

روش‌های کلاسیک بازنمایی داده‌ها

از فضای ابعاد بالا به ابعاد پایین

(ویژه دانشجویان تمام مقاطع تحصیلی مهندسی برق الکترونیک و مخابرات)



مؤلفین:

دکتر رسول حاجی زاده

(مدرس دانشگاه)

دکتر فخرالدین بشاری

(عضو هیئت علمی دانشگاه تخصصی فناوری های نوین امل)



سرشناسه	حاجی زاده، رسول، ۱۳۶۳ -
عنوان و نام پدیدآور	روش های کلاسیک بازنمایی داده ها از فضای ابعاد بالا به ابعاد پایین (ویژه دانشجویان تمام مقاطع تحصیلی مهندسين برق الکترونیک و مخابرات)/مولفان رسول حاجی زاده، فخرالدین نظری.
مشخصات نشر	آمل: دانشگاه تخصصی فناوریهای نوین آمل، ۱۳۹۷.
مشخصات ظاهری	۱۵۸ ص: مصور، جدول، نمودار.
شابک	۹۷۸-۶۲۲-۶۱۳۵-۱۴-۶
وضعیت فهرست نویسی	فیا
یادداشت	واژه نامه.
موضوع	کاهش ابعاد (آمار)
موضوع	Dimension reduction (Statistics)
موضوع	خمینه ها (ریاضیات)
موضوع	Manifolds (Mathematics)
موضوع	داده پردازی
موضوع	Electronic data processing
موضوع	تجزیه و تحلیل چند متغیره
موضوع	Multivariate analysis
شناسه افزوده	نظری، فخرالدین، ۱۳۶۳ -
رده بندی کنگره	۱۳۹۷ ع ۲۰۴۷-۲۸
رده بندی دیویی	۵۳۰/۸
شماره کتابشناسی ملی	۵۲۱۳۳۴۸

عنوان:	روش های کلاسیک بازنمایی داده ها از فضای ابعاد بالا به ابعاد پایین
مولفین:	دکتر رسول حاجی زاده، دکتر فخرالدین نظری
ویراستار:	دکتر فرید صمصامی خداداد
ناشر:	انتشارات دانشگاه تخصصی فناوریهای نوین آمل
چاپ و صحافی:	چاپ نیک
صفحه آرایي:	رضا کاوه
طراحی جلد:	رضا کاوه
نوبت چاپ:	نخست - ۱۳۹۷
شمارگان:	۱۰۰۰ نسخه
شابک:	۹۷۸-۶۲۲-۶۱۳۵-۱۴-۶
قیمت:	۱۷۰۰۰ تومان

فهرست مطالب

فصل اول.....	۱۳
۱- مفاهیم اولیه.....	۱۳
۱-۱ مقدمه.....	۱۳
۲-۱ کاهش ابعاد و اهمیت آن.....	۱۴
۳-۱ منیفلد و آموزش.....	۱۶
۴-۱ مفهوم تنک بودن و بازسازی تنک.....	۲۰
۵-۱ جمع‌بندی.....	۲۱
فصل دوم.....	۲۳
۲- روش‌های آموزش منیفلد و دسته‌بندی آنها.....	۲۳
۱-۲ مقدمه.....	۲۳
۲-۲ آموزش منیفلد PCA.....	۲۴
۳-۲ روش‌های آموزش منیفلد سراسری.....	۲۶
۱-۳-۲ آموزش منیفلد (MDS) Multidimensional Scaling.....	۲۷
۲-۳-۲ آموزش منیفلد Isomap.....	۲۹
۳-۳-۲ آموزش منیفلد (MVU) Maximum Variance Unfolding.....	۳۰
۴-۳-۲ آموزش منیفلد Kernel PCA.....	۳۲
۵-۳-۲ آموزش منیفلد (DM) Diffusion Maps.....	۳۳
۶-۳-۲ آموزش منیفلد کدکننده‌های خودکار چندلایه‌ای.....	۳۵
۴-۲ روش‌های آموزش منیفلد محلی.....	۳۷
۱-۴-۲ آموزش منیفلد (LLE) Locally Linear Embedding.....	۳۷

۳۹.....	۲-۴-۲ آموزش منیفلد Laplacian Eigenmaps(LEM)
۴۱.....	۲-۴-۲ آموزش منیفلد Hessian LLE(HLLE)
۴۲.....	۴-۴-۲ آموزش منیفلد Local Tangent Space Analysis (LTSA)
۴۴.....	۵-۲ روش‌های مبتنی بر ترکیب ویژگی‌های ساختاری محلی و سراسری
۴۴.....	۱-۵-۲ آموزش منیفلد Locally Linear Coordination (LLC)
۴۵.....	۲-۵-۲ آموزش منیفلد Manifold Charting (MC)
۴۷.....	۶-۲ جمع‌بندی
۴۹.....	فصل سوم
.....	۳. دسته‌بندی داده‌ها در فضای با ابعاد پایین، با استفاده از روش‌های آموزش منیفلد LEM, LLE.
۴۹.....	LPP و CA
۴۹.....	۱-۳ مقدمه
.....	۲-۳ بازنمایی و تدخیر بهره‌افراد در فضای با ابعاد پایین، با استفاده از روش‌های آموزش منیفلد
۵۰.....	LEM و LLE
۵۱.....	۱-۲-۳ الگوریتم آموزش منیفلد LLE
۵۴.....	۲-۲-۳ الگوریتم آموزش منیفلد LLM
۵۶.....	۳-۲-۳ بازنمایی داده‌آزمون در فضای با ابعاد پایین
۵۸.....	۴-۲-۳ نتایج شبیه‌سازی بر روی تصاویر چهار پایگاه داده ORL
۶۱.....	۳-۳ بازشناسی و بازنمایی حروف دست‌نویس فارسی با استفاده از روش آموزش منیفلد LPP...
۶۴.....	۱-۳-۳ الگوریتم آموزش منیفلد LPP
۶۶.....	۲-۳-۳ الگوریتم آموزش منیفلد PCA
۶۶.....	۳-۳-۳ نتایج شبیه‌سازی بر روی حروف دست‌نویس پایگاه داده فارسی
۶۸.....	۴-۲ جمع‌بندی
۷۱.....	فصل چهارم
۷۱.....	۴- بازنمایی منیفلد داده‌های مصنوعی و مقایسه روش‌های گوناگون آموزش منیفلد
۷۱.....	۱-۴ مقدمه
.....	۲-۴ بازنمایی داده‌های مصنوعی با استفاده از روش‌های آموزش منیفلد LLE, LEM, PCA.
۷۴.....	Diffusion Map و Isomap, MDS, LTSA, HLLE
.....	۳-۴ بررسی تغییر مقدار تعداد همسایه‌ها بر عملکرد روش‌های آموزش منیفلد محلی LLE, HLLE و
۸۰.....	LTSA
۸۳.....	۴-۴ جمع‌بندی

۸۵.....	فصل پنجم.....
۸۵.....	۵- ارزیابی کمی روش‌های آموزش منیفلد.....
۸۵.....	۵-۱ مقدمه.....
۸۶.....	۵-۲ پیشینه روش‌های ارزیابی الگوریتم‌های استخراج غیرخطی داده.....
۸۷.....	۵-۳ معیارهای ارزیابی روش‌های آموزش منیفلد.....
۸۷.....	۵-۳-۱ معیارهای ALSTD و ALED.....
۹۰.....	۵-۳-۲ معیارهای ALCD و GSCD.....
۹۱.....	۵-۴ ارزیابی روش‌های آموزش منیفلد مبتنی بر معیار GSCD.....
۹۳.....	۵-۵ نتیجه.....
۹۵.....	فصل ششم.....
۹۵.....	۶- مقدمه‌ای بر بازمایی تنک.....
۹۵.....	۶-۱ مقدمه.....
۹۸.....	۶-۲ تعاریف و مفاهیم پایه.....
۱۰۰.....	۶-۳ بازنمایی تنک یک سیگنال.....
۱۰۲.....	۶-۴ دسته‌بندی تکنیک‌های بازنمایی تنک.....
۱۰۳.....	۶-۴-۱ دسته‌بندی مساله بازنمایی تنک از نظر تحلیلی.....
۱۰۴.....	۶-۴-۲ دسته‌بندی مساله بازنمایی تنک مبتنی بر نوع محدودیت نرم.....
۱۰۸.....	۶-۵ برخی از الگوریتم‌های معرفی شده جهت حل مساله بازنمایی تنک.....
۱۰۹.....	۶-۵-۱ الگوریتم تعقیب تطبیقی.....
۱۱۱.....	۶-۵-۲ الگوریتم تعقیب تطبیقی متعامد.....
۱۱۳.....	۶-۵-۳ الگوریتم بازنمایی تنک نگاشت گرادیان (GPSR).....
۱۱۶.....	۶-۵-۴ الگوریتم بازنمایی تنک هموتوبی لاسو.....
۱۱۹.....	۶-۶ بازنمایی تنک در آموزش دیکشنری بدون نظارت KSVD.....
۱۲۲.....	۶-۷ جمع‌بندی.....
۱۲۵.....	مراجع.....
۱۳۵.....	ضمیمه.....
۱۳۵.....	ضمیمه ۱: کد متلب پیاده‌سازی روش‌های آموزش منیفلد LLE، LEM و LPP.....
.....	ضمیمه ۲: کد متلب ایجاد ۸ دسته‌داده‌ی مصنوعی Swiss Roll، Swiss Hole، Corner Planes.....
۱۳۸.....	Gaussian و 3D Clusters، Toroidal Helix، Twin Peaks، Punctured Sphere.....

Diffusion و Isomap .MDS .LTSA .HLLF .PCA	ضمیمه ۳: کد متلب روش‌های آموزش منیفلد
۱۴۲.....	Map
۱۴۸.....	ضمیمه ۴: کد متلب معیار GSCD
۱۵۱.....	ضمیمه ۵: کد متلب الگوریتم‌های روش حل مسائل بازنمایی تنک با محدودیت نرم صفر و یک...
۱۵۷.....	فهرست علائم اختصاری

پیشگفتار

امروزه با پیشرفت حسگرهای اندازه‌گیری و سامانه‌های جمع‌آوری داده‌ها داده‌ها از حجم و فضای نمایش با ابعاد بالایی برخوردار هستند. برخلاف پیشرفت روزافزون سرعت و قدرت پردازنده‌ها در تجزیه و تحلیل داده‌ها، همچنان ذخیره‌سازی داده‌های با حجم بالا، از مشکلات اساسی و مطرح بوده و انجام پردازش بر روی داده‌ها نیز پیچیدگی محاسباتی بالایی دارد. از طرفی، تقاضای افزایش سرعت پردازش و ذخیره‌سازی حجم بالایی از داده‌ها به شدت در حال افزایش است. از این رو، در ۱۵ سال اخیر، تکنیک‌های کاهش ابعاد و بازنمایی تنک جهت حل این دسته از مشکلات، بسیار مورد توجه قرار گرفته است.

بیشتر داده‌های اندازه‌گیری شده، بُعد ذاتی بسیار پایینی دارند، از جمله آنکه در فضایی با ابعاد بسیار بالا نمایش داده می‌شوند. در این کتاب، به معرفی روش‌های کاهش ابعاد مبتنی بر آموزش منیفولد پرداخته شده است. روش‌های آموزش منیفولد، از جمله تکنیک‌های کاهش ابعاد هستند که بر اساس حفظ ساختار داده‌ها، از فضای با ابعاد بالا به فضای با ابعاد پایین عمل می‌نمایند. این روش‌ها، با استخراج ویژگی‌های ساختاری داده‌ها در فضا، به تشکیل گراف‌ی از داده‌ها پرداخته و سپس به بازنمایی منیفولد داده‌ها در فضای با ابعاد پایین می‌پردازند؛ به گونه‌ای که، گراف منیفولد موجود بین داده‌ها، دست‌خوش کمترین تغییر در فضای بازنمایی با ابعاد پایین گردد.

به طور کلی، روش‌های آموزش منیفلد، به دو دسته‌ی محلی و سراسری تقسیم می‌گردند. روش‌های آموزش منیفلد محلی با بهره‌گیری از ویژگی‌های ساختاری محلی، به بازنمایی داده‌ها در فضای با ابعاد پایین می‌پردازند. روش‌های آموزش منیفلد سراسری نیز، با استخراج ویژگی‌های سراسری منیفلد داده‌ها به بازنمایی داده‌ها در فضای با ابعاد پایین می‌پردازند. در فصل دوم این کتاب، به معرفی و دسته‌بندی انواع روش‌های آموزش منیفلد پرداخته شده است.

در فصل سوم، روش‌های آموزش منیفلد LLE، LEM، LPP و PCA به صورت دقیق‌تر و مفصل‌تر، جهت به کارگیری در کاربرد دسته‌بندی داده‌ها بیان شده‌اند. در این فصل، ابتدا به کاربرد روش‌های آموزش منیفلد علی‌متداول LLE و LEM در بازنمایی و تشخیص چهره افراد پرداخته شده است، که یکی از پرکاربردترین موضوعات مطرح در پردازش تصویر است. از طرفی، یکی از مشکلات روش‌های غیرخطی آموزش منیفلد، واجد با داده‌های آزمون است که روشی جهت مواجهه با داده‌های آزمون، که نقشی در کاهش ابعاد داده‌ها ندارد، آموزش نداشته‌اند، برای دو روش آموزش منیفلد LLE و LEM معرفی می‌گردد. سپس، به بازنمایی و تشخیص چهره افراد با استفاده از روش‌های مذکور پرداخته شده و نتایج آن بیان می‌گردد. در بخش دوم این فصل، آموزش منیفلد خطی LPP و PCA توضیح داده شده و الگوریتم پیاده‌سازی آنها بیان خواهد شد. جهت بیان یکی از کاربردهای این دو تکنیک، به بازنمایی و بازشناسی حروف و ارقام دست‌نویس فارسی پرداخته شده است. این روش‌ها خطی بوده و مشکلی در برخورد با داده‌های آزمون ندارند. در این فصل، از پایگاه داده هندی، جهت انجام شبیه‌سازی‌ها و ارزیابی نتایج استفاده شده است.

در فصل چهارم، هشت منیفلد داده مصنوعی سه‌بعدی معرفی می‌شود که عموماً در ارزیابی روش‌های آموزش منیفلد به کار گرفته می‌شوند. معرفی می‌شود. این دسته از منیفلد‌های ارایه شده در فضای سه‌بعدی، می‌توانند در درک بهتر مفهوم آموزش منیفلد موثر بوده و عموماً در ارزیابی روش‌های آموزش منیفلد مختلف به کار گرفته شوند. در این فصل، به ارزیابی هشت روش آموزش منیفلد مختلف، مبتنی بر بازنمایی منیفلد داده‌های سه‌بعدی معرفی شده، و بازنمایی آنها در فضای دوبعدی خواهیم پرداخت. در ارزیابی‌های صورت گرفته، میزان پیچیدگی روش‌ها بر اساس میزان زمان مصرفی نیز با یکدیگر مقایسه شده‌اند.

همچنین معیاری کمی، جهت ارزیابی روش‌های آموزش منیفلد در فصل پنجم معرفی می‌گردد. این معیار کمی، بر اساس میزان نزدیکی داده‌ها در کنار یکدیگر، در دو فضای با ابعاد بالا و پایین بوده و می‌تواند معیاری از میزان تشابه ساختار داده‌های بازنمایی شده و داده‌های اولیه است. در این فصل، روند محاسبه این معیار توضیح داده شده و آزمایش‌هایی بر اساس این معیار صورت گرفته و با نتایج کیفی و بصری مقایسه گردیده است.

بازنمایی تنک در سال‌های اخیر، بسیار مورد توجه بوده و در کاربردهای گوناگونی مانند طبقه‌بندی، بازسازی و حذف نویز به کار گرفته شده است. در فصل ششم، به معرفی مفهوم بازنمایی تنک، و استفاده از آنها پرداخته شده است. بازنمایی تنک مبتنی بر محدودیت‌های نرم‌های صفر، یک و غیره دسته‌بندی شده و روابط حاکم بر هر یک از آنها مطرح می‌گردد. همچنین، الگوریتم‌های تعقیب تطبیقی، تقسیم طبقه متعامد، الگوریتم بازنمایی تنک مبتنی بر نگاشت گرادیان و الگوریتم هموتوپی لاسو معرفی شده است. الگوریتم آموزش دیکشنری K-SVD که یکی از کاربردهای متداول بازنمایی تنک است نیز شرح داده می‌شود.