

تمرین دوم

هدف: آشنایی با شبکه‌های خودسازمانده کوهونن

کد: کد این فعالیت را به زبان پایتون بنویسید.

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 تهیه کنید و در آن برای هر سوال، تصاویر ورودی، تصاویر خروجی و توضیحات مربوط به آن را ذکر کنید. سعی کنید توضیحات کامل و جامعی تهیه کنید.

تذکر ۱: 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.

تذکر ۲: در تمام موارد، مجموعه داده‌های ورودی را به سه قسمت آموزش (۷۰ درصد)، اعتبارسنجی^۱ (۲۰ درصد) و آزمون (۱۰ درصد) تقسیم نمایید. آموزش شبکه را روی مجموعه داده آموزشی انجام دهید و تعداد تکرارها را با استفاده از مجموعه داده اعتبارسنجی بیابید. در نهایت برای ارزیابی عملکرد مدل از مجموعه داده آزمون استفاده نمایید.

تذکر ۳: برای ارزیابی عملکرد مدل‌ها در بخش دسته‌بندی از معیارهای دقت^۲ و ماتریس درهم‌ریختگی^۳ و در بخش رگرسیون از معیار میانگین مربعات خطا^۴ استفاده نمایید.

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_HW۰۲.zip تا تاریخ ۱۳۹۹/۰۹/۱۴ ارسال نمایید. شایان ذکر است هر روز تاخیر باعث کسر ۱۰٪ نمره خواهد شد.

بخش اول

در این بخش قصد داریم عملکرد شبکه خودسازمانده کوهونن را در کاهش بعد^۵ و بصری‌سازی^۶ داده‌ها بررسی نماییم. مجموعه داده مورد استفاده، به فایل تعریف پروژه پیوست شده است. این مجموعه داده شامل داده‌های مربوط به آراء نمایندگان کنگره آمریکا به لوايح مختلف است. هر فایل موجود در این مجموعه داده، آراء هر یک از نمایندگان کنگره به یک لایحه مشخص را نمایش می‌دهند. همین‌طور در هر یک از این فایل‌ها نام هر نماینده و حزب مربوط به آن نماینده نیز ذکر شده است.

^۱ Validation

^۲ Accuracy

^۳ Confusion matrix

^۴ Mean squared error

^۵ Dimensionality reduction

^۶ Visualization

- ۱- ابتدا در فایل گزارش توضیح دهید چگونه می‌توان از یک شبکه عصبی خودسازمانده کوهونن برای نمایش داده‌های با ابعاد بالا در یک نقشه دوبعدی استفاده کرد؟ سعی کنید فرایند مورد نظر را در قالب یک شبه کد توضیح دهید.
- ۲- فرایند فوق را پیاده‌سازی کرده و داده‌های موجود در مجموعه داده ضمیمه شده را با استفاده از این مدل نمایش دهید. توجه نمایید در نقشه‌های دوبعدی تولید شده، داده‌های مربوط به نمایندگان دموکرات را با رنگ آبی و داده‌های مربوط به نمایندگان جمهوری خواه را با رنگ قرمز نمایش دهید. راهنمایی: برای این کار می‌توانید از توضیحات ارائه شده در ویدیو شماره ۱۰ کلاس تدریسار استفاده نمایید.
- ۳- نقشه‌های دو بعدی حاصل از شبکه خودسازمانده در سوال قبل را تفسیر کنید. به نظر شما اگر بخواهیم بر اساس رای هر نماینده به لویای مختلف، دموکرات یا جمهوری خواه بودن آن‌ها را تشخیص بدهیم، بهتر است از آراء نمایندگان به کدام یک از لویای موجود در مجموعه داده استفاده نماییم؟ دلیل خود را به طور کامل شرح دهید.
- ۴- با تغییر تعداد نوروهای این شبکه، سعی کنید ساختار بهینه برای بصری‌سازی داده‌ها را بیابید. نتایج تمامی آزمایشات خود را به طور کامل در فایل گزارش ذکر نمایید.

بخش دوم

- در این قسمت قصد داریم با استفاده از نقشه‌های دوبعدی حاصل از شبکه در بخش قبل، یک دسته‌بند^۷ برای تشخیص دموکرات یا جمهوری خواه بودن نمایندگان ایجاد نماییم.
- ۱- در فایل گزارش توضیح دهید خروجی شبکه خودسازمانده کوهونن چگونه می‌تواند به عنوان یک بردار ویژگی برای دسته‌بندی استفاده شود؟
 - ۲- با استفاده از ویژگی‌های استخراج شده توسط شبکه خودسازمانده پیاده‌سازی شده در بخش اول و بر اساس روش توضیح داده شده در سوال قبل، یک دسته‌بند برای مجموعه داده مورد استفاده ساخته و آموزش دهید. توجه نمایید در این تمرین باید از شبکه پرسپترون چندلایه به عنوان دسته‌بند استفاده نمایید. پارامترهای مربوط به تعداد لایه‌ها و تعداد نوروهای هر لایه را بهینه نموده و نتایج آزمایشات را به همراه ساختار بهینه شبکه در فایل گزارش به طور کامل توضیح دهید.
 - ۳- با توجه به آزمایشات انجام شده در هر دو بخش تمرین، توضیح دهید کدام یک از ساختارهای آزمایش شده از شبکه خودسازمانده کوهونن برای استخراج ویژگی از مجموعه داده مورد استفاده بهتر است؟ چرا؟

موفق باشید

^۷ Classifier