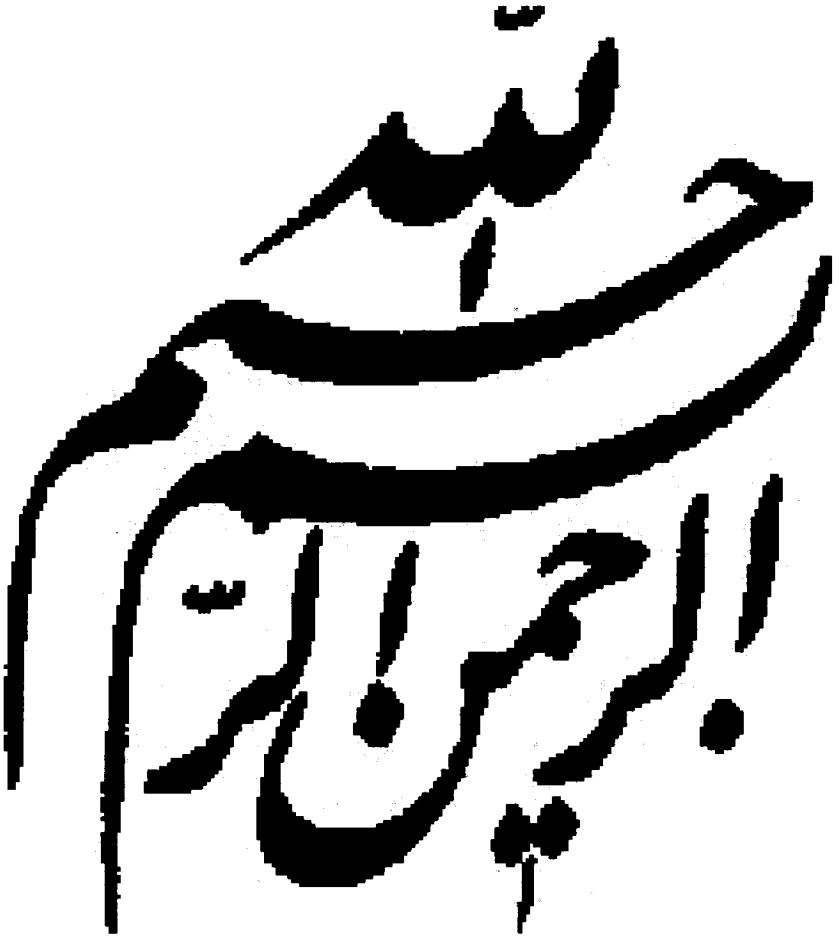
بسم الله الرحمن الرحيم

دانشگاه شهید بهشتی

دانشکده علوم ریاضی

گروه آمار

پایان‌نامه‌ی کارشناسی ارشد آمار ریاضی

عنوان

خوشه‌بندی مدل–پایه‌ای به روش بیزی

نگارش

شیما شهبازی

استاد راهنما

دکتر مجتبی خزایی

استاد مشاور

دکتر مسعود البرز

شهریور ۸۹

تاریخ .....
شماره .....
پیوست.....

دانشگاه شهید بهشتی

صورتجلسه دفاع از پایان نامه دانشجویان دوره کارشناسی ارشد

ن ۱۹۸۳۹۶۳۱۱۳ اوین
: ۲۹۹۰۱

بازگشت به مجوز دفاع شماره ۷۷۰/۷۵۸۰/د مورخ ۸۹/۶/۹ جلسه هیأت داوران ارزیابی پایان نامه **خانم شیما شهبازی** شماره شناسنامه۱۷۴صادره از **تهران** متولد۱۳۶۳دانشجوی **آمار ریاضی** دوره کارشناسی ارشد **آمار ریاضی**

با عنوان:

خوشه‌بندی مدل ـ پایه‌ای به روش بیزی

به راهنمایی:

دکتر مجتبی خزایی

طبق دعوت قبلی در تاریخ ۸۹/۶/۹ تشکیل گردید و براساس رأی هیأت داوری و با عنایت به ماده آئین نامه کارشناسی ارشد مورخ ۷۵/۱۰/۲۵ پایان نامه مزبور با نمره ۱۹٫۲۵ و درجه عالی مورد تصویب قرار گرفت .

نام استاد	مرتبه علمی	نام دانشگاه
۱) **استاد راهنما:** دکتر مجتبی خزایی	استادیار	شهیدبهشتی
۲) **استاد مشاور:** دکتر مسعود البرز	استادیار	شهیدبهشتی
۳) **داور:** دکتر محمدرضا فقیهی	استادیار	شهیدبهشتی
۴) **داور:** دکتر محمدرضا فریدروحانی	استادیار	شهیدبهشتی

نماینده تحصیلات تکمیلی:

تقدیم به

پدر و مادر عزیزم

و همه کسانی که دوستشان دارم.

قدردانی

پس از نام و یاد خدا و تشکر از او به خاطر همه‌ی مهربانی‌هایش، برخود لازم می‌دانم مراتب قدردانی خود را به استاد گرانقدر جناب آقای دکتر مجتبی خزایی، که مرا در تمامی دوره‌ی تحصیلم در دانشگاه شهید بهشتی یاری کردند، اعلام دارم. از آقای دکتر البرز نیز، به خاطر راهنمایی‌های ارزنده‌شان تشکر می‌کنم. همچنین، از اساتید گران‌قدر آقای دکتر فقیهی و آقای دکتر فریدروحانی، به سبب حضور در جمع داوران سپاسگزارم.

در پایان از همه‌ی کسانی که مرا در این راه یاری کردند، قدردانی می‌کنم.

شیما شهبازی

شهریور ۱۳۸۹

- ✓ لوح فشرده‌ی حاوی برنامه‌ها در انتهای پایان‌نامه موجود است.

چکیده

روش‌های خوشه‌بندی سلسله‌مراتبی، از انواع روش‌هایی هستند که جهت خوشه‌بندی اشیاء، براساس میزان عدم تشابه (یا تشابه) بین آن‌ها، به کار می‌روند. در این پایان‌نامه، پس از مرور مختصری بر این روش‌ها، شیوه‌های جدیدتر خوشه‌بندی اشیاء، براساس میزان عدم‌تشابه بین آن‌ها، معرفی می‌شوند. یکی از مزیت‌های این شیوه‌ها نسبت به روش‌های سلسله‌مراتبی، قابلیت آن‌ها در برآورد عدم حتمیت نتایج خوشه‌بندی است. یکی از این شیوه‌ها، خوشه‌بندی مدل—پایه‌ای اشیاء پس از برآورد پیکربندی آن‌ها در فضای اقلیدسی، به کمک روش‌های مقیاس‌بندی چندبعدی کلاسیک یا بیزی است. روش دیگر، روشی بیزی است که در آن، یافتن پیکربندی اشیاء و خوشه‌بندی آن‌ها به طور هم‌زمان انجام می‌گیرد.

به طور کلی، در روش‌های معرفی‌شده در این پایان‌نامه، فرض می‌شود که جامعه دارای چگالی نرمال آمیخته است. به این ترتیب، خوشه‌ای کردن بر پایه‌ی یک مدل احتمالی انجام می‌گیرد. همچنین، فرض می‌شود که اندازه‌های مشاهده‌شده از میزان عدم‌تشابه بین اشیاء، دارای چگالی نرمال بریده‌شده‌اند و با یک رهیافت بیزی، پیکربندی اشیاء برآورد می‌شود. برای انجام این برآورد از روش‌های مونت کارلوی زنجیره‌ی مارکوفی استفاده می‌شود.

این روش‌ها، با استفاده از مطالعات شبیه‌سازی‌شده مقایسه گردیده‌اند. نتایج حاصل از شبیه‌سازی، گویای یک پیکربندی خوب و خوشه‌بندی معقول با اندازه‌های قابل قبولی از عدم حتمیت در روش‌های جدید است. در پایان، با استفاده از این روش‌ها، یک مجموعه داده‌ی واقعی، تحلیل و نتایج آن ارائه شده است.

واژه‌های کلیدی : خوشه‌بندی سلسله‌مراتبی، عدم‌تشابه، مقیاس‌بندی چندبعدی، مدل‌های آمیخته، روش‌های مدل—پایه‌ای، روش‌های بیزی، مونت—کارلوی زنجیره‌ی مارکوفی.

پیش‌گفتار

رشد روزافزون حجم اطلاعات و داده‌ها در دنیای امروز، استخراج تفسیر و تحلیل آن‌ها را، با مشکل مواجه کرده است. بر این اساس، روش‌های چندمتغیره‌ی آماری که به کشف اطلاعات مفید از بانک‌های اطلاعاتی منجر می‌شوند، مورد توجه بسیاری از تحلیل‌گران قرار گرفته‌اند. ازجمله‌ی این روش‌ها، روش‌های خوشه‌بندی و مقیاس‌بندی چندبعدی هستند که استفاده‌ی روزافزون از آن‌ها در علوم بیولوژیک، فیزیک و علوم اجتماعی، مؤید این مطلب است.

روش‌های خوشه‌بندی، روش‌هایی برای شناسایی گروه‌های مشابه از اشیاء در بین داده‌ها هستند که با استفاده از آن‌ها، حجم وسیعی از داده‌ها، به خوبی سازمان‌دهی و خلاصه می‌شوند. به عنوان مثال، فرض کنید یک مجموعه‌ی بزرگ از صفات اندازه‌گیری شده‌ی تعدادی از بیماران افسرده، در دسترس باشد. تشخیص گروه‌های مجزا در این مجموعه، در صورت وجود، می‌تواند بیانگر انواع مختلف بیماری باشد و در انتخاب یک روش مناسب درمانی، مورد استفاده قرار بگیرد. به این ترتیب خوشه‌بندی می‌تواند نقش مهمی را در درمان ایفا کند. یکی دیگر از مثال‌های کاربرد خوشه‌بندی، خوشه‌بندی مشتریان بر اساس داده‌های برگرفته از تعامل آن‌ها با بنگاه‌های تجاری است. به وسیله‌ی خوشه‌بندی مشتریان می‌توان الگوهای رفتاری آن‌ها را کشف کرد و با تحلیل این الگوها، استراتژی‌های بنگاه را برای هر یک از خوشه‌های مشتریان تدوین کرد. از دیگر کاربردهای این روش، گروه‌بندی افراد جامعه براساس صفات مشترک آن‌ها است. بدین منظور رفتارهای متقابل افراد در جامعه، مورد بررسی قرار می‌گیرد و سپس، براساس داده‌ها و اطلاعات جمع‌آوری شده، گروه‌های مجزا تعیین می‌شوند. به این ترتیب با تقسیم جامعه به گروه‌های کوچک‌تر، مطالعه‌ی روابط اجتماعی افراد توسط جامعه‌شناسان آسان‌تر می‌شود.

در تحلیل‌های خوشه‌ای، ممکن است که با دو نوع داده سروکار پیدا کنیم. در نوع اول، داده‌ها

شامل مشاهدات یا مقادیر اندازه‌گیری شده‌ی برخی از صفات اشیاء هستند که به صورت n بردار در فضای pبعدی ثبت می‌شوند. در نوع دوم، داده‌ها شامل مشاهدات یا مقادیر اندازه‌گیری شده از میزان عدم تشابه (یا تشابه) بین اشیاءاند که به صورت یک ماتریس $n \times n$ ثبت می‌شوند. به عنوان مثال، مقادیر اندازه‌گیری شده‌ی عدم تشابه (یا تشابه) روانی بین افراد، در علوم روانشناسی از این نوع‌اند.

متناسب با انواع فوق از داده‌ها، تکنیک‌های مختلف خوشه‌بندی نیز گسترش یافته‌اند. از جمله‌ی این تکنیک‌ها، روش‌های خوشه‌بندی سنتی، مانند انواع روش‌های سلسله مراتبی و نیز روش k—میانگین و انواع روش‌های افرازکردن، هستند که در فصل اول این پایان‌نامه، مروری سریع بر آن‌ها خواهد شد. از انواع دیگر این تکنیک‌ها، روش‌های پیشرفته‌تری هستند که طی سال‌های اخیر گسترش یافته‌اند. مانند روش خوشه‌بندی مدل—پایه‌ای که توسط اسکات و سیمنز (۱۹۷۱) معرفی و توسط بنفیلد و رفتری (۱۹۹۳)، فرلی و رفتری (۱۹۹۸) و فرلی و رفتری (۲۰۰۲) تکامل یافت. در فصل دوم این پایان‌نامه، در مورد این روش و مزایای استفاده از آن، به طور مفصل بحث می‌شود.

استفاده از روش خوشه‌بندی مدل—پایه‌ای، نیاز به بردار مشاهدات دارد و در حالتی که داده‌ها به صورت ماتریس عدم تشابه بین اشیاءاند، قابل استفاده نیست. در این وضعیت، یک راهکار، می‌تواند تبدیل ماتریس عدم تشابه‌ها به یک پیکربندی، شامل n نقطه در فضای pبعدی و استفاده از روش خوشه‌بندی مدل—پایه‌ای روی پیکربندی حاصل باشد. خوشبختانه، امروزه، روش‌های مناسبی جهت یافتن پیکربندی اشیاء گسترش یافته‌اند که از آن جمله انواع روش‌های مقیاس‌بندی چندبعدی هستند. روش کلاسیک مقیاس‌بندی چندبعدی، در فصل اول معرفی می‌شود. همچنین، یک رهیافت بیزی از این روش‌ها، که توسط اه و رفتری (۲۰۰۱) ارائه شده است، در فصل سوم معرفی می‌شود.

اه و رفتری (۲۰۰۷)، با ترکیب روش خوشه‌بندی مدل—پایه‌ای و رهیافت بیزی مقیاس‌بندی چندبعدی، روشی جدید را معرفی کردند که به منظور خوشه‌بندی اشیاء، با استفاده از فواصل مشاهده‌شده‌ی بین آن‌ها، به کار می‌رود. این روش نیز، در فصل سوم معرفی می‌شود.

در پایان نتایج حاصل از انواع این روش‌ها روی داده‌های شبیه‌سازی شده و واقعی مقایسه می‌شوند. نتایج حاصل، گویای یک پیکربندی خوب و خوشه‌بندی معقول در انواع روش‌های مدل—پایه‌ای است.

فهرست مندرجات

لیست اشکال

لیست جداول