_{\cdot}^{T-1} = o_{\cdot}^{T-1}; \theta^{(k)}) \\ &= G_j^{t+1} \sum_{i \neq j} P_{ij} F_i(t) \end{aligned}$$

اگر $u \geq T - 2 - t$ باشد داریم

$$\begin{aligned} &P(S_{t+u+1} \neq j, S_{t+u-2} = j, \nu = \{\cdot, \dots, u-1\}, S_t \neq j | O_{\cdot}^{T-1} = o_{\cdot}^{T-1}; \theta^{(k)}) \\ &= \prod_{\nu=\cdot}^{T-2-t} \frac{f_j(O_{T-1-\nu})}{N_{T-1-\nu}} d_j(u) \sum_{i \neq j} P_{ij} F_i(t) \end{aligned}$$

با ترکیب اطلاعات بالا داریم

$$\begin{aligned} &P(S_{T-1-\nu} = j, \nu = \{\cdot, \dots, T-2-t\}, S_t \neq i | O_{\cdot}^{T-1} = o_{\cdot}^{T-1}; \theta^{(k)}) \\ &= \prod_{\nu=\cdot}^{T-2-t} \frac{f_j(O_{T-1-\nu})}{N_{T-1-\nu}} D_j(T-1-t) \sum_{i \neq j} P_{ij} F_i(T) \end{aligned}$$

در زمان $t = 0$ و $u \leq T - 1$ داریم

$$\begin{aligned} \gamma_t^{(k)}(j) &: = P(S_t = j | O_{\cdot}^{T-1} = o_{\cdot}^{T-1}) \\ &= \frac{P(O_{t+1}^{T-1} = o_{t+1}^{T-1} | S_t = j)}{P(O_{t+1}^{T-1} = o_{t+1}^{T-1} | O_{\cdot}^t = o_{\cdot}^t)} P(S_t = j | O_{\cdot}^t = o_{\cdot}^t) \end{aligned}$$

و

$$\begin{aligned} L_{\setminus j}(t) &:= P(S_{t+1} \neq j, S_t = j | O_{\cdot}^{T-1} = o_{\cdot}^{T-1}) \\ &= \frac{P(O_{t+1}^{T-1} = o_{t+1}^{T-1} | S_{t+1} \neq j, S_t = j)}{P(O_{t+1}^{T-1} = o_{t+1}^{T-1} | O_{\cdot}^t = o_{\cdot}^t)} P(S_{t+1} \neq j, S_t = j | O_{\cdot}^t = o_{\cdot}^t) \end{aligned}$$

همچنین

$$\begin{aligned} & P(S_u \neq j, S_{u-\nu} = j, \nu = \{1, \dots, u\} | O_{\cdot}^{T-1} = o_{\cdot}^{T-1}; \theta^{(k)}) \\ &= \frac{L_{\setminus j}(u-1)}{F_j(u-1)} \prod_{\nu=1}^u \frac{b_j(O_{u-\nu})}{N_{u-\nu}} d_j(u) \pi_j \end{aligned}$$

اگر $u \geq T-1$ باشد، داریم

$$\begin{aligned} & P(S_u \neq j, S_{u-\nu} = j, \nu = \{1, \dots, u\} | O_{\cdot}^{T-1} = o_{\cdot}^{T-1}; \theta^{(k)}) \\ &= \prod_{\nu=1}^T \frac{f_j(O_{T-\nu})}{N_{T-\nu}} d_j(u) \pi_j \end{aligned}$$

همچنین داریم

$$\begin{aligned} & \sum_{t=\cdot}^{T-2} P(S_{t+u+1} \neq j, S_{t+u-\nu} = j, \nu = \{1, \dots, u-1\}, S_t \neq j | O_{\cdot}^{T-1} = o_{\cdot}^{T-1}; \theta^{(k)}) \\ &+ P(S_u \neq j, S_{u-\nu} = j, \nu = \{1, \dots, u\} | O_{\cdot}^{T-1} = o_{\cdot}^{T-1}; \theta^{(k)}) \\ &= \sum_{t=\cdot}^{T-2} P(S_{t+1} \neq j, S_t = j | O_{\cdot}^{T-1} = o_{\cdot}^{T-1}; \theta^{(k)}) \\ &+ P(S_u \neq j, S_{u-\nu} = j | O_{\cdot}^{T-1} = o_{\cdot}^{T-1}; \theta^{(k)}) \\ &= \sum_{t=\cdot}^{T-2} P(S_{t+1} \neq j, S_t = j | O_{\cdot}^{T-1} = o_{\cdot}^{T-1}; \theta^{(k)}) \\ &+ P(S_{T-1} = j | O_{\cdot}^{T-1} = o_{\cdot}^{T-1}; \theta^{(k)}) \\ &= \sum_{t=\cdot}^{T-2} L_{\setminus j}(t) + L_{\setminus j}(T-1) \end{aligned}$$

بنابراین

$$d_j^{(k+1)}(u) = \frac{\eta_{j,u}^{(k)}}{\sum_{\nu} \eta_{j,\nu}^k} = \frac{\eta_{j,u}^k}{\sum_{t=\cdot}^{T-2} L_{\setminus j}(t) + L_{\setminus j}(T-1)}$$

و

$$f_j^{(k+1)}(y) = \frac{\sum_{t=0}^{T-1} P(O_t = y, S_t = j | O_{:}^{t-1} = o_{:}^{t-1}; \theta^{(k)})}{\sum_{t=0}^{T-1} P(S_t = j | O_{:}^{T-1} = o_{:}^{T-1}; \theta^{(k)})} = \frac{\sum_{t=0}^{T-1} L_{\lambda_j}(t) I(O_t = y)}{\sum_{t=0}^{T-1} \gamma_t^{(k)}(j)}$$

۴.۲ مدل های مختلف توزیع زمان انتظار

مدل های زمان انتظار زیر را در نظر می گیریم:

(۱) اگر توزیع زمان انتظار ناپارمتری باشد، تابع شبه درستنمایی Q_d به صورت زیر به دست می آید

$$Q_d(d_j(u) | \theta^{(k)}) = \sum_u \eta_{ju}^{(k)} \log d_j(u)$$

تحت قید $\sum_{\omega} d_j(\omega) = 1$ و با مشتق گیری از تابع لاگرانژ داریم

$$\frac{\partial}{\partial d_j(u)} \left[Q_d(d_j(u) | \theta^{(k)}) - \lambda \left(\sum_{\nu=1}^{m_j} d_j(\nu) \right) \right] = 0$$

با در نظر گرفتن قید داریم

(۲) اگر توزیع زمان انتظار هندسی با احتمال موفقیت q_j باشد. در این صورت

$$Q_d(d_j(u) | \theta^{(k)}) = \sum_{u=1}^{T-1} \eta_{ju}^{(k)} [(u-1) \log(1-q_j) + \log q_j]$$

$$\frac{\partial Q_d(d_j(u) | \theta^{(k)})}{\partial q_j} = 0$$

$$\Rightarrow \sum_{u=1}^{T-1} \eta_{ju}^{(k)} \left[\frac{-(u-1)}{(1-q_j)} + \left(\frac{1}{q_j} \right) \right] = 0$$

در نتیجه

$$q_j^{(k+1)} = \frac{\sum_{u=1}^{T-1} \eta_{ju}^{(k)}}{\sum_{u=1}^{T-1} u \eta_{ju}^{(k)}}$$

(۳) اگر توزیع زمان انتظار دو جمله ای منفی باشد، در این حالت داریم

$$d_j(u) = \binom{u-r+1}{r-1} q_j^r (1-q_j)^{u-1}$$

و تابع شبه درستتمایی به صورت زیر است

$$Q_d(d_j(u)|\theta^{(k)}) = \sum_{u=1}^{T-1} \eta_{ju}^{(k)} r \log q_j + (u-1) \log(1-q_j)$$

در نتیجه

$$q_j^{(k+1)} = \frac{r \sum_{u=1}^{T-1} \eta_{ju}^{(k)}}{\sum_{u=1}^{T-1} \eta_{ju}^{(k)} (r+u-1)}$$

۵.۲ الگوریتم ویتربی برای مدل نیمه مارکف پنهان

الگوریتم ویتربی برای مدل‌های نیمه مارکف پنهان توسط لیولژ و لوینسون در سال ۱۹۹۱ [۶]، رامش و ویلیون در سال ۱۹۹۲ [۸]، بورشتاین ۱۹۹۶ [۹] مطرح شد. همانطور که قبلاً ذکر شد، اگر $O_{1:T}$ دنباله مشاهدات باشد می‌خواهیم بهترین دنباله حالات مانند $S_{1:T}$ را به گونه‌ای بیابیم که $P(S_{1:T}|O_{1:T}; \theta)$ ماکزیمم گردد،

محمتمترین دنباله حالتی که از مشاهده اول تا مشاهده t ام را شامل شده و در حالت j ام در طی مدت زمان d پایان است را برای $1 \leq t \leq T$ به صورت زیر تعریف می کنیم

$$\begin{aligned} \delta_t(j, d) &= \max_{S_{1:t-d}} P(S_{1:t-d}, S_{t-d+1:t} = j, O_{1:t} | \theta) \\ &= \max_{i \neq j} \max_h \max_{S_{1:t-d-h}} P(S_{1:t-d-h}, S_{t-d-h+1:t-d} = i, S_{t-d+1:t} = j, O_{1:t} | \theta) \\ &= \max_{i \neq j} \max_h \delta_{t-d}(i, h) P_{(i,h)(j,d)} f_{j,d}(O_{t-d+1:t}) \end{aligned}$$

که در آن

$$P_{(i,h)(j,d)} = P_{ij} P_j(d), \quad P_{ij} = P(S_{t+1} = j | S_t = i),$$

$$P_j(d) = P(S_{[t+1:t+d]} = j | S_{t+1} = j)$$

و

$$f_{j,d}(O_{t-d+1:t}) = P(O_{t-d+1:t} | S_{t-d+1:t} = j)$$

همچنین تعریف می کنیم

$$\Psi(t-d, j, h) \equiv (t-d, i^*, h^*)$$

که i^* حالت فعلی، یعنی حالتی که مدتی است در آن توقف داریم، و h^* مدت زمان ماندن در آن حالت و t زمان پایان است.

$$(i^*, h^*) = \arg \max_{i \in S-j} \max_{h \in D} \delta_{t-d}(i, h) P_{(i,h)(j,d)}(O_{t-d+1:t})$$

بنابراین می توان دنباله حالت را با یافتن آخرین حالتی تعیین کرد که تابع درستنمایی را ماکزیمم را می کند، در حالت کلی آخرین حالت حالتی است که

$$(t_1, j_1^*, d_1^*) = \arg \max_{T+D-1 \leq t \leq T} \max_{j \in S} \max_{D \leq d \leq t-T+1} \delta_t(j, d)$$

با فرض $t_1 = T$ داریم

$$(j_1^*, d_1^*) = \arg \max_{j \in S} \max_{d \in D} \delta_t(j, d)$$

با پیگیری دنباله حالت داریم

$$(t_2, j_2^*, d_2^*) = \Psi(t_1, j_1^*, d_1^*),$$

...

$$(t_n, j_n^*, d_n^*) = \Psi(t_{n-1}, j_{n-1}^*, d_{n-1}^*)$$

اگر حالت S_1 مشخص شود، یعنی $S_1 = j_n^*$ آنگاه $(j_1^*, d_1^*), \dots, (j_n^*, d_n^*)$ دنباله برآورد شده است.

فصل ۳

تحلیل و پیش‌بینی داده‌های بورس آمریکا

در این فصل و با استفاده از داده‌های سری زمانی قیمت‌های سهام بورس آمریکا، به مقایسه مدل‌های مارکف پنهان و نیمه پنهان و قدرت پیش‌بینی آن‌ها می‌پردازیم.

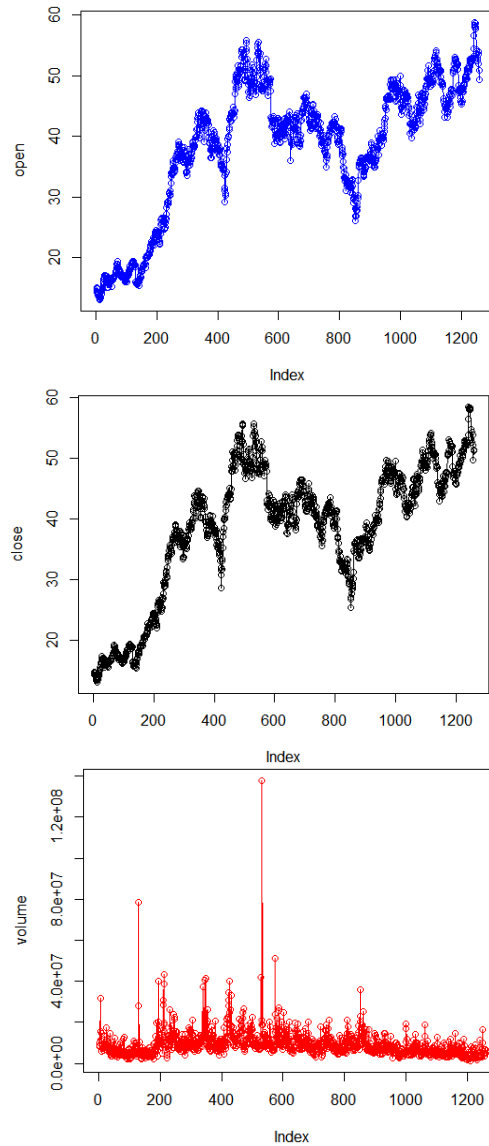
۱.۳ معرفی داده

یکی از کاربردهای مدل‌های مارکف پنهان و نیمه پنهان در مدل‌بندی برای داده‌هایی است که با مدل‌های دیگر قادر به مدل‌بندی آنها نیستیم و یا مدل‌بندی آنها از روش‌های دیگر دشوار است، در این پروژه به بررسی داده‌های مربوط به بورس سهام کشور آمریکا می‌پردازیم.

داده‌ها شامل سه متغیر، `close` و `open` و `volume` است. متغیر `open` بیانگر قیمت آغازین سهام، متغیر `close` بیانگر قیمت تمام شده سهام (آخرین قیمت سهام) و متغیر `volume` بیانگر تعداد کل معاملات انجام شده برای یک سهام است، داده‌ها در ۳ ستون و ۱۲۵۸ سطر است، نمودار سری زمانی این سه متغیر در شکل ۱.۳ داده شده است.

مساله سری زمانی در اینجا یک مساله سه متغیره است. با توجه به نمودارهای بالا متغیرهای `open` و `close` به وضوح سری‌های زمانی نایستا هستند، بنابراین ابتدا داده‌های مربوط به متغیرهای `open` و `close` را تفاضلی

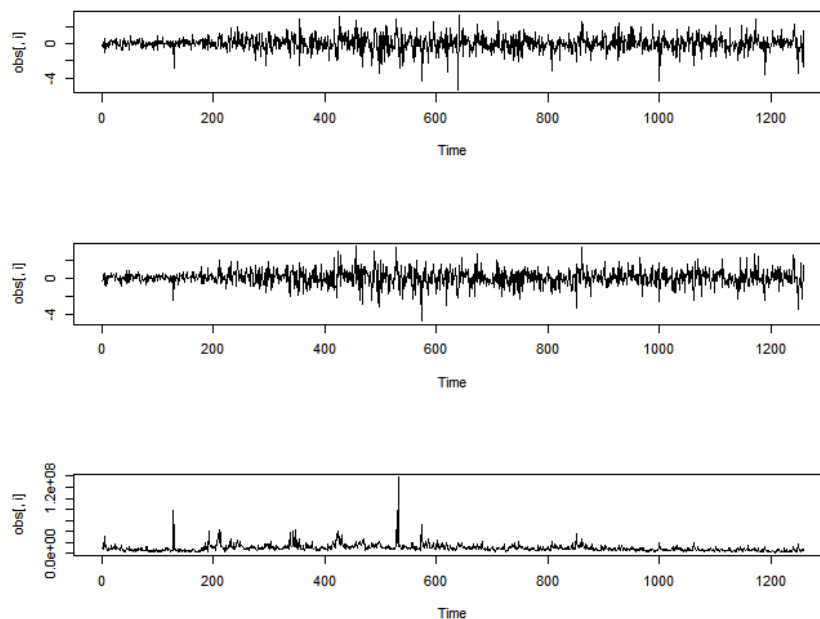
شکل ۱.۳: نمودار مربوط به متغیرهای close open و volume



می‌کنیم. نمودار سری‌های زمانی حاصل برای سه متغیر در شکل ۲.۳ آمده است.

اگرچه نمودارهای حاصل ظاهراً از ایستایی بالاتری نسبت به نمودارهای اولیه برخوردار هستند، با این حال به نظر می‌رسد در نظر گرفتن میانگین و واریانس یکسان برای این متغیرها در طول زمان کار درستی نیست. همچنین

شکل ۲.۳: نمودارهای سری‌های ایستاشده



با توجه به همبستگی سه متغیر این همبستگی نیز بایستی در تحلیل و پیش بینی مد نظر قرار بگیرند. می‌توانیم فرض کنیم یک سری زمانی سه متغیره از توزیع نرمال سه متغیره در چندین خوشه (حالت) پنهان با میانگین و واریانس متفاوت داریم و برای تحلیل و پیش بینی این داده‌ها از مدل‌های مارکف و نیمه مارکف پنهان استفاده می‌کنیم.

ابتدا برای به دست آوردن دقت پیش‌بینی مدل‌ها ۸ مشاهده آخر را به عنوان آزمایش و مابقی مشاهدات را به عنوان آموزش از هم تفکیک می‌کنیم.

```
> (n=nrow(obs))
[1] 1258
> test=obs[1251:1258,]
> train=obs[1:1250,]
```

به منظور برازش مدل مارکف پنهان و نیمه پنهان به مقدار اولیه نیاز داریم، برای به دست آوردن مقدار اولیه از الگوریتم k -میانگین که برای خوشه‌بندی داده‌های به کار می‌رود استفاده می‌کنیم، این الگوریتم با توجه به ویژگی‌ها (متغیرها) داده‌ها را خوشه‌بندی می‌کند، به گونه‌ای که در هر خوشه داده‌هایی که از نظر ویژگی‌ها شبیه هستند قرار گیرند، که در مسئله ما خوشه‌ها همان حالت‌های پنهان هستند. در مسئله مورد بررسی سه حالت پنهان در نظر می‌گیریم، بنابراین این الگوریتم را با $k = 3$ اجرا می‌کنیم.

```
> km=kmeans(train,3)
> CLUSTER=km$cluster
> CLUSTER
[1] 2 2 2 3 3 3 3 3 2 2 2 2 2 2 2 2 3 2 2 2 2 3 2 3 2 3 2 2 2 2
[31] 2 2 2 2 3 3 2 2 2 2 2 2 2 2 2 2 2 3 2 2 2 2 2 2 2 2 2 2 2 2
.....
[1231] 2 2 2 2 2 2 2 2 2 3 2 2 2 2 2 2 2 2 3 2 2
```

برای برآورد مقدار اولیه ماتریس احتمال انتقال مدل مارکف پنهان به صورت زیر عمل می‌کنیم

```
> tb=table(CLUSTER[-length(CLUSTER)],CLUSTER[-1])
> tm.hmm= tb/rowSums(tb)
> tm.hmm
      1      2      3
1 0.000000000 0.071428571 0.928571429
2 0.004376368 0.849015317 0.146608315
3 0.031152648 0.426791277 0.542056075
```

حال مقادارهای اولیه پارامترهای توزیع مشاهده‌ها را به صورت زیر به دست می‌آوریم

```

> x1=train[CLUSTER==1,]
> x2=train[CLUSTER==2,]
> x3=train[CLUSTER==3,]
> (mu1=apply(x1,2,mean))
      diff.open    diff.close    volume
5.785714e-02 -3.435714e-01  4.964579e+07
> (mu2=apply(x2,2,mean))
      diff.open    diff.close    volume
6.102186e-02  5.356393e-02  6.779791e+06
> (mu3=apply(x3,2,mean))
      diff.open    diff.close    volume
-5.627726e-02 -1.832087e-02  1.517620e+07
> (sig1=(t(x1)%*(x1)-nrow(x1)*(mu1)%*t(mu1))/nrow(x1))
      diff.open    diff.close    volume
diff.open  1.646588e+00  2.344564e+00  5.243340e+06
diff.close 2.344564e+00  4.985794e+00 -1.069250e+06

volume      5.243340e+06 -1.069250e+06  7.136902e+14
> (sig2=(t(x2)%*(x2)-nrow(x2)*(mu2)%*t(mu2))/nrow(x2))
      diff.open    diff.close    volume
diff.open  5.084249e-01  7.689882e-02  1.132792e+04
diff.close 7.689882e-02  4.138336e-01  3.694952e+04

```

```

volume      1.132792e+04 3.694952e+04 5.075468e+12
> (sig3=(t(x3)%*(x3)-nrow(x3)*(mu3)%*t(mu3))/nrow(x3))
              diff.open    diff.close    volume
diff.open    1.586832e+00  5.929527e-01 -5.978933e+05
diff.close   5.929527e-01  1.488332e+00 -7.043954e+04
volume      -5.978933e+05 -7.043954e+04  1.696431e+13

```

به منظور برازش مدل به توابع موجود در بسته نرم‌افزاری `mhsmm` نیاز داریم، بنابراین ابتدا این بسته را نصب و فراخوانی می‌کنیم. سپس یک مدل اولیه برای مدل مارکوف پنهان را به کمک تابع `hmmspec` برازش می‌دهیم.

```

> library(mhsmm)
> J=3
> initial=rep(1/J,J)
> P=tm.hmm
> b=list(mu=list(mu1,mu2,mu3),sigma=list(sig1,sig2,sig3))
> model=hmmspec(init=initial, trans=P, parms.emission=b,
dens.emission=dmvnorm.hsmm)

```

```
> model
```

Hidden Markov Model specification:

J (number of states):

3

init:

[1] 0.3333333 0.3333333 0.3333333

transition:

	1	2	3
1	0.000000000	0.071428571	0.928571429
2	0.004376368	0.849015317	0.146608315
3	0.031152648	0.426791277	0.542056075

emission:

\$mu

\$mu[[1]]

	diff.open	diff.close	volume
	5.785714e-02	-3.435714e-01	4.964579e+07

\$mu[[2]]

	diff.open	diff.close	volume
	6.102186e-02	5.356393e-02	6.779791e+06

\$mu[[3]]

	diff.open	diff.close	volume
	-5.627726e-02	-1.832087e-02	1.517620e+07

\$sigma

\$sigma[[1]]

	diff.open	diff.close	volume
diff.open	1.646588e+00	2.344564e+00	5.243340e+06
diff.close	2.344564e+00	4.985794e+00	-1.069250e+06

```
volume      5.243340e+06 -1.069250e+06  7.136902e+14
```

```
$sigma[[2]]
```

```
          diff.open  diff.close      volume
```

```
diff.open  5.084249e-01  7.689882e-02  1.132792e+04
```

```
diff.close 7.689882e-02  4.138336e-01  3.694952e+04
```

```
volume     1.132792e+04  3.694952e+04  5.075468e+12
```

```
$sigma[[3]]
```

```
          diff.open  diff.close      volume
```

```
diff.open  1.586832e+00  5.929527e-01 -5.978933e+05
```

```
diff.close 5.929527e-01  1.488332e+00 -7.043954e+04
```

```
volume     -5.978933e+05 -7.043954e+04  1.696431e+13
```

شناسه `init` یک بردار به عنوان مقدار اولیه احتمال شروع حالت‌ها است که تمام حالات را هم‌شانس در نظر می‌گیرد. شناسه `trans` ماتریس احتمال انتقال اولیه، شناسه `parms.emission` مقادیر اولیه پارامترهای مربوط به توزیع نرمال سه متغیره برای تمام حالات است و شناسه `dens.emission` مشخص کننده توزیع مشاهدات است.

برازش مدل مارکوف پنهان به کمک تابع `hmmfit` صورت می‌گیرد. ابتدا داده‌ها را در قالب استاندارد ورودی تنظیم کرده و سپس به صورت زیر مدل را برازش می‌دهیم.

```
> N=nrow(train)
```

```
> train=list(x=train,N=N)
```

```
> class(train)<-"hmm.data"
> h1 = hmmfit(train, model, mstep = mstep.mvnorm)
```

خروجی تابع بالا شامل چند زیر شی به صورت زیر است.

```
> names(h1)
[1] "model" "K" "f" "mstep" "gam" "loglik" "N" "p"
[9] "yhat"
```

شیء `yhat` مقادیر برازش یافته بر اساس مدل مارکوف پنهان، `K` تعداد خوشه‌ها، `model` مدل برازش داده شده و `loglik` دنباله مقدار لگاریتم تابع درستمایی است. مقادیر برآورد پارامترهای مدل برازش داده شده را به صورت زیر استخراج می‌کنیم.

```
> h1$model
Hidden Markov Model specification:
J (number of states):
3
init:
[1] 3.838052e-120 1.000000e+00 2.477684e-39
transition:
      [,1]      [,2]      [,3]
[1,] 0.000000000 1.157320e-64 1.0000000
[2,] 0.004693697 8.698602e-01 0.1254461
[3,] 0.045269463 1.347271e-01 0.8200034
```

emission:

\$mu

\$mu[[1]]

	diff.open	diff.close	volume
	4.880938e-01	2.158484e-01	3.488211e+07

\$mu[[2]]

	diff.open	diff.close	volume
	4.494114e-02	5.660006e-02	5.858326e+06

\$mu[[3]]

	diff.open	diff.close	volume
	-1.173812e-02	-8.935092e-03	1.180686e+07

\$sigma

\$sigma[[1]]

	diff.open	diff.close	volume
diff.open	1.801533e+00	2.380201e+00	-2.539934e+06

diff.close	2.380201e+00	4.559794e+00	-7.395035e+06
------------	--------------	--------------	---------------

volume	-2.539934e+06	-7.395035e+06	5.581022e+14
--------	---------------	---------------	--------------

\$sigma[[2]]

	diff.open	diff.close	volume
diff.open	2.694608e-01	3.518288e-02	3.927679e+04

diff.close	3.518288e-02	2.520400e-01	5.553091e+04
------------	--------------	--------------	--------------

```

volume      3.927679e+04 5.553091e+04 3.790494e+12
$sigma[[3]]
              diff.open  diff.close  volume
diff.open  1.311519e+00  3.368885e-01 -8.856034e+05
diff.close 3.368885e-01  1.083645e+00 -3.291916e+05
volume     -8.856034e+05 -3.291916e+05  1.711910e+13

```

نمودار دنباله مقدار لگاریتم تابع درستمایی را به صورت زیر رسم می‌کنیم.

```

> plot(h1$loglik, type = "b", ylab = "Log-likelihood",
      xlab = "Iteration")

```

نمودار به دست آمده در شکل ۳.۳ نشان داده شده است.

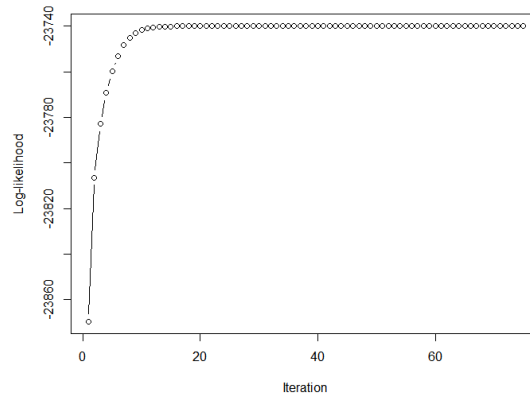
پس از برازش مدل مارکف پنهان به پیش‌بینی به کمک این مدل می‌پردازیم. از امید ریاضی توزیع مشاهدات به شرط حالت‌های پیش‌بینی شده برای پیش‌بینی استفاده می‌کنیم.

```

> (pred=sapply(predicted$s,function(i){
  h1$model$parms.emission$mu[[i]]}))
              [,1]      [,2]      [,3]      [,4]
diff.open  4.494114e-02 4.494114e-02 4.494114e-02 4.494114e-02
diff.close 5.660006e-02 5.660006e-02 5.660006e-02 5.660006e-02

```

شکل ۳.۳: نمودار مشاهدات چندگانه مدل مارکف پنهان



```
volume      5.858326e+06  5.858326e+06  5.858326e+06  5.858326e+06
```

```
           [,5]      [,6]      [,7]
```

```
diff.open   -1.173812e-02 -1.173812e-02 -1.173812e-02
```

```
diff.close  -8.935092e-03 -8.935092e-03 -8.935092e-03
```

```
volume      1.180686e+07  1.180686e+07  1.180686e+07
```

```
           [,8]
```

```
diff.open   -1.173812e-02
```

```
diff.close  -8.935092e-03
```

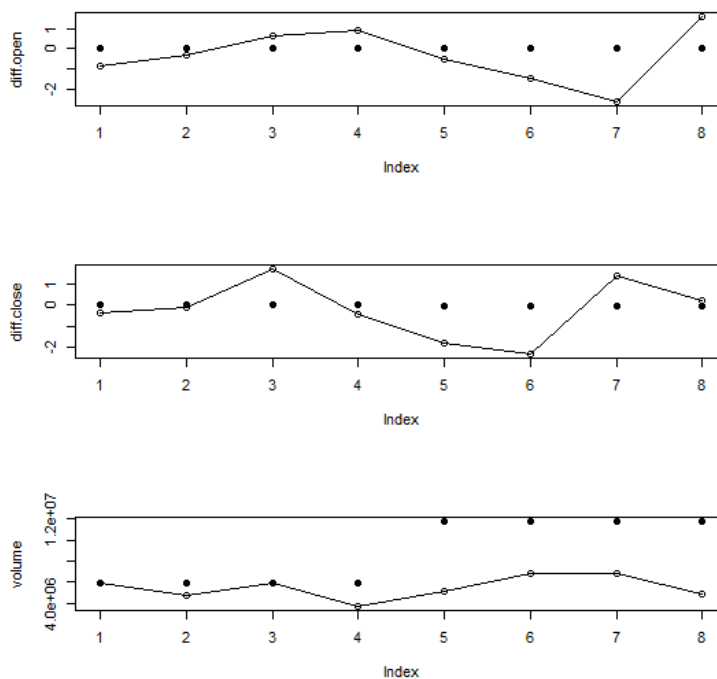
```
volume      1.180686e+07
```

در نمودار مقادیر واقعی و پیش‌بینی شده را با هم مقایسه می‌کنیم.

```
> par(mfrow=c(3,1))
> for(i in 1:3){
+ plot(test$x[,i],ylab=paste(colnames(test$x)[i]),
+ ylim=c(c(min(pred[i,],test$x[,i])),
+ max(c(pred[i,],test$x[,i]))),type="o")
+ points(pred[i,],pch=16)
+ }
```

نمودار به دست آمده در شکل ۴.۳ نشان داده شده است.

شکل ۴.۳: پیش‌بینی با مدل مارکوف پنهان



همچنین نمودار سری زمانی اصلی و پیش‌بینی‌ها را می‌توان به صورت زیر رسم کرد.

```
> par(mfrow=c(2,1))
> predopen=pred[1,]
> predopen[1]=open[1251]+pred[1,1]
> predopen=cumsum(predopen)
> plot(open[1252:1259],type="o",ylab="open")
> points(predopen,pch=16)
> #
> predclose=pred[2,]
> predclose[1]=close[1251]+pred[2,1]
> predclose=cumsum(predclose)
> plot(close[1252:1259],type="o",ylab="close")
> points(predclose,pch=16)
```

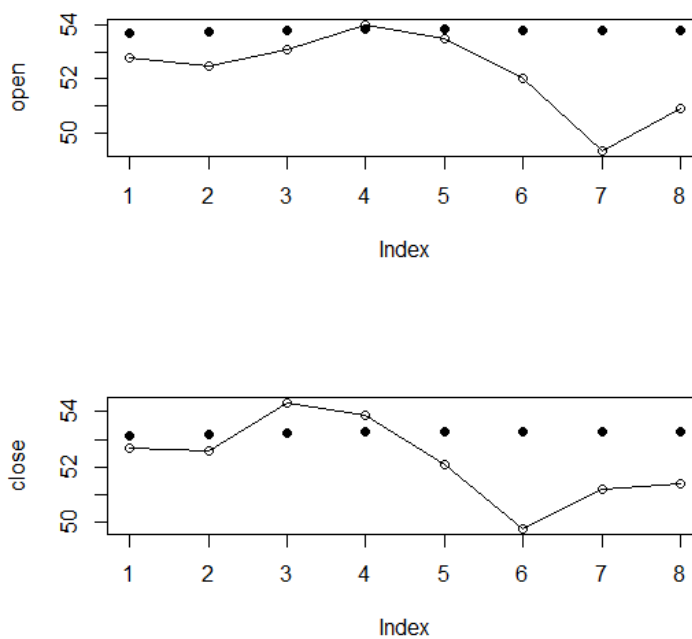
نمودار به دست آمده در شکل ۵.۳ نشان داده شده است.

برازش مدل نیمه مارکف پنهان

در برازش مدل نیمه مارکف پنهان باید به نکاتی توجه کرد:

- 1- قطر ماتریس احتمال‌های انتقال اولیه در مدل نیمه مارکف پنهان صفر است زیرا احتمال ماندن در هر حالت با توجه به توزیع زمان انتظار محاسبه می‌گردد.
- 2- توزیع زمان انتظار در مدل‌های نیمه مارکف پنهان برای مشاهدات چندگانه می‌تواند پواسون، ناپارامتری و یا گاما باشد که هر یک از توزیع‌ها ما را به نتایج مختلفی می‌رساند، بنابراین باید هر سه را امتحان کرده و آن توزیعی را که خطای پیش‌بینی آن کمتر است را به عنوان توزیع زمان انتظار برگزینیم.

شکل ۵.۳: پیش‌بینی با مدل مارکوف پنهان



بنابراین مقدار اولیه ماتریس احتمال انتقال مدل نیمه مارکوف پنهان را به صورت زیر به دست می‌آوریم.

```
> tm.hsmm=tm.hmm
> for(i in 1:3){
+ tm.hsmm[i,]=tm.hmm[i,]/tm.hmm[i,i]
+ tm.hsmm[i,i]=0
+ }
> tm.hsmm[!is.finite(tm.hsmm)]<-0.5
> tm.hsmm=tm.hsmm/rowSums(tm.hsmm)
> tm.hsmm
```

	1	2	3
1	0.00000000	0.50000000	0.50000000
2	0.02898551	0.00000000	0.97101449
3	0.06802721	0.93197279	0.00000000

با در نظر گرفتن توزیع زمان انتظار پواسون، نتایج برازش مدل نیمه مارکف به صورت زیر است.

```
> init<-rep(1/J,J)
> P=tm.hsmm
> d0 <- list(lambda = c(10, 30, 60), shift = c(10, 100, 30),
type = "poisson")

> model2 <- hsmmspec(init, P, parms.emis = b , sojourn = d0,
dens.emis = dmvnorm.hsmm)
> class(train)<-"hsmm.data"
> h.poisson <- hsmmfit(train,model2, mstep = mstep.mvnorm)
> h.poisson$model
Hidden semi-Markov Model specification:
J (number of states):
3
init:
[1] 1 0 0
transition matrix:
      [,1]      [,2]      [,3]
```

[1,] 0.0000000 0.5317355 0.4682644682

[2,] 0.9999970 0.0000000 0.0000030456

[3,] 0.7201775 0.2798225 0.0000000000

emission distribution:

\$mu

\$mu[[1]]

diff.open	diff.close	volume
-2.140010e-01	-2.312315e-01	2.032707e+07

\$mu[[2]]

diff.open	diff.close	volume
8.871645e-02	9.319223e-02	6.174394e+06

\$mu[[3]]

diff.open	diff.close	volume
1.010351e-02	7.790922e-03	1.100434e+07

\$sigma

\$sigma[[1]]

	diff.open	diff.close	volume
diff.open	2.441173e+00	8.774862e-01	1.782186e+06
diff.close	8.774862e-01	1.877218e+00	3.563820e+05
volume	1.782186e+06	3.563820e+05	3.086168e+14

\$sigma[[2]]

	diff.open	diff.close	volume
diff.open	4.454107e-01	8.929002e-02	5.238580e+04

```

diff.close 8.929002e-02 4.254154e-01 2.010346e+05
volume      5.238580e+04 2.010346e+05 6.779629e+12
$sigma[[3]]
           diff.open    diff.close      volume
diff.open  9.007781e-01  2.743801e-01 -2.572883e+05
diff.close  2.743801e-01  8.801804e-01 -2.030724e+05
volume     -2.572883e+05 -2.030724e+05  1.881994e+13
sojourn distribution:
$lambda
[1] 4.360786 56.888899 81.606084
$shift
[1] 1 1 1
$type
[1] "poisson"

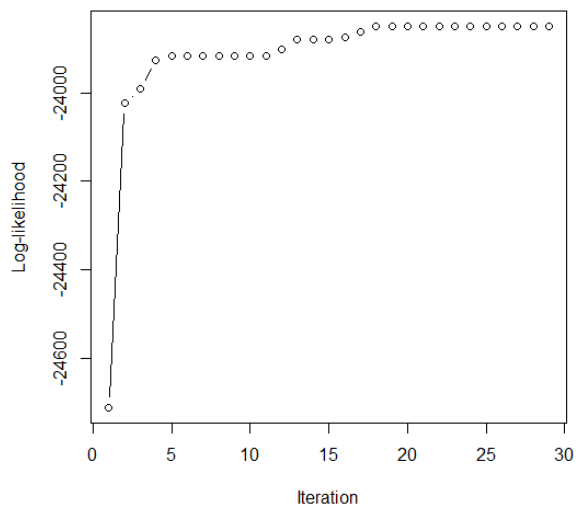
> plot(h.poisson$loglik,type = "b",
      ylab = "Log-likelihood",xlab = "Iteration")

```

نمودار لگاریتم درست‌نمایی مدل در شکل ۶.۳ نشان داده شده است. حال مانند قبل به پیش‌بینی پرداخته و

نمودار نهایی را رسم می‌کنیم.

شکل ۶.۳: نمودار دنباله لگاریتم درستنمایی مدل نیمه مارکف پنهان با زمان انتظار پواسون



```

> class(test)<-"hsmm.data"

> predicted = predict(h.poisson,test)

[1] -29.54824 -17.02058

[1] -24.86721 -43.90319 -64.60544 -85.22835 -106.31045

[6] -128.02622

[1] -141.9814 -158.3309 -175.7747 -194.7619 -217.5197 -172.8929

> pred=sapply(predicted$s,function(i){
  h.poisson$model$params$emission$mu[[i]]})

> par(mfrow=c(3,1))

> predopen=pred[1,]

> predopen[1]=open[1251]+pred[1,1]

```

```

> predopen=cumsum(predopen)
> plot(open[1252:1259],type="o",ylab="open")
> points(predopen,pch=16)
> #
> predclose=pred[2,]
> predclose[1]=close[1251]+pred[2,1]
> predclose=cumsum(predclose)
> plot(close[1252:1259],type="o",ylab="close")
> points(predclose,pch=16)
> plot(test$x[,3],ylab=paste(colnames(test$x)[3]),
+ ylim=c(c(min(pred[3,],test$x[,3])),
+ max(c(pred[3,],test$x[,3]))),type="o")
> points(pred[3,],pch=16)

```

نمودار به دست آمده در شکل ۷.۳ نشان داده شده است.

توزیع زمان انتظار ناپارامتری در شکل ۸.۳ نشان داده شده است.

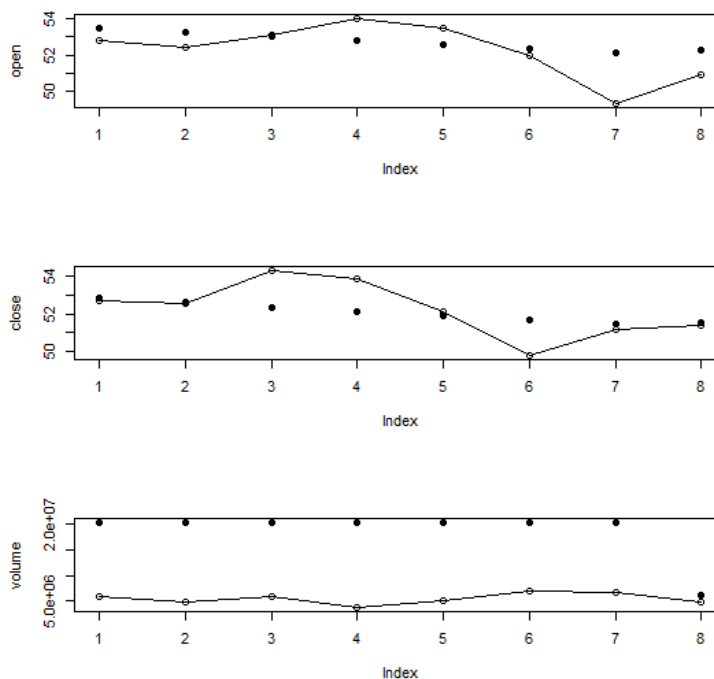
با در نظر گرفتن توزیع زمان انتظار ناپارامتری، نتایج برازش مدل نیمه مارکف به صورت زیر است.

```

> d <- cbind(dunif(1:nrow(obs), 0, 50), dunif(1:nrow(obs), 100, 175),
dunif(1:nrow(obs), 50, 130))
> start.np <- hsmmspec(init = rep(1/J, J),transition = P,parms.emis = b,
sojourn = list(d = d, type = "nonparametric"),dens.emis = dmvnorm.hsmm)
> h.np <- hsmmfit(train, start.np,
mstep = mstep.mvnorm,graphical = TRUE)

```

شکل ۷.۳: پیش‌بینی با مدل نیمه مارکوف پنهان



[1] 0.000000e+00 6.115467e-08

[1] 0.000000e+00 3.902059e-08

[1] 0.000000e+00 3.89627e-08

...

[1] 0.000000e+00 3.898307e-08

[1] 0.000000e+00 3.898309e-08

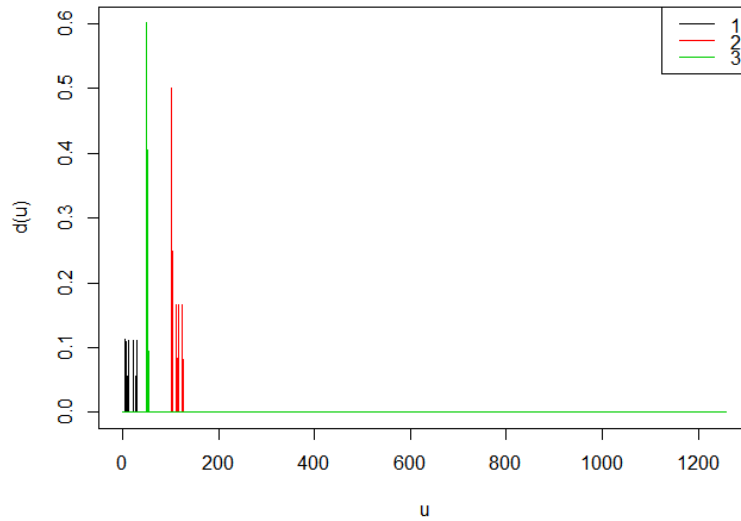
> h.np\$model

Hidden semi-Markov Model specification:

J (number of states):

!ht

شکل ۸.۳: توزیع زمان انتظار ناپارامتری



3

init:

```
[1] 1.00000e+00 0.00000e+00 5.39805e-16
```

transition matrix:

```

      [,1]      [,2]      [,3]
[1,] 0.0000000 0.3089665 6.910335e-01
[2,] 1.0000000 0.0000000 4.377643e-37
[3,] 0.7175483 0.2824517 0.000000e+00
```

emission distribution:

 μ
 $\mu[[1]]$

diff.open	diff.close	volume
-7.608076e-03	-2.000856e-02	2.643206e+07

\$mu[[2]]

diff.open	diff.close	volume
4.124495e-02	4.757230e-02	6.727200e+06

\$mu[[3]]

diff.open	diff.close	volume
2.311296e-02	1.708351e-02	1.087169e+07

\$sigma

\$sigma[[1]]

	diff.open	diff.close	volume
diff.open	2.345047e+00	1.029210e+00	3.709409e+06
diff.close	1.029210e+00	2.293415e+00	1.604939e+06
volume	3.709409e+06	1.604939e+06	4.517609e+14

\$sigma[[2]]

	diff.open	diff.close	volume
diff.open	5.823912e-01	1.197323e-01	-2.423883e+05

diff.close	1.197323e-01	5.283551e-01	-5.192972e+04
------------	--------------	--------------	---------------

volume	-2.423883e+05	-5.192972e+04	8.959282e+12
--------	---------------	---------------	--------------

\$sigma[[3]]

	diff.open	diff.close	volume
--	-----------	------------	--------

```
diff.open 9.123783e-01 2.970897e-01 -4.956716e+05
```

```
diff.close 2.970897e-01 8.508025e-01 -5.454825e+05
```

```
volume -4.956716e+05 -5.454825e+05 2.497050e+13
```

```
sojourn distribution:
```

```
$d
```

```
          [,1]          [,2]          [,3]
[1,] 3.768846e-01 0.000000e+00 0.000000e+00
[2,] 2.165410e-01 0.000000e+00 0.000000e+00
[3,] 5.158266e-15 0.000000e+00 0.000000e+00
[4,] 1.112261e-01 0.000000e+00 0.000000e+00
[5,] 9.948862e-02 0.000000e+00 0.000000e+00
```

```
.....
```

```
[1248,] 0.000000e+00 0.000000e+00 0.000000e+00
[1249,] 0.000000e+00 0.000000e+00 0.000000e+00
[1250,] 0.000000e+00 0.000000e+00 0.000000e+00
```

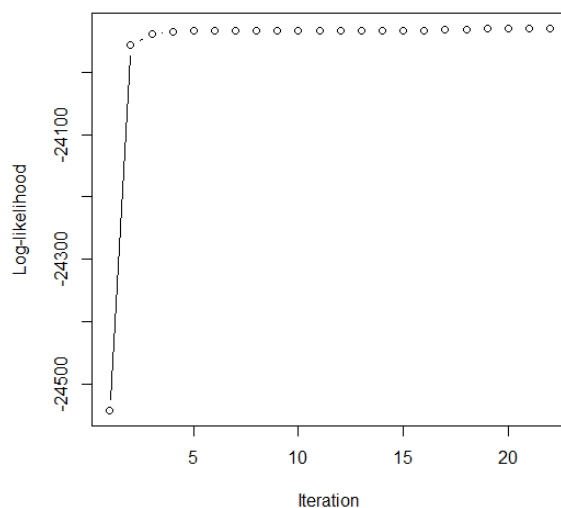
```
$type
```

```
[1] "nonparametric"
```

```
> plot(h.np$loglik,type = "b",
ylab = "Log-likelihood",xlab = "Iteration")
```

نمودار لگاریتم درستمایی مدل در شکل ۹.۳ نشان داده شده است.

شکل ۹.۳: نمودار دنباله لگاریتم درستمایی مدل نیمه مارکف پنهان با زمان انتظار ناپارامتری



حال مانند قبل به پیش‌بینی پرداخته و نمودار نهایی را رسم می‌کنیم.

```
> predicted = predict(h.np, test)
```

```
[1] -27.12138 -17.43765
```

```
[1] -21.89497 -43.33036 -96.24228 -86.92476 -108.56082
```

```
[6] -153.17145
```

```
[1] -1.000000e+300 -1.000000e+300 -1.000000e+300 -1.000000e+300
```

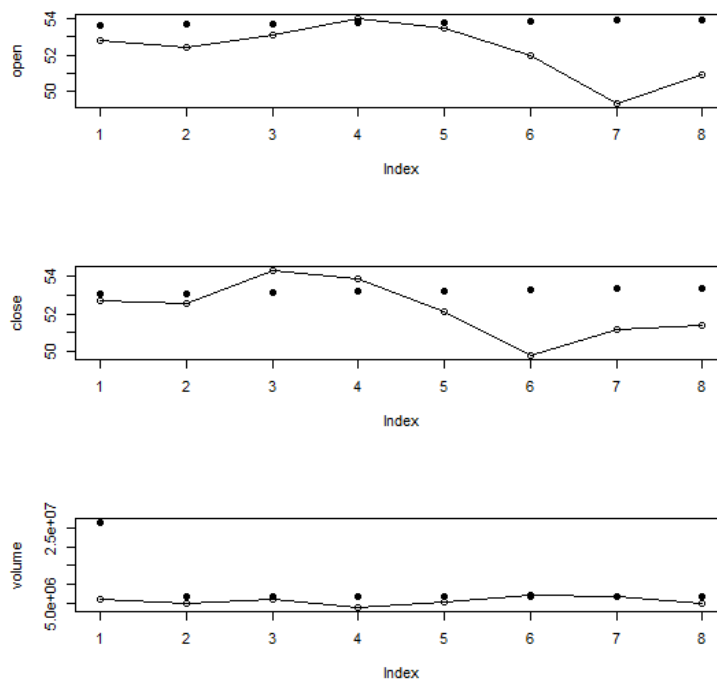
```
[5] -1.000000e+300 -1.698383e+02
```

```
> pred=sapply(predicted$s,function(i){
```

```
h.np$model$params.emission$mu[[i]])  
> par(mfrow=c(3,1))  
> predopen=pred[1,]  
> predopen[1]=open[1251]+pred[1,1]  
> predopen=cumsum(predopen)  
> plot(open[1252:1259],type="o",ylab="open")  
> points(predopen,pch=16)  
> #  
> predclose=pred[2,]  
> predclose[1]=close[1251]+pred[2,1]  
> predclose=cumsum(predclose)  
> plot(close[1252:1259],type="o",ylab="close")  
> points(predclose,pch=16)  
> plot(test$x[,3],ylab=paste(colnames(test$x)[3]),  
+ ylim=c(c(min(pred[3,],test$x[,3])),  
+ max(c(pred[3,],test$x[,3]))),type="o")  
> points(pred[3,],pch=16)
```

نمودار به دست آمده در شکل ۱۰.۳ نشان داده شده است.

شکل ۱۰.۳: پیش‌بینی با مدل نیمه مارکوف پنهان



همانطور که می‌توان دید مدل نهایی از مدل‌های دیگر پیش‌بینی به مراتب دقیق‌تری، به خصوص در ۵ زمان

اول دارد.

کتابنامه

- [1] Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The annals of mathematical statistics*, 37(6), 1554-1563.
- [2] Li, N., & Stephens, M. (2003). Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics*, 165(4), 2213-2233.
- [3] Liu, T., Lemeire, J., & Yang, L. (2014, July). Proper initialization of Hidden Markov models for industrial applications. In *2014 IEEE China summit & international conference on signal and information processing (ChinaSIP)* (pp. 490-494). IEEE.
- [4] Guédon, Y. (2005). Hidden hybrid Markov/semi-Markov chains. *Computational statistics & Data analysis*, 49(3), 663-688.
- [5] Hughes, J. P. (1997). Computing the observed information in the hidden Markov model using the EM algorithm. *Statistics & probability letters*, 32(1), 107-114.
- [6] Li, J., Najmi, A., & Gray, R. M. (2000). Image classification by a two-dimensional hidden Markov model. *IEEE transactions on signal processing*, 48(2), 517-533.
- [7] Langmead, B. *Hidden Markov Models*.

- [8] O'Connell, J., & Højsgaard, S. (2011). Hidden semi markov models for multiple observation sequences: The mhsmm package for R. *Journal of Statistical Software*, 39(4), 1-22.
- [9] Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257-286.
- [10] Yu, S. Z., & Kobayashi, H. (2003). A hidden semi-Markov model with missing data and multiple observation sequences for mobility tracking. *Signal Processing*, 83(2), 235-250.
- [11] Yu, S. Z. (2015). *Hidden Semi-Markov models: theory, algorithms and applications*. Morgan Kaufmann.
- [12] Barbu, V. S., & Limnios, N. (2009). *Semi-Markov chains and hidden semi-Markov models toward applications: their use in reliability and DNA analysis (Vol. 191)*. Springer Science & Business Media.
- [13] Bulla, J. (2006). *Application of hidden Markov models and hidden semi-Markov models to financial time series*.
- [14] Dymarski, P. (Ed.). (2011). *Hidden Markov Models: Theory and Applications*. BoD—Books on Demand.



School of Mathematics, Statistics, and Computer Science

College of Science

University of Tehran

Hidden Markov Model And Semi-Hidden Markov Model:
Theory And Applications, Using Software R

Zahra Dehghani Ghobadi

Supervisor: Dr. Seyed Morteza Amini

January 2020