

هدف: آشنایی با شبکه عصبی خودسازمانده کوهونن (SOM)

کد: کد این فعالیت را با کمک چارچوب کاری تنسورفلو و با استفاده از زبان برنامه‌نویسی پایتون بنویسید. (استفاده از چارچوب کاری تنسورفلو در این تمرین، اختیاری است)

گزارش: ملاک اصلی انجام فعالیت گزارش آن است و ارسال کد بدون گزارش فاقد ارزش است. برای این فعالیت یک فایل گزارش در قالب pdf به فرمت زیر تهیه کنید.

- **عنوان:** در این بخش بیان کنید کدامیک از موارد را بررسی می‌کنید.

- **شرایط آزمایش:** جدولی از پارامترها و تنظیمات خود را در اینجا بیان کنید و تا حد امکان روابط حاکم بر پارامترها و دلیل انتخاب تنظیمات انجام شده را شرح دهید.

- **گزارش نتایج:** در این بخش نمودارهای مورد نیاز به همراه توضیحات آن‌ها را ذکر کنید. (۱ تا ۲ پاراگراف)

- **تحلیل آزمایش:** نتایج حاصل از بررسی‌های خود را با دلایل آن به طور کامل توضیح داده و تحلیل کنید. در بیان دلایل اگر به مرجع خاصی اشاره شود بهتر است. (۲ تا ۳ پاراگراف)

تذکر: مطابق قوانین دانشگاه هر نوع کپی برداری و اشتراک کار دانشجویان غیر مجاز بوده و شدیداً برخورد خواهد شد. استفاده از کدها و توضیحات اینترنت به منظور یادگیری بلامانع است، اما کپی کردن غیرمجاز است.

راهنمایی: در صورت نیاز می‌توانید سوالات خود را در خصوص پروژه از تدریس یار درس، از طریق ایمیل زیر بپرسید.

E-mail: ann.ceit.aut@gmail.com

ارسال: فایل‌های کد و گزارش خود را در قالب یک فایل فشرده با فرمت StudentID_HW۰۳.zip تا تاریخ ۹۸/۰۲/۱۸ ارسال نمایید. شایان ذکر است هر روز تاخیر باعث کسر ۱۰٪ نمره خواهد شد.

توجه: تمامی مجموعه داده‌های مورد نیاز در هر بخش به فایل تعریف پروژه پیوست شده‌اند.

بخش اول

در این بخش قصد داریم عملکرد شبکه عصبی خودسازمانده کوهونن را در خوشه‌بندی مورد بررسی قرار دهیم. برای انجام این تمرین، ابتدا یک شبکه عصبی خودسازمانده کوهونن پیاده‌سازی کرده و سپس آزمایش‌های زیر را انجام دهید و نتایج آزمایشات را به همراه تحلیل‌های مورد نیاز مطابق با فرمت تعریف شده در فایل گزارش خود ذکر نمایید. توجه نمایید در هر بخش ابتدا ساختار شبکه را به طور کامل توضیح دهید و سپس نتایج حاصل را گزارش نمایید.

۱. عملکرد شبکه دوبعدی کوهونن در خوشه‌بندی:

در این آزمایش یک شبکه کوهونن دوبعدی را ایجاد نموده و تاثیر پارامترهای زیر را بر عملکرد آن مورد بررسی قرار دهید. برای بررسی شبکه در این بخش از مجموعه داده "۲D" استفاده نمایید. در انتها تمام داده‌ها را در یک نمودار نمایش دهید که در آن داده‌های مربوط به هر خوشه با رنگ یکسانی از خوشه‌های دیگر تفکیک شده‌اند.

• توپولوژی همسایگی

شبکه را یک بار با استفاده از توپولوژی دایره‌ای، یک بار با استفاده از توپولوژی مستطیلی و یک بار با استفاده از توپولوژی شش ضلعی مورد آزمایش قرار دهید. تاثیر هر یک از این توپولوژی‌ها را بر عملکرد شبکه بررسی نمایید. کدام توپولوژی بهترین عملکرد را داشته است؟ چرا؟

• شعاع همسایگی

شبکه را یک بار با استفاده از مرز غیر قابل تشخیص و یک بار با استفاده از مرز قابل تشخیص (با حداقل ۳ مقدار مختلف شعاع همسایگی) آموزش دهید. کدام یک از روش‌ها برای خوشه‌بندی داده‌ها مناسب‌تر است؟ چرا؟

۲. عملکرد شبکه GSOM در خوشه‌بندی:

با توجه به این که داده‌های موجود در مجموعه داده "۲D" به طور سلسله‌مراتبی قابل خوشه‌بندی هستند، از شبکه GSOM برای خوشه‌بندی این داده‌ها استفاده نمایید. ساختار اولیه شبکه و الگوریتم رشد را به طور کامل توضیح دهید. پس از اتمام آموزش عملکرد شبکه GSOM را با عملکرد شبکه دوبعدی مقایسه نمایید.

در انتها تمام داده‌ها را در یک نمودار نمایش دهید که در آن داده‌های مربوط به هر خوشه با رنگ یکسانی از خوشه‌های دیگر تفکیک شده‌اند. همین‌طور برای بررسی ساختار سلسله‌مراتبی ایجاد شده، نمودار دندروگرام حاصل را رسم نمایید. برای رسم نمودار دندروگرام می‌توانید از تابع زیر استفاده نمایید.

`Scipy.clulster.hierarchy.dendrogram()`

برای مطالعه بیشتر در مورد نحوه استفاده از این تابع به لینک زیر مراجعه نمایید.

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.dendrogram.html>

بخش دوم

یکی از کاربردهای شبکه عصبی خودسازمانده کوهونن در کاهش ابعاد مجموعه داده‌های بزرگ است. برای انجام آزمایشات این بخش از مجموعه داده "HighDim" استفاده نمایید. این مجموعه داده شامل بیش از ۶۰۰۰ داده ۱۰۲۴ بعدی است که توسط ۱۶ خوشه گاوسی تولید شده‌اند. در تمامی آزمایشات این بخش از یک شبکه عصبی پیش‌رو به عنوان دسته‌بندی کننده استفاده نمایید.

۱. داده‌های موجود در این مجموعه داده را با استفاده از یک شبکه پیش‌رو دسته‌بندی نمایید. بهترین دقت ممکن با استفاده از این دسته‌بندی کننده را به همراه ساختار شبکه گزارش نمایید.

۲. با استفاده از الگوریتم PCA^1 ابتدا ابعاد داده‌ها را کاهش دهید و سپس از بردارهای ویژگی جدید برای دسته‌بندی داده‌ها استفاده نمایید. تاثیر میزان کاهش ابعاد را بر عملکرد دسته‌بندی کننده بررسی نمایید. برای این کار حداقل ۳ مقدار مختلف برای تعداد ابعاد انتخاب شده در نظر بگیرید و بهترین دقت دسته‌بندی را به ازای هریک از مقادیر گزارش کنید. کاهش ابعاد داده‌ها، چه تاثیری بر عملکرد دسته‌بندی دارد؟ چرا؟ (می‌توانید از ابزارهای آماده برای اجرای الگوریتم PCA استفاده نمایید)

۳. تعداد ابعاد انتخاب شده را برابر با ۲ در نظر بگیرید. در یک نمودار، داده‌های دوبعدی حاصل را طوری نمایش دهید که داده‌های موجود در هر کلاس با استفاده از یک رنگ خاص از داده‌های موجود در کلاس‌های دیگر تفکیک شده باشند.

۴. این بار با استفاده از شبکه عصبی SOM ابعاد داده‌ها را کاهش دهید. آزمایش فوق را این بار با شبکه SOM تکرار کنید و نتایج را گزارش نمایید. عملکرد شبکه SOM را در کاهش ابعاد با عملکرد الگوریتم PCA مقایسه نمایید.

۵. داده‌ها را با استفاده از شبکه SOM به فضای ۲ بعدی ببرید و سپس داده‌های جدید را طوری رسم کنید که داده‌های مربوط به هر کلاس با رنگ خاصی از داده‌های کلاس‌های دیگر تفکیک شده باشند. این تصویر را با تصویر حاصل از الگوریتم PCA مقایسه نمایید. چه نتیجه‌ای می‌توان گرفت؟

نکات مربوط به پیاده‌سازی و ارزیابی

در تمام پیاده‌سازی‌ها و ارزیابی‌های انجام شده به نکات زیر توجه داشته باشید.

۱. برای ارزیابی عملکرد شبکه در خوشه‌بندی از معیار $Purity$ استفاده نمایید.
۲. در بخش اول، در یک نمودار دو بعدی، نورون‌های برنده و تعداد دفعاتی که تا کنون برنده شده‌اند را نمایش دهید. با توجه به این نمودار عملکرد شبکه را بر اساس تعداد نورون‌های مرده ایجاد شده در هر آزمایش بررسی نمایید.
۳. برنامه شما باید قادر به ذخیره‌سازی و بارگذاری بهترین شبکه در هر یک از آزمایش‌ها باشد. در زمان تحویل تمرین باید بتوانید نتایج گزارش شده در هر آزمایش را با بارگذاری شبکه مربوطه که از قبل آموزش دیده و ذخیره شده است، بدون آموزش دوباره آن، به دست بیاورید.

موفق باشید

احمد اسدی

¹ Principal Component Analysis