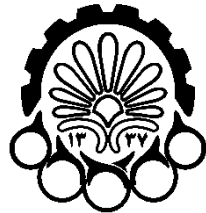


به نام خداوند بخشنده و مهربان



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

درس هوش محاسباتی

گزارش پروژه پایانی

استاد درس: دکتر عباد زاده

نام دانشجو:

روزبه قاسمی ۹۵۳۱۴۲۴

دی ۱۳۹۸

در این گزارش به بررسی سه بخش زیر بر روی سه مجموعه می پردازیم:

بخش اول- بررسی تاثیر مقادیر مختلف پارامتر شامل تاثیر تعداد مراکز خوشه ها و شعاع خوشه ها (مركزها)

بخش دوم- مقدار شعاع مرکز $\lambda = 0.1$ و تعداد خوشه ها زیاد به عنوان مثال ۴۰

بخش سوم- مقدار شعاع مرکز $\lambda = 1$ و تعداد خوشه ها برابر ۳

در هر یک از بخش ها نتایج هر سه مجموعه داده را گزارش و مورد تحلیل قرار می دهیم.

مشخصات سه مجموعه داده مورد استفاده در این پروژه بصورت زیر است:

- مجموعه داده 2clstrain1200 شامل ۱۲۰۰ داده می باشد که در دو کلاس قرار می گیرند.
- مجموعه داده 4clstrain1200 شامل ۱۲۰۰ داده می باشد که در چهار کلاس قرار می گیرند.
- مجموعه داده 5clstrain1500 شامل ۱۵۰۰ داده می باشد که در پنج کلاس قرار می گیرند.

حل بخش اول

در جدول ۱ مقادیر دقت بدست آمده بر روی دادههای آموزش را به ازای مقادیر مختلف پارامترها بر روی مجموعه داده 2clstrain1200 گزارش نمودیم.

جدول ۱: دقت مجموعه آموزش بر روی مجموعه داده 2clstrain1200

تعداد خوشه ها					شعاع مرکز
25	15	8	4	2	
0.925	0.982	0.969	0.951	0.74	0.02
0.984	0.979	0.975	0.964	0.608	0.2
0.981	0.986	0.969	0.964	0.493	1.5
0.982	0.977	0.963	0.945	0.504	10

بررسی جدول ۱ با ثابت در نظر گرفتن تعداد خوشه ها و تغییر شعاع مرکز نشان می دهد که با افزایش شعاع مرکز دقت آموزش در بعضی از حالتها تعداد خوشه شروع به افزایش می نماید. پس از اینکه به بالاترین دقت در هر شعاع مرکز رسیدیم دقت شروع به کاهش می نماید و دیگر افزایش شعاع نمی تواند بهبود دقت را نتیجه دهد. به عنوان مثال زمانی که تعداد خوشه را ۴ یا ۸ در نظر بگیریم با افزایش شعاع مرکز دقت در حال افزایش است و پس از دست یافتن به بالاترین دقت آموزش، افزایش شعاع منجر به کاهش دقت آموزش می شود.

حال اگر شعاع مرکز را در جدول ۱ ثابت در نظر بگیریم و تعداد خوشه ها تغییر نماید مشاهده می نماییم که با افزایش تعداد مراکز خوشه ها دقت آموزش در تمام روش ها شروع به افزایش می نماید و پس از اینکه به بالاترین دقت آموزش رسیدیم، افزایش تعداد خوشه منجر به کاهش دقت می گردد. با توجه به جدول ۱

بهترین مقدار دقت آموزش مقدار ۰,۹۸۶ است که با تعداد ۱۵ خوشه و شعاع مرکز ۱,۵ بدست آمد که در جدول ۲ مشاهده می نماییم که با در نظر گرفتن این مقادیر پارامتر، دقت تست برابر ۰,۹۷۸ خواهد بود در حالیکه از بررسی جدول ۲ مشاهده می نماییم که بالاترین مقدار دقت تست ۰,۹۹۲ است که به ازای تعداد ۱۵ خوشه و شعاع مرکز ۰,۲ بدست آمد. در جدول ۲ نیز مشاهده می شود که با افزایش شعاع مرکز با ثابت فرض کردن تعداد خوشه (۴ و ۸ و ۱۵) دقت تست شروع به افزایش می نماید و سپس با رسیدن به بالاترین مقدار دقت با این تعداد خوشه دقت تست با افزایش شعاع کاهش می یابد. در حالت تعداد ۲ خوشه افزایش شعاع مرکز منجر به کاهش دقت تست می گردد و در حالت تعداد ۲۵ خوشه افزایش شعاع منجر به افزایش دقت تست می شود.

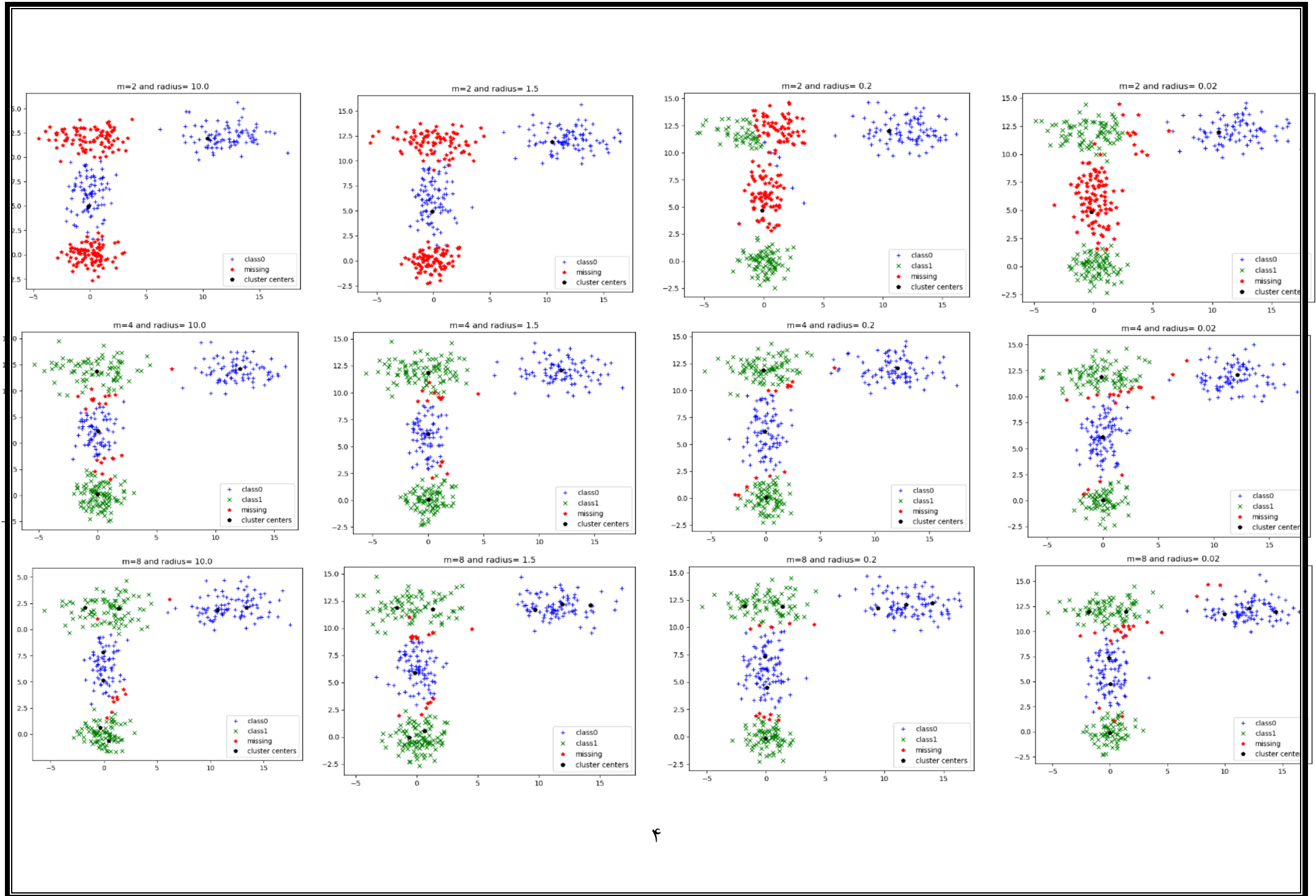
جدول ۲: دقت مجموعه تست بر روی مجموعه داده 2c1strain1200

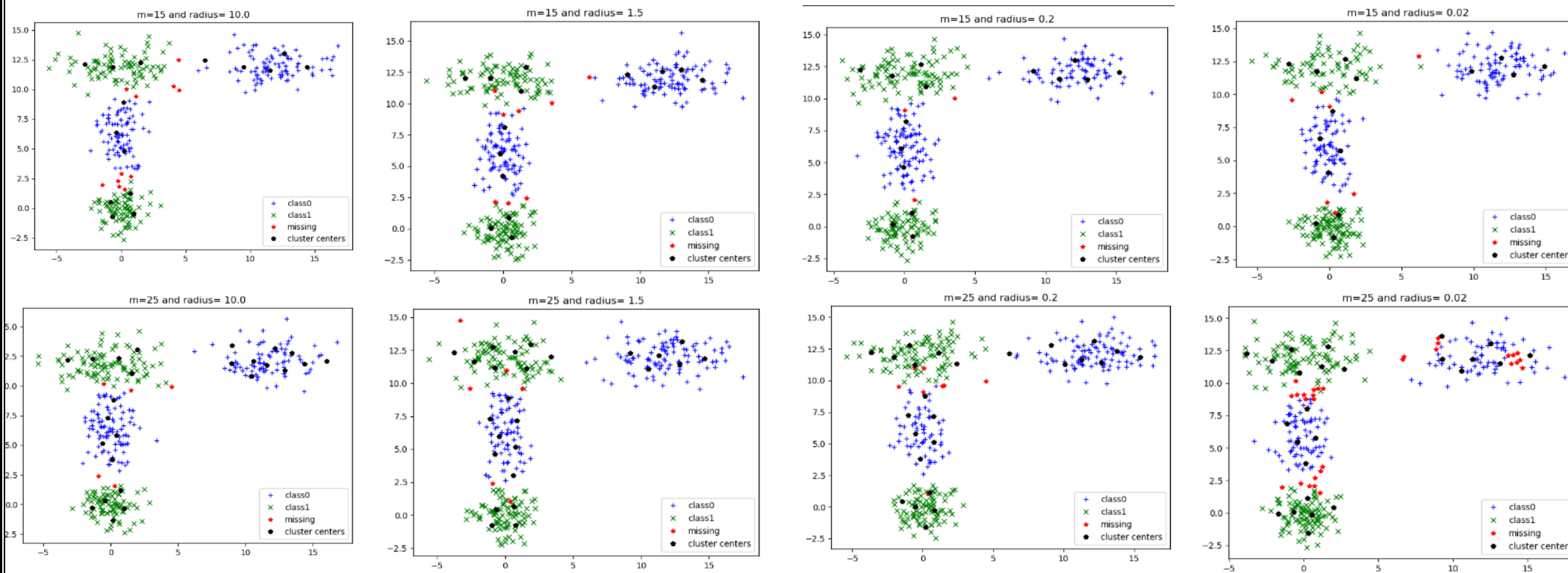
شعاع مرکز	تعداد خوشه ها				
	2	4	8	15	25
0.02	0.689	0.947	0.942	0.981	0.917
0.2	0.619	0.964	0.969	0.992	0.978
1.5	0.517	0.961	0.958	0.978	0.983
10	0.492	0.939	0.972	0.969	0.986

در شکل ۱ داده های تستی که بدرستی برچسب شان تخمین زده شده است، داده های تستی که برچسب شان به اشتباه تخمین زده شده است را به همراه مراکز خوشه های بدست آمده نشان دادیم. همانطور که در شکل مشاهده می نماییم در حالت تعداد ۲ خوشه با افزایش شعاع تا اندازه ۰,۲ تنها روش قادر است داده های تست از هر دو کلاس را تشخیص دهد و از شعاع ۱,۵ به بعد روش دیگر قادر نیست داده های تست یک کلاس را تخمین بزند. از بررسی تعداد خوشه ۲ متوجه می شویم که هنگامیکه شعاع مقدارش کم است با هر مرکز خوشه قادر به تشخیص داده های یکی از کلاس ها هستند که با افزایش شعاع وضعیت بگونه ای تغییر می نماید که با دو مرکز خوشه فقط قادر به تشخیص داده های یک کلاس خواهیم بود.

با افزایش تعداد خوشه ها از ۲ به ۴ همانطور که در جدول ۱ و ۲ دیدیم دقت آموزش و تست بشدت افزایش یافت که شکل ۱ نشان می دهد که با این تعداد خوشه بخاطر اینکه داده های ۲ کلاس در ۴ ناحیه قرار گرفته است حالا هر ناحیه یک مرکز خوشه را به خود اختصاص داد که بخاطر همین این افزایش شدید دقت رخ داد. بطور کلی با در نظر گرفتن ۴ خوشه یا بیشتر حالا هر یک از ۴ ناحیه ای که داده های دو کلاس قرار دارند حداقل یک مرکز خوشه بر رویشان قرار گرفته بخاطر همین دقت در حالت بیشتر مساوی ۴ خوشه بشدت بیشتر از حالت ۲ خوشه است.

با توجه به جدول ۱ و ۲ و شکل ۱ مشاهده می شود که برای داده های دو کلاسی که در ۴ ناحیه قرار گرفته اند حداقل تعداد خوشه ها را باید ۴ در نظر گرفت تا از هر کلاس بتوانیم تعدادی داده را بدرستی تشخیص دهیم.





شکل ۱: برچسب داده‌های تست بعد از تخمین و مراکز خوشه‌های بدست آمده بر روی مجموعه داده 2c1strain1200.

در جدول ۳ مقادیر دقت بدست آمده بر روی داده‌های آموزش را به ازای مقادیر مختلف پارامترها بر روی مجموعه داده 4clstrain1200 گزارش نمودیم.

جدول ۳: دقت مجموعه آموزش بر روی مجموعه داده 4clstrain1200

تعداد خوشه ها					شعاع مرکز
25	15	8	4	2	
0.973	0.977	0.974	0.969	0.549	0.02
0.981	0.969	0.971	0.961	0.501	0.2
0.980	0.969	0.964	0.951	0.490	1.5
0.976	0.975	0.955	0.938	0.484	10

جدول ۳ را در نظر بگیرید در صورتیکه شعاع مرکز را ثابت در نظر بگیریم و تعداد خوشه ها فرق نماید همانطور که در جدول مشاهده می شود با افزایش تعداد خوشه ها دقت آموزش افزایش می یابد (البته در بعضی جاها کاهش داشتیم و مجدد افزایش که این مسئله می تواند ناشی از این باشد که مجموعه آموزش و تست تصادفی در هر حالت ایجاد شد). در صورتیکه تعداد خوشه ها را ثابت در نظر بگیریم و شعاع مرکز تغییر نماید همانطور که در جدول مشاهده می گردد در حالتی که تعداد خوشه ها ۲، ۴ و ۸ است با افزایش شعاع مرکز دقت آموزش کاهش می یابد و در حالتی که تعداد خوشه ها ۱۵ است دقت آموزش با افزایش شعاع مرکز، ابتدا کاهش یافته و سپس ثابت بود و مجددا شروع به افزایش نمود و در حالتی که تعداد خوشه ها ۲۵ است دقت آموزش با افزایش شعاع ابتدا افزایش یافت و پس از رسیدن به بالاترین مقدار دقت، افزایش شعاع مرکز منجر به کاهش دقت آموزش گردید.

با توجه به جدول ۳ مشاهده می گردد بالاترین مقدار دقت آموزش به ازای تعداد ۲۵ خوشه و شعاع مرکز ۰٫۲، بدست آمد که برابر ۰٫۹۸۱ بود و با توجه به جدول ۴ مشاهده می گردد به ازای این مقادیر پارامتر دقت تست برابر ۰٫۹۶۱ است در حالیکه بالاترین مقدار دقت تست برابر ۰٫۹۸۱ است که به ازای ۱۵ خوشه و شعاع مرکز ۱۰ بدست آمد. با توجه به جدول ۳ و ۴ مشاهده می شود که کمترین میزان دقت به ازای تعداد خوشه ۲ در تمام مقادیر شعاع مرکز بدست آمده است که این مسئله بیانگر این است که این تعداد خوشه برای این مجموعه داده که شامل ۴ کلاس است تعداد مناسبی نمی باشد.

در جدول ۴ با ثابت در نظر گرفتن تعداد خوشه ها و متغیر بودن شعاع مرکز مشاهده می شود دقت در حالتی که تعداد خوشه برابر ۴ و ۸ است با افزایش شعاع ابتدا دقت تست افزایش یافت و سپس شروع به کاهش نمود. در حالتی که تعداد خوشه ۲ و ۲۵ است با افزایش شعاع دقت تست ابتدا کاهش یافت و سپس شروع به افزایش نمود و در حالت ۱۵ خوشه با افزایش شعاع دقت تست نیز افزایش یافت.

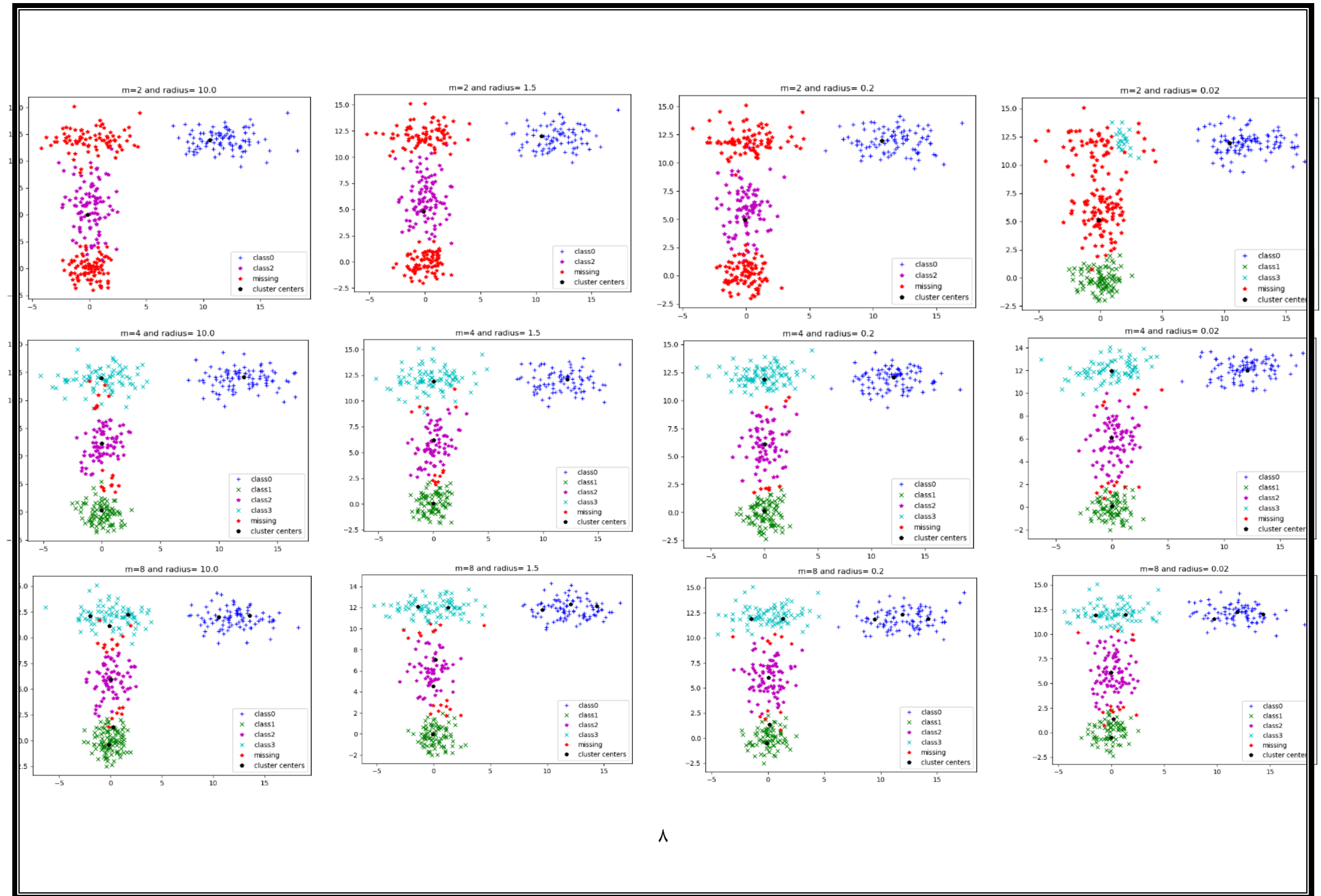
در شکل ۲ مراکز خوشه ها و برچسب تخمینی داده‌های تست نشان داده شده است. در حالتی که تعداد خوشه برابر ۲ است در شعاع مرکز ۰٫۲ از ۴ کلاس موجود فقط تعدادی از داده‌های ۳ کلاس را توانستیم بدرستی تشخیص دهیم و در شعاع های مرکز بیشتر ۰٫۲، تنها توانستیم تعدادی از داده‌های دو کلاس از ۴ کلاس را درست تشخیص دهیم و با دقت در شکل برای حالتی که تعداد ۲ خوشه متوجه می شویم که زمانی که شعاع

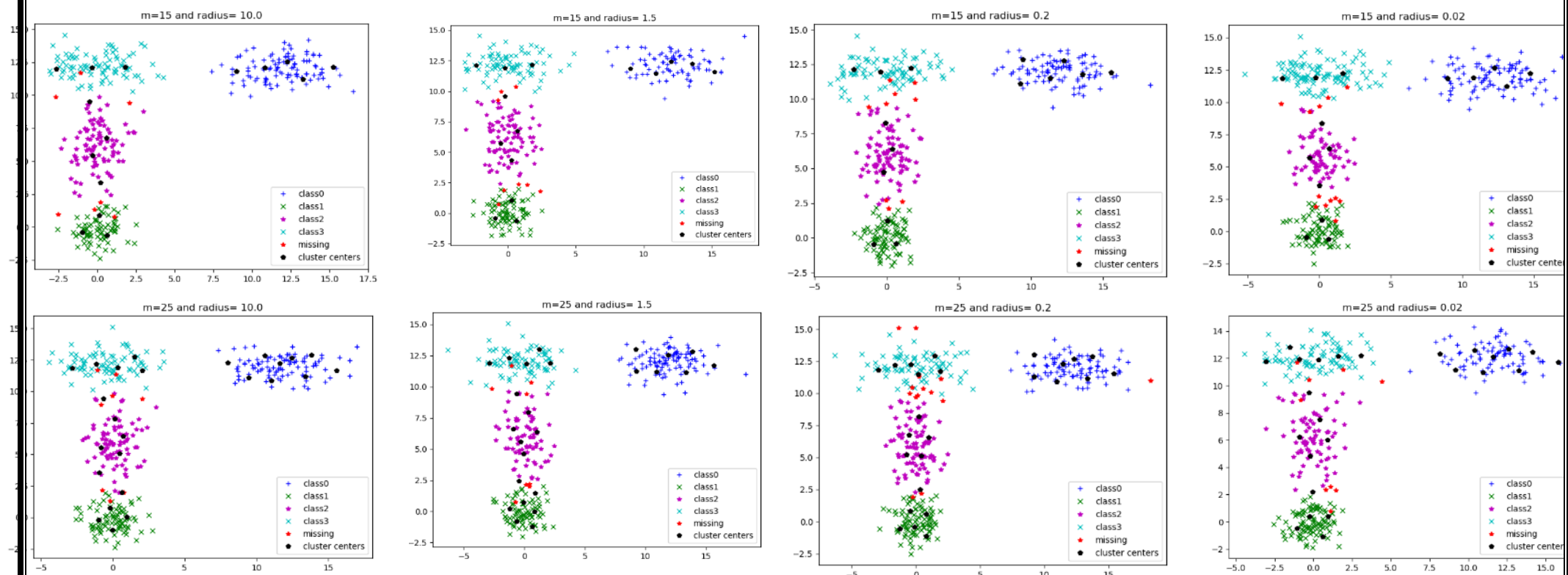
برابر ۰,۰۲ است دو مرکز خوشه در وسط داده های دو کلاس قرار گرفته که اکثر داده های قرار گرفته در نزدیکی یک مرکز خوشه را برای یک کلاس را بدرستی تشخیص داده و داده های قرار گرفته در نزدیکی مرکز خوشه دیگر را اصلا نتوانستیم برچسب شان را درست تشخیص دهیم و همچنین داده های دو کلاس دیگر را که از این دو مرکز خوشه دور هستند تنها تعدادی را توانستیم درست تشخیص دهیم. همچنین در شعاع های بالاتری از ۰,۰۲ مشاهده می شود که تنها توانستیم داده های نزدیک دو مرکز خوشه را برای دو کلاس بدرستی تشخیص دهیم. این مشاهدات بیانگر نامناسب بودن تعداد ۲ خوشه برای داده هایی است که شامل ۴ کلاس است. با بررسی شکل مشاهده می شود زمانیکه تعداد خوشه ها برابر ۴ یا بیشتر از آن است دیگر قادریم از هر کلاس تعدادی از داده های را بدرستی تشخیص دهیم و با این تعداد خوشه ها از هر کلاس حتما تعدادی داده از هر کلاس بدرستی تشخیص داده می شود در صورتیکه بر روی داده های هر کلاس حتما حداقل یک مرکز خوشه قرار گیرند (در بعضی اجراها نتایجی داشتیم که تعداد خوشه ۴ یا بیشتر بود بر روی داده های یک کلاس اصلا مرکز خوشه ای قرار نگرفت که این منجر شد که دقت بشدت پایین بیاید و داده های یک کلاس را حتی نتوانستیم بدرستی تشخیص دهیم که از آوردن این نتایج دیگر خودداری کردیم).

جدول ۴: دقت مجموعه تست بر روی مجموعه داده 4clstrain1200

شعاع مرکز	تعداد خوشه ها				
	25	15	8	4	2
0.02	0.975	0.967	0.967	0.961	0.583
0.2	0.961	0.975	0.969	0.975	0.497
1.5	0.978	0.978	0.95	0.964	0.522
10	0.978	0.981	0.953	0.95	0.635

با توجه به نتایج جدول ۳ و ۴ و شکل ۲ به این نتیجه دست می یابیم که حداقل تعداد ۴ خوشه برای داده های ۴ کلاسی که داده هایش در ۴ ناحیه قرار گرفته باید مدنظر قرار داد تا بتوان از هر کلاس تعدادی داده را بدرستی تشخیص داد





شکل ۲: برچسب داده‌های تست بعد از تخمین و مراکز خوشه‌های بدست آمده بر روی مجموعه داده 4clstrain1200.

در جدول ۵ مقادیر دقت بدست آمده بر روی داده‌های آموزش را به ازای مقادیر مختلف پارامترها بر روی مجموعه داده 5clstrain1500 گزارش نمودیم.

جدول ۵: دقت مجموعه آموزش بر روی مجموعه داده 5clstrain1500

تعداد خوشه ها					شعاع مرکز
27	16	9	5	4	
0.951	0.956	0.954	0.948	0.789	0.02
0.959	0.963	0.951	0.949	0.806	0.2
0.965	0.957	0.952	0.949	0.767	1.5
0.955	0.956	0.943	0.922	0.756	10

از بررسی جدول ۵ با ثابت نگه داشتن تعداد خوشه ها و متغیر در نظر گرفتن شعاع مرکز مشاهده می شود در حالتیهایی که تعداد خوشه ها برابر ۴ و ۵ و ۱۶ و ۲۷ است با افزایش شعاع مرکز دقت آموزش نیز افزایش مییابد و پس از رسیدن به بالاترین میزان دقت آموزش، افزایش شعاع مرکز منجر به کاهش دقت میگردد. در صورتیکه شعاع مرکز را ثابت در نظر بگیریم و تعداد خوشه متغیر باشد مشاهده می شود با افزایش تعداد مراکز خوشه ها دقت در تمام شعاع شروع به افزایش می نماید و پس از اینکه به بالاترین دقت آموزش رسیدیم، افزایش تعداد خوشه ها منجر به کاهش دقت آموزش شدند. مشاهده می شود که بالاترین دقت آموزش برابر ۰٫۹۶۵ است که به ازای ۲۷ خوشه و شعاع مرکز ۱٫۵ بدست آمد که با این مقادیر پارامتر با مراجعه به جدول ۶ مشاهده می شود دقت تست برابر ۰٫۹۴۹ است در صورتیکه به ازای تعداد ۲۷ خوشه با مرکز شعاع ۱۰ دقت تست به بالاترین مقدار که برابر ۰٫۹۷۶ است دست یافت.

در جدول ۶ مشاهده می شود که با ثابت در نظر گرفتن شعاع مرکز، با افزایش تعداد خوشه ها در تمام شعاع مرکزها دقت شروع به افزایش نمود که در بعضی موارد پس از این به بالاترین دقت تست دست یافتیم افزایش تعداد خوشه ها منجر به کاهش دقت تست شد. با ثابت نگه داشتن تعداد مراکز خوشه ها و متغیر در نظر گرفتن شعاع مرکز در جدول ۶ مشاهده می شود در حالتیهایی که تعداد خوشه برابر ۴ و ۹ و ۲۷ است با افزایش شعاع دقت تست نیز افزایش یافته و پس از رسیدن به بالاترین دقت تست، افزایش شعاع دیگر نمی تواند دقت تست را افزایش دهد و سبب کاهش دقت می گردد. در حالت ۵ خوشه افزایش شعاع منجر به کاهش دقت میشود و حالت ۱۶ مرکز خوشه افزایش شعاع ابتدا سبب کاهش دقت و سپس باعث افزایش دقت شد و در پایان مجدد به کاهش دقت گردید.

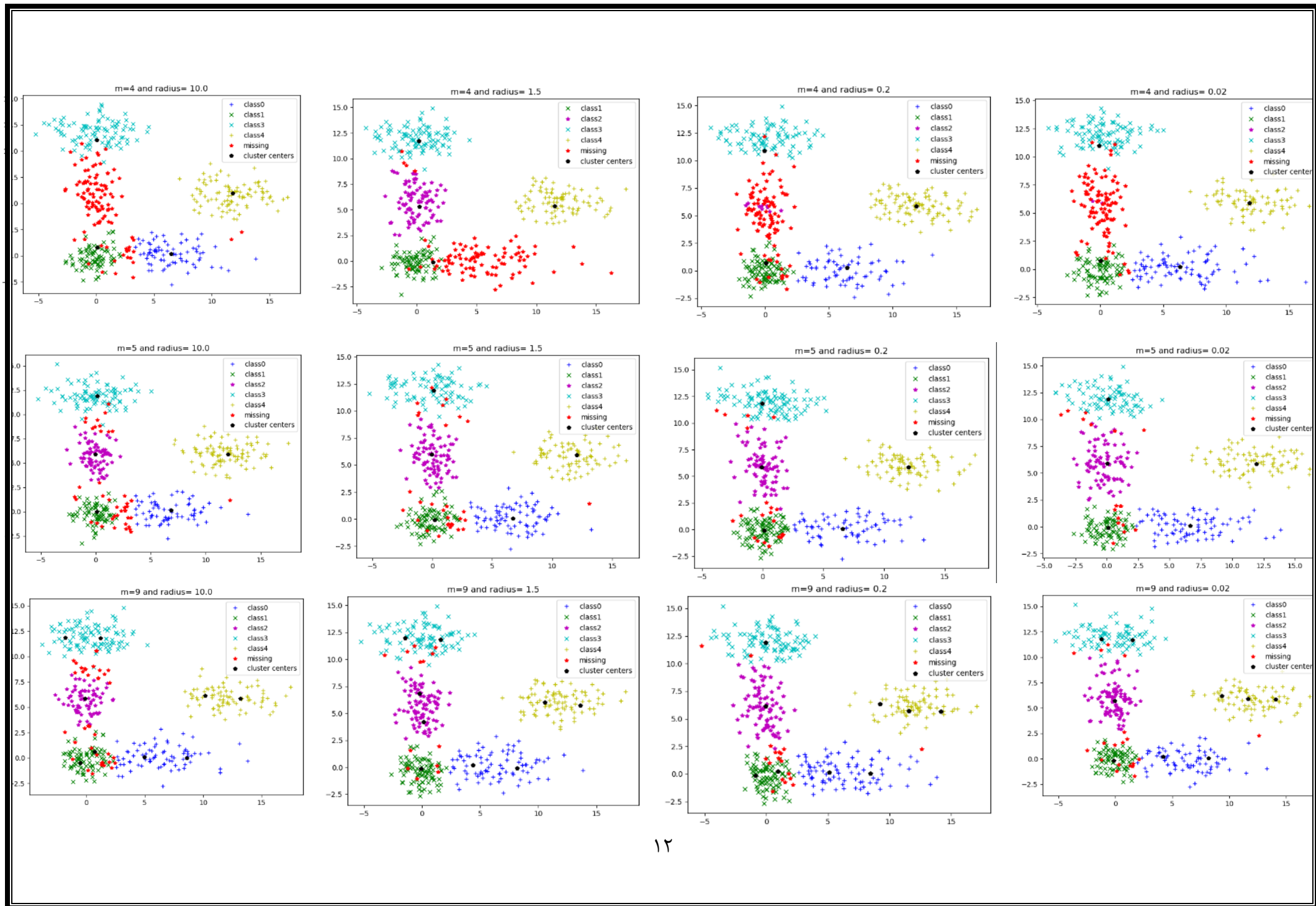
در شکل ۳ مراکز خوشه ها و برچسب داده‌های تست بعد از تخمین نشان داده شده است. همانطور که در شکل مشاهده می شود در حالتی که تعداد خوشه ها برابر ۴ است و شعاع کمتر مساوی ۰٫۲ است می توانیم از هر کلاس تعدادی داده را بدرستی تشخیص دهیم و در صورتیکه شعاع بیشتر از ۰٫۲ است از ۵ کلاس موجود تنها قادریم ۴ کلاس را بدرستی تشخیص دهیم. همانطور که در شکل مشاهده می گردد با تعداد خوشه

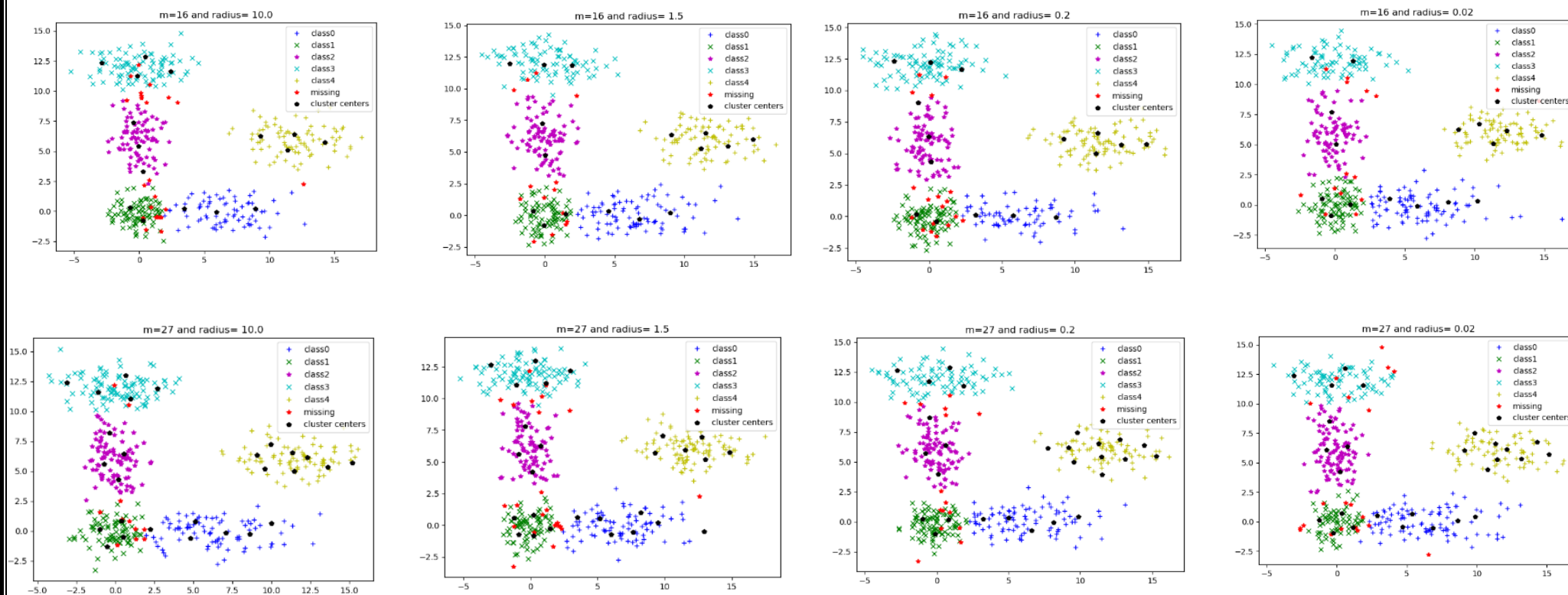
۵ و بیشتر از آن با هر مقدار شعاع مرکز قادریم از هر کلاس تعدادی از داده‌ها را بدرستی تشخیص دهیم و با تعدادخوشه‌های بیشتر از ۴ بر روی داده‌های هر کلاس یک مرکز خوشه قرار گرفته که با توجه به جدول ۵ و ۶ و این مشاهده می‌توان نتیجه گرفت که قرار گرفتن حداقل یک مرکز خوشه بر روی داده‌های هر کلاس می‌تواند دقت را به طور قابل توجهی بهبود دهد.

جدول ۶: دقت مجموعه تست بر روی مجموعه داده Scistrain1500

شعاع مرکز	تعداد خوشه‌ها				
	4	5	9	16	
0.02	0.733	0.962	0.958	0.969	0.956
0.2	0.784	0.953	0.960	0.960	0.967
1.5	0.787	0.933	0.971	0.969	0.949
10	0.76	0.907	0.918	0.949	0.976

با توجه به جدول ۵ و ۶ و شکل ۳ برای داده‌های ۵ کلاسی که در ۵ ناحیه قرار گرفته‌اند در صورتیکه بخواهیم به ازای هر مقدار از شعاع بتوانیم از هر کلاس تعدادی داده را بدرستی تشخیص دهیم لازم است تعداد مراکز خوشه را حداقل برابر ۵ قرار دهیم و همچنین هر مرکز خوشه باید روی داده‌های یک کلاس حتما قرار گیرد.





شکل ۳: برچسب داده‌های تست بعد از تخمین و مراکز خوشه‌های بدست آمده بر روی مجموعه داده 5clstrain1500.

حل بخش دوم

در این بخش تاثیر تعداد زیاد مرکز خوشه بر روی مجموعه داده‌های مختلف با ثابت در نظر گرفتن شعاع مرکز برابر ۰,۱ مورد بررسی قرار خواهد گرفت.

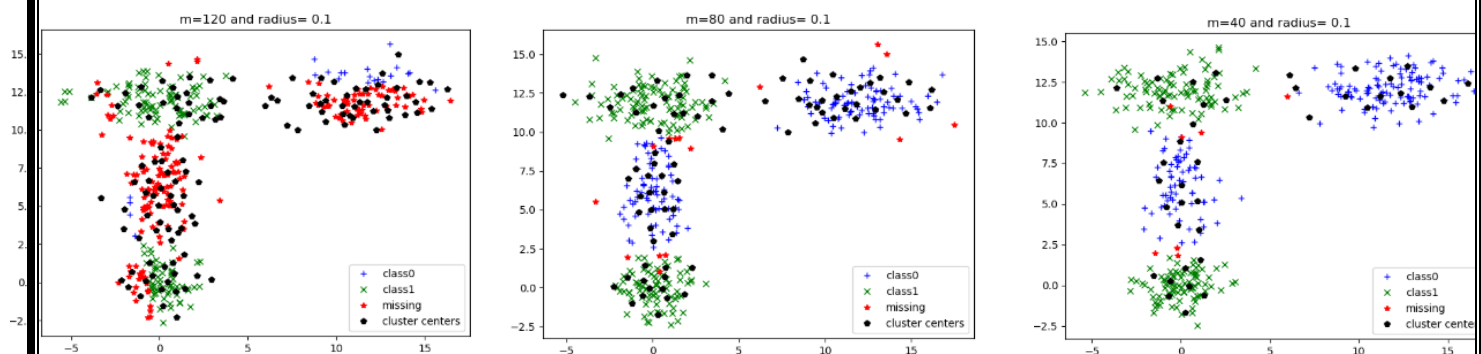
با توجه به جدول ۷ مشاهده می شود با افزایش تعداد مراکز خوشه ها، دقت آموزش کاهش می یابد که با افزایش تعداد مراکز خوشه ها دیگر مراکز خوشه ها بر روی داده‌هایی که خیلی دور از داده‌ها هستند قرار گرفته که سبب منحرف شدن شبکه می شود و در جدول ۸ نیز مشاهده می شود با افزایش تعداد مراکز خوشه ها دقت تست نیز کاهش یافت و در شکل ۴ مشاهده می کنیم در حالتی که تعداد مراکز خوشه ۱۲۰ است این مراکز خوشه در تمام نواحی که داده‌ها پراکنده شده اند پخش شده و حتی در نواحی که داده خیلی کمی وجود دارد. افزایش تعداد مراکز خوشه ها سبب منحرف شدن شبکه و صرف هزینه زیاد مانند افزایش زمان محاسبات شبکه می شود در حالیکه با تعداد خوشه ها کمتر توانسته بودیم به دقت آموزش و تست بیشتر دست یابیم

جدول ۷: دقت مجموعه آموزش در مجموعه داده 2clstrain1200 با شعاع مرکز ۰,۱

تعداد خوشه ها			دقت آموزش
120	80	40	
0.462	0.982	0.987	

جدول ۸: دقت مجموعه تست در مجموعه داده 2clstrain1200 با شعاع مرکز ۰,۱

تعداد خوشه ها			دقت تست
120	80	40	
0.478	0.958	0.981	



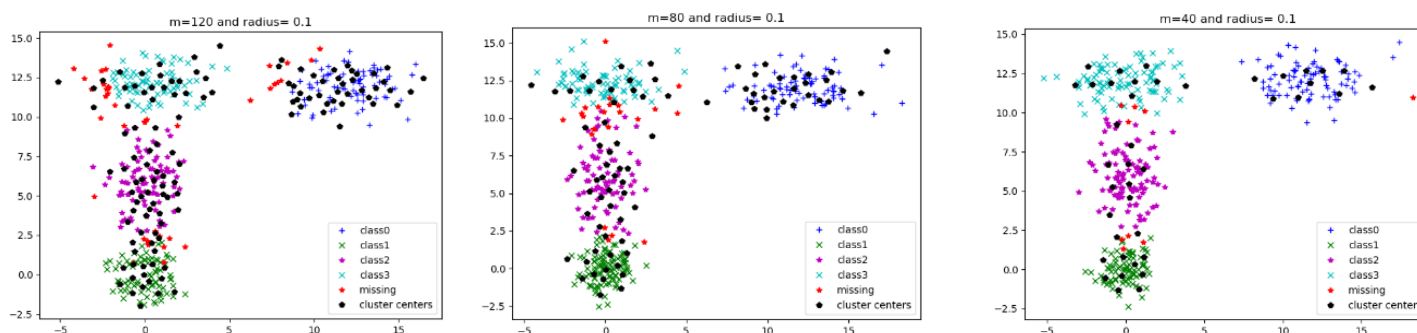
شکل ۴: برچسب داده‌های تست بعد از تخمین و مراکز خوشه های بدست آمده بر روی مجموعه داده 2clstrain1200

جدول ۹: دقت مجموعه آموزش در مجموعه داده 4clstrain1200 با شعاع مرکز ۰,۱

تعداد خوشه ها			دقت آموزش
120	80	40	
0.918	0.965	0.976	

جدول ۱۰: دقت مجموعه تست در مجموعه داده 4clstrain1200 با شعاع مرکز ۰,۱

تعداد خوشه ها			دقت تست
120	80	40	
0.89	0.939	0.975	



شکل ۵: برچسب داده‌های تست بعد از تخمین و مراکز خوشه‌های بدست آمده بر روی مجموعه داده 4clstrain1200

با توجه به جدول ۹ مشاهده می شود با افزایش تعداد مراکز خوشه ها، دقت آموزش کاهش می یابد که با افزایش تعداد مراکز خوشه ها دیگر مراکز خوشه ها بر روی داده‌هایی که خیلی دور از داده‌ها هستند قرار گرفته که سبب منحرف شدن شبکه می شود و در جدول ۱۰ نیز مشاهده می شود با افزایش تعداد مراکز خوشه ها دقت تست نیز کاهش یافت و در شکل ۴ مشاهده می کنیم در حالتی که تعداد مراکز خوشه ۱۲۰ است این مراکز خوشه در تمام نواحی که داده‌ها پراکنده شده اند پخش شده و حتی در نواحی که داده خیلی کمی وجود دارد.

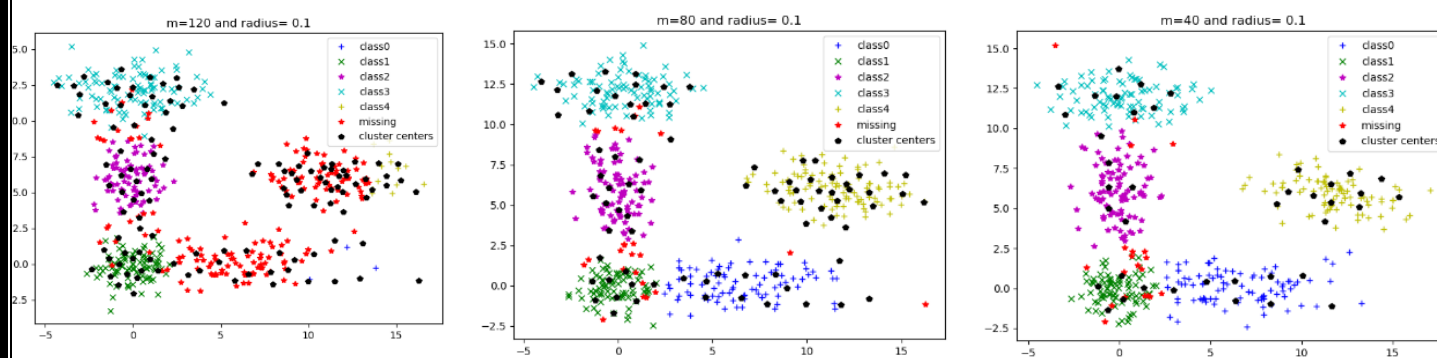
با توجه به جدول ۱۱ مشاهده می شود با افزایش تعداد مراکز خوشه ها، دقت آموزش کاهش می یابد که با افزایش تعداد مراکز خوشه ها دیگر مراکز خوشه ها بر روی داده‌هایی که خیلی دور از داده‌ها هستند قرار گرفته که سبب منحرف شدن شبکه می شود و در جدول ۱۲ نیز مشاهده می شود با افزایش تعداد مراکز خوشه ها دقت تست نیز کاهش یافت و در شکل ۴ مشاهده می کنیم در حالتی که تعداد مراکز خوشه ۱۲۰ است این مراکز خوشه در تمام نواحی که داده‌ها پراکنده شده اند پخش شده و حتی در نواحی که داده خیلی کمی وجود دارد.

جدول ۱۱: دقت مجموعه آموزش در مجموعه داده 5clstrain1500 با شعاع مرکز ۰,۱

تعداد خوشه ها			دقت آموزش
120	80	40	
0.597	0.957	0.966	

جدول ۱۲: دقت مجموعه تست در مجموعه داده 4clstrain1500 با شعاع مرکز ۰,۱

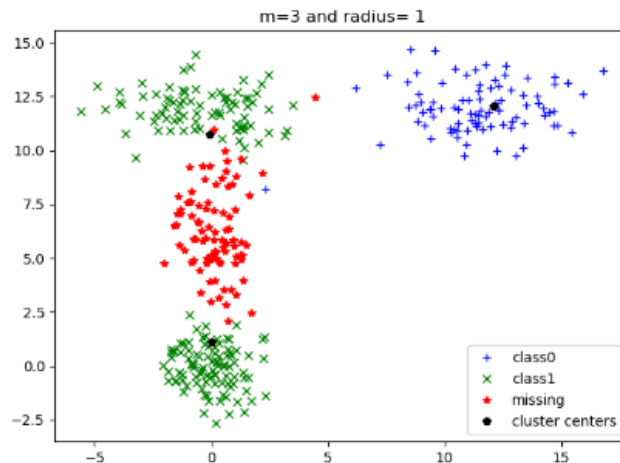
تعداد خوشه ها			دقت تست
120	80	40	
0.625	0.947	0.956	



شکل ۶: برچسب داده‌های تست بعد از تخمین و مراکز خوشه‌های بدست آمده بر روی مجموعه داده 5clstrain1500

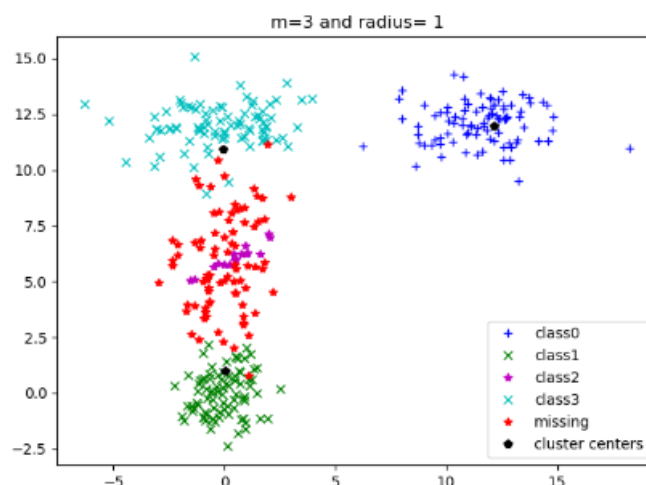
حل بخش سوم

در صورتیکه تعداد خوشه‌ها برابر ۳ و شعاع مرکز برابر ۱ باشد در مجموعه داده 2clstrain1200 دقت آموزش برابر ۰,۷۵ و دقت تست برابر ۰,۷۴۷ است. مراکز خوشه‌ها و برچسب داده‌های تست بعد از تخمین در شکل ۷ نشان داده شده است. با توجه به دقت آموزش و تست و شکل ۷ و همچنین بررسی بخش‌های قبلی سوال در این مجموعه داده متوجه می‌شویم که بالینکه داده‌ها متعلق به دو کلاس می‌باشند اما از آنجاییکه داده‌های این دو کلاس در ۴ ناحیه مختلف قرار گرفته‌اند، در نظر گرفتن این تعداد خوشه نمی‌تواند دقت خوبی را چه در حالت آموزش و چه در حالت تست در مقایسه با تعداد خوشه‌های ۴ یا بیشتر بدست آورد.



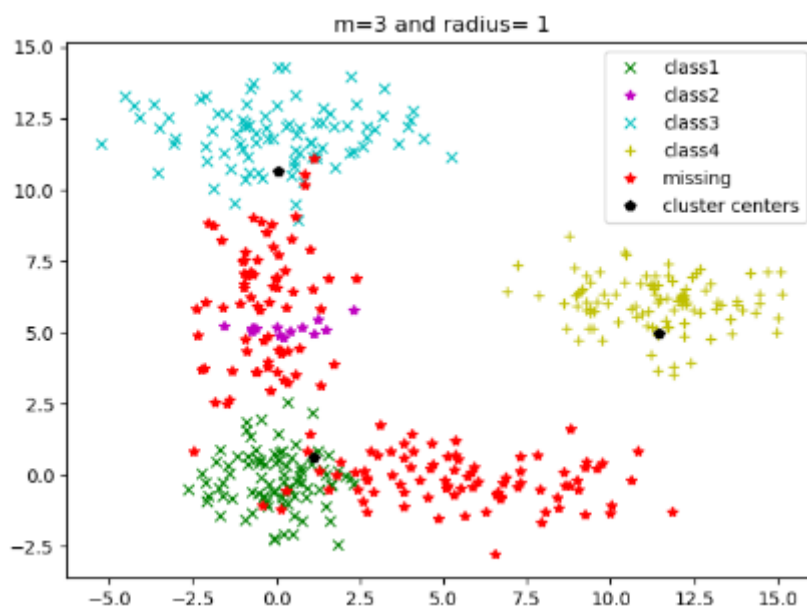
شکل ۷: مراکز خوشه ها و برجسب داده‌های تست بعد از تخمین در مجموعه داده 2clstrain1200

در صورتیکه تعداد خوشه ها برابر ۳ و شعاع مرکز برابر ۱ باشد در مجموعه داده 4clstrain1200 دقت آموزش برابر ۰,۸۰۵ و دقت تست برابر ۰,۷۷۲ است. مراکز خوشه ها و برجسب داده‌های تست بعد از تخمین در شکل ۸ نشان داده شده است. با توجه به دقت آموزش و تست و شکل ۸ و همچنین بررسی بخش های قبلی سوال در این مجموعه داده متوجه می شویم که با اینکه داده‌ها متعلق به چهار کلاس می باشند اما از آنجاییکه داده‌های این ۴ کلاس در ۴ ناحیه مختلف قرار گرفته اند، در نظر گرفتن این تعداد خوشه نمی تواند دقت خوبی را چه در حالت آموزش و چه در حالت تست در مقایسه با تعداد خوشه های ۴ یا بیشتر بدست آورد. همانوط که در شکل ۸ مشاهده می کنیم با این تعداد مرکز خوشه داده‌های یکی از کلاسها را که بین دو مرکز خوشه قرار گرفته و روی داده های آن کلاس مرکز خوشه قرار نگرفته است را نتوانستیم چندان درست تشخیص دهیم و داده های سه کلاس دیگر که مرکز خوشه بر روی داده های کلاسشان قرار گرفته را توانستیم اکثرا درست تشخیص دهیم. همانطور که قبلا بیان شد در این مجموعه داده باید حداقل ۴ مرکز خوشه در نظر گرفت که هر یک از این ۴ مرکز خوشه بر روی داده‌های یکی از ۴ کلاس قرار گیرند تا در این وضعیت بتوانیم به دقت های نسبتا خوبی دست یابیم.



شکل ۸: مراکز خوشه ها و برجسب داده‌های تست بعد از تخمین در مجموعه داده 4clstrain1200

در صورتیکه تعداد خوشه ها برابر ۳ و شعاع مرکز برابر ۱ باشد در مجموعه داده 5clstrain1500 دقت آموزش برابر ۰,۶۲۶ و دقت تست برابر ۰,۶۴۹ است. مراکز خوشه ها و برچسب داده‌های تست بعد از تخمین در شکل ۹ نشان داده شده است. با توجه به دقت آموزش و تست و شکل ۹ و همچنین بررسی بخش های قبلی سوال در این مجموعه داده متوجه می شویم که با اینکه داده‌ها متعلق به ۵ کلاس می باشند اما از آنجائیکه داده‌های این ۵ کلاس در ۵ ناحیه مختلف قرار گرفته اند، در نظر گرفتن این تعداد خوشه نمی تواند دقت خوبی را چه در حالت آموزش و چه در حالت تست در مقایسه با تعداد خوشه های ۵ یا بیشتر بدست آورد. با توجه به شکل ۹ متوجه می شویم با این تعداد خوشه داده‌های سه کلاس را توانستیم بخوبی تشخیص دهیم که بر روی داده‌های کلسشان یک مرکز خوشه قرار گرفته است و داده‌های یکی از کلاسها را که اصلا نتوانستیم بدرستی تشخیص دهیم و داده‌های کلاسی را که بین دو مرکز خوشه قرار گرفته و رنگ بنفش دارد را خیلی کم توانستیم تشخیص دهیم. همانطور که قبلا بیان شد در این مجموعه داده باید حداقل ۵ مرکز خوشه در نظر گرفت که هر یک از این ۵ مرکز خوشه بر روی داده‌های یکی از ۵ کلاس قرار گیرند تا در این وضعیت بتوانیم به دقت های نسبتا خوبی دست یابیم.



شکل ۹: مراکز خوشه ها و برچسب داده‌های تست بعد از تخمین در مجموعه داده 5clstrain1500

در قسمت آخر از ما خواسته شد تا برای خوشه هایمان مرزبندی کنیم. در این بخش طبق روشی که در صورت پروژه ذکر شد، مرزبندی برای خوشه هایمان در یک شکل جدید بدست می آید که چند تا از آن ها به شک زیر است:

