

پیشنهاد می‌شود پیش از آغاز پیاده‌سازی یک بار تعریف پروژه را تا انتها بخوانید.

هدف از این پروژه بکارگیری شبکه خودسازمانده برای خوشه‌بندی اسناد است. شبکه خود سازمانده باید در حالت‌های یک بعدی، دوبعدی و سه بعدی پیاده‌سازی گردد. در تمام پیاده‌سازی‌ها باید نرخ یادگیری بر اساس مراحل آموزش تنظیم گردد.^۱ برای این پروژه از مجموعه داده [20news groups](#) استفاده کنید. برای خوشه‌بندی اسناد موجود در این مجموعه داده، ابتدا برای هر سند، یک بردار ویژگی مبتنی بر مدل کیسه کلمات^۲ استخراج نمایید. برای استخراج ویژگی‌ها می‌توانید از ابزارهای آماده استفاده کنید. سپس با استفاده از بردارهای ویژگی استخراج شده شبکه را برای خوشه‌بندی آموزش داده و شبکه آموزش دیده را ذخیره کنید. از شبکه ذخیره شده برای خوشه‌بندی داده‌های تست استفاده کرده و مدل را بر این اساس ارزیابی نمایید. برای ارزیابی مدل از معیارهای RI ^۳، $purity$ و معیار F استفاده کنید.^۴

$$purity(\Omega, \mathbb{C}) = \frac{1}{N} \sum_k \max_j |\omega_k \cap c_j|$$

where $\Omega = \{\omega_1, \omega_2, \dots, \omega_K\}$ is the set of clusters and $\mathbb{C} = \{c_1, c_2, \dots, c_J\}$ is the set of classes.

$$RI = \frac{TP + TN}{TP + FP + FN + TN}$$

$$P = \frac{TP}{TP + FP} \quad R = \frac{TP}{TP + FN} \quad F_\beta = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}$$

$$\beta = 1$$

¹ Adaptive Learning Rate

² Bag of words

³ Rand index

^۴ برای اطلاع بیشتر می‌توانید از لینک زیر استفاده کنید.

موارد قابل بررسی:

- معماری های مختلف (یک بعدی، دو بعدی و سه بعدی) را بر اساس معیارهای ارزیابی مقایسه کرده و نتایج را تحلیل کنید.
- نتایج خوشه بندی حاصل از این شبکه و خوشه بندی kmeans را مقایسه نمایید. میتوانید از تابع آماده kmeans در کتابخانه های موجود استفاده کنید.
- شبکه GSOM را پیاده سازی کرده و عملکرد آن را با شبکه فوق مقایسه نمایید.
- در این پروژه ارائه بهترین مدل (از نظر معیارهای ارزیابی) ضروری است.

پیاده سازی و واسط کاربری:

کد این پروژه را در محیط متلب و یا پایتون بنویسید. در پیاده سازی ها از هیچ یک از ابزارهای آماده مربوط به شبکه های عصبی، نمی توانید استفاده کنید. شرایط زیر را در پیاده سازی و واسط کاربری در نظر بگیرید:

- می بایست امکان تنظیم پارامترهای متفاوت الگوریتم های پیاده سازی شده وجود داشته باشد. این پارامترها به شرح زیرند:
 - تعداد لایه های کوهونن و تعداد نوروں های هر لایه.
 - پارامتر های نوع فاصله برای تعیین نوروں برنده.
 - پارامتر های تعیین نوع فاصله: قابل تشخیص و غیر قابل تشخیص، توپولوژی های متفاوت دایره، مستطیل و شش ضلعی.
- نحوه تقسیم داده ها به مجموعه های آموزشی، ارزیابی و آزمایشی در واسط کاربری قابل تنظیم باشد.
- برنامه شما بایستی قادر به ذخیره شبکه آموزش دیده و بازیابی یک شبکه از پیش آموزش دیده باشد. به عبارت دیگر بایستی بتوانید وزن های شبکه را ذخیره و بازیابی کنید. ذخیره بهترین مدلی که در فرآیند آموزش به آن رسیده اید برای بررسی پروژه در زمان تحویل حضوری ضروری است.
- نمودارهای حاصل از اجرا باید به صورت گام به گام در طول آموزش نمایش داده شود. همچنین نمودارها باید قابلیت ذخیره سازی داشته باشد. این امکان باید وجود داشته باشد که بتوان نتایج نمودارهای یک اجرا را با اجراهای دیگر مقایسه کرد.
- شرط پایان آموزش قابل انتخاب باشد. آموزش شبکه را می توان پس از تعداد مشخصی اپک آموزشی و یا با رسیدن به خطای معینی متوقف کرد.

گزارش:

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

- عنوان: در این بخش بیان کنید پاسخ کدام یک از موارد را بررسی می کنید.
- 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.
- نتیجه انجام آزمایش: در این بخش نمودارهای مورد نیاز به همراه توضیحات آن ها را ذکر کنید. (1 تا 2 پاراگراف)
- نتیجه گیری: نتایج حاصل از بررسی های خود را با دلایل آن به طور کامل توضیح دهید. در بیان دلایل اگر به مرجع خاصی اشاره شود بهتر است. (2 تا 3 پاراگراف)

نحوه تحویل:

فایل‌های کد و گزارش خود را در قالب یک فایل فشرده به شکل Pr3_StdNum.rar نامگذاری نمایید (مانند Pr3_94131052.rar) و در سایت دروس دانشکده کامپیوتر بارگذاری کنید. مهلت انجام پروژه تا سه شنبه 97/2/18 است.

مطابق قوانین دانشگاه هر نوع کپی‌برداری و اشتراک کار دانشجویان غیرمجاز بوده و شدیداً برخورد خواهد شد. استفاده از کدها و توضیحات اینترنت به‌منظور یادگیری تشویق می‌شود اما کپی کردن غیرمجاز است.

در صورت نیاز می‌توانید سؤالات خود را در خصوص صورت پروژه از تدریس یاران درس، مهندس اصلانزاده و مهندس اسدی، در فرومی که برای آن در سایت دروس دانشکده ساخته‌شده مطرح کنید.