

به نام خدا



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

درس شبکه‌های عصبی
استاد صفابخش

تمرین پنجم

علیرضا مازوچی

۴۰۰۱۳۱۰۷۵

سوال ۱

ابتدا نکات کلی در مورد پیاده‌سازی را بیان می‌کنم:

- برای پیاده‌سازی هر دو شبکه به پیاده‌سازی یک سلول SimpleRNN در گیت‌هاب مراجعه شده است و برخی از مقداردهی‌های اولیه و ایده‌های کلی از آن گرفته شده است.
- با توجه به کم بودن داده‌ها تعداد دسته‌ها بر روی ۱۶ تنظیم شده است.
- برای آموزش هر دو شبکه اجازه داده شده است تا حداکثر ۲۵۰ گام آموزش انجام بگیرد و برای جلوگیری از بیش برآزش از یک کالبد EarlyStopping استفاده شده است.

نتایج برای شبکه Elman در جدول زیر ذکر شده است:

ردیف	تعداد نورون مخفی	تعداد گام آموزش	دقت مجموعه آموزش	دقت مجموعه اعتبارسنجی	دقت مجموعه ارزیابی
۱	۸	۱۴۱	٪۸۶/۷۳	٪۸۰/۳۴	٪۸۹/۲۵
۲	۳۲	۱۹۳	٪۹۷/۸۴	٪۹۷/۳۸	٪۹۵/۷۰
۳	۶۴	۹۸	٪۹۵/۶۸	٪۸۹/۱۳	٪۹۰/۳۲
۴	۱۲۸	۴۵	٪۹۳/۸۳	٪۹۳/۴۸	٪۹۰/۳۲
۵	۵۱۲	۲۹	٪۸۹/۸۱	٪۸۲/۶۱	٪۸۸/۱۷
۶	۲۰۴۸	۴۶	٪۹۹/۳۸	٪۹۵/۶۵	٪۹۶/۷۷

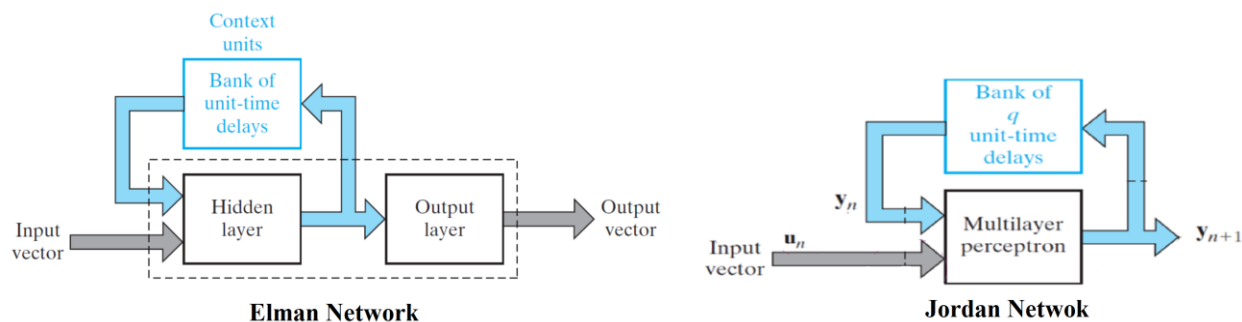
در بین تنظیمات انجام‌شده بهترین دقت ارزیابی را شبکه با ۲۰۴۸ نورون داشته است. البته شبکه با ۳۲ نورون هم توانسته است دقت نزدیکی را ثبت کند. به طور کلی اگر تعداد نورون‌ها خیلی کم باشد (مثلا ۸ نورون) شبکه دچار کم‌برآزش می‌شود و نمی‌تواند دقت‌های خوبی را داشته باشد. از طرفی اگر تعداد نورون‌ها زیاد باشد استعداد بیش‌برآزش وجود دارد ولی با توجه به کالبد استفاده‌شده آموزش خاتمه پیدا می‌کند؛ بدین ترتیب جلوی بیش‌برآزش تا حد زیادی گرفته می‌شود و به طور جانبی آموزش هم زودتر تمام می‌شود. پس تعداد گام آموزش برای شبکه با تعداد سلول بیشتر، کمتر است.

برای شبکه Jordan به طور مشابه نتایج در جدول زیر آورده شده است:

ردیف	تعداد نورون مخفی	تعداد گام آموزش	دقت مجموعه آموزش	دقت مجموعه اعتبارسنجی	دقت مجموعه ارزیابی
۱	۸	۳۳	۷۵/۰۰٪	۷۶/۰۹٪	۸۱/۷۲٪
۲	۳۲	۱۵۷	۹۵/۳۷٪	۹۱/۳۰٪	۹۴/۶۲٪
۳	۶۴	۷۴	۹۱/۰۵٪	۷۳/۹۱٪	۹۱/۴۰٪
۴	۱۲۸	۱۱۸	۹۷/۲۲٪	۸۶/۹۶٪	۸۹/۲۵٪
۵	۵۱۲	۲۷	۸۴/۸۸٪	۶۵/۲۲٪	۸۶/۰۲٪
۶	۲۰۴۸	۱۶	۷۹/۹۴٪	۷۶/۰۹٪	۸۱/۷۲٪

برای این شبکه هم تعداد نورون بسیار کم مطابق انتظار خوب نیست و شبکه نمی‌تواند آموزش ببیند؛ در کنار آن می‌بینیم که تعداد نورون زیاد هم باعث عدم یادگیری و خاتمه زودهنگام آموزش می‌شود. بهترین حالت برای این شبکه با ۳۲ نورون رخ می‌دهد.

به طور کلی شبکه Elman دقت‌های بهتری از شبکه Jordan دارد. اما تعداد گام و زمان اجرای Jordan مناسب‌تر است. برای فهمیدن چرایی این موضوع باید تفاوت این دو شبکه را درک کنیم. در تصویر زیر معماری دو شبکه آورده شده است:



همانطور که مشخص است در Elman خروجی لایه مخفی وارد شبکه در لحظه بعد می‌شود، اما در Jordan خروجی نهایی شبکه وارد شبکه در لحظه بعد می‌شود؛ حال توجه کنید که مسئله ما یک مسئله دو کلاسه است پس خروجی دو بعدی است

درحالی که خروجی لایه مخفی به تعداد نورون‌ها ابعاد دارد؛ یعنی اگر ۳۲ نورون داشته باشیم، خروجی ۳۲ بعدی وارد شبکه در لحظه بعد می‌شود. با این اوصاف طبیعی است که Elman می‌تواند دانش بیشتری را از لحظات قبل در اختیار بگیرد و منطقی است که دقت بیشتری داشته باشد و در ازای آن پیچیدگی بیشتر و زمان اجرای بیشتر داشته باشد. پیچیدگی کمتر Jordan در عین حال که باعث دقت کمتر می‌شود، برای تنظیمات با تعداد نورون مخفی بیشتر به طرز شدیدتری منجر به بیش‌برازش می‌شود.

با توجه به نتایج دریافت‌شده و تحلیل ارائه‌شده واضحاً شبکه Elman برای مجموع‌داده سوال مناسب‌تر به نظر می‌آید.

سوال ۲

برای شبکه‌های این سوال، ابتدا هر یک از زیرشبکه‌ها با داده‌های مربوط به خود (تمام یا بخشی از داده‌ها) آموزش پیدا کرده است؛ سپس خروجی آن‌ها به یک لایه میانگین‌گیر داده شده است. جزئیات آموزش زیرشبکه‌ها کاملاً مشابه سوال ۱ است.

مجموعه داده یکسان

برای شبکه Elman و در حالتی که کل داده‌ها را در اختیار تمام زیر مدل‌ها قرار بدهیم، نتایج زیر را به ازای برخی از تنظیمات بدست آوردم:

تعداد نوروں زیرمدل‌ها	دقت مجموعه آموزش	دقت مجموعه اعتبارسنجی	دقت مجموعه ارزیابی
۸، ۸، ۸، ۸، ۸	۸۸/۵۸٪	۷۱/۷۴٪	۸۴/۹۵٪
۳۲، ۳۲، ۳۲، ۳۲، ۳۲	۹۸/۷۷٪	۹۷/۸۳٪	۹۶/۷۷٪
۲۵۶، ۲۵۶، ۲۵۶، ۲۵۶، ۲۵۶	۹۸/۷۷٪	۹۷/۸۳٪	۹۵/۷۰٪
۲۰۴۸، ۲۰۴۸، ۲۰۴۸، ۲۰۴۸، ۲۰۴۸	۹۹/۶۹٪	۱۰۰/۰۰٪	۹۷/۸۵٪
۳۲، ۶۴، ۱۲۸، ۵۱۲، ۲۰۴۸	۹۸/۴۶٪	۹۷/۸۳٪	۹۴/۶۲٪
۳۲، ۲۰۴۸	۱۰۰/۰۰٪	۱۰۰/۰۰٪	۱۰۰/۰۰٪

همانطور که مشخص است نتایج بسیار خوبی از ترکیب زیرشبکه‌های Elman بدست آمده است به گونه‌ای که ترکیب یک شبکه ساده ۳۲ نورونه و ۲۰۴۸ نورونه توانسته است به دقت کامل روی هر سه مجموعه برسد!

در مقایسه‌ای دیگر دقت روی مجموعه ارزیابی در حالت تکی و گروهی از مدل‌های یکسان را در جدول زیر آورده‌ام برای این جدول هم می‌بینیم که برای شبکه‌ها با تعداد نوروں معقول افزایش دقت رخ داده است.

تعداد نوروں مخفی	دقت ارزیابی تک مدل	دقت ارزیابی مدل گروهی
۸	۸۹/۲۵٪	۸۴/۹۵٪
۳۲	۹۵/۷۰٪	۹۶/۷۷٪
۲۰۴۸	۹۶/۷۷٪	۹۷/۸۵٪

برای شبکه Jordan نتایج در جدول زیر آورده شده است. برای Jordan با توجه به نتایج تک مدل تصمیم گرفتیم از زیرمدل‌های ساده‌تر استفاده کنیم:

تعداد نورون زیرمدل‌ها	دقت مجموعه آموزش	دقت مجموعه اعتبارسنجی	دقت مجموعه ارزیابی
۸، ۸، ۸، ۸، ۸	۸۴/۵۷٪	۷۱/۷۴٪	۸۷/۱۰٪
۳۲، ۳۲، ۳۲، ۳۲، ۳۲	۹۱/۰۵٪	۷۶/۰۹٪	۹۲/۴۷٪
۱۲۸، ۱۲۸، ۱۲۸، ۱۲۸، ۱۲۸	۹۶/۶۰٪	۹۱/۳۰٪	۹۶/۷۷٪
۲۵۶، ۲۵۶، ۲۵۶، ۲۵۶، ۲۵۶	۹۳/۸۳٪	۸۴/۷۸٪	۸۸/۱۷٪
۸، ۳۲، ۱۲۸، ۲۵۶، ۵۱۲	۹۷/۲۲٪	۸۴/۷۸٪	۹۴/۶۲٪
۳۲، ۶۴	۹۱/۶۷٪	۸۰/۴۳٪	۹۱/۴۰٪

برای شبکه Jordan هم مشاهده می‌شود که دقت‌ها به طور کلی افزایش داشته است. به طور مشابه شبکه تشکیل شده از یک مدل Jordan و گروهی از زیرمدل‌های Jordan با تنظیم یکسان را هم می‌توانیم مقایسه کنیم. در این مورد هم افزایش دقت به طور کلی مشهود است.

تعداد نورون مخفی	دقت ارزیابی تک مدل	دقت ارزیابی مدل گروهی
۸	۸۱/۷۲٪	۸۷/۱۰٪
۳۲	۹۴/۶۲٪	۹۲/۴۷٪
۱۲۸	۸۹/۲۵٪	۹۶/۷۷٪

تا اینجا دیدیم که هم شبکه Elman و هم شبکه Jordan در حالت گروهی از حالت تکی بهتر بوده است. حال اگر بخواهیم دو شبکه گروهی Elman و Jordan را با یکدیگر مقایسه کنیم، می‌توان دید که شبکه Elman نه تنها در حالت تک مدله موفق‌تر بوده است، بلکه در حالت گروهی هم نتایج بهتری از حالت گروهی Jordan داشته است.

برپایه نتایجی که تا به اینجا بدست آوردیم شبکه‌های ترکیبی از Elman و Jordan را ایجاد کرده‌ام. سعی شده است به طور کلی ترکیباتی بررسی شود که احتمال می‌رود نتایج بهتری را داشته باشد:

تعداد نورون زیرمدل‌های Elman	تعداد نورون زیرمدل‌های Jordan	دقت مجموعه آموزش	دقت مجموعه اعتبارسنجی	دقت مجموعه ارزیابی
۳۲، ۲۰۴۸	۱۲۸، ۱۲۸، ۱۲۸	۹۵.۹۹٪	۸۹/۱۳٪	۹۰/۳۲٪
۳۲، ۳۲، ۳۲	۳۲، ۳۲	۱۰۰/۰۰٪	۱۰۰/۰۰٪	۹۸/۹۲٪
۲۰۴۸	۳۲	۹۸/۱۵٪	۱۰۰/۰۰٪	۹۳/۵۵٪
۳۲، ۲۰۴۸	۳۲	۹۷/۸۴٪	۹۷/۸۳٪	۹۴/۶۲٪
۱۲۸، ۵۱۲، ۲۰۴۸	۶۴، ۳۲	۹۵/۶۸٪	۹۵/۶۵٪	۹۲/۴۷٪

در این ترکیبات هم دقت‌های خوبی مشاهده می‌شود. به طور خاص شبکه‌ای متشکل از زیرمدل‌های ۳۲ نورونه Elman و Jordan به دقت کامل روی مجموعه آموزش و اعتبارسنجی و دقت نزدیک به ۹۹٪ روی مجموعه ارزیابی رسیده است که دقت بسیار خوبی است.

مجموعه داده متفاوت و تقسیم‌شده

برای این قسمت سوال مجموعه‌داده را به قسمت‌های مساوی تقسیم کردم و به زیرشبکه‌ها سپردم. ابتدا دقت شبکه‌ها با بهترین تنظیم از قسمت قبل را با رویکرد جدید بدست آوردم:

تعداد نورون زیرمدل‌های Elman	تعداد نورون زیرمدل‌های Jordan	دقت مجموعه آموزش	دقت مجموعه اعتبارسنجی	دقت مجموعه ارزیابی
۳۲، ۳۲، ۳۲	۳۲، ۳۲	۸۱.۱۷٪	۶۷/۳۹٪	۸۸/۱۷٪
۲۰۴۸	۳۲	۸۹/۵۱٪	۸۲/۶۱٪	۹۰/۳۲٪
-	۱۲۸، ۱۲۸، ۱۲۸، ۱۲۸، ۱۲۸	۷۹/۶۳٪	۶۷/۳۹٪	۷۴/۱۹٪
۱۲۸، ۳۲	-	۸۴/۲۶٪	۶۷/۳۹٪	۸۷/۱۰٪

ناگفته پیداست که دقت‌های جدید کاهش شدیدی پیدا کرده‌اند. شبکه‌هایی که بعضاً دقت کامل داشته‌اند اینجا نهایتاً به دقت ۹۰٪ رسیده‌اند. بد بودن این نتایج وقتی واضح‌تر می‌شود که بدانیم مجموعه داده سوال نامتوازن است و دقت ۷۰ تا ۸۰ درصد یک دقت بدیهی است! علت کاهش دقت به نظر من به اندازه مجموعه داده بر می‌گردد. برای مسئله فعلی داده‌ها موجود واقعاً کم است. طبیعی است که تقسیم این داده‌ها به مثلاً پنج قسمت باعث می‌شود تا یک مدل شبکه عصبی نتواند به خوبی آموزش ببیند. در عین حال تقسیم داده‌ها می‌تواند باعث کاهش بیش‌برازش در حالت کلی شود ولی در این مسئله و با تدابیر اندیشیده شده مشکل بیش‌برازش چندان بروز پیدا نکرد و این حسن کاربرد زیادی ندارد.

نهایتاً برای پاسخگویی به پرسش مطرح شده مبنی بر تاثیر افزایش تعداد زیرشبکه‌ها آزمایشی دیگر را انجام دادم که نتایج آن در جدول زیر آورده شده است. برای سادگی و ثابت بودن سایر عوامل تمام زیرشبکه‌ها از نوع Elman با ۳۲ نورون در نظر گرفته شده است. (اولین سطر از سوال یک آورده شده است و منطقاً یک زیرشبکه با تمام داده‌ها است):

تعداد زیرشبکه Elman با ۳۲ نورون	دقت مجموعه آموزش	دقت مجموعه اعتبارسنجی	دقت مجموعه ارزیابی
۱	۹۷/۸۴٪	۹۷/۸۳٪	۹۵/۷۰٪
۲	۸۶/۴۲٪	۷۸/۲۶٪	۸۸/۱۷٪
۴	۸۰/۸۶٪	۶۷/۳۹٪	۷۹/۵۷٪
۸	۷۸/۷۰٪	۶۷/۳۹٪	۸۶/۰۲٪
۱۶	۷۹/۳۲٪	۷۱/۷۴٪	۸۰/۵۶٪

مطابق انتظار و بر اساس دلایل ارائه شده به طور کلی افزایش تعداد زیرمدل و کوچک کردن مجموعه داده باعث کاهش دقت می‌شود.