

نقشه خودسازمانده کوهونن

سوال ۱

هدف اصلی این تمرین، بررسی عملکرد شبکه‌های خودسازمانده کوهونن^۱ در کاهش ابعاد مجموعه‌های داده و خوشه‌بندی آن‌ها است. با انجام آزمایشات این تمرین و بررسی تاثیر هر یک از پارامترها بر عملکرد مدل، درک خوبی از جنبه‌های مختلف عملکردی این مدل، حاصل می‌شود.

مجموعه داده

در این تمرین، از مجموعه داده Bank Marketing استفاده می‌شود. این مجموعه داده اطلاعات مربوط به کمپین‌های تبلیغاتی یک بانک را جمع‌آوری کرده و مشخص کرده که آیا مشتریان هدف این تبلیغات، اقدام به افتتاح سپرده نزد این بانک کرده‌اند یا خیر. اطلاعات بیشتر در خصوص این مجموعه داده و لینک‌های دریافت آن را می‌توانید از طریق آدرس زیر مشاهده کنید.

<https://archive.ics.uci.edu/dataset/222/bank+marketing>

سوالات

ساختار کلی شبکه خودسازمانده را پیاده‌سازی کرده و به سوالات زیر پاسخ دهید. توجه نمایید، تطابق تحلیل آزمایشات با دانسته‌های شما، حائز اهمیت است؛ لذا، انتخاب پارامترهای مناسب و طراحی صحیح آزمایشات بر عهده شماست.

۱ - در مرحله اول از این شبکه برای کاهش ابعاد مجموعه داده بهره بگیرید. ابتدا در فایل گزارش خود، روش‌های کاهش بعد با بکارگیری شبکه خودسازمانده را بطور مختصر توضیح داده سپس با انجام آزمایشات این مجموعه داده را یک‌بار به ۸ بعد، یک‌بار به ۴ بعد و یک‌بار دیگر به ۲ بعد کاهش دهید. در هر حالت، با آموزش یک دسته‌بند روی ابعاد جدید، عملکرد کاهش بعد را ارزیابی و مقایسه نمایید. نتایج آزمایشات کاهش بعد را با آموزش دسته‌بند بدون کاهش بعد نیز مقایسه نمایید.

۲ - توزیع داده‌های موجود در مجموعه داده را با استفاده از شبکه خودسازمانده در یک نقشه ۲ بعدی ترسیم کنید. بر اساس نقشه تولید شده، چه برداشت‌هایی در خصوص مجموعه داده دارید؟ بطور کامل توزیع این مجموعه داده را بر اساس نتیجه حاصل، تحلیل نمایید.

^۱ Self Organizing Map (SOM)

۳ - پارامترهای کلیدی شبکه خودسازمانده مانند اندازه شبکه (تعداد نرون‌ها)، نرخ یادگیری و تعداد تکرارهای آموزش را تغییر دهید و تاثیر هرکدام را بر کیفیت کاهش بعد بررسی کنید. تحلیل‌های خود را به کمک نمودارها یا جدول‌های مقایسه‌ای ارائه دهید.

۴ - شبکه خودسازمانده را برای خوشه‌بندی داده‌ها بدون استفاده از برچسب‌ها به کار ببرید. خوشه‌های حاصل را با استفاده از برچسب‌های واقعی (افتتاح حساب یا عدم افتتاح حساب) تحلیل کرده و کیفیت خوشه‌بندی را ارزیابی نمایید.

۵ - روش شبکه خودسازمانده را با یکی دیگر از روش‌های مرسوم کاهش بعد مانند PCA یا t-SNE مقایسه کنید. تحلیل خود را هم از نظر تفسیرپذیری نتایج و هم از منظر عملکرد دسته‌بندی پس از کاهش بعد ارائه دهید.

۶ - (سوال امتیازی^۲) یک مدل سلسله‌مراتبی از شبکه‌های خودسازمانده برای کاهش ابعاد مجموعه داده به دو بُعد پیشنهاد دهید. در این مدل، انتظار می‌رود لایه‌های بالاتر شبکه بتوانند نواحی با تراکم داده‌ای بالاتر را با دقت بیشتری مدل‌سازی کنند. به‌طور مشخص، مدل باید قادر باشد در نواحی‌ای که نرون‌های بیشتری فعال می‌شوند، جزئیات بیشتری را یاد بگیرد و ساختار داده‌ها را بهتر نمایش دهد. معماری پیشنهادی خود را توضیح داده و نحوه عملکرد آن را در مقایسه با یک شبکه خودسازمانده سنتی تحلیل کنید.

معیارهای ارزیابی

برای تحلیل دقیق‌تر نتایج حاصل از کاهش بعد و خوشه‌بندی، از معیارهای زیر استفاده نمایید:

- Accuracy: نرخ صحت پیش‌بینی مدل دسته‌بندی پس از کاهش بعد.
- Adjusted Rand Index (ARI): برای ارزیابی کیفیت خوشه‌بندی نسبت به برچسب‌های واقعی.
- Silhouette Score: برای ارزیابی فاصله بین خوشه‌ها و یکنواختی داخلی خوشه‌ها.
- Confusion Matrix: تحلیل خطاهای مدل دسته‌بندی.
- Training Time: زمان مورد نیاز برای آموزش شبکه خودسازمانده تحت تنظیمات مختلف.

^۲ پاسخ دادن به این سوال، اجباری نیست. در صورتی که پاسخ مناسبی به این سوال بدهید، از نمره امتیازی مازاد بر نمره تمرین بهره‌مند خواهید شد.

در انجام تمرینات به نکات زیر توجه فرمایید.

۱ - پیاده‌سازی‌های کامپیوتری را به زبان برنامه‌نویسی پایتون و با بهره‌گیری از چارچوب کاری PyTorch انجام دهید.

۲ - بخش عمده‌ای از نمره تمرینات به گزارش تمرین اختصاص دارد و ارسال برنامه‌ها بدون گزارش فاقد ارزش است. در تهیه گزارش دقت نمایید که تمام اطلاعات، تصاویر و نمودارهای مورد نیاز برای اثبات پاسخ‌ها مبتنی بر آزمایشات خواسته شده در تمرین، بطور کامل و دقیق ذکر شده باشند.

۳ - مطابق قوانین دانشگاه هر نوع کپی‌برداری و اشتراک کار دانشجویان غیرمجاز است. در صورت مشاهده چنین مواردی، با طرفین شدیداً برخورد خواهد شد.

۴ - استفاده از کدها و توضیحات اینترنت یا کدها و توضیحات تولید شده با مدل‌های هوش مصنوعی به منظور یادگیری بلامانع است، اما کپی‌برداری و انجام تمرینات توسط این ابزارها غیرمجاز است. در صورتی که از چنین ابزارهایی بهره می‌گیرید، حتماً به تمام جزئیات و نکات مرتبط با پاسخ‌ها مسلط باشید، در غیر این‌صورت نمره کل تمرین را از دست خواهید داد.

۵ - مجموعه داده‌های مورد استفاده را به جز در مواردی که صریحاً در صورت سوال ذکر شده باشد، حتماً قبل از استفاده بطور تصادفی به سه بخش آموزشی (۷۰ درصد)، آزمون (۲۰ درصد) و اعتبارسنجی (۱۰ درصد) تقسیم نمایید.

۶ - در صورت نیاز می‌توانید سوالات خود را در خصوص پروژه از تدریس‌یارهای درس، به ایمیل زیر ارسال نموده یا در گروه «بله» مطرح نمایید.

ann.ceit.aut@gmail.com

۷ - فایل‌های کد و گزارش خود را در قالب یک فایل فشرده با فرمت StudentID_HW05.zip تا تاریخ ۱۴۰۴/۰۲/۰۵ فقط در بخش مربوطه در سایت درس بارگذاری نمایید. توجه نمایید، هر روز تاخیر منجر به کسر ۱۰ درصد از نمره پروژه می‌شود.