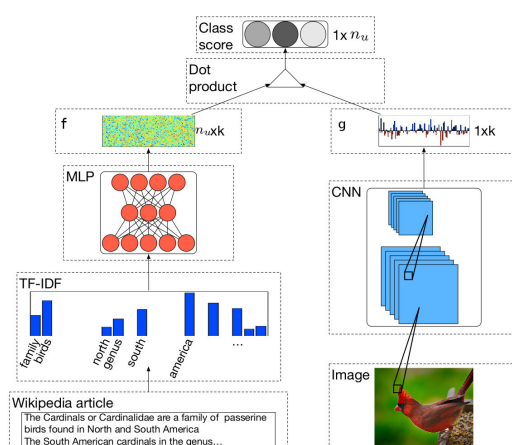


و اجبار می‌کند که تمامی نمونه‌های دیده‌شده باید در طرف منفی دسته‌بند خطی  $w$  قرار گیرند. نویسندگان این پژوهش، روش خود را با استفاده از تکنیک هسته  $\gamma^*$  برای دسته‌بندهای غیرخطی نیز توسعه داده‌اند [۳۶].

## ۸-۲ نگاشت به یک فضای میانی

در برخی روش‌ها هر دوی نگاشت‌های  $\phi$  و  $\theta$ ، معرفی شده در ابتدای فصل با توجه به داده‌ها یاد گرفته می‌شوند و در نتیجه فضای مشترک مورد استفاده نه فضای تصاویر و نه فضای توصیف‌هاست؛ بلکه فضای ثالثی است. این فضای میانی در برخی از روش‌ها یک فضای با بعد کمتر است و تعبیر معنایی برای آن موجود نیست. در برخی روش‌های دیگر، فضای میانی را با بعد  $n_g$  یعنی تعداد دسته‌های دیده شده در نظر گرفته‌اند و تعبیر معنایی برای آن ارائه شده است. این فضای میانی بر اساس توصیف دسته‌ها و نمونه‌های دیده نشده بر اساس شباهت آن‌ها با دسته‌های دیده شده استوار است.



شکل ۲-۲: شبکه مورد استفاده برای یادگیری توام نگاشت تصاویر و توصیف‌ها که یک شبکه عصبی ژرف با دو ورودی است. ورودی اول از نوع تصویر است و ابتدا با یک شبکه پیچشی سپس با چند لایه چگال به فضایی  $k$ -بعدی می‌رود. ورودی دوم که یک مقاله از ویکی‌پدیای انگلیسی است پس از تبدیل به نمایش برداری به صورت  $tf-idf$  با چند لایه با اتصالات چگال پردازش شده و به فضایی  $k$ -بعدی می‌رود. در نهایت امتیاز تعلق تصویر به دسته‌ی متن با ضرب داخلی این دو نگاشت تعیین می‌شود [۱].

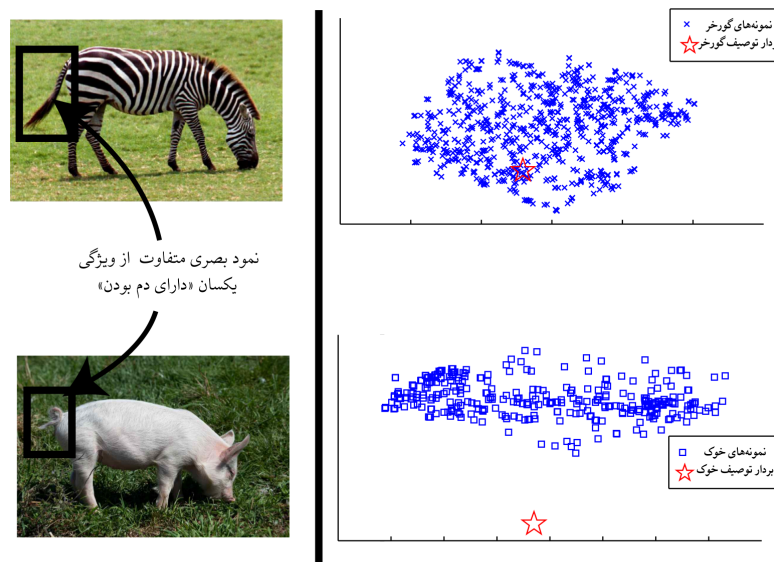
در [۱] از شبکه‌های عصبی ژرف برای یادگیری توام نگاشت‌های  $\phi$  و  $\theta$  استفاده شده است. نمای کلی شبکه مورد استفاده در این روش در تصویر ۲-۲ نشان داده شده است. توصیف‌های متنی و ویژگی‌های بصری دو ورودی جداگانه به چنین شبکه‌ای هستند که ابتدا به صورت جداگانه با یک یا چند لایه‌ی با اتصالات کامل به یک فضای مشترک نگاشته

<sup>\*</sup>kernel trick

## ۹-۲ روش‌های نیمه‌نظارتی

در این بخش به بررسی روش‌های نیمه‌نظارتی می‌پردازیم. این روش‌ها از نظر نوع نگاشت‌های مورد استفاده در یکی از دسته‌های قبلی قابل بیان بودند ولی با توجه به این که روش پیشنهادی ما نیز نیمه‌نظارتی است، برای پررنگ‌تر شدن نحوه‌های استفاده از داده‌های آزمون در جریان آموزش این دسته را به طور جداگانه مورد بررسی قرار می‌دهیم.

در [۳۹] برای نخستین بار مشکل جابجایی دامنه<sup>۴۰</sup> معرفی شد. این مشکل که در شکل ۲-۳ قابل مشاهده است به متفاوت بودن خواص صفت‌ها برای دسته‌های مختلف اشاره می‌کند. برای مثال صفت راه‌راه بودن برای دو حیوان گورخر و ببر از نظر بصری خواص متفاوتی دارد و یادگیری یک دسته‌بند برای تشخیص راه‌راه بودن با استفاده از تصاویر گورخر در تشخیص وجود و یا عدم وجود این صفت در تصویر ببر ضعیف خواهد بود.

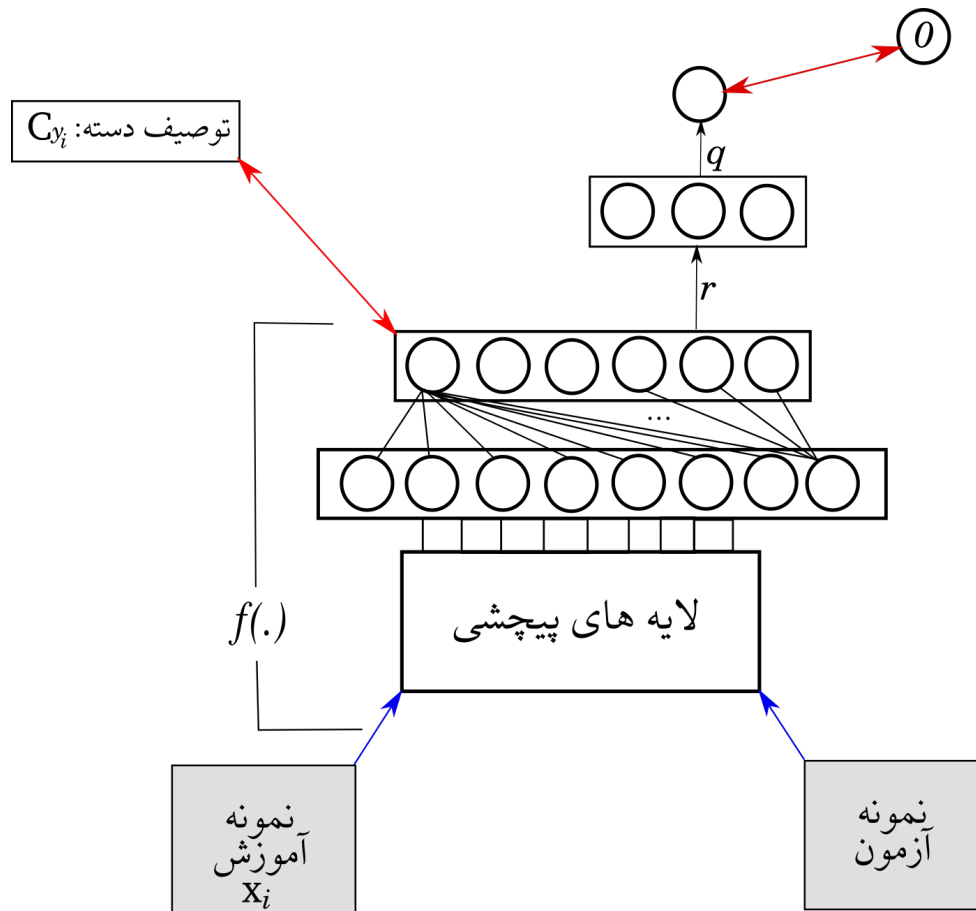


شکل ۲-۳: مشکل جابجایی دامنه بین دو دسته‌ی دیده شده (گورخر) و دیده نشده (خوک) نمایش داده شده است. صفت یکسان «دارای دم بودن» در این دو دسته دارای دو نمود بصری متفاوت است (سمت چپ) و نگاشت یادگرفته شده برای بردن این صفت به فضای مشترک برای دسته‌ی دیده نشده عمل‌کرد ضعیف‌تری نسبت به دسته‌ی دیده شده به نمایش می‌گذارد (سمت راست) [۳۹].

در [۳۹] برای حل این مشکل دو تکنیک به کار گرفته شده است. ابتدا یافتن نمایش مشترک برای سه دامنه‌ی تصاویر، بردار صفت و بردار نام دسته‌ها به صورت توأمان با استفاده از CCA<sup>۴۱</sup> [۴۰] و سپس برچسب‌گذاری داده‌های بدون

<sup>۴۰</sup>Domain shift problem

<sup>۴۱</sup>Canonical Correlation Analysis

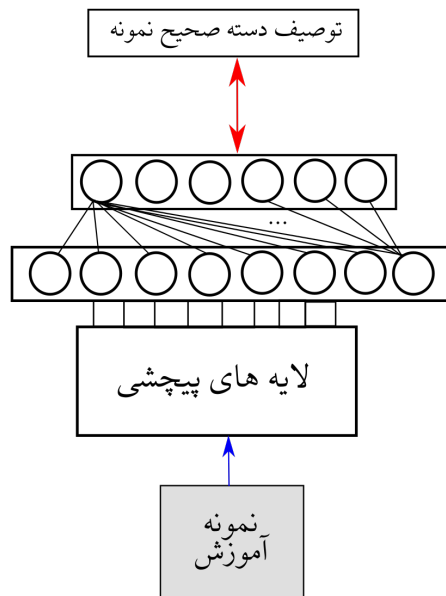


شکل ۳-۲: ساختار شبکه چند وظیفه‌ای پیشنهادی. فلش‌های آبی رنگ ورودی‌های شبکه را نشان می‌دهند و فلش‌های قرمز رنگ مقایسه خروجی شبکه با خروجی مورد انتظار را. خطوط سیاه رنگ اتصالات شبکه را نشان می‌دهند. زیر شبکه‌ی برگرفته شده از شبکه vgg و یک لایه‌ی با اتصالات چگال اضافه شده بین دو ورودی مشترک هستند. لایه‌های  $r$  و  $q$  مخصوص نمونه‌های آزمون هستند. خروجی لایه‌ی  $q$  همواره با مقدار صفر مقایسه می‌شود.

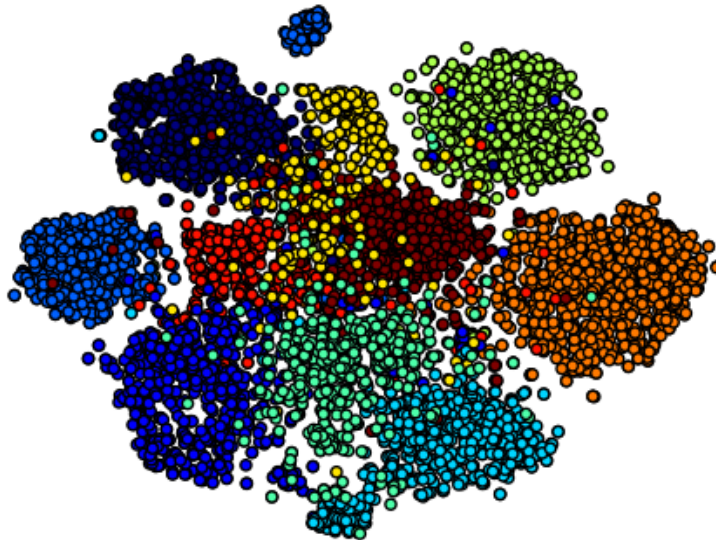
است که در حین یادگیری نگاشت اجبار شود که حاصل نگاشت یک نمونه‌ی آزمون به نوعی نزدیک به نگاشت توصیف یکی از دسته‌های آزمون باشد. همان طور که در بخش ۲-۹ بیان شد، چنین راه‌حلی در [۴۲] استفاده شده است. معیار نزدیکی نگاشت‌ها در آن روش یک امتیاز پیشین از شباهت هر نمونه‌ی آزمون با دسته‌های دیده نشده است که توسط یک روش دیگر استخراج شده می‌شود. یعنی ابتدا یک روش دسته‌بندی احتمالی که در آن پژوهش روش IAP [۱۳] برای این کار انتخاب شده بود، به صورت مستقل روی مجموعه داده‌گان اجرا شده و احتمال‌هایی که برای انتساب هر نمونه به دسته‌های آزمون از آن روش بدست می‌آید بعنوان وزن‌های شباهت در نظر گرفته می‌شود و فاصله هر توصیف پیش‌بینی شده برای هر نمونه با توصیف دسته‌های آزمون متناسب با این وزن‌های شباهت جریمه می‌شود. ما در این بخش یک روش

## ۳-۲-۳ یک مدل پایه برای مقایسه

برای روشن شدن تاثیر استفاده از اطلاعات بدون نظارت نمونه‌های آزمون در یادگیری بهتر نگاشت، قصد داریم در فصل آتی مدل ارائه شده را با یک مدل ساده برای پیش‌بینی صفت مقایسه کنیم که در این‌جا این مدل پایه را معرفی می‌کنیم. در این مدل ساده تنها از لایه‌های با اتصالات کامل بعد از استخراج ویژگی با لایه‌های پیچشی، برای پیش‌بینی صفت استفاده می‌کنیم. ساختار این مدل در تصویر ۳-۳ نمایش داده شده است. در این شبکه از یک یا چند لایه با اتصالات کامل بعد از لایه‌های پیچشی استفاده می‌شود. مشابه حالت قبل تابع فعال‌سازی برای مجموعه داده‌گانی که مقادیر توصیف دسته‌هایشان دودویی است تابع سیگموید، و برای مجموعه داده‌گانی که مقادیر بردارهای توصیف در آن‌ها مقادیر دلخواه حقیقی است تابع ReLU استفاده شده است. ابعاد لایه‌های با اتصالات کامل پایانی الزاماً برابر تعداد ابعاد بردارهای توصیف است و برای سایر لایه با اتصالات کامل نیز همین تعداد ابعاد انتخاب شده است. مقایسه نتایج دقت دسته‌بندی بین مدل قبلی و این مدل در بخش ۴-۴ نشان‌دهنده تاثیر مثبت استفاده از اطلاعات بدون نظارت موجود در نمونه‌های آزمون است که باعث بهبود حداقل ۱۰ درصدی دقت دسته‌بندی می‌شود.



شکل ۳-۳: ساختار شبکه پایه. فلش آبی رنگ ورودی‌های شبکه را نشان می‌دهند و فلش‌های قرمز رنگ مقایسه خروجی شبکه با خروجی مورد انتظار را.



شکل ۳-۴: نمایش دوبعدی بوسیله  $t$ -SNE برای ده دسته‌ی آزمون از مجموعه داده‌گان AWA با ده رنگ متفاوت نشان داده شده است. درستی فرض قابل خوشه‌بندی در تصویر مشخص است، یعنی ویژگی‌های استخراج شده با استفاده از شبکه‌های ژرف توانایی ایجاد تمایز بالا میان دسته‌ها را دارا هستند و نمونه‌های هر دسته نیز نزدیک به یکدیگر هستند.

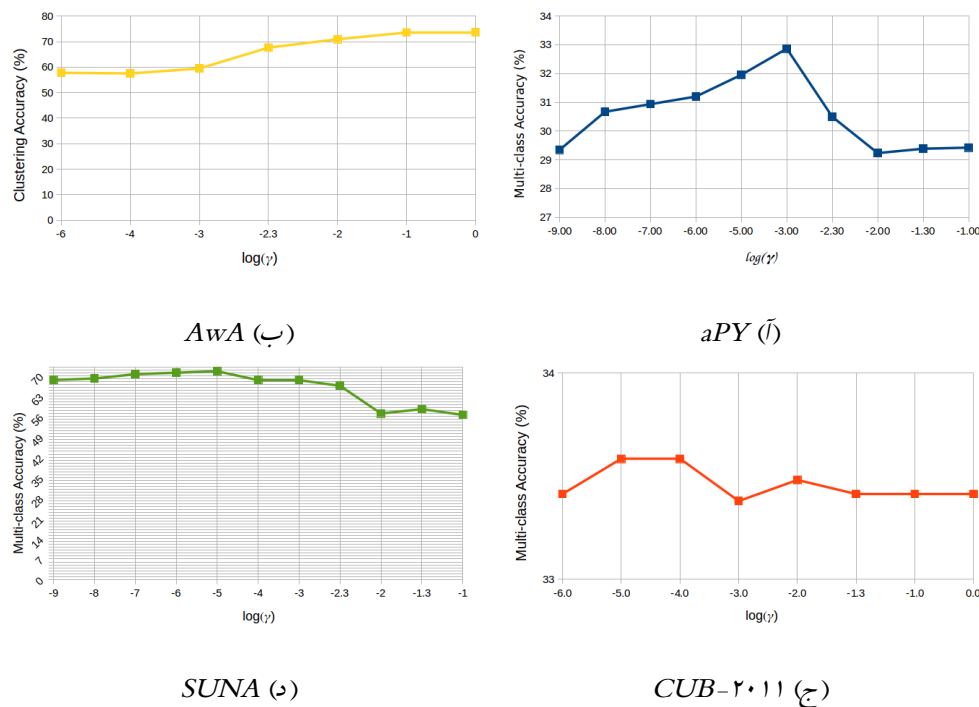
### ۳-۳ تابع مطابقت مبتنی بر خوشه‌بندی

در اکثر روش‌های پیشین که در فصل ۲ مرور شد، تابع مطابقت میان تصاویر و توصیف‌ها برای اختصاص برچسب به داده‌های آزمون بر اساس فاصله کمینه یا ضرب داخلی بیشینه در یک فضای مشترک محاسبه می‌شد. استثنای این موضوع، استفاده از روش انتشار برچسب در [۳۹] و [۴۲] و همچنین پیش‌بینی مستقیم برچسب‌ها در [۴۱] و [۴۵] هستند.

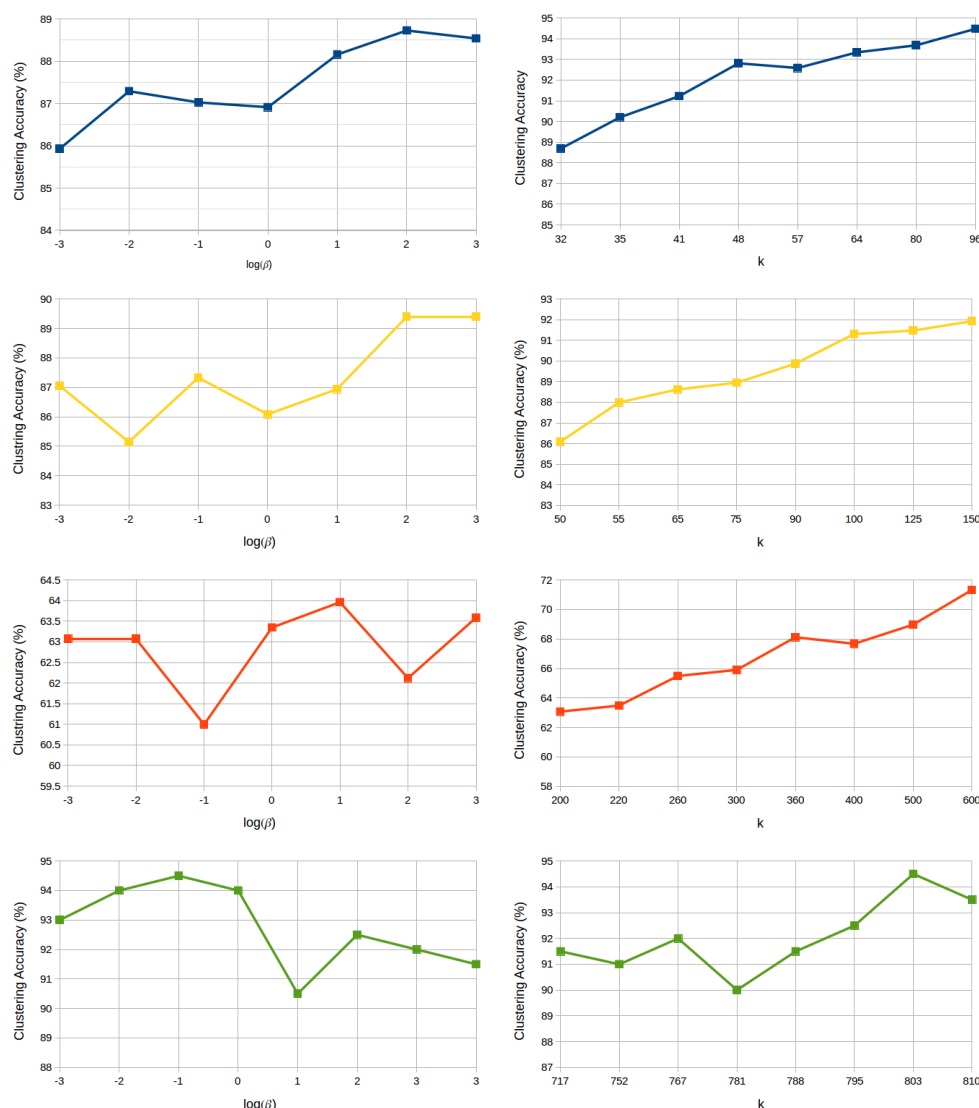
در این بخش ما یک تابع مطابقت جدید بر اساس یک خوشه‌بندی روی داده‌های دسته‌های دیده نشده، تعریف می‌کنیم. اگر فضای نمایش تصاویر دارای این خاصیت باشد که دسته‌های مختلف به صورت خوشه‌های مجزا باشند، استفاده از خوشه‌بندی برای انتساب برچسب از نظر شهودی توجیه‌پذیر است. با توجه به نمایش غنی بوجود آمده برای تصاویر توسط شبکه‌های ژرف این فرض در بسیاری از موارد برقرار است. برای نمونه، نمایش  $t$ -SNE نمونه‌های آزمون مجموعه داده‌های AWA در تصویر ۳-۴ نشان داده شده است و برقراری فرض قابل خوشه‌بندی بودن در آن قابل مشاهده است. این ادعا با

## ۴-۴-۲ تحلیل پارامتر

در این بخش به تحلیل تاثیر پارامتر  $\gamma$  در رابطه (۲-۳) می‌پردازیم. این پارامتر وزن جمله‌ی اضافه شده به شبکه‌ی پایه که برای تضمین شباهت خروجی شبکه روی نمونه‌های آزمون به بردار توصیف یکی از دسته‌های آزمون به کار می‌رود و مقدار آن در جریان آموزش با اعتبارسنجی تعیین می‌شود. تاثیر مقدار این پارامتر بر دقت نهایی دسته‌بندی در تصویر ۱-۴ آمده است. از سیر صعودی ابتدای این نمودار تاثیر مثبت وجود این جمله نسبت به حالتی که از آن استفاده نشود مشخص است. از طرفی محل بیشینه شدن آن، اهمیت آن نسبت به جمله‌ی دیگر نشان می‌دهد که باید با وزنی حدود  $\frac{1}{1000}$  جمله دیگر یعنی خطای پیش‌بینی صحیح صفات برای نمونه‌های آزمون وجود داشته باشد و در غیر این صورت باعث کاهش دقت خواهد شد.



شکل ۱-۴: میزان دقت دسته‌بندی چند دسته‌ای در شبکه چندوظیفه‌ای ارائه شده (نسخه یک لایه) بر حسب  $\log_{10}$  پارامتر  $\gamma$  در معادله (۲-۳).



شکل ۴-۲: تاثیر پارامترهای روش خوشه‌بندی نیمه‌نظارتی. سمت چپ: نتیجه دقت خوشه‌بندی با تغییر تعداد خوشه‌های در نظر

گرفته شده. سمت راست: نتیجه دقت دسته‌بندی چند خوشه‌بندی بر حسب مقادیر پارامتر  $\beta$  در رابطه (۳-۱۲).

برای راحتی مقایسه محور عمودی همه‌ی نمودارها با بازه‌های یک درصدی تقسیم‌بندی شده‌اند.

سطر اول (آبی‌رنگ): مجموعه دادگان aPY. سطر دوم (زرد رنگ): مجموعه دادگان AWA. سطر سوم (قرمز رنگ): مجموعه

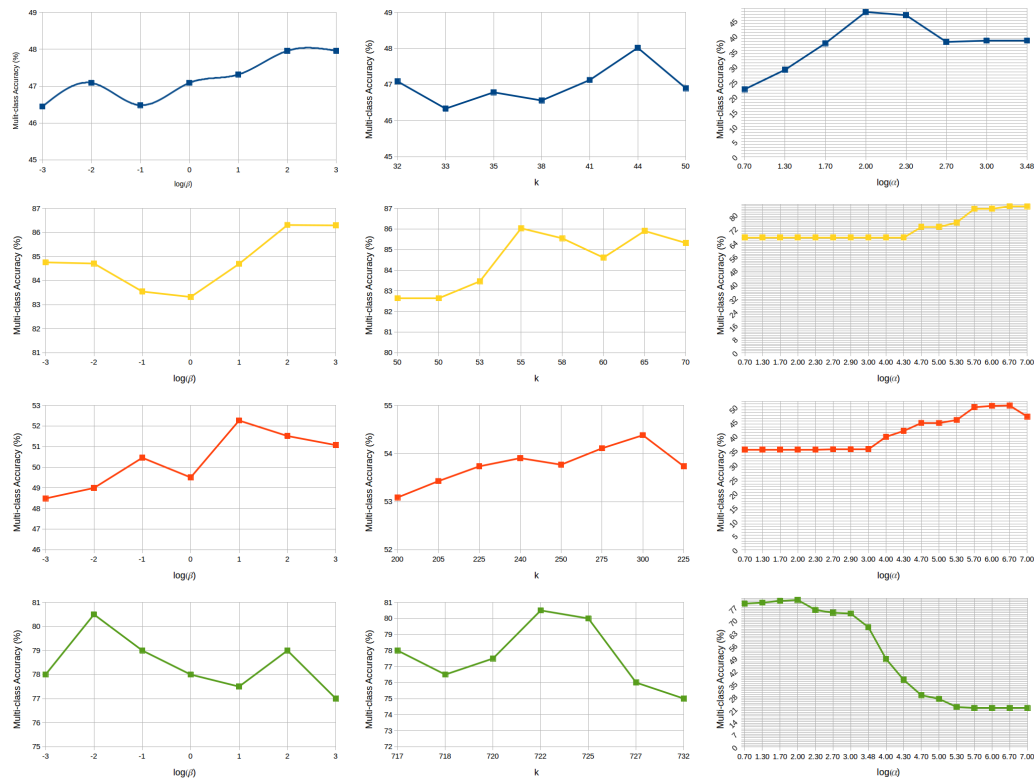
دادگان CUB-2011. سطر چهارم (سبز رنگ): مجموعه دادگان SUNA.

در رابطه (۳-۱۸) تاثیر قابل توجهی بر دقت دسته‌بندی نهایی دارد، در نتیجه ما مقدار این پارامتر را با استفاده از روند

اعتبارسنجی شرح داده شده در بخش ۴-۲ تنظیم کرده‌ایم. از طرف دیگر مشاهده می‌شود تعداد خوشه‌ها در خوشه‌بندی

نیمه‌نظارتی ارائه شده تاثیر قابل توجهی بر دقت دسته‌بندی ندارد، در نتیجه برای سادگی و کاهش زمان روند آموزش ما این

پارامتر را همان‌طور که در بخش ۳-۴ شرح داده شد با استفاده از یک قاعده سرانگشتی بر حسب تعداد دسته‌ها تعیین می‌کنیم که تعداد خوشه‌ها برای هر مجموعه داده‌گان برابر  $k = n_s + 2n_u$  در نظر گرفته می‌شود.

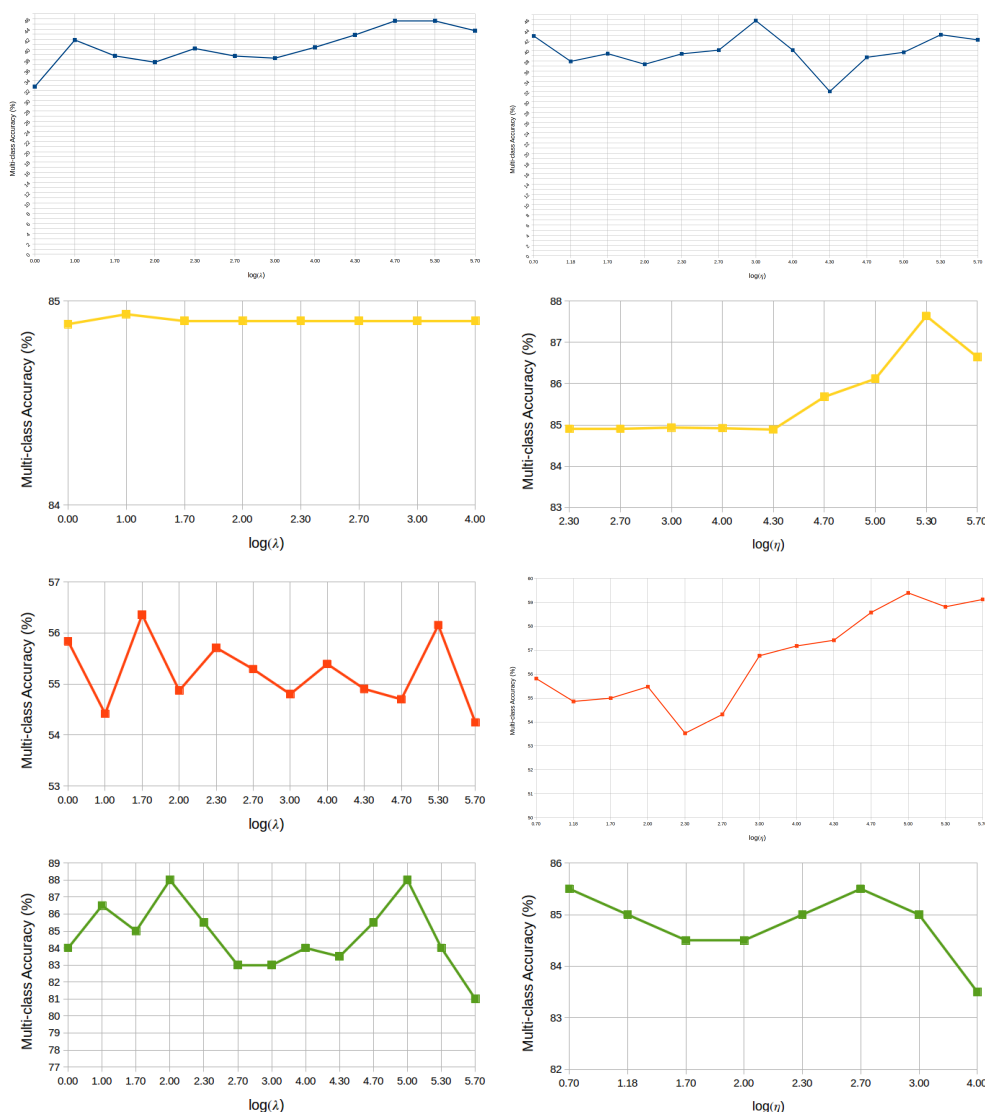


شکل ۳-۴: تاثیر پارامترهای روش دسته‌بندی با خوشه‌بندی نیمه‌نظارتی. سمت چپ: نتیجه دقت دسته‌بندی چند دسته‌ای بدست آمده بر حسب پارامتر  $\alpha$  در رابطه (۳-۱۸) که اهمیت جمله منظم‌سازی را نشان می‌دهد. همان‌طور که مشاهده می‌شود، عمل‌کرد روش به این پارامتر حساس است. وسط: نتیجه دقت دسته‌بندی چند دسته‌ای بدست آمده بر حسب تعداد خوشه‌ها در خوشه‌بندی نیمه‌نظارتی. با توجه مقیاس این نمودار مشخص می‌شود که دقت حاصل شده حساسیت کمی نسبت به این پارامتر دارد. سمت راست: نتیجه دقت دسته‌بندی چنددسته‌بای بر حسب پارامتر  $\beta$  در خوشه‌بندی نیمه‌نظارتی (رابطه (۳-۱۲)).

برای راحتی مقایسه محور عمودی همه‌ی نمودارها با بازه‌های یک درصدی تقسیم‌بندی شده‌اند.

سطر اول (آبی‌رنگ): مجموعه دادگان aPY. سطر دوم (زرد رنگ): مجموعه دادگان AW. سطر سوم (قرمز رنگ): مجموعه دادگان CUB-2011. سطر چهارم (سبز رنگ): مجموعه دادگان SUNA.

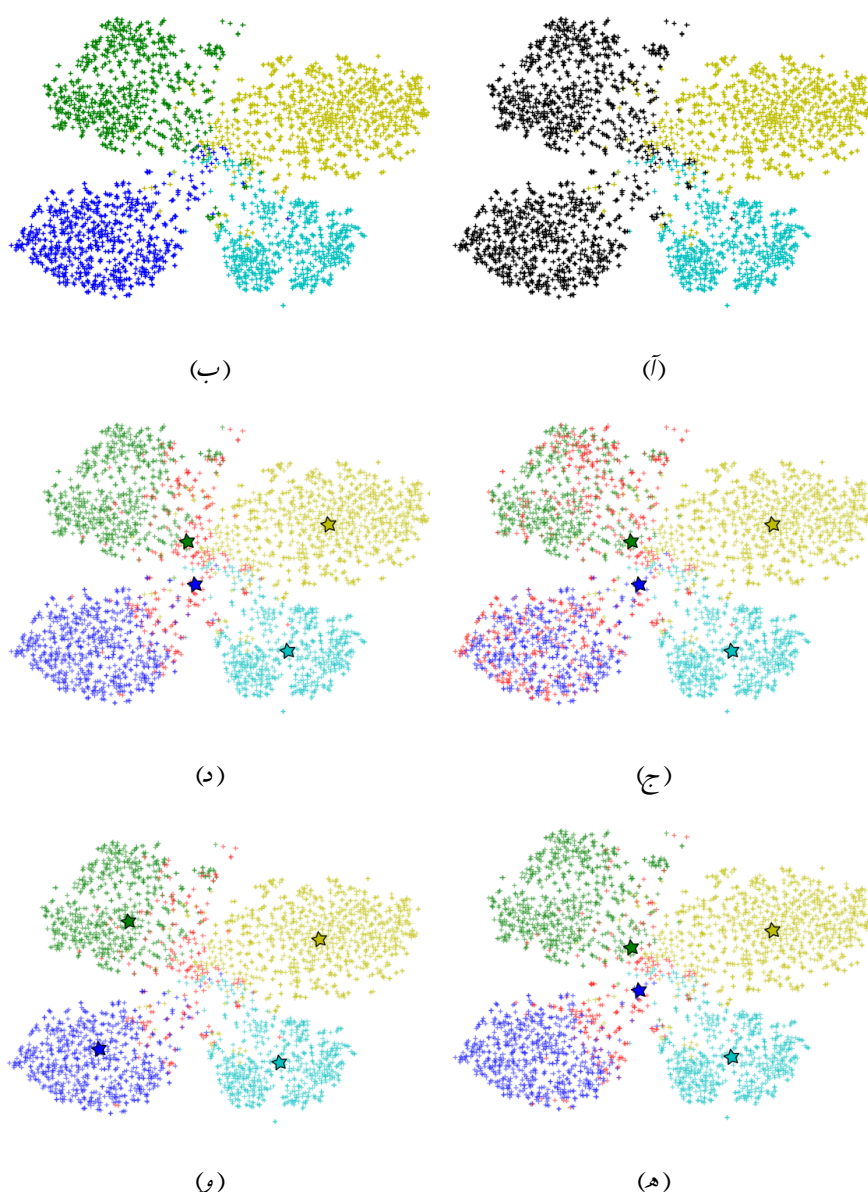




شکل ۴-۴: تاثیر پارامترهای روش یادگیری نگاشت و خوشه‌بندی نیمه‌نظارتی برای مجموعه داده‌گان مختلف. سمت چپ: نتیجه دقت دسته‌بندی چند دسته‌ای بدست آمده بر حسب پارامتر  $\gamma$  در رابطه (۳-۲۰) که اهمیت جمله منظم‌سازی را نشان می‌دهد. سمت راست: نتیجه دقت دسته‌بندی چند دسته‌ای بدست آمده بر حسب مقادیر پارامتر  $\lambda$  در رابطه (۳-۲۰) برای راحتی مقایسه محور عمودی همه‌ی نمودارها با بازه‌های یک درصدی تقسیم‌بندی شده‌اند.

سطر اول (آبی‌رنگ): مجموعه داده‌گان aPY. سطر دوم (زرد رنگ): مجموعه داده‌گان AWA. سطر سوم (قرمز رنگ): مجموعه داده‌گان CUB-2011. سطر چهارم (سبز رنگ): مجموعه داده‌گان SUNA.

است که توانایی جداسازی دسته‌ها در آن بسیار پایین‌تر از ویژگی‌های بدست آمده از شبکه‌های عصبی ژرف است که در روش‌های پیشنهادی ما مورد استفاده قرار گرفته است. روش‌های [۲۸، ۳۱] با استفاده از توابع هزینه‌ی بیشترین حاشیه



شکل ۴-۵: نمایش دوبعدی چهار دسته از مجموعه دادگان AWA با استفاده از نگاشت  $t$ -SNE، دو دسته‌ی دیده شده شامل بزگوزن (فیروزه‌ای) خرس گریزلی (زرد) و دو دسته‌ی دیده نشده شامپانزه (آبی) و پاندا (سبز). تصاویر با نماد بعلاوه و نگاشت توصیف دسته‌ها در فضای تصاویر با ستاره نشان داده شده است. در تصاویر  $b$  تا  $f$  نقطه‌های قرمز نمونه‌هایی که را نشان می‌دهد که دسته‌ای به جز چهار دسته‌ی موجود در شکل برای آن‌ها پیش‌بینی شده است.  $\bar{A}$  دسته‌های دیده شده با برجسب صحیح و دیده‌نشده با رنگ مشکی (ب) نمایش برجسب صحیح برای تمامی دسته‌ها (ج) توصیف‌ها با نگاشت (۳-۱۸) به فضای تصاویر برده شده‌اند و دسته‌بندی با دسته‌بند نزدیک‌ترین همسایه انجام شده است. (د) نگاشت مانند حالت قبل و دسته‌بندی با تابع مطابقت پیشنهادی به همراه خوشه‌بند  $k$ -means (ه) نگاشت مانند حالت قبل و دسته‌بندی با تابع مطابقت پیشنهادی به همراه خوشه‌بند نیمه‌نظارتی پیشنهاد شده (و) دسته‌بندی و نگاشت با استفاده از روش پیشنهادی برای یادگیری نگاشت و خوشه‌بندی توام