

# تکلیف دوم درس شبکه های عصبی

وحید ملکی، پوریا دادستان

۱۴۰۴ آذر ۱۷

## ۱ مقدمه و تعریف مسئله

در این بخش از تمرین، یک معماری نوین برای شبکه عصبی پیشنهاد شده است که همزمان سه ویژگی پیشرفته را دارد:

۱. انعطاف‌پذیری (**Flexibility**): استفاده از توابع فعال‌ساز ترکیبی با پارامترهای یادگیرنده.

۲. رفتار عاطفی (**Emotional Behavior**): استفاده از تاریخچه خطای برای تعديل فرآیند بهروزرسانی وزن‌ها.

۳. آموزش پیشرفته: مقایسه الگوریتم Adam (مرتبه اول) و Levenberg-Marquardt (مرتبه دوم).

هدف نهایی، بررسی کارایی این ساختار بر روی دو دسته مسئله رگرسیون (سری زمانی) و طبقه‌بندی (داده‌های آماری) است.

## ۲ مدل ریاضی و روابط حاکم

### ۱۰۲ توابع فعال‌ساز منعطف ( $F_1$ و $F_2$ )

برخلاف شبکه‌های کلاسیک، در این مدل هر نورون دارای تابع فعال‌سازی است که ترکیبی خطی از چهار تابع پایه می‌باشد. پارامترهای شکل دهنده و ضرایب ترکیب، همگی در حین آموزش بهینه می‌شوند. تابع فعال‌ساز نوع اول ( $F_1$ ): کامل ترین فرم (شامل بخش خطی):

$$F_1(x) = \alpha \cdot \sigma(\theta_1 x) + \beta \cdot \tanh(\theta_2 x) + \gamma \cdot \text{ReLU}(\theta_3 x) + \delta \cdot (\theta_4 x) \quad (1)$$

تابع فعال‌ساز نوع دوم ( $F_2$ ): فرم فشرده (بدون بخش خطی):

$$F_2(x) = \alpha \cdot \sigma(\theta_1 x) + \beta \cdot \tanh(\theta_2 x) + \gamma \cdot \text{ReLU}(\theta_3 x) \quad (2)$$

در اینجا  $\sigma(z) = \frac{1}{1+e^{-z}}$  است. مجموعه پارامترهای آموزش‌پذیر برای هر نورون عبارتند از:

### ۲۰۲ تزریق منطق عاطفی (**Emotional Logic**)

در شبکه‌های عصبی استاندارد، هدف کمینه‌سازی خطای لحظه‌ای  $e(k)$  است. اما در شبکه عاطفی پیشنهادی، یک سیگال عاطفی  $r(k)$  تعریف می‌شود که به تغییرات خطای نیز حساس است:

$$e(k) = y_{target}(k) - y_{pred}(k) \quad (3)$$

$$\Delta e(k) = e(k) - e(k-1) \quad (4)$$

$$r(k) = k_1 e(k) + k_2 \Delta e(k) \quad (5)$$

ضرایب  $k_1$  و  $k_2$  میزان اهمیت خطای لحظه‌ای و روند تغییرات خطای را تعیین می‌کنند (در این پیاده‌سازی  $k_1 = 0.7$ ,  $k_2 = 0.3$ ).

## ۳۰۲ روابط پیشرو (Forward Propagation)

در این شبکه، محاسبات لایه به لایه انجام می‌شود. فرض کنیم  $A^{(l-1)} = X$  بردار ورودی به لایه  $l$  ام باشد (که است). خروجی خالص ( $A$ ) و خروجی فعال شده ( $net$ ) برای هر نورون  $j$  در لایه  $l$  به صورت زیر محاسبه می‌شود:

$$net_j^{(l)} = \sum_i w_{ji}^{(l)} A_i^{(l-1)} + b_j^{(l)} \quad (6)$$

سپس تابع فعال‌ساز ترکیبی  $F_1$  (یا  $F_2$ ) روی  $net$  اعمال می‌شود. رابطه کلی برای کامل ترین حالت ( $F_1$ ) عبارت است از:

$$\begin{aligned} A_j^{(l)} = F(net_j^{(l)}, \Omega_j^{(l)}) &= \alpha_j^{(l)} \sigma(\theta_{1,j}^{(l)} net_j^{(l)}) \\ &\quad + \beta_j^{(l)} \tanh(\theta_{2,j}^{(l)} net_j^{(l)}) \\ &\quad + \gamma_j^{(l)} \text{ReLU}(\theta_{3,j}^{(l)} net_j^{(l)}) \\ &\quad + \delta_j^{(l)} (\theta_{4,j}^{(l)} net_j^{(l)}) \end{aligned} \quad (7)$$

که در آن  $\Omega_j^{(l)}$  مجموعه پارامترهای آموزش‌پذیر تابع فعال‌ساز برای نورون  $j$  در لایه  $l$  است.

## ۴۰۲ روابط پسرو و اصلاح گرادیان (Backpropagation)

هدف فرآیند پسرو، محاسبه گرادیان تابع هزینه نسبت به تمامی پارامترهای شبکه ( $\Theta = \{W, b, \Omega\}$ ) جهت استفاده در الگوریتم LM یا Adam است.

### ۱۰۴۰۲۰ تابع هزینه عاطفی و قاعده زنجیره‌ای اصلاح شده

در شبکه‌های عصبی عاطفی، تابع هزینه بر اساس سیگال عاطفی  $r$  تعریف می‌شود، نه صرفاً خطای خام  $e$ .

$$E(k) = \frac{1}{2} r(k)^2, \quad r(k) = k_1 e(k) + k_2 [e(k) - e(k-1)] \quad (8)$$

برای محاسبه گرادیان نسبت به یک پارامتر دخواه  $\Theta$ ، از قاعده زنجیره‌ای استفاده می‌کنیم:

$$\frac{\partial E}{\partial \Theta} = \frac{\partial E}{\partial r} \cdot \frac{\partial r}{\partial e} \cdot \frac{\partial e}{\partial \Theta} = r(k) \cdot (k_1 + k_2) \cdot \frac{\partial e}{\partial \Theta} \quad (9)$$

از آنجا که  $r(k) = k_1 e(k) + k_2 [e(k) - e(k-1)]$  داریم  $\frac{\partial r}{\partial \Theta} = -\frac{\partial e}{\partial \Theta} = -y_d - y_{net}$ . بنابراین، تمام مشتقات استاندارد شبکه در ضریب  $k_1 + k_2$  ضرب می‌شوند تا رفتار عاطفی را مدل‌سازی کنند.

## ۲۰۴۰۲ مشتقات پارامترهای انعطاف‌پذیر

برای هر نورون، مشتق خروجی فعال‌ساز ( $A$ ) نسبت به پارامترهای شکل دهنده آن ( $\Omega$ ) باید محاسبه شود. این مشتقات ستون‌های ماتریس ژاکوبین را تشکیل می‌دهند.  
 (الف) مشتق نسبت به ضرایب ترکیب ( $\alpha, \beta, \gamma, \delta$ ):

$$\frac{\partial A}{\partial \alpha} = \sigma(\theta_1 net) \quad (10)$$

$$\frac{\partial A}{\partial \beta} = \tanh(\theta_2 net) \quad (11)$$

$$\frac{\partial A}{\partial \gamma} = \text{ReLU}(\theta_3 net) \quad (12)$$

$$\frac{\partial A}{\partial \delta} = \theta_4 net \quad (13)$$

(ب) مشتق نسبت به پارامترهای درونی ( $\theta_1, \theta_2, \theta_3, \theta_4$ ): در اینجا نیاز به مشتق‌گیری از تابع نسبت به آرگومان داخلی داریم:

$$\frac{\partial A}{\partial \theta_1} = \alpha \cdot net \cdot \sigma(\theta_1 net)[1 - \sigma(\theta_1 net)] \quad (14)$$

$$\frac{\partial A}{\partial \theta_2} = \beta \cdot net \cdot [1 - \tanh^2(\theta_2 net)] \quad (15)$$

$$\frac{\partial A}{\partial \theta_3} = \gamma \cdot net \cdot H(\theta_3 net) \quad (\text{Step Heaviside H:}) \quad (16)$$

$$\frac{\partial A}{\partial \theta_4} = \delta \cdot net \quad (17)$$

## ۳۰۴۰۲ حساسیت لایه‌ها (Sensitivities)

برای انتشار خطابه لایه‌های عقب‌تر، مشتق خروجی نسبت به ورودی خالص ( $net$ ) محاسبه می‌شود:

$$F'(net) = \frac{\partial A}{\partial net} = \alpha \theta_1 \sigma' + \beta \theta_2 \operatorname{sech}^2 + \gamma \theta_3 H + \delta \theta_4 \quad (18)$$

اگر  $\delta^{(l)}$  بردار حساسیت لایه  $l$  باشد، رابطه بازنگشتنی عبارت است از:

$$\delta^{(l-1)} = (W^{(l)})^T \delta^{(l)} \odot F'(net^{(l-1)}) \quad (19)$$

که در نهایت گرادیان وزن‌ها برابر  $\frac{\partial E}{\partial W^{(l)}}$  خواهد بود.

## ۳ تحلیل تتابع تجربی

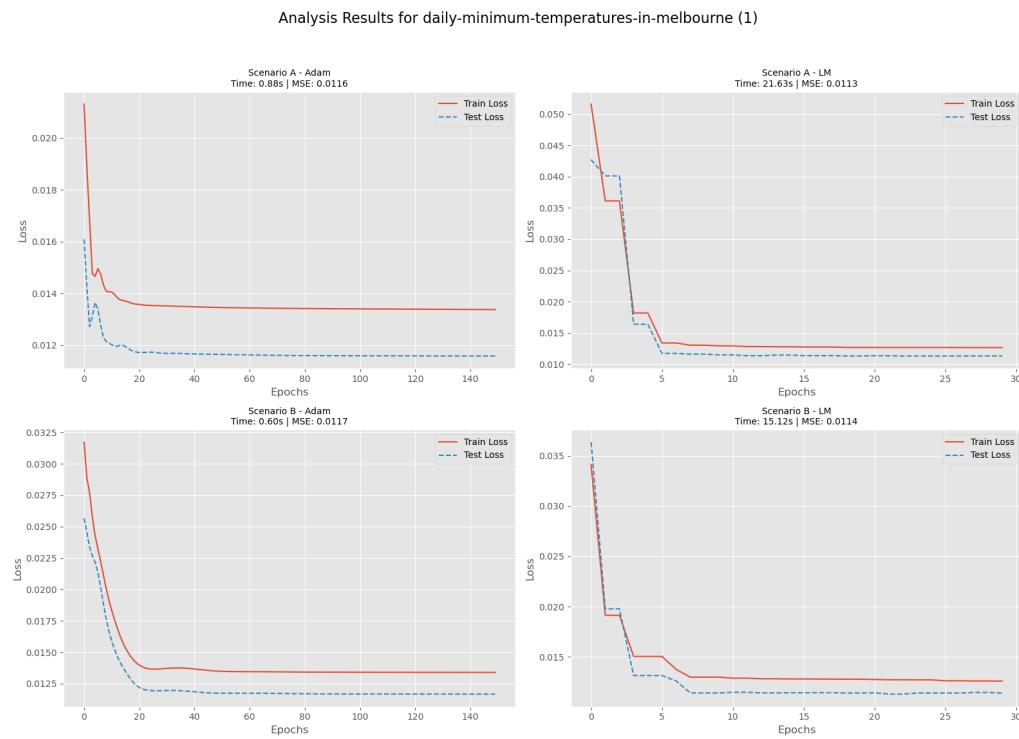
آزمایش‌ها بر روی ۴ دیتاست انجام شد. دو سناریویی معماری بررسی گردید:

• سناریو A: تمام لایه‌ها از تابع کامل  $F_1$ .

• سناریو B: لایه مخفی  $F_2$  و لایه خروجی خطی (Pureline).

## ۱۰۳ تحلیل مسائل رگرسیون (سری زمانی) ۱۰۱۰۳ ۰۱ دیتاست دمای ملبورن (Melbourne Temperature)

هدف: پیش‌بینی دما بر اساس ۵ روز گذشته.



شکل ۱: روند آموزش در دیتاست ملبورن

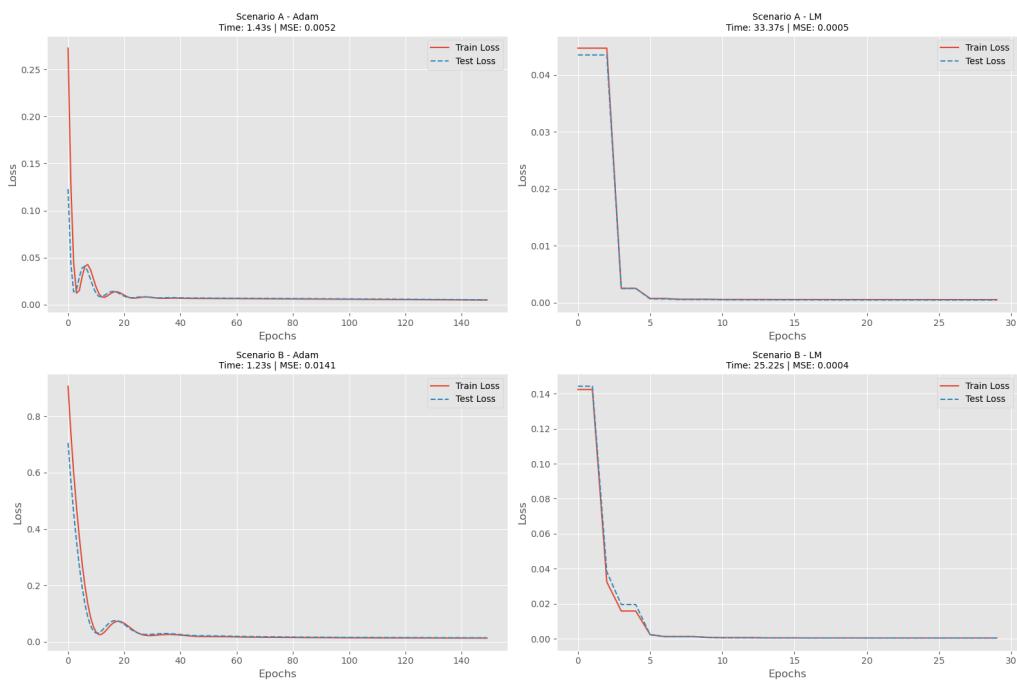
### تحلیل نتایج:

- دقت نهایی: الگوریتم LM در سناریوی A به کمترین خطأ ( $MSE = 0.0113$ ) دست یافت. الگوریتم Adam نیز بسیار نزدیک عمل کرد ( $MSE = 0.0116$ ).
- زمان اجرا: تفاوت اصلی در زمان اجراست. Adam تنها در ۰.۸۸ ثانیه اجرا شد، در حالی که LM حدود ۲۱.۶ ثانیه زمان برد.
- نتیجه‌گیری: برای دیتاست‌های ساده و هموار مانند دما، استفاده از Adam به صرفه‌تر است زیرا با دقت مشابه، ۲۰ برابر سریع‌تر است.

## ۲۰۱۰۳ ۰۲ دیتاست نظارت تصویری (Video Surveillance)

هدف: دنبال کردن الگوهای پیکسلی در فریم‌های ویدیو.

#### Analysis Results for Video surveillance Dataset



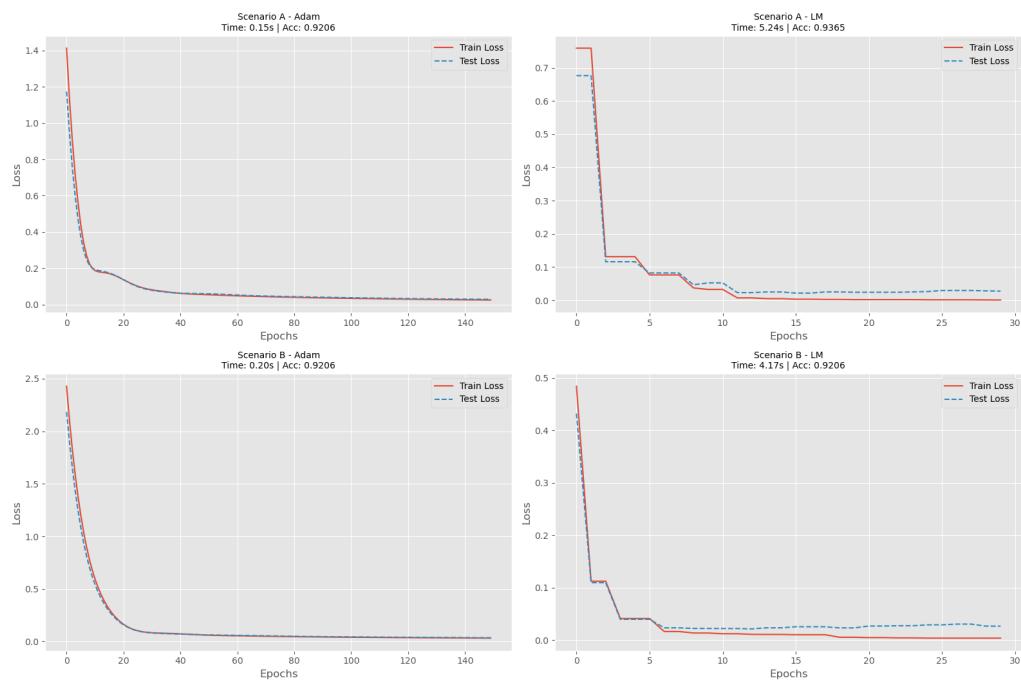
شکل ۲: مقایسه عملکرد LM و Adam در دیتاست ویدیو

#### تحلیل نتایج (نقطه قوت LM):

- برتری قاطع LM: در سناریوی B، الگوریتم LM به خطای خیره‌کننده 0.0004 رسید. این در حالی است که Adam در هرین حالت خطای 0.0052 را ثبت کرد (بیش از ۱۰ برابر اختلاف).
- تحلیل عاطفی-انعطاف‌پذیر: ترکیب خروجی خطی (سناریوی B) با بهینه‌ساز مرتبه دوم، به شبکه اجازه داد تا تغییرات ظریف در سیگال ویدیو را مدل‌سازی کند. مکانیزم عاطفی نیز به پایداری همگرایی در این سطح خطای پایین کمک کرده است.

### ۲۰۳ تحلیل مسائل طبقه‌بندی ۱۰۲۰۳ دیتاست دانه‌ها (Seeds) هدف: تشخیص نوع گندم (۳ کلاس).

### Analysis Results for classification-seeds



**شکل ۳:** دقت دسته‌بندی در دیتاست Seeds

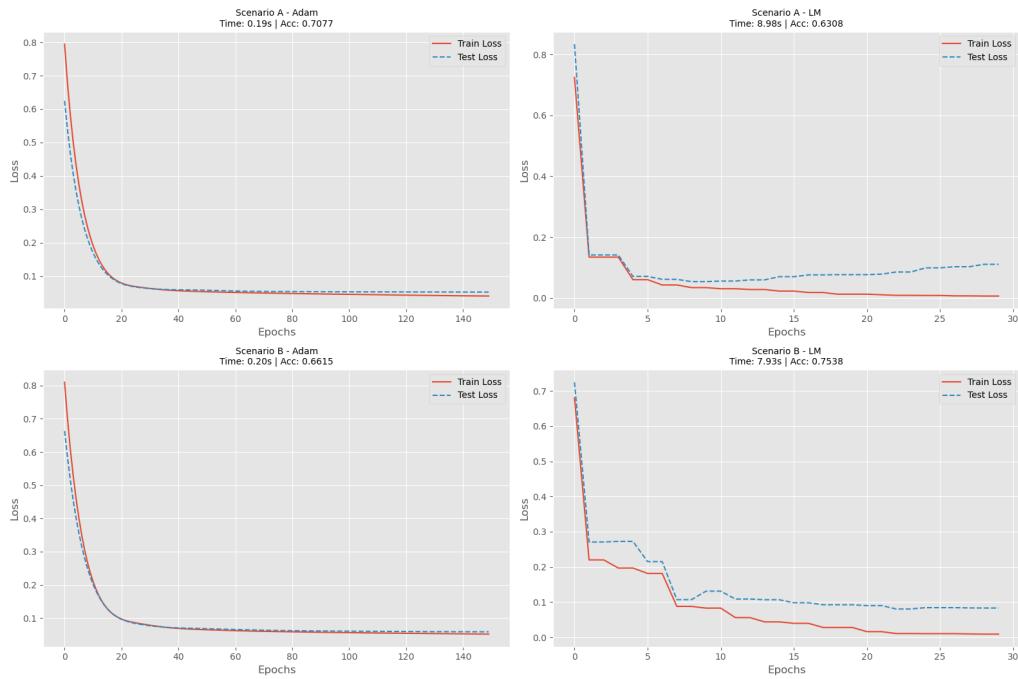
تحلیل:

- بهترین دقت توسط LM در سناریوی A بدست آمد (93.65%).
- الگوریتم Adam در هر دو سناریو روی عدد 92.06% ثابت ماند. این نشان می‌دهد LM توانسته است مرزهای تصمیم‌گیری دقیق‌تری را با تنظیم پارامترهای انعطاف‌پذیر پیدا کند.

### ۲۰۲۰۳ دیتاست شیشه (Glass)

هدف: تشخیص نوع شیشه (دیتاست دشوار و نامتوازن).

### Analysis Results for classification-glass



شکل ۴: تأثیر سناریوی B در بهبود دقت دیتاست Glass

### تحلیل مهم:

- شکست سناریوی A: سناریوی A با الگوریتم LM تنها به دقت 63.08% رسید. دلیل این امر احتمالاً Overfitting یا گیر افتادن در بهینه محلی به دلیل پیچیدگی بیش از حد توابع فعال‌ساز در قمای لایه‌هاست.
- پیروزی سناریوی B: با ساده‌سازی لایه خروجی به Pureline (سناریوی ،B) دقت LM به طور چشمگیری افزایش یافت و به 75.38% رسید که بهترین نتیجه بین تمام مدل‌های است. این نشان می‌دهد برای داده‌های نویزی یا نامتوارن، ترکیب لایه مخفی منعطف با خروجی خطی، تعمیم‌پذیری بهتری دارد.

## ۴ نتیجه‌گیری نهایی

جدول زیر خلاصه بهترین نتایج بدست آمده در این آزمایش را نشان می‌دهد:

جدول ۱: خلاصه عملکرد نهایی مدل پیشنهادی

Score	Metric	Scenario Best	Optimizer Best	Dataset
0.0113	MSE	A	LM	(Reg) Melbourne
0.0004	MSE	B	LM	(Reg) Video
93.65%	Accuracy	A	LM	(Cls) Seeds
75.38%	Accuracy	B	LM	(Cls) Glass

جمع‌بندی: پیاده‌سازی شبکه عاطفی-انعطاف‌پذیر نشان داد که استفاده از سیگکال عاطفی به همراه توابع فعال‌ساز ترکیبی، قدرت یادگیری شبکه را افزایش می‌دهد. اگرچه الگوریتم Adam سریع‌تر است، اما الگوریتم LM (به

خصوص در سناریوی (B) برای دستیابی به دقت‌های بسیار بالا و خطای نزدیک به صفر در مسائل رگرسیون پیچیده،  
بی‌رقیب است.