

شناسایی آماری الگو

(جلد دوم)

یادگیری بدون نظارت

دکتر زهره عظیمی (استادیار دانشگاه شیراز)

مهندس احسان احمدی (دانشجوی دکتری دانشگاه شیراز)

انتشارات نوید شیراز

پیشگفتار

شناسایی الگو از مهمترین مباحث هوش مصنوعی می باشد که تلاش می کند ساختارها و الگوهای یک مسئله را در محیط را بیاموزد و از آن در مواجهه با وقایع جدید از همان گونه استفاده کند. تئوریتهای شناسایی الگو تابع نگاشت ورودی به خروجی مطلوب را بر اساس داده های آماری یاد می گیرند و برای داده های جدید خروجی متناظر را پیش بینی می کنند. کتاب حاضر به منظور کمک به مباحث شناسایی الگو می پردازد و تلاش دارد تا به صورت ساده و روان مبانی پایه ای را بیان کند. در جلد اول کتاب، اصول مقدماتی یادگیری نظارتی مورد بررسی و مطالعه قرار داده شد. در یک سیستم شناسایی الگو که از یادگیری نظارتی استفاده می کند از اطلاعات بیرونی برای برچسب کلاس الگوها برای آموزش بهره می گیرد.

این جلد از کتاب به مباحث یادگیری بدون نظارت اختصاص داده شده است. در یادگیری بدون نظارت اطلاعات خاصی درباره ی الگوها وجود ندارد و هدف کشف ساختار و ماهیت داده ها بر اساس شباهت ها و عدم شباهت های بین آن ها می باشد. در سیستم عمده ی شناسایی الگو که با الگوهای بدون برچسب سر و کار دارند تخمین چگالی و خوشه بندی می یابند. این دو شاخه ی شناسایی الگو ارتباط تنگاتنگی با یکدیگر دارند. شاخه ی تخمین چگالی، ساختار و ماهیت درونی داده ها را با یک تابع چگالی احتمال نمایان می کند. فصل اول این کتاب به تخمین پارامتریک چگالی احتمال می پردازد، به طوری که برای داده ها یک تابع احتمال فرض کرده و سپس پارامترهای آن را تخمین خواهد زد. این بحث در علم شناسایی الگو تحت عنوان تخمین پارامتر نیز مطرح می شود. فصل دوم نیز به تخمین چگالی غیر پارامتریک اختصاص دارد که در آن تابع چگالی احتمال داده ها به طور مستقیم از روی داده ها استخراج می شود و مدلی از پیش برای آن تعیین نمی شود. تخمین چگالی با استفاده از هسته ها مهمترین

روش این شاخه به شمار می‌رود. در فصل سوم و چهارم خوشه‌بندی داده‌ها که اصلی‌ترین مبحث یادگیری بدون نظارت محسوب می‌شود مورد مطالعه قرار می‌گیرد. خوشه‌بندی به تقسیم داده‌ها به گروه‌هایی به نام خوشه اطلاق می‌شود. این مبحث نیز در دو فصل پارامتریک و غیرپارامتریک گنجانده شده است. در خوشه‌بندی پارامتریک یک مدل ترکیبی برای توزیع خوشه‌ها در نظر گرفته می‌شود و پارامترهای مدل تخمین زده خواهد شد. مدل ترکیبی گوسی و الگوریتم امید ریاضی-بیشینه‌سازی از مسائل عمده‌ی فصل سه می‌باشند. در نهایت در فصل چهارم خوشه‌بندی غیر پارامتریک آورده شده است که داده‌ها در آن بر اساس میزان شباهت‌ها دسته‌های مختلف قرار می‌گیرند. الگوریتم‌های خوشه‌بندی بر اساس مرکز ثقل، بر اساس مراکز و الگوریتم‌های سلسله‌مراتبی در این فصل معرفی می‌شوند.

همان‌طور که مشخص است این کتاب برای دانشجویان گرایش هوش مصنوعی (مهندسی و علوم کامپیوتر) طراحی شده است ولی مباحث آن قابل استفاده‌ی همه‌ی رشته‌هایی است که از مفاهیم و اصول الگوریتم‌های شناسایی الگو و هوش مصنوعی در کاربردهای خود استفاده می‌کنند. لذا این کتاب می‌تواند برای رشته‌های مهندسی، رشته‌ی ریاضی کاربردی، اقتصاد، علوم انسانی، زمین‌شناسی و غیره که به داده‌ها سروکار دارند و هدف آن‌ها پیشبینی یک خروجی مطلوب می‌باشد مفید است. همچنین این کتاب می‌تواند توسط دانشجویان مهندسی پزشکی و پزشکی که داده‌های بهداشتی، پزشکی و مالی می‌کنند مورد استفاده قرار گیرد.

باعث افتخار می‌باشد که اساتید و دانشجویان محترم نظرات و پیشنهادات خود را با ما در میان بگذارند. از آدرس ایمیل pattern.recognition.chinazu@gmail.com می‌توانید برای تماس با ما استفاده کنید.

فهرست مطالب

د	فهرست اختصاها	۱
ش	علامه راسخ	۱
۱	تخمین چگالی پارامتری	۱
۱	۱.۱ مقدمه	۱
۳	۲.۱ تخمین بیشینه گرایی	۳
۴	۱.۲.۱ تخمین میانگین تابع چگالی گوسی با واریانس از قبل تعیین شده	۴
۶	۲.۲.۱ تخمین میانگین و واریانس تابع چگالی گوسی	۶
۹	۳.۲.۱ ارزیابی تخمین پارامترها	۹
۱۲	۳.۱ تخمین بیشینه گرایی احتمال پسین	۱۲
۱۴	۱.۳.۱ تخمین پارامتر توزیع برنولی با دانش اولیه توزیع بتا	۱۴
۱۶	۴.۱ تخمین بیز	۱۶
۱۷	۱.۴.۱ دانش اولیه‌ی مزدوج	۱۷
۱۸	۲.۴.۱ توزیع‌های مزدوج گوسی	۱۸
۲۱	۳.۴.۱ مقایسه روش‌های بیشینه گرایی، درست‌نمایی، بیشینه گرایی احتمال پسین و بیز	۲۱
۲۴	۵.۱ خلاصه‌ی فصل	۲۴
۲۷	۶.۱ تمرینات	۲۷
۲۹	مراجع	۲۹

۲ تخمین چگالی غیر پارامتریک ۳۱

۳۱	۱.۲	مقدمه
۳۱	۲.۲	تخمین چگالی با روش هیستوگرام
۳۳	۳.۲	تخمین چگالی با استفاده از هسته ها
۳۴	۱.۳.۲	پنجره‌ی پارزن
۳۶	۲.۳.۲	هسته‌های هموار
۳۷	۳.۳.۲	انتخاب پهنای باند
۴۰	۴.۳.۲	نکاتی چند درباره‌ی تخمین چگالی چند بعدی
۴۴	۴.۲	تخمین چگالی با روش نزدیکترین همسایه
۴۶	۲	خوشه‌ی فصل
۴۹	۶.۲	تمرینات
۵۲		

۵۵ مراجع

۵۷ ۳ خوشه‌بندی پارامتریک

۵۷	۱.۳	مقدمه
۵۷	۲.۳	خوشه‌بندی مدل ترکیبی
۵۸	۳.۳	روش امید ریاضی-بیشینه‌سازی در قالب
۶۲	۴.۳	روش امید ریاضی-بیشینه‌سازی
۶۶	۵.۳	خوشه‌بندی مدل ترکیبی گوسی
۶۹	۶.۳	خلاصه‌ی فصل
۷۳	۷.۳	تمرینات
۷۶		

۷۷ مراجع

۷۹ ۴ خوشه‌بندی غیر پارامتریک

۷۹	۱.۴	مقدمه
۷۹	۲.۴	خوشه‌بندی مبتنی بر مرکز ثقل
۸۰	۱.۲.۴	الگوریتم خوشه‌بندی k میانگین
۸۱		

۲.۲.۴	ارتباط الگوریتم‌های k میانگین و امید ریاضی-بیشینه‌سازی	۸۳
۳.۲.۴	مزایا و معایب الگوریتم k میانگین	۸۶
۴.۲.۴	الگوریتم‌های دیگر مبتنی بر مرکز ثقل	۸۷
۳.۴	خوشه‌بندی مبتنی بر چگالی	۸۸
۱.۳.۴	الگوریتم خوشه‌بندی DBSCAN	۸۸
۲.۳.۴	نکات قوت و ضعف الگوریتم DBSCAN	۹۰
۹۲	خوشه‌بندی سلسله‌مراتی	۹۲
۱.۴.۴	خوشه‌بندی سلسله‌مراتی ادغام‌کننده	۹۳
۲.۴.۴	خوشه‌بندی سلسله‌مراتی تقسیم‌کننده	۹۶
۲.۴.۴	پارامترهای خوشه‌ها	۹۷
۵.۴	خلاصه‌ی فصل	۱۰۲
۶.۴	تمرینات	۱۰۶

مراجع

۱۰۷

آ سفید کردن داده

۱۰۹

۱.۱ مقدمه

۱۰۹

۲.۱ ناهمبسته کردن داده

۱۱۰

۳.۱ سفید کردن داده

۱۱۱

ب معیارهای مجاورت

۱۱۵

۱.ب مقدمه

۱۱۵

۲.ب معیارهای عدم شباهت

۱۱۵

۳.ب معیارهای شباهت

۱۱۸

پ شکل‌های کتاب با کیفیت بالا

۱۱۹

نمایه

۱۳۷

لیست شکل‌ها

۶	تخمین پارامتر میانگین μ یک تابع چگالی گوسی	۱.۱
۸	تخمین پارامترهای میانگین و واریانس یک تابع چگالی گوسی	۲.۱
۱۵	تخمین بیشینه احتمال پسین از پارامتر توزیع برنولی	۳.۱
۱۷	تخمین پارامتر میانگین برای تابع گوسی با واریانس از قبل مشخص	۴.۱
۲۲	توزیع احتمال پسین $p(x x_0)$ به ازای مجموعه داده‌های مختلف X	۵.۱
۳۵	تخمین تابع چگالی احتمال روش هیستوگرام	۱.۲
۳۷	پنجره‌ی پارزن	۲.۲
۳۸	تخمین تابع چگالی احتمال به روش پارزن	۳.۲
۴۰	تخمین تابع چگالی احتمال با استفاده از تابع هسته گوسی	۴.۲
۴۳	تأثیر انتخاب پارامتر هموارسازی h هسته در میانگین و واریانس	۵.۲
۴۷	تخمین چگالی احتمال با روش k نزدیک‌ترین همسایه	۶.۲
۶۳	مجموعه داده‌ی کامل بدست آمده از پرتاب سکه‌ها	۱.۳
۶۴	مجموعه داده‌ی ناکامل بدست آمده از پرتاب سکه‌ها	۲.۳
۶۵	مراحل انجام الگوریتم امید ریاضی-بیشینه‌سازی برای آزمایش سکه‌ها	۳.۳
۶۸	مراحل E و M در الگوریتم امید ریاضی-بیشینه‌سازی	۴.۳
۷۲	خوشه‌بندی پارامتریک با استفاده از مدل ترکیبی گوسی	۵.۳
۸۴	خوشه‌بندی با استفاده از الگوریتم k -میانگین	۱.۴
۸۵	مقایسه‌ی الگوریتم‌های k -میانگین و امید ریاضی-بیشینه‌سازی	۲.۴
۸۶	خوشه‌بندی داده‌های دارای توزیع غیر کروی با استفاده از الگوریتم k -میانگین	۳.۴

- ۴.۴ مفاهیم مربوط به الگوریتم DBSCAN ۹۱
- ۵.۴ خوشه‌بندی با استفاده از الگوریتم DBSCAN ۹۳
- ۶.۴ رده‌بندی حیوانات به عنوان نمونه‌ای از خوشه‌بندی سلسله‌مراتبی ۹۴
- ۷.۴ خوشه‌بندی سلسله‌مراتبی ادغام‌کننده با الگوریتم‌های اتصال ۹۶
۱. پ. ۱ (شکل ۱.۱) تخمین پارامتر میانگین μ یک تابع چگالی گوسی ۱۲۱
۲. پ. ۲ (شکل ۲.۱) تخمین پارامترهای میانگین و واریانس تابع چگالی گوسی ۱۲۲
۳. پ. ۳ (شکل ۳.۱) تخمین بیشینه‌گر احتمال پسین از پارامتر توزیع برنولی ۱۲۳
۴. پ. ۴ تخمین بیز پارامتر میانگین برای تابع گوسی با واریانس معلوم (شکل ۴.۱) ۱۲۴
۵. پ. ۵ (شکل ۵.۱) توزیع احتمال پسین $P(\mu|X)$ ۱۲۵
۶. پ. ۶ تخمین تابع چگالی احتمال به روش هیستوگرام (شکل ۱.۲) ۱۲۶
۷. پ. ۷ (شکل ۷.۱) تخمین تابع چگالی احتمال به روش پنجره‌ی پارزن ۱۲۷
۸. پ. ۸ (شکل ۸.۱) تخمین تابع چگالی احتمال با استفاده از تابع هسته‌ی گوسی ۱۲۸
۹. پ. ۹ (شکل ۵.۲) تاثیر پارامتر هموارسازی h هسته در بایاس و واریانس ۱۲۹
۱۰. پ. ۱۰ (شکل ۶.۲) تخمین چگالی احتمال با روش h نزدیک‌ترین همسایه ۱۳۰
۱۱. پ. ۱۱ (شکل ۴.۳) مراحل M و E در الگوریتم امید ریاضی-بیشینه‌سازی ۱۳۱
۱۲. پ. ۱۲ (شکل ۵.۳) خوشه‌بندی پارتریک با استفاده از مدل ترکیبی گوسی ۱۳۲
۱۳. پ. ۱۳ (شکل ۱.۴) خوشه‌بندی با استفاده از الگوریتم k -میانگین ۱۳۳
۱۴. پ. ۱۴ (شکل ۲.۴) مقایسه الگوریتم‌های k -میانگین و امید ریاضی-بیشینه‌سازی ۱۳۴
۱۵. پ. ۱۵ (شکل ۳.۴) خوشه‌بندی داده‌های با توزیع ترکیبی پارتریک-گوسی با الگوریتم k -میانگین ۱۳۵
۱۶. پ. ۱۶ (شکل ۵.۴) خوشه‌بندی با استفاده از الگوریتم DBSCAN ۱۳۶

لیست الگوریتم‌ها

۸۳	الگوریتم خوشه‌بندی k -میانگین	۱
۹۲	الگوریتم خوشه‌بندی DBSCAN	۲

www.ketab.ir