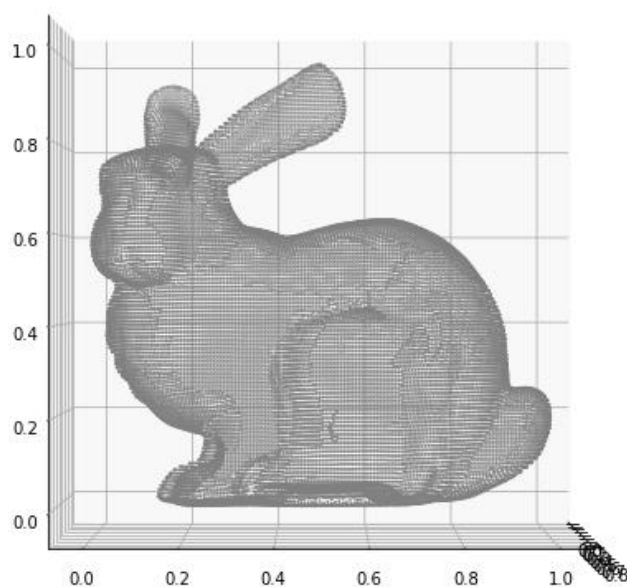


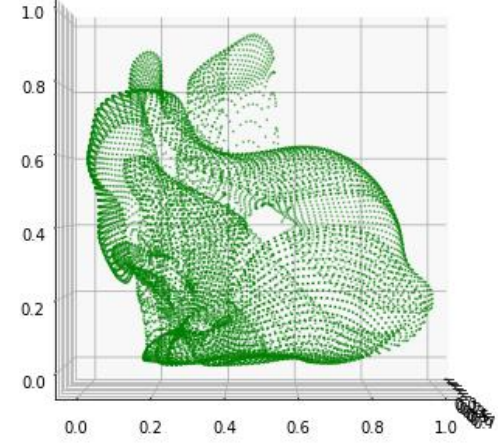
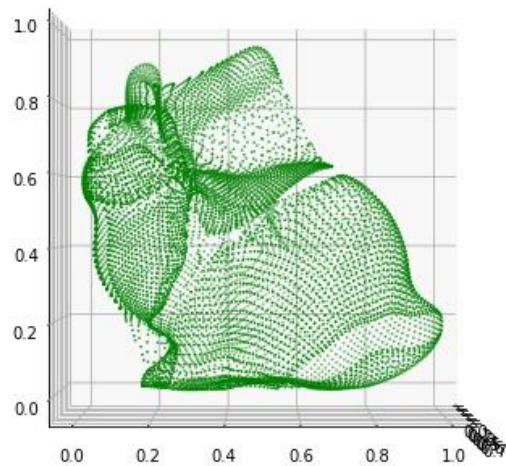
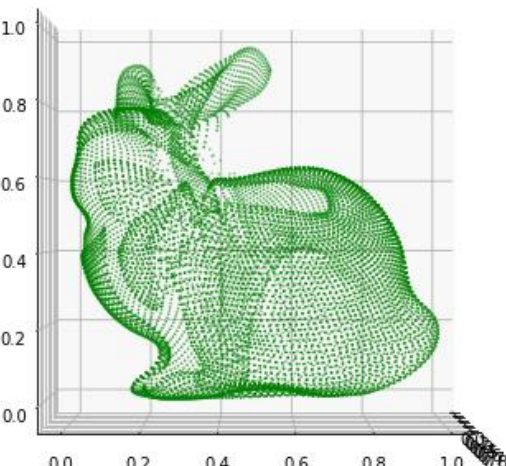
گزایش پیاده‌سازی و پاسخ به سوالات

سوال ۱:

الف) ابتدا داده‌ها خوانده شده و در فضای سه بعدی نمایش داده شده است که خروجی آن به صورت زیر است:



در ادامه با استفاده از کتابخانه minisom شبکه SOM ساخته شده است که در آن با آزمون و خطا به مقدار نرخ یادگیری ۰.۱ و سیگمای ۶ رسیده‌ایم که مقدار سیگما انتشار همسایه‌های نورون برنده را نشان می‌دهد. همچنین طبق خواسته سوال از همسایگی مربعی استفاده شده است. در ادامه با آموزش مدل ساخته شده بر روی داده‌ها تا تعداد ایپاک‌های مختلف خروجی را در هر ایپاک به دست می‌آوریم که تصاویر خروجی هر ایپاک در ادامه دیده می‌شود:

توزیع به دست آمده	خطا	شماره ایپاک
	۰.۰۱۴	۱
	۰.۰۱۲	۲
	۰.۰۱۰	۳

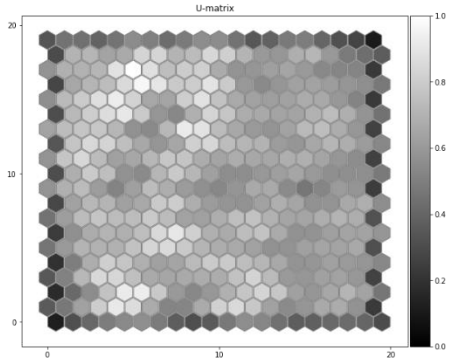
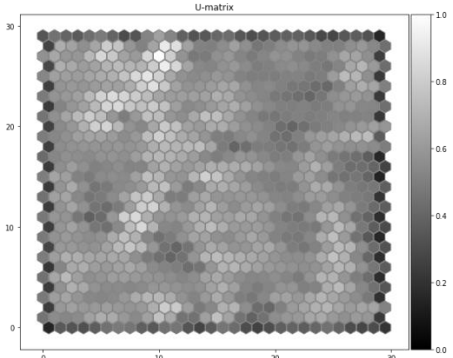
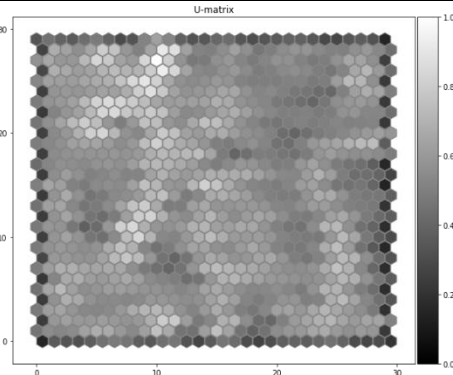
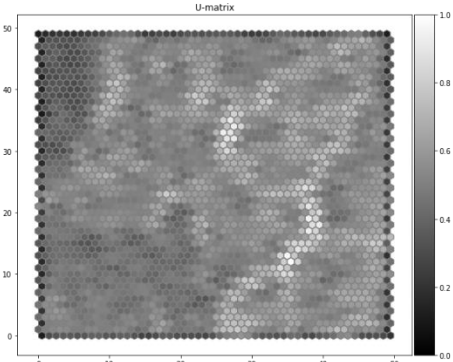
ب) در کاربرد بازسازی سطح، در انتها نورون‌هایی در فضا در کنار یک‌دیگر قرار می‌گیرند که مقدار بردارهای وزنی آن‌ها مشابه باشد. در نتیجه در این حالت در لبه‌های شکل حاصل، به ازای برنده شدن یک نورون، نورون‌هایی که فاصله کمی با بردار نورون برنده دارند به عنوان همسایه‌ی آن در نظر گرفته می‌شوند که آپدیت وزن‌ها برای این نورون‌ها هم باید انجام شود.

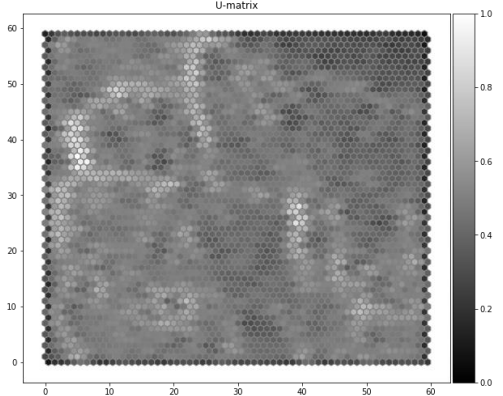
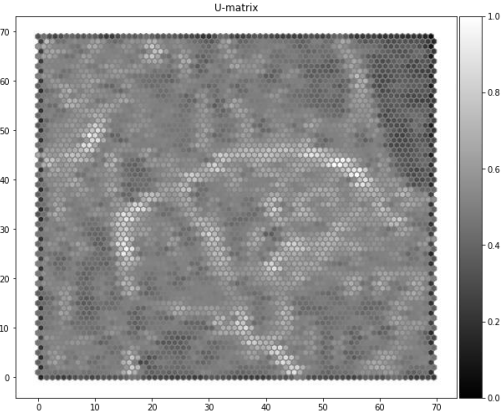
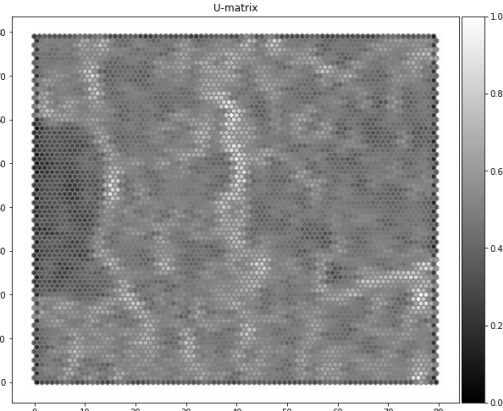
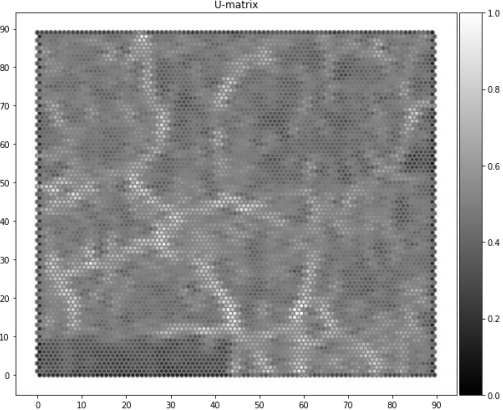
سوال ۲:

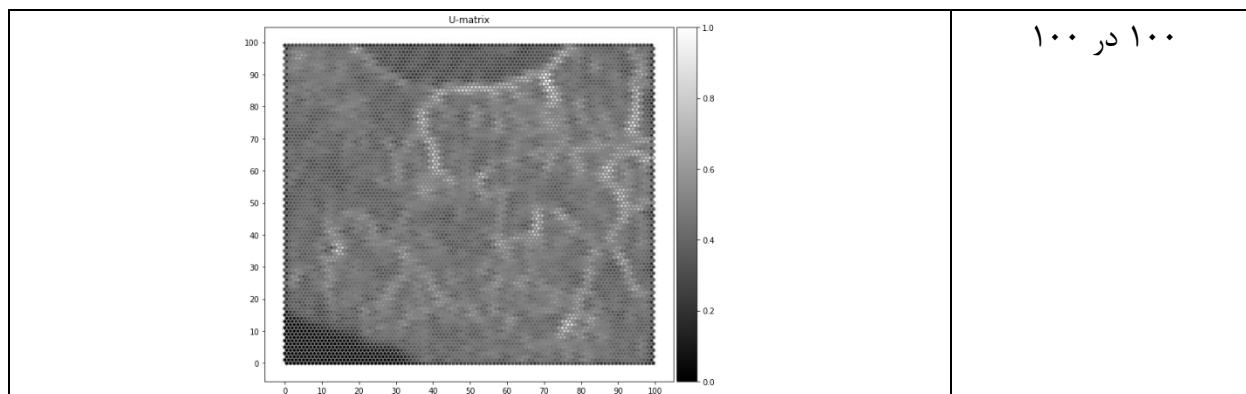
الف) دو خروجی شبکه SOM به صورت زیر است که با توجه به توضیحات آن مورد اول که ماتریس یو است برای خوشه‌بندی مناسب می‌باشد:

- اگر فاصله هر سلول از سلول‌های همسایه‌ش را محاسبه کرده و میانگین بگیریم و عدد به دست آمده را به صورت رنگی برای سلول مرکزی نشان دهیم به ماتریس یو می‌رسیم. در این ماتریس در جاهایی که عدد سلول‌ها کمتر است نشان دهنده فاصله کم سلول از سلول‌های اطرافش است و تراکم بالا در آن ناحیه را نشان می‌دهد و در جاهایی که عدد مربوط به سلول زیاد است نشان دهنده تفاوت بین آن سلول و سلول‌های مجاورش است که پراکندگی در آن ناحیه را نشان می‌دهد. در نتیجه با بررسی کلی این ماتریس می‌توان نواحی پرتراکم که به وسیله نواحی کم‌تراکم از هم جدا شده‌اند را به دست آورد که همان خوشه‌های مربوط به داده‌ها هستند.
- اگر پس از آموزش مدل SOM به ازای هر یک از داده‌ها نورون برنده را مشخص کنیم و برچسب مربوط به آن داده را برای آن نورون در نظر بگیریم و در انتها برای هر سلول پرتکرارترین برچسب را برچسب آن سلول فرض کنیم، به توزیعی از برچسب‌های داده‌ها در سلول‌های شبکه می‌رسیم. این خروجی به خوبی تراکم داده‌های مربوط به یک برچسب را در قسمت‌های مختلف سلول‌های شبکه نشان می‌دهد. واضح است که به دلیل نیاز به برچسب داده‌ها این روش برای خوشه‌بندی مناسب نمی‌باشد.

ب) برای یافتن ابعاد مناسب برای مدل از ابعاد کوچک به بزرگ مدل را بر روی داده‌ها آموزش می‌دهیم و خروجی ماتریس-یو را بررسی می‌کنیم. خروجی‌های به دست آمده به صورت زیر است:

خروجی ماتریس-یو	ابعاد مدل
	۲۰ در ۲۰
	۳۰ در ۳۰
	۴۰ در ۴۰
	۵۰ در ۵۰

 <p>A U-matrix plot for a 60x60 grid. The plot shows a noisy pattern of points with a color scale on the right ranging from 0.0 (black) to 1.0 (white). The axes are labeled from 0 to 60.</p>	۶۰ در ۶۰
 <p>A U-matrix plot for a 70x70 grid. The plot shows a noisy pattern of points with a color scale on the right ranging from 0.0 (black) to 1.0 (white). The axes are labeled from 0 to 70.</p>	۷۰ در ۷۰
 <p>A U-matrix plot for an 80x80 grid. The plot shows a noisy pattern of points with a color scale on the right ranging from 0.0 (black) to 1.0 (white). The axes are labeled from 0 to 80.</p>	۸۰ در ۸۰
 <p>A U-matrix plot for a 90x90 grid. The plot shows a noisy pattern of points with a color scale on the right ranging from 0.0 (black) to 1.0 (white). The axes are labeled from 0 to 90.</p>	<u>۹۰ در ۹۰</u>

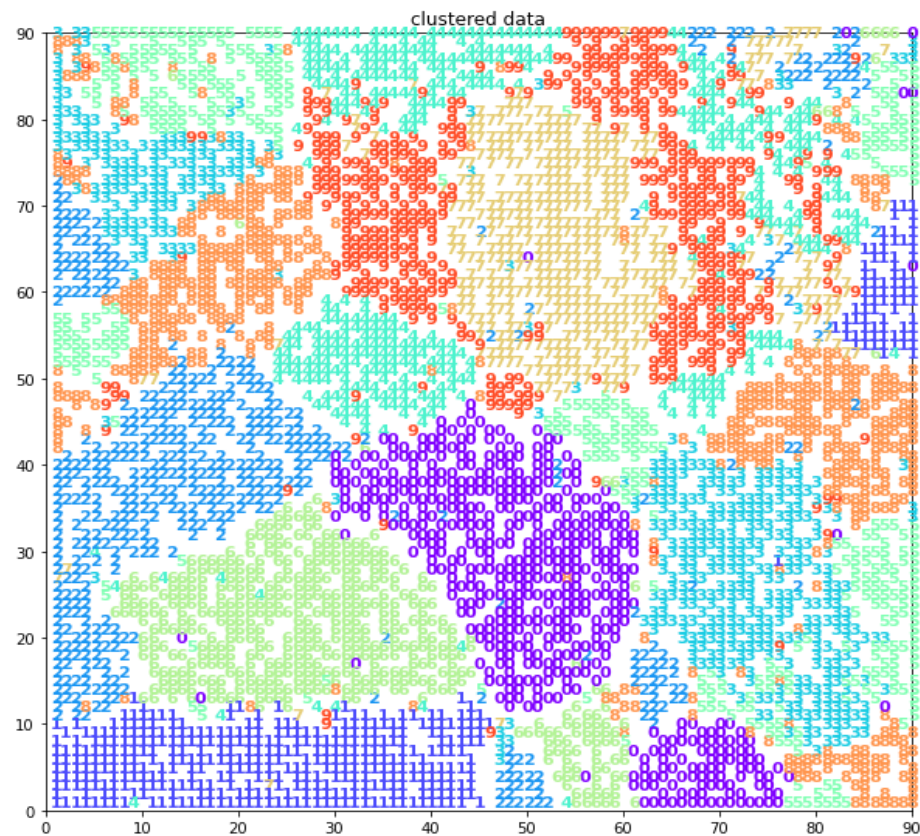


با بررسی خروجی‌های به دست آمده دیده می‌شود که در ابعاد پایین، رگه‌های سفید رنگ که نشان دهنده نقاط کم تراکم و مرز بین خوشه‌ها است وضوح کمی دارند و با افزایش ابعاد این وضوح بهتر شده است و مدل توانسته به خوبی داده‌ها را خوشه‌بندی کند. ولی در ابعاد بسیار بالای ۱۰۰ در ۱۰۰ علاوه بر آن که زمان آموزش مدل نیز بسیار افزایش می‌یابد دیده می‌شود که خوشه‌ها پراکنده‌تر شده‌اند و در نتیجه شدت رنگ قسمت‌های تیره کمتر شده است. با توجه به این توضیحات ابعاد مناسب برای مدل را ۹۰ در ۹۰ انتخاب می‌کنیم.

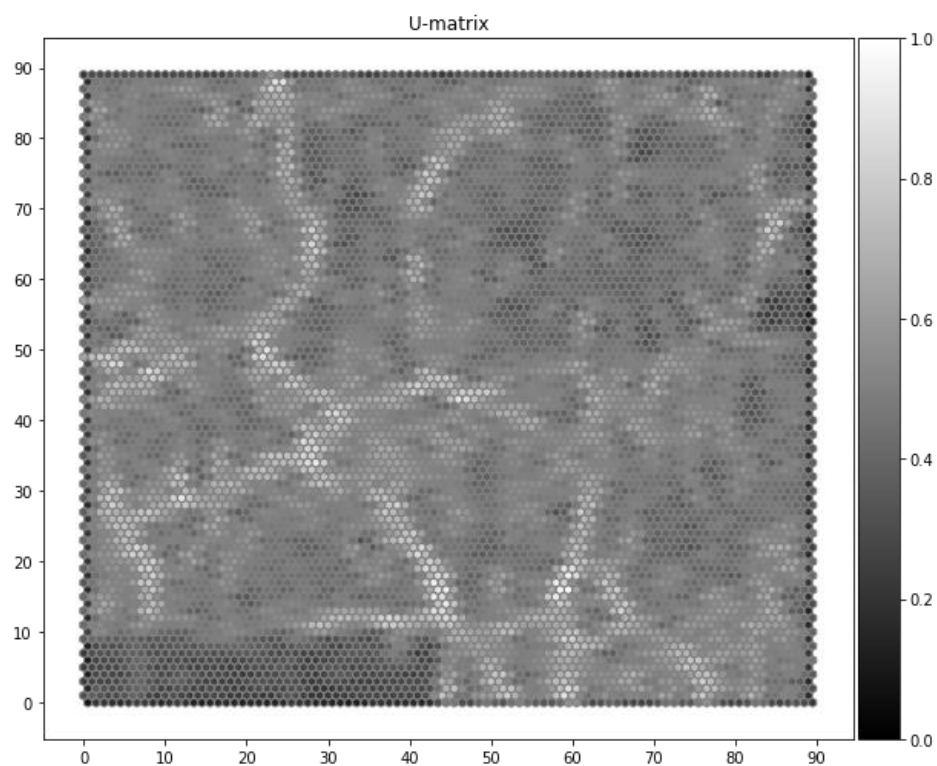
ج) برای تعیین ابعاد مناسب برای شبکه روش‌هایی ارائه شده است که به عنوان مثال گفته می‌شود تعداد سلول‌های شبکه ۵ برابر رادیکال تعداد کل نمونه‌های موجود باشد. ولی به طور کلی تعداد دقیقی برای یک کاربرد خاص مطرح نشده است و باید با آزمون و خطا به ابعاد مناسب برای شبکه رسید.

ولی در مدل‌های جدید شبکه SOM مانند GSOM که به معنی growing self-organizing map می‌باشد برای حل این مشکل سعی شده است ابعاد شبکه SOM از قبل تعیین نشود و به مرور و در طی آموزش مدل این شبکه رشد یابد و بر حسب نیاز بزرگ‌تر شود. نحوه این کار هم به این صورت است که ابتدا مدل با تعداد کمی مانند ۴ سلول شروع می‌شود و به مرور سعی می‌کند با انتخاب یکی از سلول‌های مرزی شبکه، آن سلول را در جهت‌هایی که امکان رشد دارد، رشد دهد و در نتیجه شبکه در آن جهت گسترش می‌یابد. همچنین برای کنترل مقدار این رشد پارامتر Spread Factor (SF) وجود دارد که قابل تنظیم می‌باشد.

د) پس از آموزش مدل، هر یک از داده‌ها را به مدل می‌دهیم و نوروں برنده را مشخص می‌کنیم و برچسب آن داده را برای آن سلول در نظر می‌گیریم و در انتها برای هر سلول پرتکرارترین برچسب را برچسب آن سلول فرض می‌کنیم. پس از این فرایند مشخص می‌شود که داده‌های مربوط به هر یک از برچسب‌ها در کدام قسمت شبکه قرار گرفته‌اند. خروجی این کار برای مدل آموزش دیده به صورت زیر است:



خروجی ماتریس-یو مربوط به این شبکه هم به صورت زیر است:



با مقایسه محل خوشه‌ها در تصویر مربوط به برچسب داده‌ها و قسمت‌های تیره و پرتراکم در ماتریس-یو، دیده می‌شود که رگه‌های سفید رنگ در ماتریس-یو به خوبی توانسته‌اند مرز خوشه‌ها را نشان دهند و قسمت‌های تیره پرتراکم با هر یک از خوشه‌ها در تصویر اول مطابقت دارد.