بسم الله الرحمن الرحيم

دانشگاه تربیت مدرس

دانشکده علوم ریاضی

بسمه تعالی

تأییدیه اعضای هیأت داوران حاضر در جلسه دفاع از پایان‌نامه کارشناسی‌ارشد

اعضای هیأت داوران نسخهٔ نهایی پایان‌نامه آقای **مهدی بیگی‌اسفتانج** رشته **آمار** به شماره دانشجویی ۹۰۵۲۰۲۱۰۰۲ تحت عنوان: «**معرفی خانواده‌های پارامتری از توزیع‌های احتمال برای داده‌های تابعی با استفاده از توابع مفصل**» را در تاریخ ۱۳۹۲/۱۲/۱۱ از نظر فرم و محتوا بررسی نموده و آن را برای اخذ درجهٔ کارشناسی‌ارشد مورد تأیید قرار دادند.

اعضای هیأت داوران	نام و نام‌خانوادگی	رتبه علمی	امضاء
۱- استاد راهنما	دکتر محمدرضا فریدروحانی	استادیار	
۲- استاد مشاور	دکتر محسن محمدزاده	استاد	
۳- استاد ناظر داخلی	دکتر مجید جعفری‌خالدی	استادیار	
۴- استاد ناظر خارجی	دکتر صدیقه شمس	دانشیار	
۵- نماینده تحصیلات تکمیلی	دکتر مجید جعفری‌خالدی	استادیار	

آیین نامه چاپ پایان‌نامه (رساله)های دانشجویان دانشگاه تربیت مدرس

نظر به اینکه چاپ و انتشار پایان نامه (رساله)های تحصیلی دانشجویان دانشگاه تربیت مدرس، مبین بخشی از فعالیتهای علمی - پژوهشی دانشگاه است بنابراین به منظور آگاهی و رعایت حقوق دانشگاه،دانش آموختگان این دانشگاه نسبت به رعایت موارد ذیل متعهد می‌شوند:

ماده ۱: در صورت اقدام به چاپ پایان نامه (رساله)ی خود، مراتب را قبلاً به طور کتبی به «دفتر نشر آثارعلمی» دانشگاه اطلاع دهد.

ماده ۲: در صفحه سوم کتاب (پس از برگ شناسنامه) عبارت ذیل را چاپ کند:

«کتاب حاضر، حاصل پایان نامه کارشناسی ارشد/ رساله دکتری نگارنده در رشته

است که در سال   در دانشکده

دانشگاه تربیت مدرس به راهنمایی سرکار خانم/جناب آقای دکتر   ، مشاوره

سرکار خانم/جناب آقای دکتر   و مشاوره سرکار خانم/جناب آقای دکتر

از آن دفاع شده است.»

ماده ۳: به منظور جبران بخشی از هزینه‌های انتشارات دانشگاه، تعداد یک درصد شمارگان کتاب (در هر نوبت‌چاپ) را به «دفتر نشر آثارعلمی» دانشگاه اهدا کند. دانشگاه می‌تواند مازاد نیاز خود را به نفع مرکز نشر درمعرض فروش قرار دهد.

ماده ۴: در صورت عدم رعایت ماده ۳، ۵۰٪ بهای شمارگان چاپ شده را به عنوان خسارت به دانشگاه تربیت‌مدرس، تأدیه کند.

ماده ۵: دانشجو تعهد و قبول می کند در صورت خودداری از پرداخت بهای خسارت، دانشگاه می‌تواند خسارت مذکور را از طریق مراجع قضایی مطالبه و وصول کند؛ به علاوه به دانشگاه حق می‌دهد به منظور استیفای حقوق خود، از طریق دادگاه، معادل وجه مذکور در ماده ۴ را از محل توقیف کتابهای عرضه شده نگارنده برای فروش، تامین نماید.

ماده ۶: اینجانب **مهدی بیگی اسفتانج** دانشجوی رشته **آمارریاضی** مقطع **کارشناسی‌ارشد** تعهد فوق وضمانت اجرایی آن را قبول کرده، به آن ملتزم می شوم.

نام و نام خانوادگی: مهدی بیگی اسفتانج

تاریخ و امضا: ۹۲/۱۲/۱۷

آیین‌نامه حق مالکیت مادی و معنوی در مورد نتایج پژوهشهای علمی دانشگاه تربیت مدرس

مقدمه: با عنایت به سیاست‌های پژوهشی و فناوری دانشگاه در راستای تحقق عدالت و کرامت انسانها که لازمه شکوفایی علمی و فنی است و رعایت حقوق مادی و معنوی دانشگاه و پژوهشگران، لازم است اعضای هیأت علمی، دانشجویان، دانش‌آموختگان و دیگر همکاران طرح، در مورد نتایج پژوهشهای علمی که تحت عناوین پایان‌نامه، رساله و طرحهای تحقیقاتی با هماهنگی دانشگاه انجام شده است، موارد زیر را رعایت نمایند:

ماده ۱- حق نشر و تکثیر پایان نامه/ رساله و درآمدهای حاصل از آنها متعلق به دانشگاه می باشد ولی حقوق معنوی پدید آورندگان محفوظ خواهد بود.

ماده ۲- انتشار مقاله یا مقالات مستخرج از پایان‌نامه/ رساله به صورت چاپ در نشریات علمی و یا ارائه در مجامع علمی باید به نام دانشگاه بوده و با تایید استاد راهنمای اصلی، یکی از اساتید راهنما، مشاور و یا دانشجو مسئول مکاتبات مقاله باشد. ولی مسئولیت علمی مقاله مستخرج از پایان نامه و رساله به عهده اساتید راهنما و دانشجو می باشد.

تبصره: در مقالاتی که پس از دانش‌آموختگی بصورت ترکیبی از اطلاعات جدید و نتایج حاصل از پایان‌نامه/ رساله نیز منتشر می‌شود نیز باید نام دانشگاه درج شود.

ماده ۳- انتشار کتاب، نرم افزار و یا آثار ویژه (اثری هنری مانند فیلم، عکس، نقاشی و نمایشنامه) حاصل از نتایج پایان‌نامه/ رساله و تمامی طرحهای تحقیقاتی کلیه واحدهای دانشگاه اعم از دانشکده ها، مراکز تحقیقاتی، پژوهشکده ها، پارک علم و فناوری و دیگر واحدها باید با مجوز کتبی صادره از معاونت پژوهشی دانشگاه و براساس آئین نامه های مصوب انجام شود.

ماده ۴- ثبت اختراع و تدوین دانش فنی و یا ارائه یافته ها در جشنواره‌های ملی، منطقه‌ای و بین‌المللی که حاصل نتایج مستخرج از پایان‌نامه/ رساله و تمامی طرحهای تحقیقاتی دانشگاه باید با هماهنگی استاد راهنما یا مجری طرح از طریق معاونت پژوهشی دانشگاه انجام گیرد.

ماده ۵- این آیین‌نامه در ۵ ماده و یک تبصره در تاریخ ۸۷/۴/۱ در شورای پژوهشی و در تاریخ ۸۷/۴/۲۳ در هیأت رئیسه دانشگاه به تایید رسید و در جلسه مورخ ۸۷/۷/۱۵ شورای دانشگاه به تصویب رسیده و از تاریخ تصویب در شورای دانشگاه لازم‌الاجرا است.

«اینجانب **مهدی بیگی اسفتانج** دانشجوی رشته **آمارریاضی** ورودی **نیمسال اول** سال تحصیلی ۹۰-۹۱ مقطع **کارشناسی ارشد** دانشکده **علوم ریاضی** متعهد می شوم کلیه نکات مندرج در آئین نامه حق مالکیت مادی و معنوی در مورد نتایج پژوهش های علمی دانشگاه تربیت مدرس را در انتشار یافته های علمی مستخرج از پایان نامه / رساله تحصیلی خود رعایت نمایم. در صورت تخلف از مفاد آئین نامه فوق الاشعار به دانشگاه وکالت و نمایندگی می دهم که از طرف اینجانب نسبت به لغو امتیاز اختراع بنام بنده و یا هر گونه امتیاز دیگر و تغییر آن به نام دانشگاه اقدام نماید. ضمناً نسبت به جبران فوری ضرر و زیان حاصله بر اساس برآورد دانشگاه اقدام خواهم نمود و بدینوسیله حق هر گونه اعتراض را از خود سلب نمودم»

امضا:

تاریخ: ۹۲/۱۲/۱۷

دانشگاه تربیت مدرس

دانشکده علوم ریاضی

پایان نامه دوره کارشناسی ارشد آمار

معرفی خانواده‌های پارامتری از توزیع‌های احتمال برای داده‌های تابعی با استفاده از توابع مفصل

توسط

مهدی بیگی

استاد راهنما

دکتر محمدرضا فرید روحانی

استاد مشاور

دکتر محسن محمدزاده

اسفند ۱۳۹۲

قدردانی

سپاس و ستایش معبود یگانه را که پرتو الطاف بی‌شمارش بر لحظه لحظهٔ زندگی‌ام ساطع و آشکار است. حمد و ثنا می‌گزارم او را که فکرت و اندیشه را در بستر روحم روان ساخت و بهره‌گیری از خوان گستردهٔ دانش اساتیدم را نصیب و روزی‌ام گردانید.

امتنان و سپاس می‌گزارم تلاشها، زحمات و راهنمایی‌های ظریف، ارزشمند و بی‌شائبهٔ اساتید فرزانه و گرانمایه‌ام، جناب دکترمحمدرضا فرید روحانی و جناب دکتر محسن محمدزاده را که با حمیت و جدیت، مرا به دقت، اندیشه، درک و تعمق وامی‌داشتند.

مهدی بیگی

دی ۱۳۹۲

تقدیم به بهترین واژگان حیات

پدر و مادر مهربانم

چکیده

داده‌هایی که ذاتاً تابعی پیوسته از یک متغیر دیگر مانند زمان، باشند داده‌های تابعی نامیده می‌شوند. خانواده‌های توزیع احتمال پارامتری ابزار حیاتی برای مدل‌بندی احتمالاتی و تحلیل داده‌ها محسوب می‌شوند. با این وجود در تحلیل داده‌های تابعی این خانواده از توزیع‌ها کمتر مورد مطالعه قرار گرفته است. در این پایان‌نامه یک خانواده توزیع احتمال پارامتری برای داده‌های تابعی مورد بررسی قرار گرفته است. توزیع احتمال مورد نظر با استفادهٔ همزمان امیانگین شبه‌حسابی و تابع مفصل ارشمیدسی به‌دست می‌آید. علاوه بر این، با استفاده از مشتق گاتو، تابع چگالی متناظر با آن نیز تعیین می‌شود. سپس این توزیع احتمال برای خوشه‌بندی داده‌های تابعی شبیه‌سازی شده بر اساس روش تجزیه آمیخته، و داده‌های واقعی به کار گرفته می‌شود. در انتها بحث و نتیجه‌گیری ارائه می‌شود.

واژه‌های کلیدی : میانگین شبه‌حسابی، تابع مفصل ارشمیدسی، آنالیز داده‌های تابعی، مشتق گاتو، تجزیه آمیخته.

فهرست مندرجات

لیست اشکال

فصل ۱

تعاریف و مفاهیم اولیه

رشد روز افزون حجم داده‌های ذخیره شده در پایگاه اطلاعاتی و تولید داده‌های با ماهیت تابعی نیاز به ابزارهای ریاضی برای مدل‌بندی این گونه داده‌ها را افزایش داده است. در واقع با پیشرفت فناوری‌های ثبت و ذخیره‌سازی داده‌ها، دیگر هیچ محدودیتی برای حجم اطلاعات ذخیره شده وجود ندارد و جداول داده‌ها هم از نظر تعداد واحدهای ذخیره و ثبت شده و هم از نقطه نظر ابعاد واحدهای مطالعاتی، به شدت گسترش یافته‌اند.

از آنجا که روش‌های کلاسیک تحلیل داده‌ها برای داده‌های حجیم که در آنها تعداد واحدها کوچکتر از بعد متغیر مورد نظر است، مناسب نیستند لذا مطالعه و بررسی روش‌هایی برای خلاصه‌سازی داده‌های زیاد به گونه‌ای که کمترین میزان اطلاعات از دست برود، بیش از پیش ضرورت می‌یابد.

می‌توان تابع توزیع تجمعی را به عنوان تلخیصی کامل و کارا از داده‌های به‌دست آمده در هریک از گروه‌های مورد مطالعه در نظر گرفت. اگرچه ساخت چنین تلخیصی برای هر گروه به آسانی صورت می‌گیرد اما ضروری است تا ابزارهای تحلیل داده مناسبی برای این نوع داده‌ها بسط و تعمیم داده شود. از این‌رو در بخشی از این پایان‌نامه به معرفی نوعی از داده‌ها به عنوان داده‌های نمادی پرداخته و

خوشه‌بندی آنها مورد مطالعه قرار می‌گیرد.

از طرفی توابع توزیع پیوسته را اگر چه می‌توان به عنوان داده‌های نمادی در نظر گرفت، اما می‌توانند به عنوان داده‌های تابعی نیز منظور گردند. از این‌رو در بخش دیگری از پایان نامه در مورد ساخت تابع توزیع احتمال داده‌های تابعی و استفاده از آن به منظور تحلیل این نوع داده‌ها به ویژه خوشه‌بندی آنها می‌پردازیم.

در فصل اول با مفاهیمی همچون مشتق گاتو، تابع مفصل و داده‌های تابعی به اختصار آشنا می‌شویم. در فصل دوم با داده‌های نمادی و نیز یکی از روش‌های خوشه‌بندی این نوع داده‌ها با عنوان تجزیه آمیخته، که روشی مبتنی بر تابع توزیع احتمال است، آشنا خواهیم شد.

در فصل سوم به منظور ساخت تابع توزیع احتمال برای داده‌های تابعی ابتدا براساس توابع توزیع متناهی بعد حاصل از توابع مفصل، تابع توزیعی را معرفی کرده و نقاط ضعف آن مورد بررسی قرار می‌گیرد. سپس برای اجتناب از نقایص بیان شده در روش مزبور از مفهوم میانگین شبه‌حسابی به همراه تابع مفصل ارشمیدسی برای ساخت تابع توزیع تابعی استفاده می‌کنیم. در نهایت با استفاده از مشتق گاتو، تابع چگالی گاتوی متناظر با تابع توزیع تابعی ساخته شده را به‌دست می‌آوریم.

در فصل چهارم برای مجموعه‌ی داده شبیه‌سازی شده، تابع توزیع و تابع چگالی گاتو را به‌دست می‌آوریم. سپس با استفاده از تابع چگالی مزبور داده‌های شبیه‌سازی شده را به روش تجزیه آمیخته خوشه‌بندی می‌کنیم. نتایج حاصل نشان می‌دهد که تابع توزیع احتمالاتی بررسی شده در این پایان نامه به صورت مفیدی قابل استفاده در روش‌هایی همچون تجزیه آمیخته در تحلیل داده‌های تابعی هستند.

۱.۱ مشتق گاتو

از آنجا که در این پایان‌نامه با توابع چگالی متغیرهای نامتناهی بعد در فضای توابع انتگرال‌پذیر مربعی [۱] روی زیرمجموعه‌ای از R سروکار داریم، برای به‌دست آوردن این توابع چگالی به تعمیمی از مشتق سوئی [۲] به نام مشتق گاتو [۳] نیاز است، که در ادامه به اختصار معرفی می‌شود.

یک تابع دو متغیره $f(x,y)$ از R^2 به R را در نظر بگیرید، که می‌تواند در تمام جهت‌های موجود بر روی صفحه xy دارای تغییرات باشد. برای بررسی آهنگ تغییرات این تابع در هریک از جهت‌ها، ابتدا ساده‌ترین حالت، یعنی دو جهت x ـ مثبت و y ـ مثبت را در نظر می‌گیریم.
آهنگ تغییرات تابع f در هرکدام از این مسیرها را مشتق جزئی [۴] تابع f به ترتیب نسبت به x و y نامیده و با $\frac{\partial f}{\partial x}$ و $\frac{\partial f}{\partial y}$ نشان می‌دهند، که هریک از آن‌ها به صورت

$$\frac{\partial f}{\partial x} = \lim_{\triangle x \to 0} \frac{f(x+\triangle x, y) - f(x,y)}{\triangle x} \tag{۱.۱.۱}$$

تعریف می‌شود.

برای تعریف مشتق گاتو ابتدا مروری خلاصه بر فضای باناخ [۵] به عنوان یکی از فضاهای مجرد، انجام می‌دهیم. فرض کنید $\mathcal{X}$ مجموعه‌ای ناتهی باشد که تحت دو عملگر جمع و ضرب اسکالر بسته باشد، یعنی

$$\forall x_1, x_2 \in \mathcal{X}; \quad x_1 + x_2 \in \mathcal{X}$$

[۱] Square integrable functions

[۲] Directional derivative

[۳] Gateaux derivative

[۴] Partial derivative

[۵] Banach space

و

$$\forall c \in R, \forall x \in \mathcal{X}; \quad c.x \in \mathcal{X}$$

آنگاه فضای خطی حقیقی را به صورت زیر تعریف می‌کنیم.

تعریف ۱.۱.۱ : مجموعه $\mathcal{X}$ با دو عملگر جمع و ضرب اسکالر، فضای خطی حقیقی[۶] نامیده می‌شود اگر شرایط زیر برقرار باشند:

۱) $\forall x_1, x_2 \in \mathcal{X}; \quad x_1 + x_2 = x_2 + x_1$.

۲) $\forall x_1, x_2, x_3 \in \mathcal{X}; \quad (x_1 + x_2) + x_3 = x_1 + (x_2 + x_3)$.

۳) یک عنصر یکتای خنثی در $\mathcal{X}$ وجود دارد که با نماد $o_{\mathcal{X}}$ نمایش داده می‌شود و برای هر $x \in \mathcal{X}$ رابطه $x + o_{\mathcal{X}} = x$ برقرار است.

۴) برای هر $x \in \mathcal{X}$ عنصر یکتایی در $\mathcal{X}$ وجود دارد که با نماد $-x$ نمایش می‌دهیم به طوری که $x + (-x) = o_{\mathcal{X}}$

۵) $\forall c \in R, \forall x_1, x_2 \in \mathcal{X}; \quad c.(x_1 + x_2) = c.x_1 + c.x_2$

۶) $\forall c_1, c_2 \in R, \forall x_1 \in \mathcal{X}; \quad (c_1 + c_2).x = c_1.x + c_2.x$

۷) $\forall c_1, c_2 \in R, \forall x_1 \in \mathcal{X}; \quad c_1.(c_2.x) = (c_1 \times c_2).x$

۸) $\forall x_1 \in \mathcal{X}; \quad 1.x = x$

---

[۶] Real linear space

ذکر این نکته ضروری است که اعضای یک فضای خطی می‌تواند ماتریس، تابع یا بردارهای n بعدی عضو R^n باشند. بنابراین نباید طبق آنچه که به اشتباه مرسوم شده است، فضای برداری را معادل با فضای خطی بدانیم بلکه فضای برداری شامل بردارهای n بعدی در واقع یک فضای خطی است.

تعریف ۲.۱.۱ : فرض کنید $\mathcal{X}$ و $\mathcal{Y}$ دو فضای خطی باشند. تابع $L : \mathcal{X} \rightarrow \mathcal{Y}$ را عملگر خطی گوییم هرگاه:

الف — L جمعی باشد، یعنی $\forall x, y \in \mathcal{X}; \quad L(x+y) = L(x) + L(y)$

ب — L همگن[۷] باشد، یعنی $\forall c \in R, \forall x \in \mathcal{X}; \quad L(c.x) = c.L(x)$

برای تعریف فضای خطی نرم‌دار ابتدا به معرفی مفهوم نرم[۸] در فضای خطی می‌پردازیم.

تعریف ۳.۱.۱ : فرض کنید $\mathcal{X}$ فضای خطی حقیقی، $x \in \mathcal{X}$ و $\|x\|$ نشان دهنده تابعی باشد از $\mathcal{X}$ به R به طوری که

۱) به ازاء هر $x \in \mathcal{X}$ ، $\|x\| \geq ۰$

۲) اگر $\|x\| = ۰$ آنگاه $x = ۰$

۳) به ازاء هر $x \in \mathcal{X}$ و هر $c \in R$ ، $\|c.x\| = |c|.\|x\|$

۴) به ازاء هر x, y در $\mathcal{X}$ ، $\|x+y\| \leq \|x\| + \|y\|$

آنگاه تابع $\|x\|$ را یک نرم در فضای خطی $\mathcal{X}$ می‌نامیم.

تعریف ۴.۱.۱ : اگر فضای خطی حقیقی $\mathcal{X}$ را به یک نرم مجهز کنیم آن را فضای خطی نرم‌دار یا ساده‌تر فضای نرم‌دار می‌گوییم .

---

[۷] Homogenous

[۸] Norm

قضیه ۱.۱.۱ : فرض کنید $\mathcal{X}$ فضای خطی نرم‌دار با نرم $\|.\|$ باشد. در این صورت $\mathcal{X}$ با تابع

$$d : \mathcal{X} \times \mathcal{X} \to R$$

و با ضابطهٔ

$$d(x, y) = \|x - y\|$$

به عنوان متر، فضای متریک است.

اینک می‌توان تمام مفاهیمی را که در فضاهای متریک بیان می‌شود، در مورد فضاهای خطی نرم‌دار نیز بیان کرد. به عنوان مثال مفاهیم دنباله و همگرایی دنباله‌ها را می‌توان نام برد.

تعریف ۵.۱.۱ : اگر $\{x_n\}$ یک دنباله در فضای خطی نرم‌دار $\mathcal{X}$ باشد، گوییم این دنباله به $x \in \mathcal{X}$ همگرا است هرگاه

$$\lim_{n\to\infty} \|x_n - x\| = 0$$

از طرفی دنباله $\{y_n\}$ در فضای خطی نرم‌دار $\mathcal{X}$ را کوشی[۹] نامیم هرگاه Nایی وجود داشته باشد که به ازای هر $m, n \geq N$ رابطه‌ی

$$\lim_{n,m\to\infty} \|y_n - y_m\| = 0$$

برقرار باشد.

یک فضای خطی نرم‌دار$\mathcal{X}$ را کامل گوییم هر گاه هر دنباله کوشی در آن فضا به عنصری مانند $x \in \mathcal{X}$ همگرا باشد.

---
[۹] Cauchy sequence