

سلام الغفران



دانشگاه تهران  
مرکز تحقیقات بیوشیمی-بیوفیزیک

## عنوان پایان نامه

نگارش  
نام و نام خانوادگی دانشجو

استاد(های) راهنما  
نام و نام خانوادگی استاد(های) راهنما

استاد مشاور  
در صورت وجود

پایان نامه برای دریافت درجه دکترا  
در رشته بیوانفورماتیک

تاریخ

تقدیم

♦ ♦ ♦ ♦

سپاسگزارى...

سپاس از...

و بهمين سپاس از ا... ..

نام و نام خانوادگى  
فصل سال

# تعهدنامه اصالت اثر

اینجانب ..... متعهد می شوم که مطالب مندرج در این پایان نامه/رساله حاصل کار پژوهشی اینجانب است و به دستاوردهای پژوهشی دیگران که در این پژوهش از آنها استفاده شده است، مطابق مقررات ارجاع و در فهرست منابع و مآخذ ذکر گردیده است. این پایان نامه/رساله قبلاً برای احراز هیچ مدرک هم سطح یا بالاتر ارائه نشده است. در صورت اثبات تخلف (در هر زمان) مدرک تحصیلی صادر شده توسط دانشگاه از اعتبار ساقط خواهد شد. کلیه حقوق مادی و معنوی این اثر متعلق به دانشکده ریاضی، آمار و علوم کامپیوتر دانشکدگان علوم دانشگاه تهران می باشد.

نام و نام خانوادگی امضا

# حق مالکیت معنوی

حق مالکیت معنوی این اثر متعلق به دانشگاه تهران می باشد. استفاده از مطالب این پایان نامه در فعالیتهای تحقیقاتی با ذکر منبع بلامانع می باشد. در صورت استفاده تجاری، مانند چاپ این پایان نامه، هماهنگی لازم و اجازه کتبی از دانشگاه و نگارنده پایان نامه الزامی می باشد.

# خروجی های پایان نامه

1. paper1
2. paper2

## چکیده

چکیده پایان‌نامه باید شامل نتایج اصلی مورد مطالعه باشد. توجه نمایید که چکیده با پیشگفتار تفاوت دارد و محلی برای طرح مبانی یا مفاهیم مقدماتی نیست. چکیده تنها باید شامل نتایج و کار اصلی انجام شده در پایان‌نامه باشد. طول چکیده معمولاً بیشتر از یک یا دو پاراگراف نیست.

**کلمات کلیدی:** کلمات اصلی که در موضوع و متن پایان‌نامه استفاده شده‌اند و ترجیحاً در چکیده و عنوان نیستند که با کاما از هم جدا شده باشند



# پیشگفتار

پیشگفتار، فصلی از پایان‌نامه است که معمولاً شامل بخش یا زیربخش نیست و حدود یک یا دو صفحه است. در آن، مقدمه‌ای به زبان ساده و عاری از فرمول‌بندی ریاضی، از مساله مورد مطالعه در پایان‌نامه ارائه می‌شود. همچنین در پیشگفتار است که می‌توان خواننده را با تاریخچه‌ای مختصر از تلاش‌هایی که برای حل مساله مورد مطالعه در پایان‌نامه شده است آشنا نمود و نیز تلاش نمود تا خواننده اهمیت کار انجام شده در پایان‌نامه را دریابد. این قسمت، گرچه در نگاه نخست به ظاهر بسیار ساده می‌نماید، اما در حقیقت یکی از مهمترین قسمت‌های پایان‌نامه است، زیرا بازتاب دهنده دانسته‌ها و فهم دانشجو از کلیت مساله است. توصیه می‌کنیم که نوشتن این قسمت را به آخر موکول نمایید!

یک نکته مهم: اگر پایان‌نامه بر اساس یک یا چند مقاله نوشته شده است، آنگاه از دانشجو انتظار می‌رود که مراجع اصلی خود را در پیشگفتار معرفی نماید.

معمولاً بخش آخر پیشگفتار به ارائه یک نمایه کلی از پایان‌نامه اختصاص می‌یابد و نگارنده در پاراگراف‌های مجزا، به معرفی کوتاهی از فصل‌بندی و کار انجام شده در هر فصل پایان‌نامه می‌پردازد.

# فهرست مطالب

یک	چکیده
یک	پیشگفتار
دو	فهرست مطالب
سه	فهرست شکل‌ها
چهار	فهرست جداول
۱	۱ مفاهیم مقدماتی
۲	۱.۱ بخش .....
۵	۲ عنوان فصل
۶	۳ عنوان فصل
۷	۴ عنوان فصل
۸	۵ نتیجه‌گیری و جمع‌بندی
۸	۱.۵ جمع‌بندی .....
۸	۲.۵ پیشنهادات .....
۱۱	مراجع
۱۲	پیوست: کدهای R
۱۴	واژه‌نامه فارسی - انگلیسی

# فهرست شکل‌ها

۲	۱.۱	مصورسازی گرافی یک مدل مارکوف پنهان
	۲.۱	پایه B-اسپلاین درجه سه با استفاده از نه گره با فاصله‌های برابر در بازه
۳		(0, 1)
۴	۳.۱	نمایی از معادله بازگشتی تابع پایه B-اسپلاین درجه ۴

# فهرست جداول

۱.۴	جدول مقایسه توزیع گسیل ناپارامتری و نرمال و نرمال آمیخته در مدل
۷	مارکوف پنهان .....

# فصل ۱

## مفاهیم مقدماتی

فصل اول پایان‌نامه به بیان تعاریف و مفاهیم اولیه اختصاص دارد. معمولاً نمادگذاری و معرفی کلیدواژه‌های به کار رفته در پایان‌نامه در این فصل انجام می‌پذیرد و خواننده در صورت عدم آشنایی با مفهومی، می‌تواند به این فصل مراجعه نماید.

توجه کنید که لزومی به اثبات همه مطالب و قضیه‌های مقدماتی، در این فصل، نیست و اصولاً چنین کاری شیوه مرسوم نیست. معمولاً انتخاب و ارجاع خواننده به یک کتاب معتبر یا هر مرجع معتبر دیگری بسیار مناسب‌تر است. چنین فصلی، جایی مناسب برای قضیه‌های «کلاسیک» درباره مساله مورد مطالعه است. توجه کنید که یک قضیه می‌تواند صورت‌های مختلفی داشته باشد و بهتر است از آن صورت مورد استفاده خود در این قسمت استفاده نمایید!

تعداد صفحات متوسط برای هر فصل بین ۱۵ تا ۲۵ صفحه و تعداد صفحات معمول برای پایان‌نامه بین ۷۰ تا ۱۰۰ صفحه است. به طور معمول پایان‌نامه‌های استاندارد دارای ۴ یا ۵ فصل هستند که در فصل اول مقدمات، در فصل‌های دوم و سوم مطالب نظری اصلی و در فصل چهارم برای رشته‌هایی نظیر آمار، آزمایش‌های عددی و تحلیل داده‌ها آورده می‌شود. فصل پنجم می‌تواند یک فصل کوتاه دو یا سه صفحه‌ای شامل جمع‌بندی، نتیجه‌گیری و پیشنهادات آینده باشد.

برای زیرنویس معمولاً از دستوری به این صورت<sup>۱</sup> استفاده می‌شود. تعداد زیرنویس‌ها نبایستی خیلی زیاد باشد و بهتر است تنها به عنوان‌های تخصصی زیرنویس داد و مابقی واژگان را در بخش واژه‌نامه فارسی به انگلیسی در انتهای پایان‌نامه آورد. فرمول‌های داخل متن به عنوان نمونه به صورت  $\{Y_t\}_{t=1}^T$  و فرمول‌های بین خط‌ها به عنوان نمونه به صورت زیر

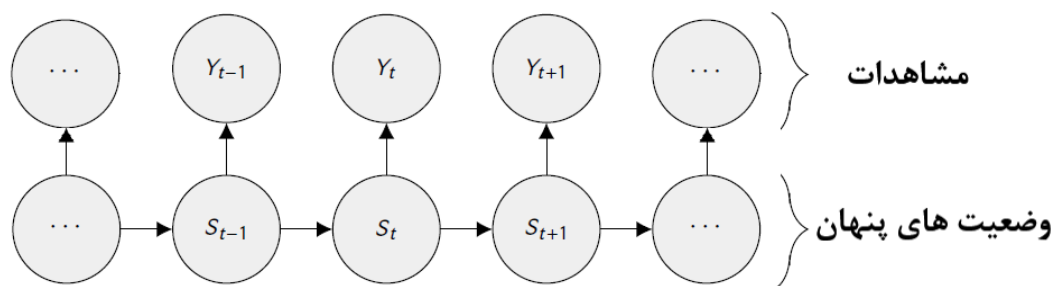
<sup>1</sup>A footnote

می آیند

$$\Pr(\mathbf{s}, \mathbf{y}) = \Pr(s_1) \Pr(y_1 | s_1) \prod_{t=2}^T \Pr(s_t | s_1, \dots, s_{t-1}) \Pr(y_t | y_1, \dots, y_{t-1}, s_1, \dots, s_t) \quad (1.1)$$

و برای ارجاع دادن به این فرمول از (۱.۱) استفاده می شود. رعایت نیم فاصله ها در تمامی متن توصیه می شود. همچنین نقطه و کاما بایستی چسبیده به انتهای کلمه ماقبل و با فاصله از ابتدای کلمه بعد باشند.

برای شروع یک پاراگراف جدید حتما بایستی یک خط فاصله خالی بین پاراگراف قبلی و بعدی بگذارید. برای آوردن شکل به صورت شکل ۱.۱ عمل می کنیم.



شکل ۱.۱: مصورسازی گرافی یک مدل مارکوف پنهان

توجه کنید که اگر شکل را از منبعی به امانت گرفته اید بایستی حتما در زیرنویس شکل به آن منبع ارجاع دهید. برای ایجاد یک بخش جدید به صورت زیر عمل می کنیم.

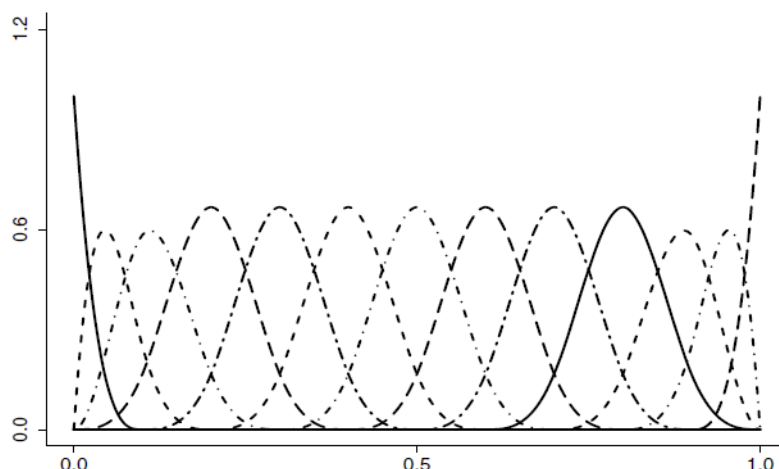
## ۱.۱ بخش

همینطور از نماد دو نقطه در فارسی تنها برای لیستی به صورت زیر استفاده می شود و قبل از فرمول ها استفاده از دونقطه توصیه نمی شود:

۱. مورد اول

۲. مورد دوم ...

برای ایجاد یک فرمول چند خطی می توانید به صورت زیر عمل کنید



شکل ۲.۱: پایه B-اسپلاین درجه سه با استفاده از نه گره با فاصله‌های برابر در بازه  $(0, 1)$

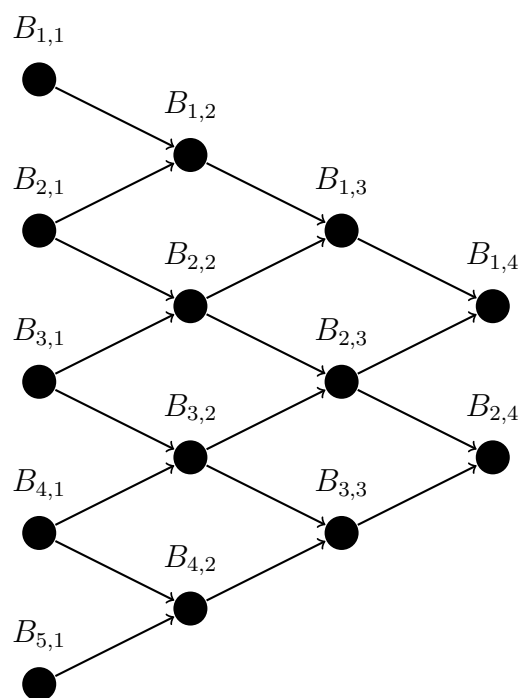
$$\begin{aligned} A &= B \\ &= C \end{aligned} \quad (2.1)$$

برای ارجاع دادن به یک مرجع در متن به عنوان نمونه می‌توان به صورت [۱۴، ۱۶] عمل کرد.

**تعریف ۱.۱.** عدد طبیعی  $p$ ،  $p \neq 1$ ، را اول گوییم، هرگاه به غیر از خودش و یک مقسوم علیه دیگری نداشته باشد.

در این بخش یک شکل دیگر نیز برای نمونه می‌آوریم تا فهرست شکل‌ها کمی پربارتر باشد.

گاهی نیاز دارید تا یک شکل با کیفیت را در خود لایک ایجاد کنید. نمونه‌ای از این کار در شکل ۳.۱ آورده شده است. توجه کنید که ارجاع به فرمول‌ها با پرانتز و ارجاع به شکل‌ها و جدول‌ها بدون پرانتز است.



شکل ۳.۱: نمایی از معادله بازگشتی تابع پایه B-اسپلاین درجه ۴



## فصل ۲

# عنوان فصل

در این فصل به مطالب اصلی نظری پایان‌نامه پرداخته می‌شود. ممکن است در این فصل هنوز به مدل‌های اصلی و ایده‌های اصلی مقاله اصلی پایان‌نامه اشاره نشود، ولی دیگر مفاهیم و تعاریف اولیه در این فصل گفته نمی‌شوند و بیش‌تر این فصل می‌تواند در برگیرنده قضایا و مطالب پژوهشی و یا شرح مدل‌های قبلی و روش‌های پیشین باشند که قرار است با روش نظری مقاله اصلی مقایسه شوند. البته این مساله در رشته‌های گوناگون می‌تواند متفاوت باشد ولی در کل شباهت‌هایی بین ساختارهای گوناگون در رشته‌های گوناگون از این منظر وجود دارد.

## فصل ۳

### عنوان فصل

فصل سوم معمولا اوج پایان نامه و پرداختن به مطالب اصلی مقاله اصلی پایان نامه است.

## فصل ۴

# عنوان فصل

در این فصل به مطالب تکمیلی و اثبات‌های قضایای اصلی و یا کاربردها پرداخته می‌شود. در رشته‌هایی نظیر آمار، این فصل بیش‌تر به تحلیل داده‌ها، شبیه‌سازی و یا مثال‌های عددی می‌پردازد و می‌تواند عنوانی مانند ”آزمایش‌های عددی و تحلیل داده‌ها” داشته باشد. در این فصل معمولاً جدول‌هایی نیز آورده می‌شود که نمونه‌ای از آن در جدول ۱.۴ آورده شده است. این جدول‌ها در فهرست جدول‌ها آورده می‌شوند.

توزیع گسیل	همگنی (%)	AIC	BIC
ناپارامتری	۸۰، ۸۸/۹، ۸۶/۲	۶۴۱۹/۰۶۳	۶۸۰۵/۵۲۵
نرمال	۵۸/۳، ۹۹/۱، ۸۱/۱	۷۵۶۰/۳۴۶	۷۶۶۱/۱۶۲
نرمال آمیخته	۵۸/۳، ۹۹/۱، ۸۱/۱	۷۵۹۳/۳۷۶	۷۷۹۵/۰۰۸

جدول ۱.۴: جدول مقایسه توزیع گسیل ناپارامتری و نرمال و نرمال آمیخته در مدل مارکوف پنهان

انتظار می‌رود در این فصل تمامی جنبه‌های عددی روش‌های ارائه شده بررسی شده و این روش‌ها با روش‌های دیگر بر اساس ملاک‌های گوناگون مقایسه شوند.

## فصل ۵

# نتیجه‌گیری و جمع‌بندی

در این فصل، ابتدا به جمع‌بندی مباحث مطرح شده در مورد روش‌های مورد مطالعه می‌پردازیم. سپس برای پژوهش‌های آینده پیشنهادهایی را مطرح خواهیم کرد که می‌تواند برای پژوهش‌گران علاقه‌مند به این حوزه برای توسعه روش‌های جدیدتر یا استفاده از روش‌های موجود در چالش‌های کاربردی دیگر ارائه دهد.

### ۱.۵ جمع‌بندی

در این بخش یک جمع‌بندی از مطالب پایان‌نامه و نتایج ارائه می‌شود.

### ۲.۵ پیشنهادات

در ادامه برخی پیشنهادات برای مطالعات آینده ارائه می‌شود.

# مراجع

- [1] M. Amini, A. Bayat, and R. Salehian. hhsmm: an r package for hidden hybrid markov/semi-markov models. *Computational Statistics*, pages 1–53, 2022.
- [2] H. J. Bierens. The nadaraya-watson kernel regression function estimator. 1988.
- [3] E. Cantoni, J. M. Flemming, and E. Ronchetti. Variable selection in additive models by non-negative garrote. *Statistical modelling*, 11(3):237–252, 2011.
- [4] J. G. De Gooijer, G. E. Henter, and A. Yuan. Kernel-based hidden markov conditional densities. *Computational Statistics & Data Analysis*, 169:107431, 2022.
- [5] P. H. Eilers and B. D. Marx. Flexible smoothing with b-splines and penalties. *Statistical science*, 11(2):89–121, 1996.
- [6] S. Frühwirth-Schnatter and S. Frühwirth-Schnatter. *Finite mixture and Markov switching models*, volume 796. Springer, 2006.
- [7] S. M. Goldfeld and R. E. Quandt. A markov model for switching regressions. *Journal of Econometrics*, 1:3–16, 1973.

- [8] J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the econometric society*, pages 357–384, 1989.
- [9] G. Kauermann. A note on smoothing parameter selection for penalized spline smoothing. *Journal of statistical planning and inference*, 127(1-2):53–69, 2005.
- [10] G. Kauermann, C. Schellhase, and D. Ruppert. Flexible copula density estimation with penalized hierarchical b-splines. *Scandinavian Journal of Statistics*, 40(4):685–705, 2013.
- [11] C.-J. Kim, J. Piger, and R. Startz. Estimation of markov regime-switching regression models with endogenous switching. *Journal of Econometrics*, 143(2):263–273, 2008.
- [12] R. Langrock, T. Adam, V. Leos-Barajas, S. Mews, D. L. Miller, and Y. P. Papastamatiou. Spline-based nonparametric inference in general state-switching models. *Statistica Neerlandica*, 72(3):179–200, 2018.
- [13] R. Langrock, T. Kneib, R. Glennie, and T. Michelot. Markov-switching generalized additive models. *Statistics and Computing*, 27(1):259–270, 2017.
- [14] R. Langrock, T. Kneib, A. Sohn, and S. L. DeRuiter. Nonparametric inference in hidden markov models using p-splines. *Biometrics*, 71(2):520–528, 2015.
- [15] G. Marra and S. N. Wood. Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis*, 55(7):2372–2387, 2011.
- [16] J. Pohle, R. Langrock, F. M. van Beest, and N. M. Schmidt. Selecting the number of states in hidden markov models: pragmatic solutions illustrated using animal move-

ment. *Journal of Agricultural, Biological and Environmental Statistics*, 22(3):270–293, 2017.

- [17] P. Wang and M. L. Puterman. Markov poisson regression models for discrete time series. part 1: Methodology. *Journal of Applied Statistics*, 26(7):855–869, 1999.

# پیوست: کدهای R

## Nonparametric M-step

```
1 rm(list = ls())
2 library(psych)
3 library(cpr)
4 library(MASS)
5 nonpar_mstep = function (x, wt,
6   control = list(K = 5, lambda0 = 0.5))
7 {
8   defcon <- list(K = 5, lambda0 = 0.5)
9   control <- modifyList(defcon, control)
10  K <- control$K
11  lambda0 <- control$lambda0
12  nstate <- ncol(wt)
13  emission <- list(coef = list())
14  lambda <- numeric(nstate)
15  d <- ncol(x)
16  n <- nrow(x)
17  tryCatch({
18    a <- matrix(0, nrow = n, ncol = K^d)
19    if (object.size(a) > 1.8e+09)
20      warning("The dimension of the data or
21  the degree of the spline is large!
22  \n\t\tThis will result in a very slow progress!")
23    rm(a)
24  }, error = function(cond) {
25    stop("The dimension of the data or
26  the degree of the spline is too large!
27  \n\t\tThere is no enough memory for fitting!
28  Try another emission distribution.")
29  })
30  basis = btensor(lapply(1:d, function(i) x[, i]), df = K,
31    bknots = lapply(1:d, function(i)
32      c(min(x[, i]) - 0.01,
33        max(x[, i]) + 0.01)))
34  for (j in 1:nstate) {
35    lambda[j] <- lambda0
36    mloglike_lambda0 <- function(beta) {
37      dbeta <- diff(beta, differences = 2)
```



```

38     omega <- exp(beta)/sum(exp(beta))
39     loglike <- t(wt[, j]) %*% log(basis %*% omega) -
40       lambda0/2 * sum(dbeta^2)
41     return(-loglike)
42   }
43   start <- runif(K^d)
44   suppressWarnings(fit <- nlm(mloglike_lambda0, start,
45     hessian = T))
46   H_lambda0 <- -fit$hessian
47   difference <- 1
48   eps <- 1e-06
49   cntr <- 1
50   beta_hat <- list(rep(1, K))
51   while (difference > eps) {
52     mloglike <- function(beta) {
53       dbeta <- diff(beta, differences = 2)
54       omega <- exp(beta)/sum(exp(beta))
55       inf_index <- which(is.infinite(log(basis %*%
56         omega)))
57       loglike <- t(wt[, j]) %*% log(basis %*% omega) -
58         lambda[j]/2 * sum(dbeta^2)
59       return(-loglike)
60     }
61     start <- runif(K^d)
62     suppressWarnings(fit <- nlm(mloglike, start, hessian = T))
63     H <- -fit$hessian
64     beta_hat[[cntr + 1]] <- fit$estimate
65     df_lambda <- tr(ginv(H) %*% H_lambda0)
66     dbeta <- diff(beta_hat[[cntr + 1]], differences = 2)
67     lambda[j] <- (df_lambda - d)/(sum(dbeta^2))
68     difference <- sum(beta_hat[[cntr + 1]] - beta_hat[[cntr]])
69     cntr <- cntr + 1
70   }
71   emission$coef[[j]] <- exp(beta_hat[[cntr]])/
72     sum(exp(beta_hat[[cntr]]))
73 }
74 emission
75 }

```

# واژه‌نامه فارسی - انگلیسی

Backward Algorithm	الگوریتم پسرو
Step Function	تابع پله‌ای
Transition Probability Matrix	ماتریس احتمال انتقال

### **Abstract**

The English abstract should match the persion one once traslated.

**Keywords:** The english keywords should match the persian ones once translated.



University of Tehran  
College of Science  
School of Mathematics, Statistics, and Computer Science

Insert the english title of the dissertation

**By**

Insert your name and surname here

**Supervisor(s)**

Insert the name(s) of supervisor(s) here

**Advisor**

Insert the name(s) of Advisor here

A thesis submitted to graduate studies office in partial fulfillment of the requirements for the  
degree of master of science in Mathematics and Applications / Applied Mathematics / Statistics /  
Computer Science

Date