

به نام خدا



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

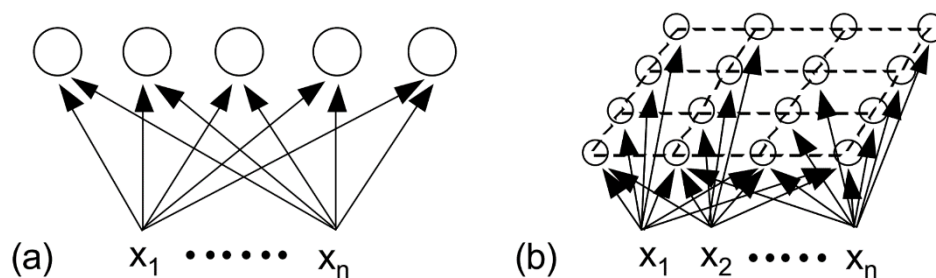
درس شبکه‌های عصبی
استاد صفابخش

تمرین سوم

علیرضا مازوچی
۴۰۰۱۳۱۰۷۵

سوال ۱

در مدل SOM که توسط کوهونن ارائه شده است معماری شبکه از دو لایه تشکیل شده است؛ یک لایه ورودی از نورون‌های که تعداد آن برابر با ابعاد داده‌های مسئله است و یک لایه خروجی که تعداد نورون‌های آن برابر با تعداد خوشه‌های اولیه است که ما آن را مدنظر داریم. نورون‌های لایه خروجی می‌تواند به صورت یک بعدی و به فرم یک خط باشند و می‌توانند به صورت دو بعدی و به فرم یک گرید قرار گرفته باشند. در تصویر زیر معماری شبکه برای هر دو حالت مذکور آورده شده است:



یادگیری SOM مطابق با یادگیری رقابتی است؛ در این حالت داده‌های ورودی یکی یکی به شبکه داده می‌شوند. برای هر داده، یک نورون خروجی دارای کمترین فاصله با داده ورودی است. این نورون اصطلاحاً نورون برنده نامیده می‌شود. تعدادی از نورون‌ها در مجاورت این نورون برنده هستند. خود نورون برنده و همسایه‌های آن در هنگام ورودی داده بروز می‌شوند و وزن مابقی نورون‌ها هیچ تغییری نخواهد کرد. این تغییر در جهتی است که نورون برنده و همسایه‌های آن را به داده ورودی نزدیک‌تر کند. منطقاً انتظار داریم نورونی که به نورون برنده نزدیک‌تر است بیشتر تغییر کند و بالعکس.

مطابق با الگوریتم‌های یادگیری ماشین، تمایل داریم که در ابتدا متناسب با یک نرخ یادگیری بالا تغییرات شدید باشد و در انتها این تغییرات اندک باشد. همچنین در ابتدا تمایل داریم محدوده همسایگی گسترده‌تری داشته باشیم و به مرور این محدوده کوچک و کوچک‌تر شود. در فرمول زیر نحوه بروزرسانی نورون‌ها آورده شده است:

$$w_j(t) = w_j(t - 1) + \beta(t) * NS(d, t) * (x(t) - w_j(t - 1))$$

در این رابطه t زمان و j اندیس هر نورون خروجی است. $w_j(t - 1)$ وزن سابق نورون j و $w_j(t)$ وزن بروز شده آن است. $\beta(t)$ نرخ یادگیری است که تابع زمان است و

انتظار می‌رود با گذشت زمان کمتر شود. $NS(d, t)$ اصطلاحاً قدرت همسایگی است و مشخص می‌کند تا چه محدوده‌ای نورون‌ها همسایه یک نورون برنده هستند. قدرت همسایگی به صورت کلی تابع زمان و فاصله با نورون برنده است. طبیعتاً انتظار می‌رود با افزایش زمان و یا فاصله این مقدار کم شود تا بروز شدن یک نورون معقول باشد. $x(t)$ مقدار ورودی در زمان t ام است و نهایتاً $(x(t) - w_j(t - 1))$ جهتی است که یک نورون را به ورودی لحظه t ام نزدیک می‌کند. طبیعتاً هرچه ضرایب پشت این مقدار بیشتر باشد وزن نورون به داده نزدیک‌تر می‌شود.

پس از آنکه شبکه آموزش یافت هر نورون به ازای برخی از داده‌ها برنده می‌شود و به نوعی مسئول آن دسته از داده‌ها می‌شود. در اصل این نورون مرکز خوشه‌ای از داده‌هایی است که مسئولش شده است. هر داده جدید هم بسته به آنکه به کدام نورون یا مرکز خوشه نزدیک است متعلق به آن نورون و خوشه متناظر خواهد بود. اگر نورون‌ها به صورت یک‌بعدی در شبکه باشند، هر داده یک شماره خوشه دارد و اگر دو بعدی باشد، هر داده دو مختصات که بیانگر مختصات نورون برنده آن است را خواهد داشت. پس بدین ترتیب کاهش ابعاد داریم و فضای ورودی فارغ از ابعاد آن به یک فضای یک یا دو بعدی تبدیل می‌شود. همچنین می‌توانیم به جای نگهداری مختصات نورون برنده یک بردار از مقادیر تعلق داده به هر یک از نورون‌های خروجی ایجاد کنیم. این بردار می‌توان one-hot باشد و می‌تواند دارای مقادیر مختلف باشد. در این صورت تعداد ابعاد برابر با تعداد نورون‌های لایه خروجی می‌شود.

سوال ۲

نورون مرده نورونی است که در طول یادگیری هیچ‌گاه برنده نشده است. این نورون‌ها هیچ نقشی در بهبود نتایج خوشه‌بندی ندارد و بود و نبود آن باعث نمی‌شود تا داده‌ها به خوشه‌های بیشتری تقسیم شوند. لذا در این شرایط بی‌جهت منابع حافظه و پردازشی هدر می‌رود. چراکه باید در هر حال وزن‌های آن را نگه داشت و برای هر داده هم باید فاصله نورون تا داده را سنجید.

نورون مرده در شرایطی به وجود می‌آید که آن نورون از داده‌های ورودی فاصله داشته باشد. لذا یک راه حل آن است که مقداردهی اولیه نورون‌ها کاملاً تصادفی نباشد و مقدار برخی از داده‌ها به وزن‌ها نسبت داده شود. به این شکل اغلب نورون‌ها در گام اول یک بار برنده می‌شوند. به عنوان ایده‌ای بهتر می‌توان انتخاب داده‌ها به عنوان مقدار اولیه وزن‌ها را به گونه‌ای انتخاب کرد که چگالی نورون‌ها در فضای ورودی تقریباً یکنواخت باشد. به عنوان مثال شاید ایده‌ی مقداردهی اولیه KMeans++ ایده خوبی باشد.

در مقالات هم به روش‌های دیگری اشاره کرده‌اند؛ مثلاً می‌توان بعد از چندین گام آموزش لیست نورون‌های مرده را درآورد و به ازای هر کدام نزدیک‌ترین داده موجود را پیدا کرد و آن نورون را به گونه‌ای آموزش داد که به سمت آن داده حرکت کند. بدین ترتیب نورون‌های مرده به سمت داده‌ها حرکت می‌کنند. یا به عنوان راه دیگری می‌توان در ابتدا محدوده همسایگی نورون برنده را خیلی بزرگ در نظر گرفت تا نورون‌هایی که با بقیه فاصله زیادی دارند هم به مرور به سمت محل‌های پرداده حرکت کنند.^۱

¹http://www.dcs.fmph.uniba.sk/diplomovky/registracia/getfile.php/master_thesis_novak.pdf?id=443&fid=788&type=application%2Fpdf

سوال ۳

با توجه به اینکه مجموعه داده در حال حاضر به دو قسمت آموزش و آزمون شکسته شده است، قست آزمون را بدون تغییر گذاشتم ولی تصمیم گرفتم بخشی از قسمت آزمون را به اعتبارسنجی تخصیص بدهم. من ۱۲/۵٪ آموزش را به اعتبارسنجی و مابقی را برای آموزش جدید در نظر گرفتم. بدین ترتیب نسبت ۱ به ۷ آموزش به اعتبارسنجی که در تذکر ۲ به آن اشاره شده است رعایت می شود. نهایتا باید ذکر کنم که داده ها به صورت تصادفی شافل شده اند.

سوال ۴

برای پیدا کردن مقادیر بهینه در دو مرحله کار را انجام داده ام: ابتدا نرخ یادگیری را برای یک شبکه پیش فرض تعیین می کنم و سپس با ثابت کردن نرخ یادگیری تعدادی شبکه مختلف با تعداد لایه و تعداد نورون مختلف در هر لایه را بررسی کرده ام. طبیعی است که در اجراهای مختلف دقت های مختلف حاصل می شود ولی من اعداد مربوط به آخرین اجرا را در نظر گرفته ام.

شبکه ای که برای قسمت اول استفاده شده است دارای سه لایه مخفی با ابعاد به ترتیب ۱۲۸، ۶۴ و ۳۲ نورون است. مقادیر مختلف مربوط به نرخ یادگیری بررسی شده به همراه صحت آن ها در جدول زیر آورده شده است:

ردیف	نرخ یادگیری	صحت آموزش	صحت اعتبارسنجی
۱	۰/۰۰۰۱	۹۸/۸۳	۹۷/۷۱
۲	۰/۰۰۰۵	۹۸/۸۵	۹۷/۸۲
۳	۰/۰۰۱	۹۷/۹۵	۹۶/۵۲
۴	۰/۰۰۳	۹۸/۸۰	۹۸/۰۴
۵	۰/۰۰۵	۹۸/۴۰	۹۷/۱۷
۶	۰/۰۱	۹۸/۷۷	۹۷/۹۳
۷	۰/۱	۹۸/۵۷	۹۷/۹۳

با توجه به آنکه بهترین صحت برای نرخ یادگیری 0.003 حاصل شد، در ادامه این نرخ یادگیری ملاک خواهد بود. برای مرحله دوم پیدا کردن پارامتر بهینه از تنظیمات موجود در جدول زیر استفاده کردیم. در این جدول صحت آموزش و اعتبارسنجی برای هر تنظیم آورده شده است:

ردیف	تعداد لایه مخفی	تعداد نورون لایه مخفی	صحت آموزش	صحت اعتبارسنجی
۱	۰	-	۹۸/۸۲	۹۷/۹۳
۲	۱	۳۲	۹۸/۹۶	۹۷/۹۳
۳	۱	۶۴	۹۸/۸۷	۹۷/۸۲
۴	۱	۱۲۸	۹۸/۸۷	۹۸/۰۴
۵	۲	۳۲-۳۲	۹۸/۸۷	۹۸/۲۶
۶	۲	۶۴-۳۲	۹۸/۴۰	۹۷/۰۶
۷	۲	۱۲۸-۱۲۸	۹۷/۲۸	۹۶/۶۳
۸	۳	۶۴-۳۲-۳۲	۹۸/۴۹	۹۷/۷۱
۹	۳	۶۴-۶۴-۶۴	۹۸/۱۵	۹۷/۳۹
۱۰	۳	۱۲۸-۶۴-۳۲	۹۸/۲۶	۹۷/۲۸
۱۱	۴	۱۲۸-۱۲۸-۶۴-۶۴	۱۸/۷۶	۱۸/۱۷
۱۲	۴	۱۲۸-۶۴-۳۲-۱۶	۵۲/۴۵	۵۰/۸۲
۱۳	۴	۱۲۸-۶۴-۳۲-۶۴	۱۸/۷۶	۱۸/۱۷

با توجه به نتایج بدست آمده به نظر می‌رسد یک شبکه با دو لایه مخفی که هر کدام ۳۲ نورون داشته باشند مناسب‌ترین تنظیم باشد.

سوال ۵

پیش از هر چیز لازم است تا نکاتی را درباره پیاده‌سازی خودم بیان کنم:

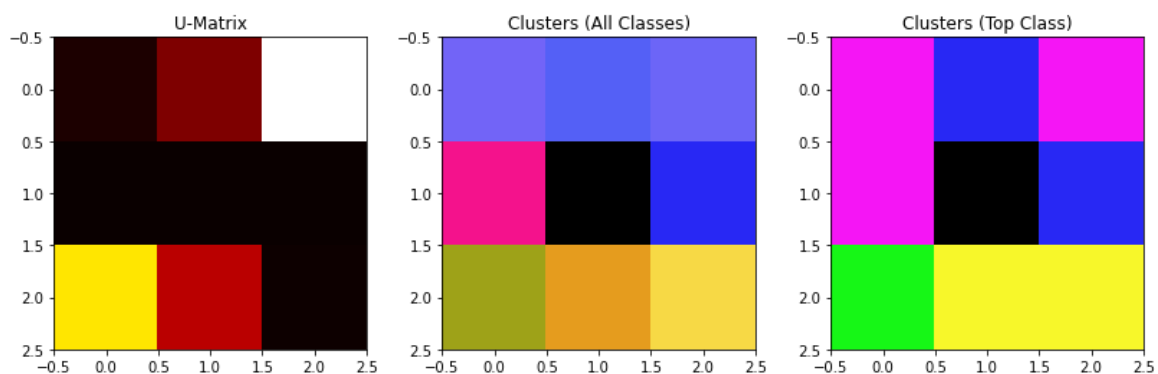
- در مدل من شعاع همسایگی به مرور زمان کاهش می‌یابد. باتوجه به آنکه این مورد در صورت سوال درخواست نشده است، برای تمام آزمایشات در هر گام مقدار شعاع ۵٪ کاهش می‌یابد.
- در مدل من حداکثر ۱۰۰ گام آموزش ادامه پیدا می‌کند. اگر اختلاف وزن‌های دو گام متوالی به کمتر از یک واحد برسد، آموزش متوقف می‌شود.
- نرخ یادگیری مطابق با درخواست سوال به صورت خطی کاهش می‌یابد. در پیاده‌سازی من در آخرین گام مقدار نرخ یادگیری برابر با ۰ است و مقدار نرخ یادگیری برای اولین گام به عنوان ورودی به مدل داده می‌شود. باتوجه به اینکه در آزمایشات من همواره تعداد گام به روی ۱۰۰ تنظیم می‌شود، می‌توان نتیجه گرفت ضریب کاهش نرخ یادگیری برابر با یک صدم نرخ یادگیری اولیه خواهد بود.
- برای درک بهتر از خروجی، علاوه بر U-Matrix که چگالی داده در نوروں‌های مختلف را نشان می‌دهد از دو نمودار دیگر هم استفاده کرده‌ام. در این دو نمودار برای هر یک از شش برچسب، یک رنگ در نظر گرفته می‌شود. در یک نمودار رنگ برچسب غالب به نوروں داده می‌شود و در نمودار دیگر میانگین رنگ برچسب‌های داده‌های نوروں ترسیم می‌شود. در هر دو نمودار رنگ سیاه به معنای نوروں مرده است. نهایتاً باید ذکر کنم که در U-Matrix رنگ روشن به معنای تراکم بیشتر و رنگ تیره به معنای تراکم کمتر است.
- برای کاهش تعداد نوروں مرده و داشتن وضعیت اولیه مناسب، وزن اولیه هر نوروں را به تصادف از نمونه‌های آموزشی موجود انتخاب کردم.

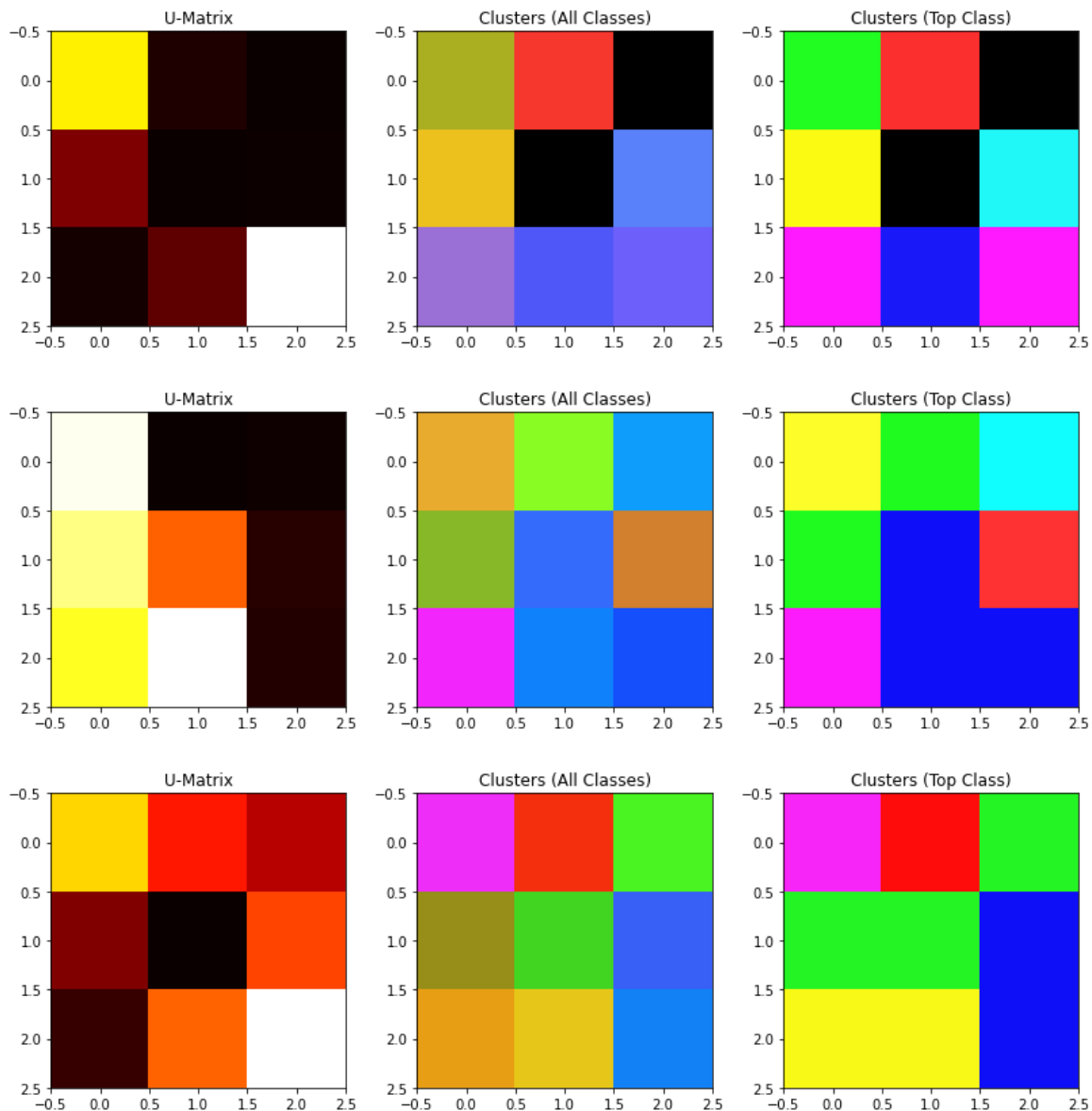
برای بررسی پارامترهای بهینه مقادیر موجود در جدول زیر را بررسی کردم:

ردیف	ابعاد لایه خروجی	شعاع همسایگی اولیه	نرخ یادگیری اولیه
۱	۳*۳	۳	۰/۰۰۱

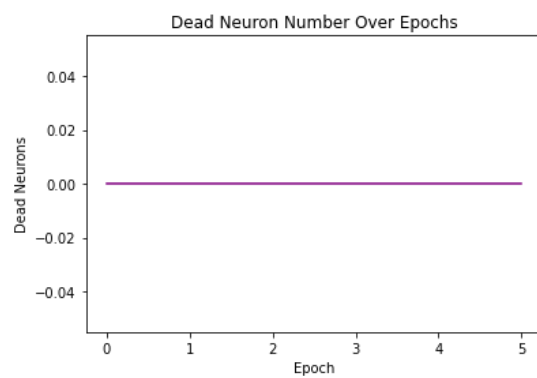
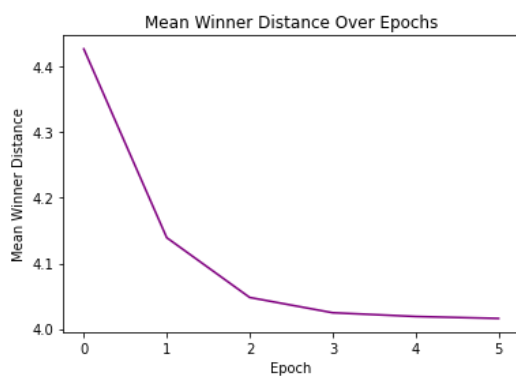
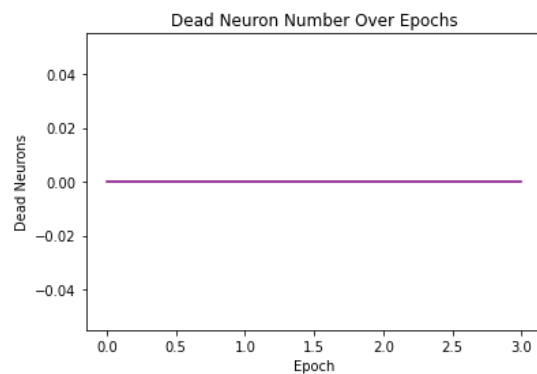
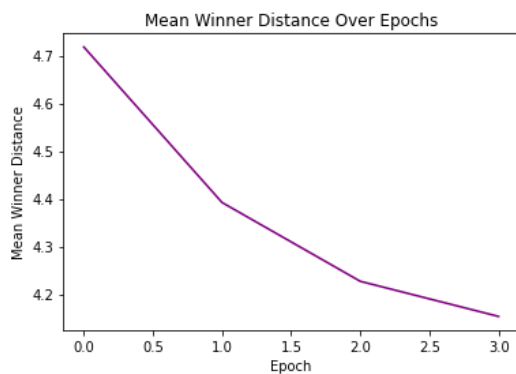
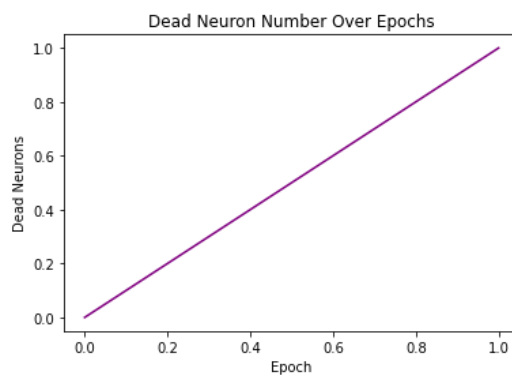
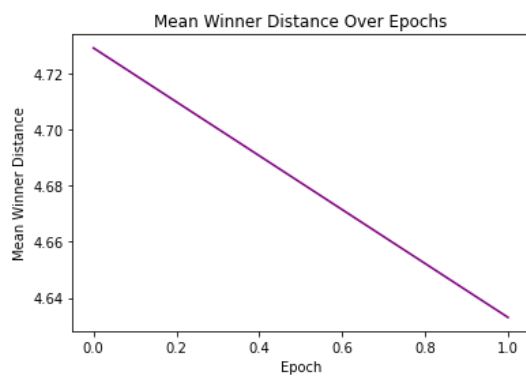
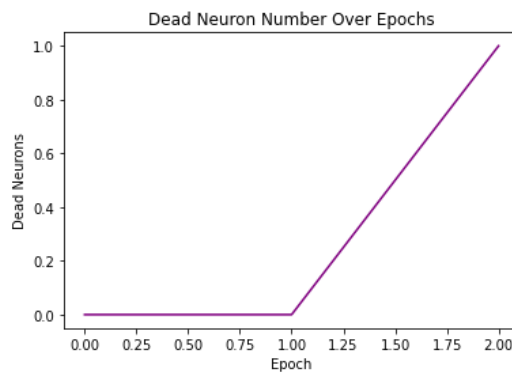
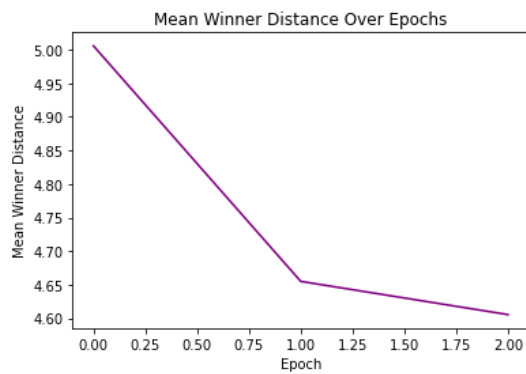
۰/۰۱	۳	۳*۳	۲
۰/۰۰۱	۱	۳*۳	۳
۰/۰۱	۱	۳*۳	۴
۰/۰۰۱	۵	۵*۵	۵
۰/۰۱	۵	۵*۵	۶
۰/۰۰۱	۳	۵*۵	۷
۰/۰۱	۳	۵*۵	۸
۰/۰۰۱	۱۰	۱۰*۱۰	۹
۰/۰۱	۱۰	۱۰*۱۰	۱۰
۰/۰۰۱	۳	۱۰*۱۰	۱۱
۰/۰۱	۳	۱۰*۱۰	۱۲
۰/۰۰۱	۲۰	۲۰*۲۰	۱۳
۰/۰۱	۲۰	۲۰*۲۰	۱۴
۰/۰۰۱	۵	۲۰*۲۰	۱۵
۰/۰۱	۵	۲۰*۲۰	۱۶

برای بررسی منظم‌تر نتایج مربوط به شبکه‌های با ابعاد یکسان را باهم بررسی خواهیم کرد. ابتدا شبکه‌های ۳*۳ را بررسی می‌کنیم. نمودارهای مربوط به لایه خروجی چهار تنظیم اول به ترتیب آورده شده است:



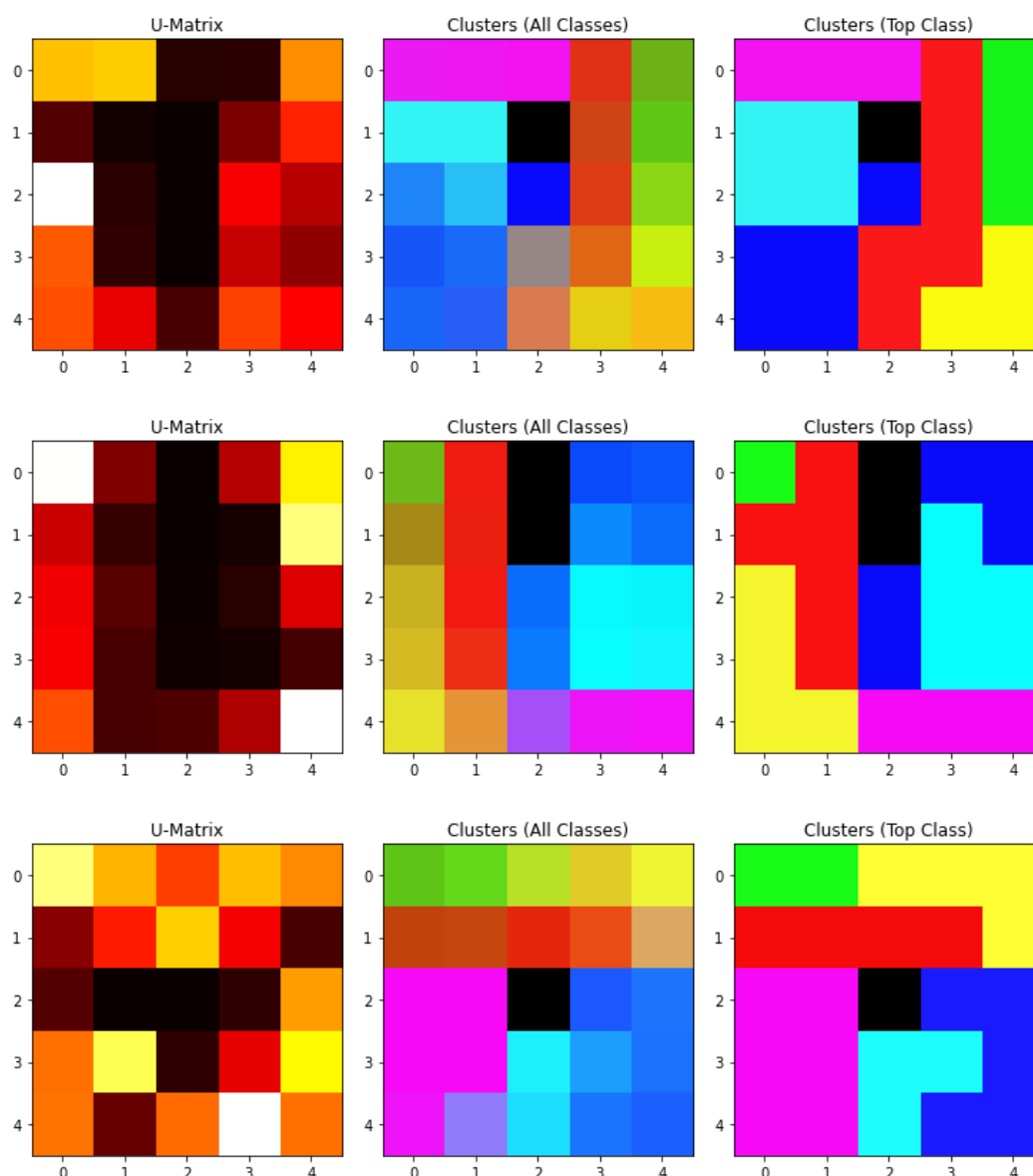


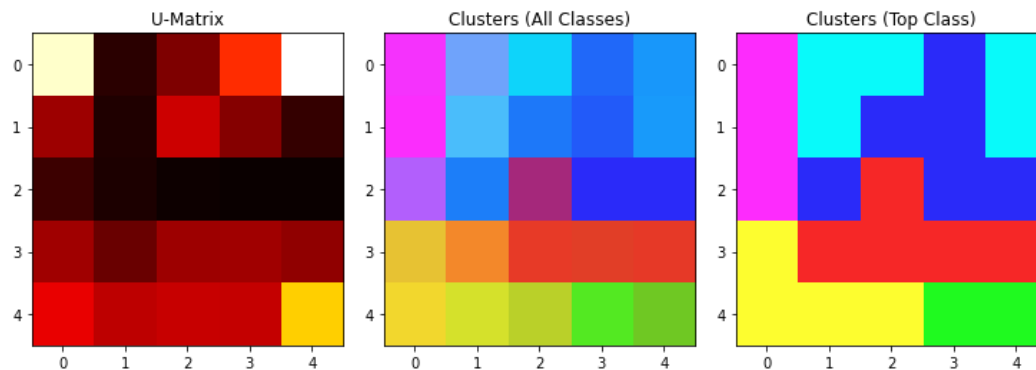
سایر نمودارهای درخواست شده هم به ترتیب در ادامه آورده شده است. اما باید قبل از آن اشاره کنم که برای ابعاد 3×3 همگرایی به سرعت رخ می‌دهد و تعداد گام‌ها خیلی کم است:



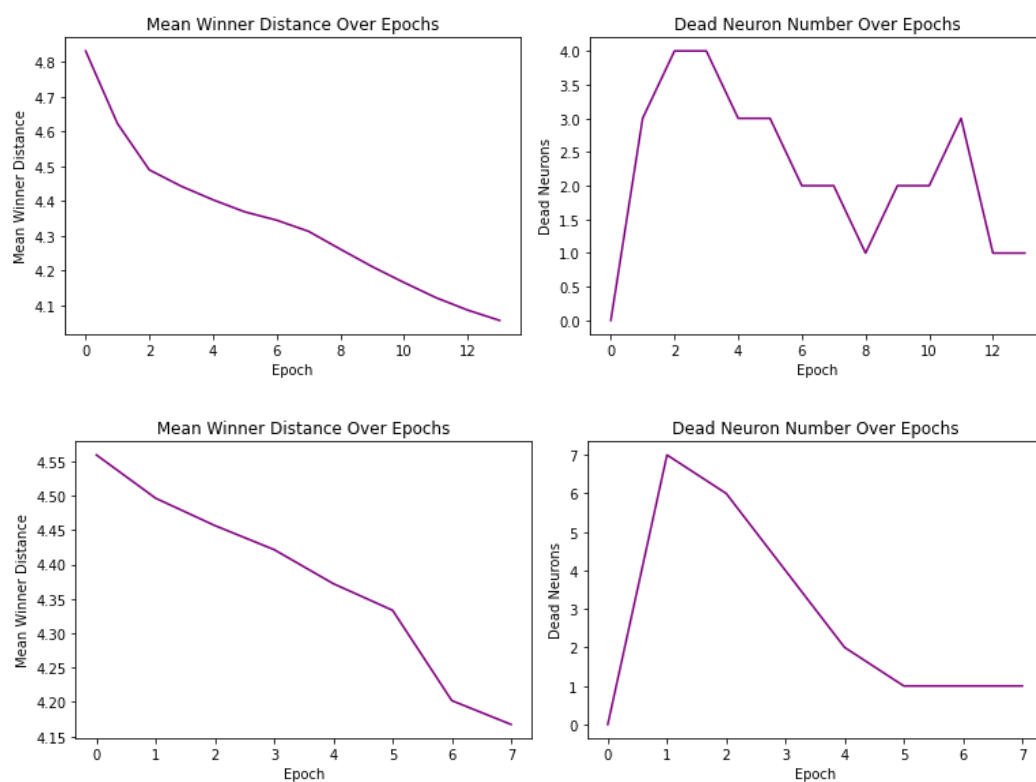
با بررسی نتایج فوق به نظر می‌رسد دو تنظیم اول (شعاع همسایگی ۳) به دلیل وجود نورون‌های مرده تنظیمات خوبی نباشند؛ چراکه در این صورت سایر نورون‌ها به سختی می‌توانند خاصیت تمایزگری داشته باشند. از بین دو تنظیم دیگر به نظر آخرین تنظیم (نرخ یادگیری ۰/۰۱) توانسته است چگالی نورون‌ها را یکنواخت‌تر و خروجی‌های خالص‌تری ایجاد کند. البته واضح است که هیچ کدام از این چهار تنظیم خروجی مطلوبی ندارند.

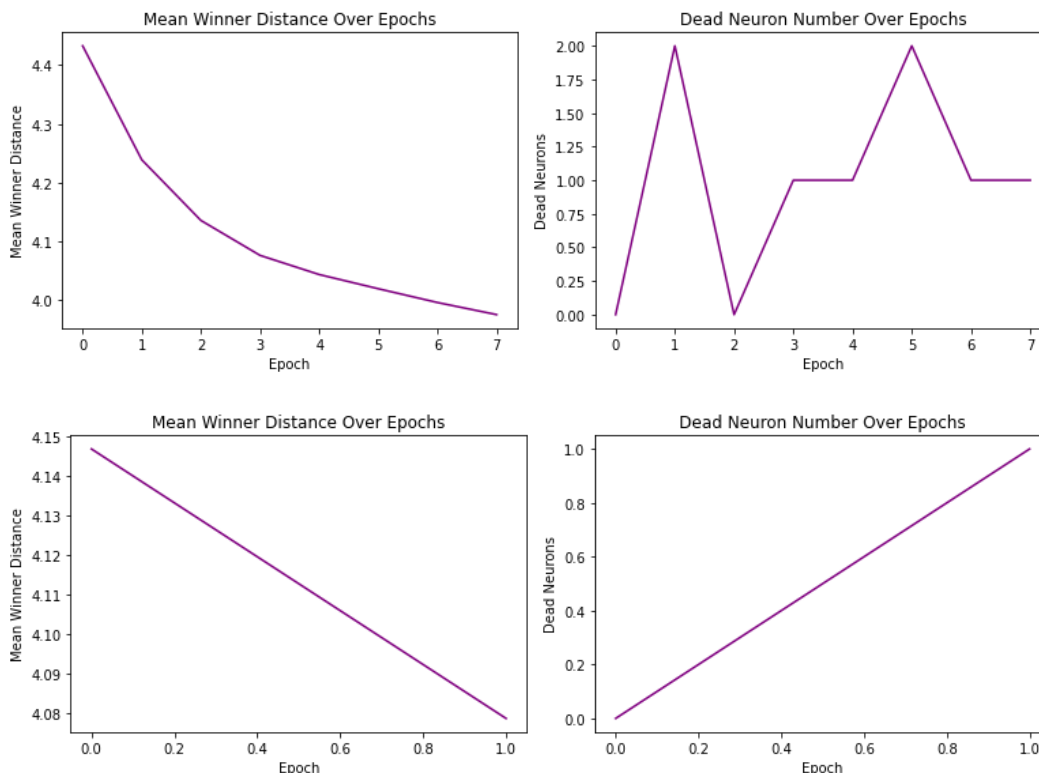
حال نتایج مربوط به چهار تنظیم شبکه‌های SOM با اندازه خروجی ۵*۵ را به ترتیب بررسی می‌کنیم:





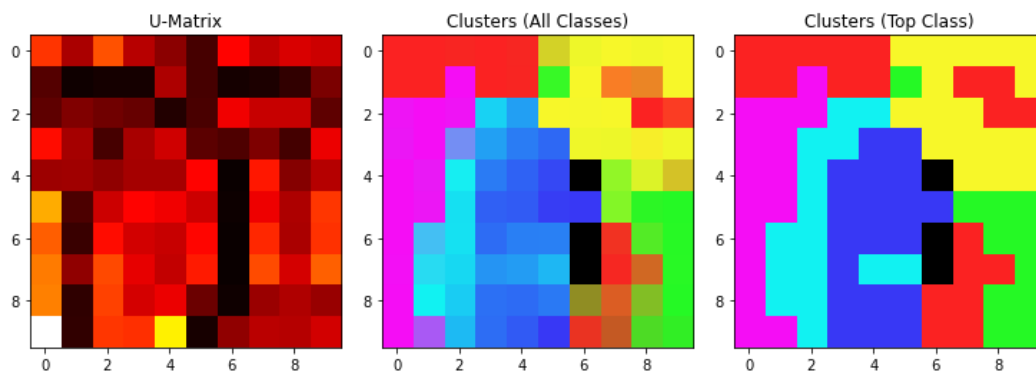
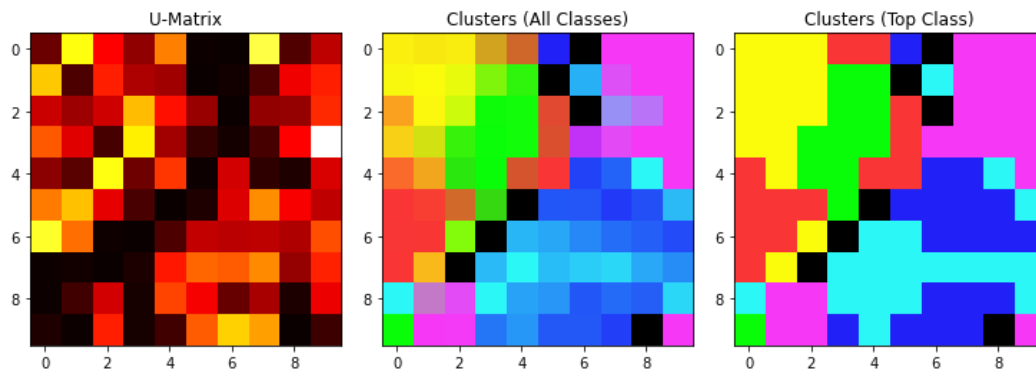
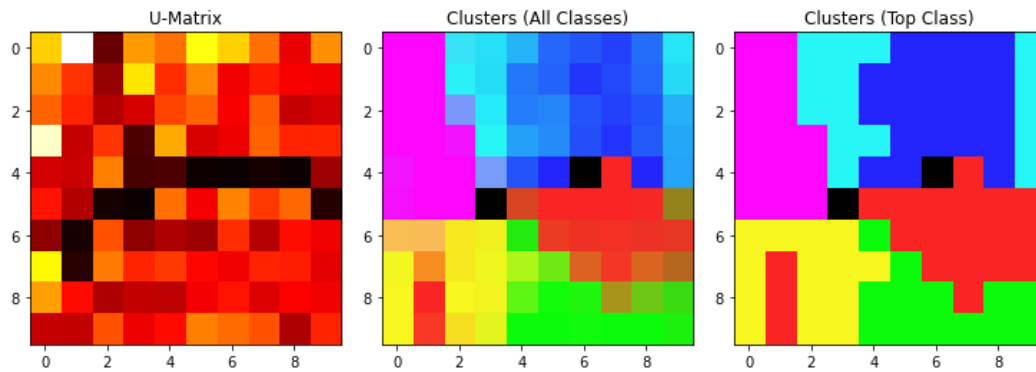
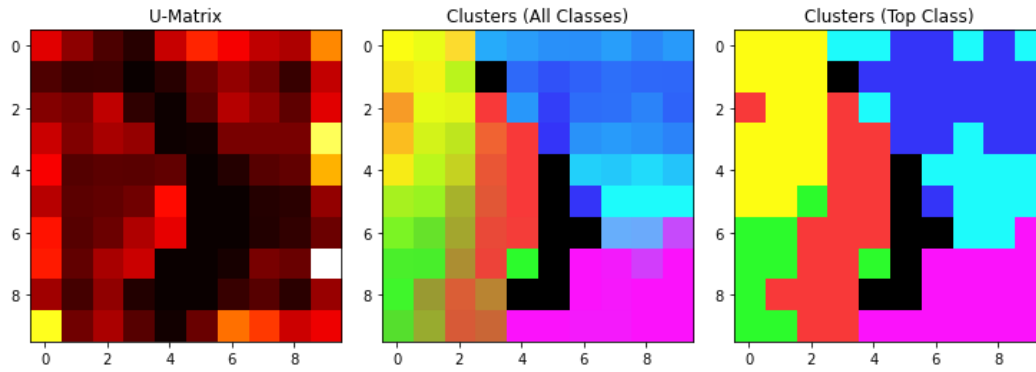
سایر نمودارها برای چهار تنظیم مربوط به شبکه‌های ۵*۵ به ترتیب در ادامه آورده شده است:

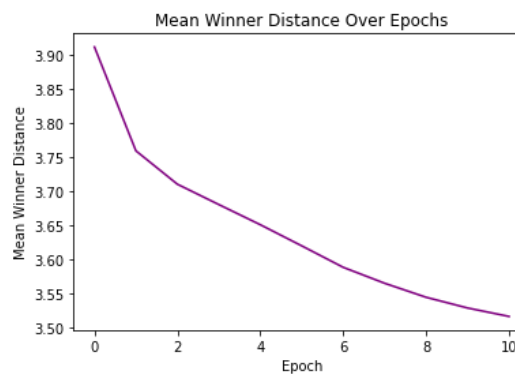
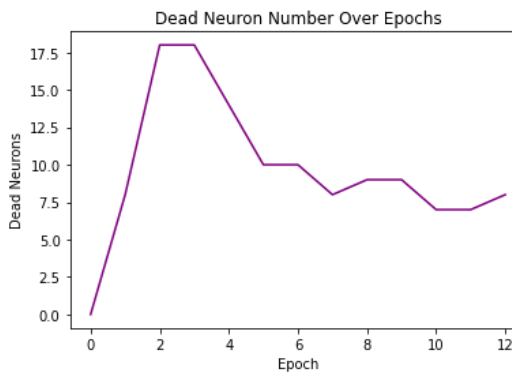
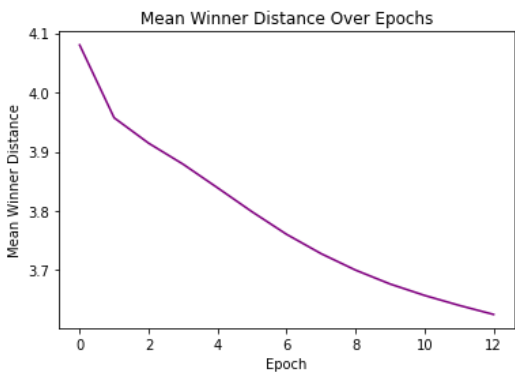
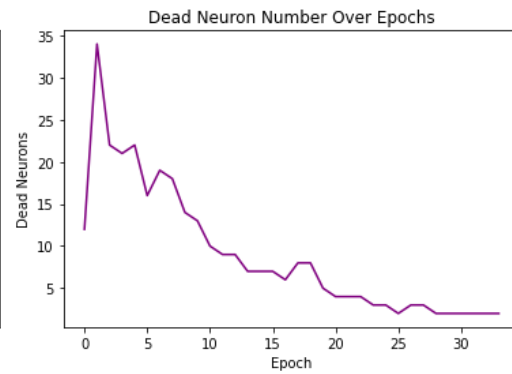
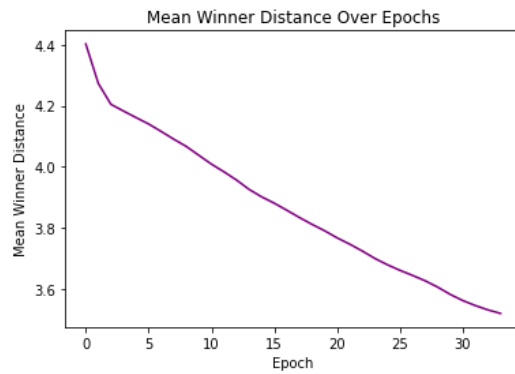
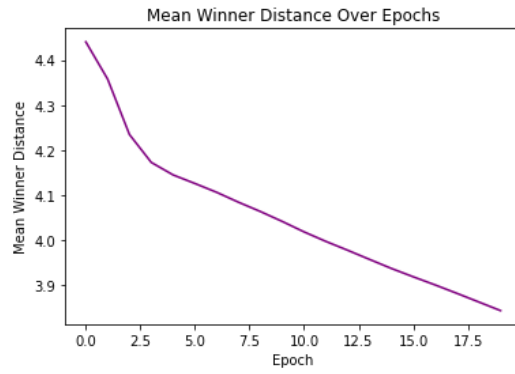




با بررسی نمودارهای فوق در می‌یابیم که نوروها مرده در تمام شبکه‌ها در انتهای کار مناسب شده است. در نمودارهای U-Matrix مخصوصا نمودار تنظیم آخر یک شکاف و فاصله بین سه کلاس آبی پررنگ، آبی کم‌رنگ و صورتی با سه کلاس زرد، سبز و قرمز دیده می‌شود. همچنین با تطبیق نمودار خوشه‌ها با U-Matrix به نظر می‌رسد کلاس صورتی از همه بهتر جدا شده است. پس با شبکه‌های فعلی یک جدایی اولیه بدست آمده است ولی مشخصا کافی نیست و باید با ابعاد بیشتر خروجی کار را پیش برد. نهایتا آنکه هر چهار تنظیم نتایج نسبتا مشابهی را داشتند ولی اگر قرار به انتخاب باشد به نظرم تنظیم آخر (شعاع همسایگی ۳ و نرخ یادگیری ۰/۰۱) مناسب‌تر است.

در نمودارهای زیر نتایج مربوط به چهار تنظیم شبکه‌های ۱۰*۱۰ به ترتیب ارائه خواهد شد:



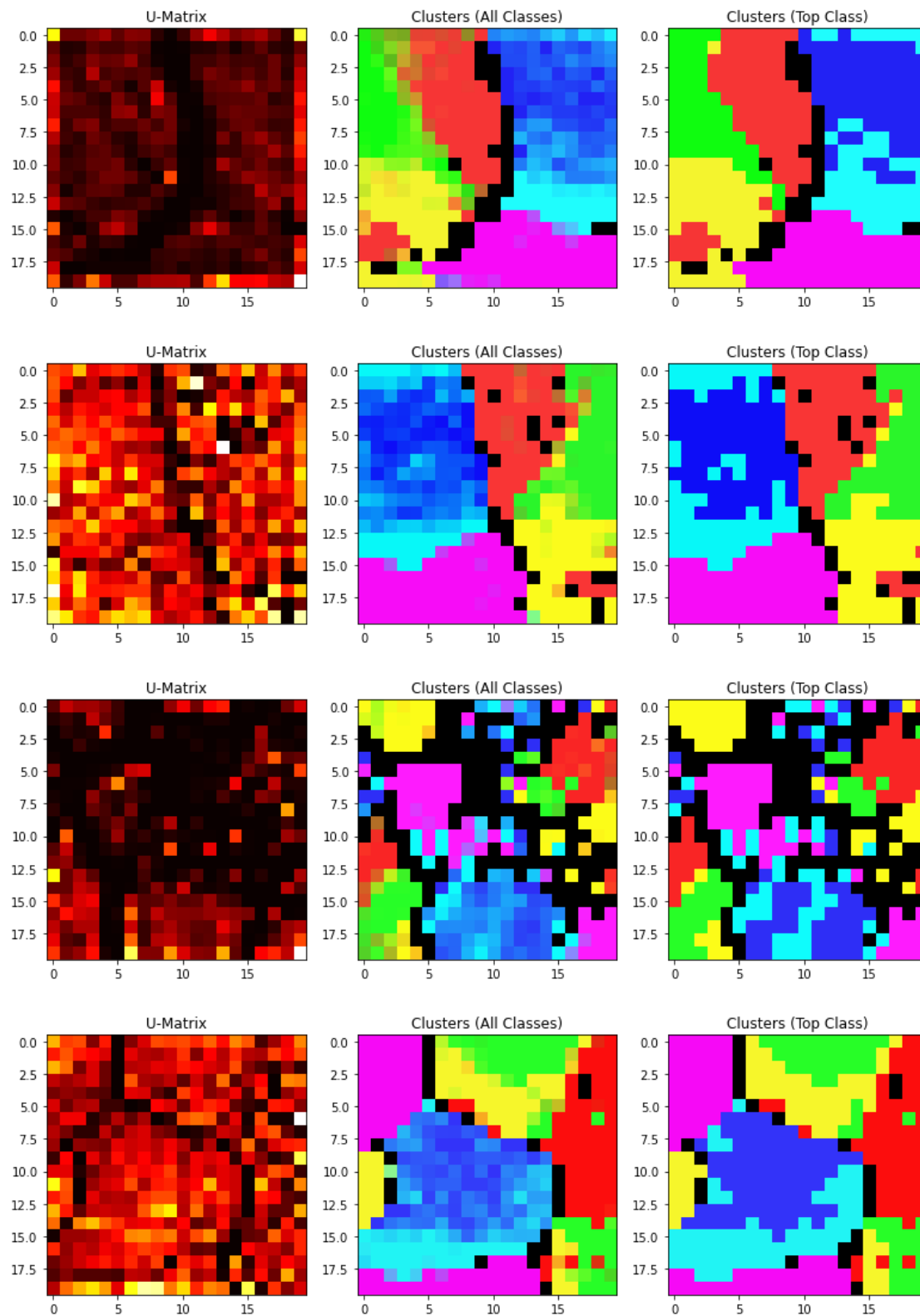


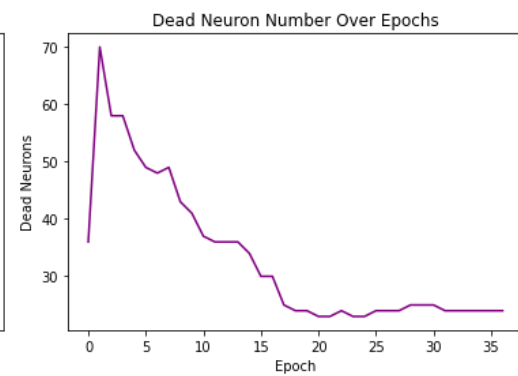
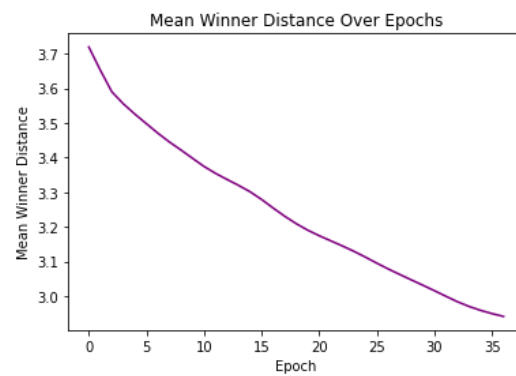
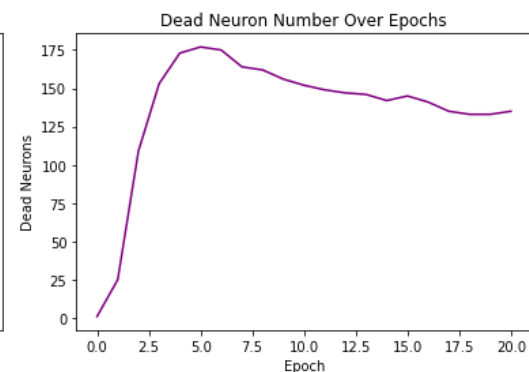
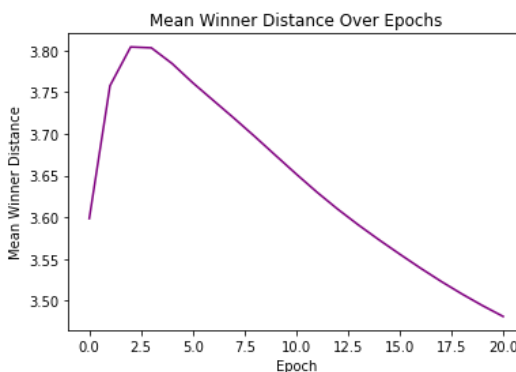
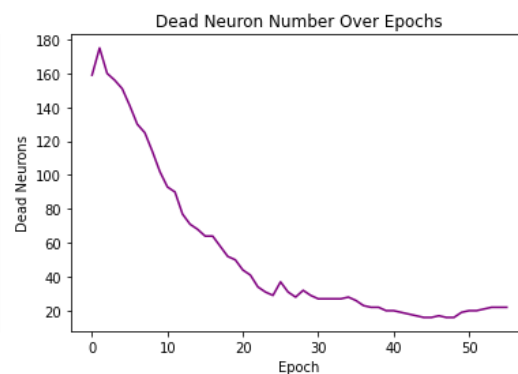
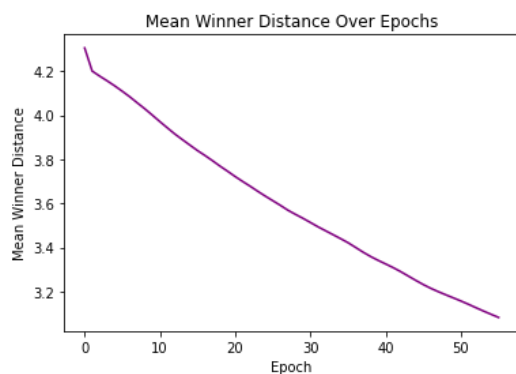
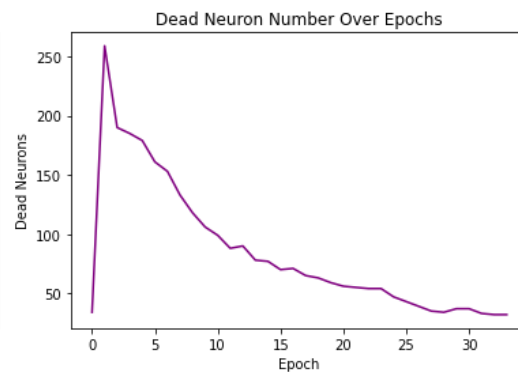
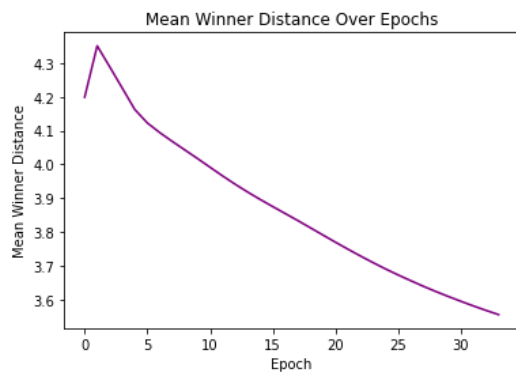
تعداد نورون مرده تمامی شبکه‌ها به نسبت قابل قبول است اما در تنظیم دوم و چهارم (نرخ یادگیری ۰/۰۱) تعداد کمتر از دو تنظیم دیگر بوده است. نکته‌ی دیگری که در تمام شبکه‌ها می‌توان ملاحظه کرد آن است که تعداد نورون مرده بعد از یک شوک اولیه به مرور کاهش پیدا کرده است که از این بابت خوب است. علت شوک اولیه این است که در ابتدای آموزش وزن هر نورون را روی یکی از داده‌ها تنظیم می‌کنیم که این باعث می‌شود تعداد نورون مرده ابتدایی به صورت مصنوعی کم باشد.

همانند شبکه‌های ۵*۵ مشاهده می‌شود که سه کلاس آبی کم‌رنگ، آبی پررنگ و صورتی از سه کلاس دیگر فاصله دارند. در این شبکه‌ها کلاس صورتی هم تقریباً جدا شده است. دو کلاس آبی با هم تداخل دارند و سه کلاس دیگر هم دو به دو مخلوط شده‌اند. در مجموع دو تنظیم اول (شعاع همسایگی ۱۰) نسبت به دو تنظیم دوم (شعاع همسایگی ۳) مطلوب‌تر به نظر می‌رسد. مثلاً کلاس قرمز در دو تنظیم دوم پخش شده است یا در تنظیم سوم حتی کلاس صورتی پخش شده است. بین دو تنظیم اول، تنظیم دوم (شعاع همسایگی ۱۰ و نرخ یادگیری ۰/۰۱) مناسب‌تر است.

این مورد در ابتدا با چیزی که در دو سری شبکه با ابعاد کوچکتر قبل دیده شد متفاوت است. در دو سری شبکه پیشین شعاع همسایگی اولیه کم نتایج بهتری داشت و در اینجا شعاع همسایگی اولیه زیاد. علت این امر به تعداد گام اجرا بر می‌گردد. در شبکه‌های کوچک به دلیل همگرایی سریع الگوریتم زود خاتمه پیدا می‌کرد و در شبکه‌های بزرگ الگوریتم دیرتر همگرا می‌شود. پس برای شبکه‌های بزرگ و به دلیل خاصیت کوچک شدن شعاع همسایگی این امکان وجود دارد که در گام‌های ابتدایی با شعاع بالا و در گام‌های انتهایی با شعاع کم کار پیش برود. در این صورت بهتر است که در ابتدا شعاع زیاد در نظر گرفته شود تا هم شعاع زیاد و هم شعاع کم در حین آموزش استفاده شود. اما در شبکه‌های کوچک تنها می‌توان یکی از اندازه حدودی شعاع را تجربه کرد که در این صورت شعاع کم مطلوب‌تر خواهد بود.

نهایتاً نوبت به بررسی شبکه‌ها با ابعاد خروجی ۲۰*۲۰ می‌رسد. به ترتیب برای چهار تنظیم مختلف نتایج در ادامه آورده شده است:



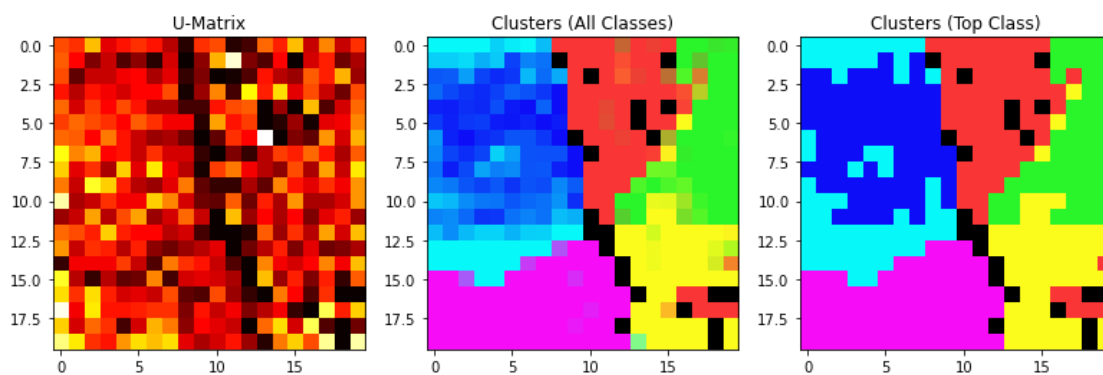
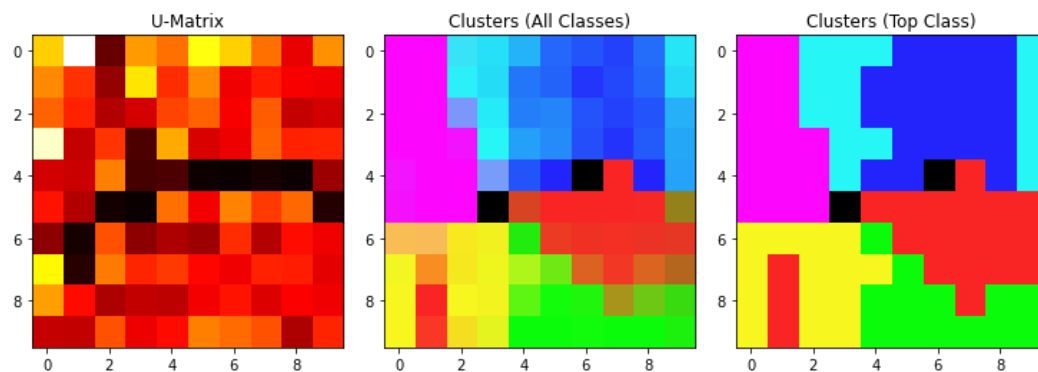
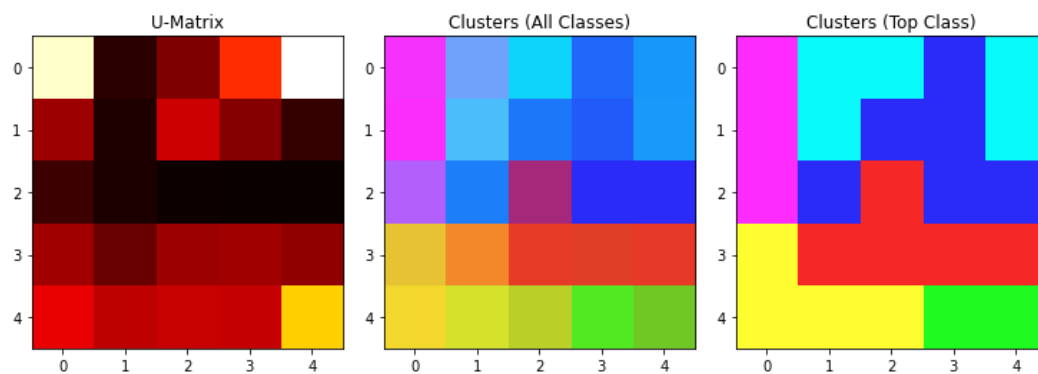
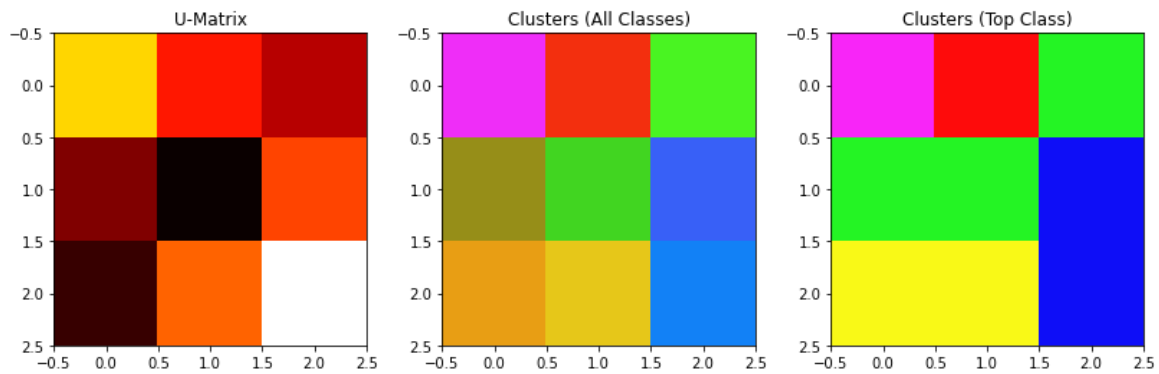


با بررسی نمودارهای فوق در می‌یابیم که تعداد نورون مرده برای تنظیم اول و سوم (نرخ یادگیری ۰/۰۰۱) و مخصوصاً تنظیم سوم زیاد است. این میزان ولی برای دو تنظیم دیگر قابل قبول است که این نشان می‌دهد از این منظر ابعاد لایه خروجی مشکلی ندارد. در نمودار U-Matrix تنظیم اول و سوم هم نواحی تاریک زیادی دیده می‌شود که خود نشان از عدم تخصیص مناسب داده‌ها به مراکز خوشه‌ها است. نحوه کاهش نورون‌های مرده مانند شبکه‌های ۱۰*۱۰ است. تنها برای تنظیم سوم می‌بینیم که این مقدار به خوبی کاهش پیدا نکرده است. خروجی دو تنظیم آخر (شعاع همسایگی ۵) همچون قسمت قبل نسبت به دو تنظیم اول (شعاع همسایگی ۲۰) بدتر است. علت مشابه شبکه‌های ۱۰*۱۰ است. از بین دو تنظیم اول تنظیم دوم بهتر است؛ چراکه هم نورون مرده کمتری دارد و هم خوشه‌های بهتری.

در ادامه خروجی تنظیم دوم (شعاع همسایگی ۲۰ و نرخ یادگیری ۰/۰۱) را بیشتر بررسی می‌کنیم. در این مدل کلاس صورتی تقریباً به طور کامل جدا شده است. کلاس‌های قرمز، سبز و زرد تقریباً جدا شده‌اند و هر کدام از این‌ها به غیر از قرمز تنها در یک قسمت از نقشه قرار گرفته‌اند که از این بابت بهبود خیلی خوبی نسبت به شبکه‌های ۱۰*۱۰ است؛ اما معضلی که حل نشده باقی ماند مخلوط شدن دو کلاس آبی است. اگرچه بخش مهمی از داده‌های این دو کلاس خالص شده‌اند ولی نمی‌توان قسمت‌های ترکیبی را نادیده گرفت.

با توجه به نتایج کسب شده برای سوال بعد از یک شبکه SOM با ابعاد ۲۰*۲۰، نرخ یادگیری اولیه ۰/۰۱ و شعاع همسایگی اولیه ۲۰ استفاده می‌شود. قبل از رفتن به سوال بعد شاید جالب باشد که نتایج بهترین تنظیم هر سری را در کنار هم ببینیم:

در این نمودارها جدا شدن داده کلاس‌های مختلف به خوبی محسوس است. فقط کافی است که برخی از نمودارها چرخانده شود.



سوال ۶

برای کاهش ابعاد و تبدیل ویژگی‌های خام ۵۶۱ تایی به ویژگی‌های SOM مطابق با توضیحات سوال ۱ می‌توانیم دو روش را استفاده کنیم:

- ویژگی‌های ساده: ساده‌ترین راه آن است که به هر داده مختصات نوروں برنده را نظیر کنیم. در این صورت ابعاد هر داده به ۲ کاهش پیدا می‌کند.
- ویژگی‌های پیچیده: راه دیگر آن است که برای هر داده میزان تعلق به هر نوروں خروجی را بدست بیاوریم. برای حصول این ویژگی‌ها معکوس فاصله هر داده تا هر یک از نوروں‌های خروجی را بدست می‌آوریم. در این صورت تعداد ویژگی‌ها برابر با تعداد نوروں لایه خروجی می‌شود؛ برای مدل SOM ما که قصد داریم از لایه خروجی 20×20 استفاده کنیم این تعداد ۴۰۰ خواهد بود.

برای تحلیل بهتر ترکیب SOM و دسته‌بند MLP، برای هر دو حالت فوق نتایج را بدست می‌آوریم. نحوه تنظیم دقیق پارامترهای MLP کاملاً مشابه با سوال ۴ است. جدول مقادیر صحت به ازای نرخ یادگیری‌های بررسی‌شده برای حالت ویژگی‌های پیچیده در ادامه آورده شده است:

ردیف	نرخ یادگیری	صحت آموزش	صحت اعتبارسنجی
۱	۰/۰۰۰۱	۹۰/۲۴	۸۹/۱۲
۲	۰/۰۰۰۵	۹۲/۰۳	۹۰/۶۴
۳	۰/۰۰۱	۹۲/۲۹	۹۰/۲۱
۴	۰/۰۰۳	۹۱/۰۶	۸۹/۷۷
۵	۰/۰۰۵	۹۲/۰۱	۹۰/۷۵
۶	۰/۰۱	۹۲/۳۴	۸۹/۹۹
۷	۰/۱	۹۱/۲۲	۹۰/۱۰

با توجه به مقادیر بدست آمده نرخ یادگیری ۰/۰۰۵ از مابقی مناسب‌تر است. در جدول بعد مقادیر صحت به ازای تنظیمات مختلف شبکه آورده شده است:

ردیف	تعداد لایه مخفی	تعداد نورون لایه مخفی	صحت آموزش	صحت اعتبارسنجی
۱	۰	-	۸۸/۶۸	۸۸/۴۷
۲	۱	۳۲	۹۱/۵۵	۸۹/۷۷
۳	۱	۶۴	۹۲/۲۱	۹۰/۴۲
۴	۱	۱۲۸	۹۱/۷۹	۹۰/۱۰
۵	۲	۳۲-۳۲	۹۲/۱۷	۹۰/۶۴
۶	۲	۶۴-۳۲	۹۲/۸۲	۹۱/۱۹
۷	۲	۱۲۸-۱۲۸	۹۱/۷۰	۹۰/۱۰
۸	۳	۶۴-۳۲-۳۲	۹۱/۱۲	۹۰/۲۱
۹	۳	۶۴-۶۴-۶۴	۹۰/۷۵	۸۸/۶۸
۱۰	۳	۱۲۸-۶۴-۳۲	۹۲/۲۰	۹۰/۸۶
۱۱	۴	۱۲۸-۱۲۸-۶۴-۶۴	۹۱/۴۵	۸۹/۶۶
۱۲	۴	۱۲۸-۶۴-۳۲-۱۶	۹۱/۵۷	۸۹/۹۹
۱۳	۴	۱۲۸-۶۴-۳۲-۶۴	۹۱/۹۶	۹۰/۴۲

با توجه به مقادیر بدست آمده به نظر می‌رسد شبکه با دو لایه مخفی با تعداد نورون ۶۴ و ۳۲ نورون مناسب باشد.

حال از ویژگی‌های ساده SOM استفاده می‌کنیم. مقادیر صحت به ازای نرخ‌های یادگیری مختلف به شرح زیر است:

ردیف	نرخ یادگیری	صحت آموزش	صحت اعتبارسنجی
۱	۰/۰۰۰۱	۸۵/۰۹	۸۵/۲۰
۲	۰/۰۰۰۵	۸۴/۷۳	۸۵/۹۶
۳	۰/۰۰۱	۸۵/۱۵	۸۶/۴۰
۴	۰/۰۰۳	۸۴/۲۸	۸۴/۹۸
۵	۰/۰۰۵	۸۴/۶۳	۸۴/۹۸
۶	۰/۰۱	۸۱/۹۷	۸۳/۰۳
۷	۰/۱	۸۴/۹۵	۸۵/۷۵

با توجه به صحت‌های حاصل‌شده به نظر می‌رسد ۰/۰۰۱ نرخ یادگیری خوبی باشد. برای این نرخ یادگیری و شبکه‌های مختلف صحت‌های زیر حاصل شد:

ردیف	تعداد لایه مخفی	تعداد نورون لایه مخفی	صحت آموزش	صحت اعتبارسنجی
۱	۰	-	۷۱/۱۵	۷۰/۶۲
۲	۱	۳۲	۸۱/۷۳	۸۱/۶۱
۳	۱	۶۴	۸۲/۹۵	۸۳/۵۷
۴	۱	۱۲۸	۸۳/۳۷	۸۳/۲۴
۵	۲	۳۲-۳۲	۸۳/۹۴	۸۴/۸۷
۶	۲	۶۴-۳۲	۸۵/۸۵	۸۶/۱۸
۷	۲	۱۲۸-۱۲۸	۸۶/۴۰	۸۶/۸۳
۸	۳	۶۴-۳۲-۳۲	۸۴/۱۹	۸۴/۸۷
۹	۳	۶۴-۶۴-۶۴	۸۶/۱۳	۸۶/۲۹
۱۰	۳	۱۲۸-۶۴-۳۲	۸۶/۳۰	۸۶/۴۰
۱۱	۴	۱۲۸-۱۲۸-۶۴-۶۴	۸۸/۰۸	۸۹/۱۲
۱۲	۴	۱۲۸-۶۴-۳۲-۱۶	۸۴/۵۳	۸۴/۹۸
۱۳	۴	۱۲۸-۶۴-۳۲-۶۴	۸۷/۲۱	۸۶/۹۴

از بین شبکه‌های بررسی‌شده شبکه با چهار لایه مخفی و تعداد نورون ۱۲۸، ۶۴ و ۶۴ بهترین نتایج را داشته و انتخاب می‌شود.

حال که پارامتر بهینه برای دسته‌بندهای مختلف بدست آمده است می‌توان یک مقایسه جامع از دسته‌بندها داشت. در جدول زیر صحت به ازای مجموعه داده‌های مختلف به همراه تعداد ویژگی مورد استفاده برای دسته‌بند MLP، دسته‌بند MLP با ویژگی‌های پیچیده SOM و دسته‌بند MLP با ویژگی‌های ساده SOM آورده شده است:

مدل	تعداد ویژگی	صحت آموزش	صحت اعتبارسنجی	صحت آزمون
MLP	۵۶۱	۹۹/۰۴	۹۷/۸۲	۹۵/۴۹
SOM (Complex) + MLP	۴۰۰	۹۱/۴۰	۸۹/۵۵	۸۷/۴۱
SOM (Simple) + MLP	۲	۸۳/۷۹	۸۵/۵۳	۸۲/۹۷

همانطور که مشخص است SOM باعث بهبود دقت نشده است. دسته‌بند MLP اولیه می‌توانست به راحتی به صحت بالای ۹۵٪ برسد و از این نظر جای بهبود کمی وجود داشت. از طرفی در مدل‌های SOM هم تعداد ویژگی‌ها کمتر شده است و هم حتی در خروجی بهترین مدل‌های ما کلاس‌ها به طور کامل از هم جدا نمی‌شدند؛ طبیعتاً در این شرایط کاهش صحت دور اصلاً دور از انتظار نبود.

با تمام این‌ها SOM در زمینه کاهش ابعاد موفق بوده است. در حالتی که برای هر داده تنها نوروں برنده را مشخص کنیم و ابعاد را از ۵۶۱ به ۲ کاهش دهیم، صحت از ۹۵/۴۹٪ به ۸۲/۹۷٪ کاهش می‌یابد که به نظر مناسب می‌آید و واضح است که بخش عمده دانش توسط SOM حفظ شده است؛ استفاده از ویژگی‌های پیچیده SOM ابعاد را کاهش جدی نمی‌دهد ولی خطا را زیاد می‌کند که به نظرم مناسب نیست.