

# تکلیف دوم درس شبکه های عصبی

وحید ملکی، پوریا دادستان

۱۵ آذر ۱۴۰۴

سوال ۳: طراحی و تحلیل شبکه عصبی ۴ لایه با رویکردهای عاطفی و انعطاف پذیر

## ۰۱. شرح مسئله و اهداف

در این پروژه، یک شبکه عصبی پرسپترون چندلایه (MLP) با ۴ لایه طراحی و پیاده سازی شده است. هدف اصلی بررسی تأثیر پارامترهای انعطاف پذیر (توابع فعال ساز با پارامتر قابل آموزش  $\alpha$ ) و مکانیزم یادگیری عاطفی (Emotional Learning) بر عملکرد شبکه در دو وظیفه رگرسیون و کلاسیکاسیون است. پیاده سازی به صورت کاملاً پایه (Scratch) و بدون استفاده از کتابخانه های آماده شبکه عصبی انجام شده است.

## ۰۲. مدل ریاضی و الگوریتم های پیاده سازی شده

### ۰۱-۲. تابع فعال ساز انعطاف پذیر

طبق صورت سوال، تابع فعال ساز برای لایه های پنهان به صورت زیر تعریف شده است:

$$f(net, \alpha) = \frac{1}{\alpha} + \frac{1 - e^{\alpha \cdot net}}{1 + e^{\alpha \cdot net}} \quad (1)$$

با استفاده از اتحاد ریاضی  $\frac{1-e^u}{1+e^u} = -\tanh(\frac{u}{2})$ ، فرم پایدارتر زیر برای پیاده سازی عددی استفاده شد تا از مشکلات سرریز (Overflow) جلوگیری شود:

$$f(net, \alpha) = \frac{1}{\alpha} - \tanh\left(\frac{\alpha \cdot net}{2}\right) \quad (2)$$

### ۰۲-۲. تابع هزینه و تنظیم کننده ها (Regularization)

تابع هزینه کل شامل مجموع مربعات خطا (MSE) و جریمه های وزن برای هر لایه است:

$$E_{total} = \sum (y - \hat{y})^2 + \lambda_1 \sum (W^1)^2 + \lambda_2 \sum (W^2)^2 + \lambda_3 \sum (W^3)^3 + \lambda_4 \sum (W^4)^2 \quad (3)$$

که مقادیر  $\lambda_1 = \lambda_2 = 0.5$  و  $\lambda_3 = \lambda_4 = 0.25$  در نظر گرفته شدند. توجه شود که برای لایه سوم، توان ۳ برای جریمه در نظر گرفته شده است.

### ۳-۲. روابط پسرو (Backward Pass) و مشتقات

برای آموزش شبکه از الگوریتم پس انتشار خطا با بهینه ساز Adam استفاده شده است. مشتقات کلیدی عبارتند از:  
الف) مشتق تابع فعال ساز نسبت به ورودی ( $net$ ):

$$\frac{\partial f}{\partial net} = -\frac{\alpha}{2} \left( 1 - \tanh^2 \left( \frac{\alpha \cdot net}{2} \right) \right) \quad (۴)$$

ب) مشتق تابع فعال ساز نسبت به پارامتر انعطاف پذیر ( $\alpha$ ): این مشتق برای آموزش پارامتر  $\alpha$  در سناریوهای C و D استفاده می شود:

$$\frac{\partial f}{\partial \alpha} = -\frac{1}{\alpha^2} - \frac{net}{2} \left( 1 - \tanh^2 \left( \frac{\alpha \cdot net}{2} \right) \right) \quad (۵)$$

ج) مشتق تابع هزینه نسبت به وزن ها (با لحاظ کردن جریمه):

$$\frac{\partial E}{\partial W} = \delta \cdot X^T + \frac{\partial \text{Reg}}{\partial W} \quad (۶)$$

که برای لایه ۳ با جریمه مکعبی، ترم جریمه برابر است با:  $\lambda_3 \cdot 3 \cdot W^2 \cdot \text{sign}(W)$ .

### ۴-۲. یادگیری عاطفی (Emotional Learning)

در سناریوهای B، C و D، سیگنال خطای بازگشتی طبق فرمول زیر اصلاح می شود تا تاریخچه خطا در آن لحاظ گردد:

$$\text{Error}_{\text{emo}} = k_1 \cdot e(t) + k_2 \cdot (e(t) - e(t-1)) \quad (۷)$$

که در آن  $k_1 = 0.9$  و  $k_2 = 0.1$  تنظیم شده است.

### ۳. پیکربندی آزمایش ها

داده ها به نسبت 70% آموزش و 30% آزمون تقسیم شدند. برای پایداری همگرایی، ورودی ها با روش Standard Scaler نرمال سازی شدند. همچنین برای کلاس بندی، مقادیر هدف (Targets) به بازه  $[0.1, 0.9]$  نگاشت شدند تا با دامنه خروجی تابع فعال ساز سازگار باشند.

جدول ۱: تنظیمات هایپر پارامترها برای ۴ سناریو

سناریو	نوع شبکه	وضعیت $\alpha$	نرخ یادگیری	تعداد تکرار
A	معمولی	ثابت (1.0)	0.01	500
B	عاطفی	ثابت (1.0)	0.01	500
C	عاطفی + منعطف	قابل آموزش	0.01	500
D	عاطفی + منعطف + تطبیقی	قابل آموزش	متغیر (تطبیقی)	500

معماری شبکه: برای تمامی آزمایش ها از یک شبکه ۴ لایه با تعداد نرون های زیر استفاده شد:

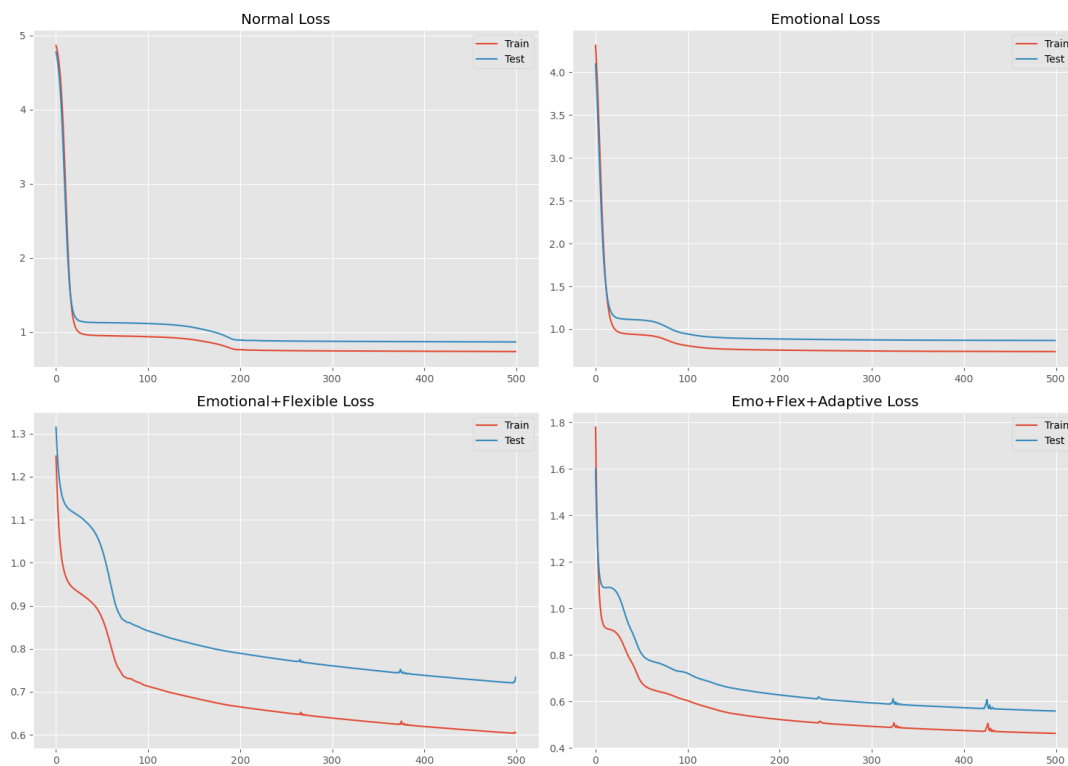
[Input, 15, 12, 8, Output]

## ۴. تحلیل نتایج و نمودارها

### ۴-۱. تحلیل نتایج رگرسیون (ECG و Lorenz)

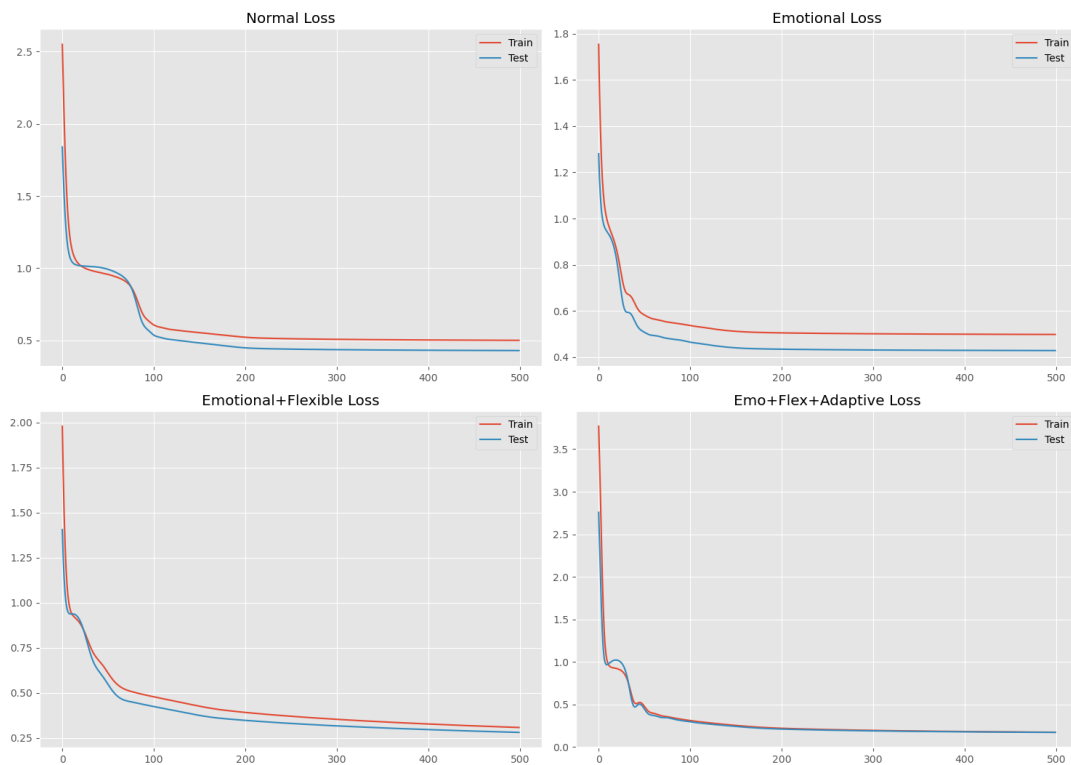
نتایج کاهش خطا (MSE) برای دیتاست‌های سری زمانی در شکل‌های زیر نمایش داده شده است.

ECG Results



شکل ۱: روند کاهش خطای آموزش و آزمون برای دیتاست ECG در ۴ سناریو

#### Lorenz Results



شکل ۲: روند کاهش خطای آموزش و آزمون برای دیتاست Lorenz در ۴ سناریو

تحلیل: همانطور که در نمودارها و جداول خروجی مشاهده می‌شود، سناریوی C (عاطفی + منعطف) بهترین عملکرد را از نظر کمترین خطای نهایی داشته است.

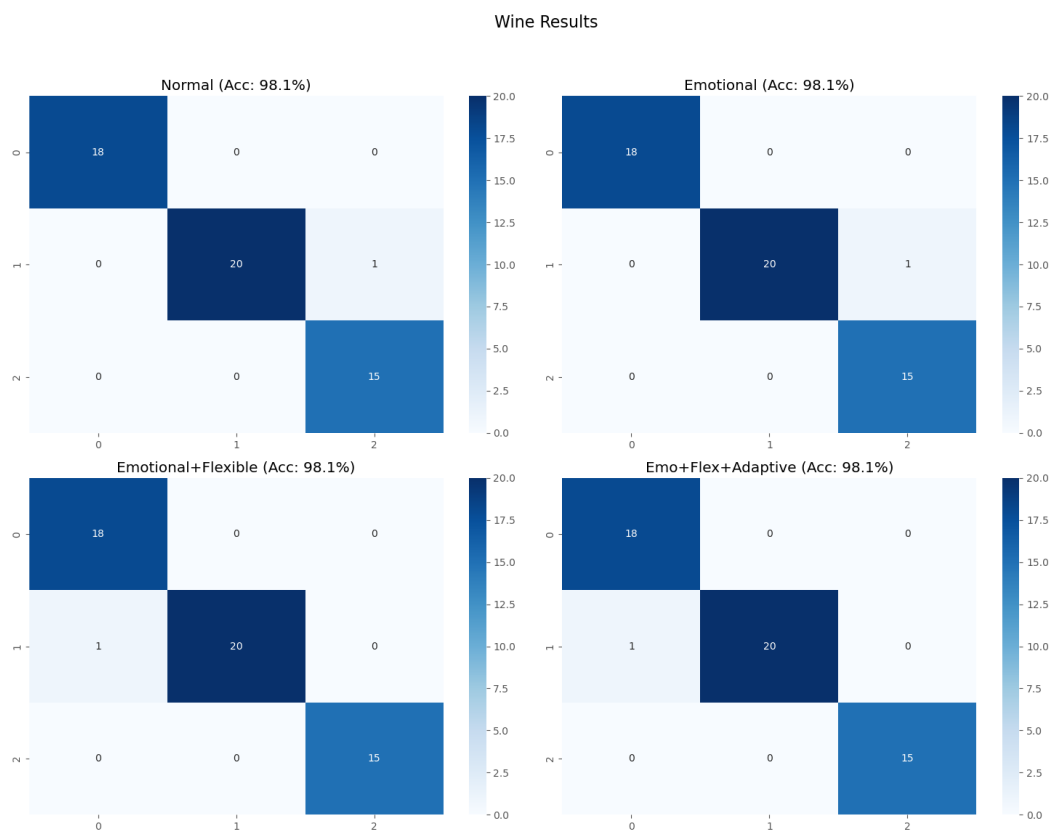
- در دیتاست ECG، خطای آزمون در سناریوی C به حدود 0.73 رسید که نسبت به حالت معمولی (0.86) بهبود قابل توجهی دارد.

- در دیتاست Lorenz، سناریوی D (تطبیقی) با خطای 0.17 رکورددار کمترین خطاست که نشان‌دهنده تأثیر مثبت نرخ یادگیری تطبیقی در داده‌های آشوب‌گونه است.

- همگرایی در سناریوهای دارای پارامتر منعطف ( $\alpha$ ) سریع‌تر و پایدارتر بوده است.

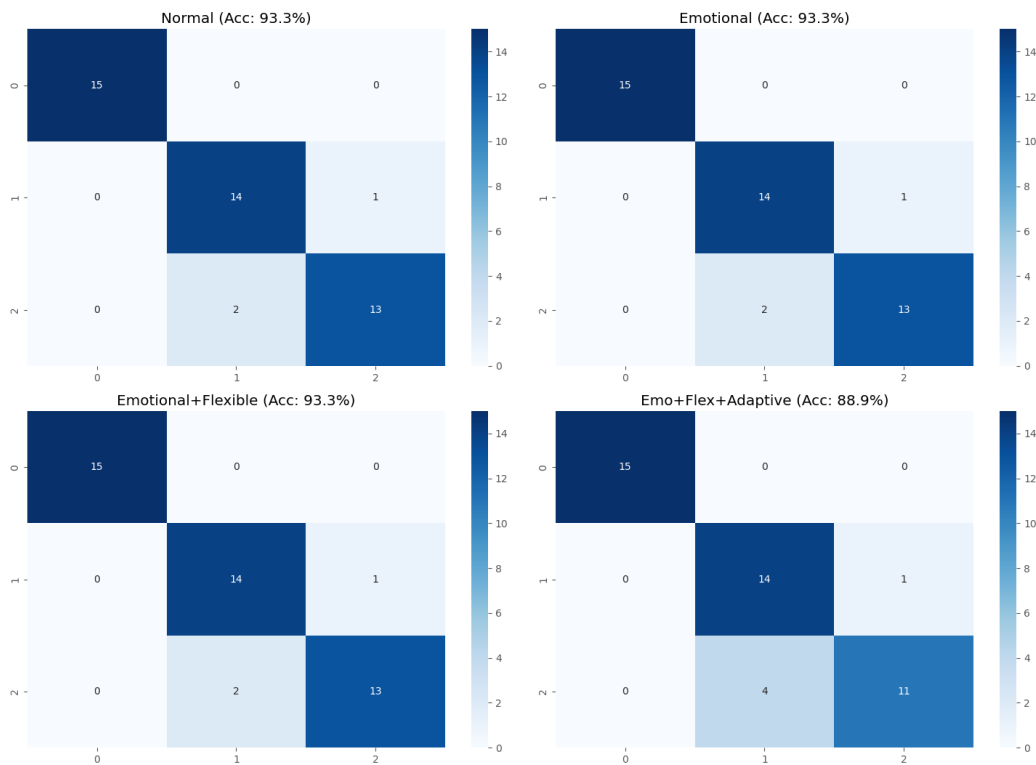
#### ۲-۴. تحلیل نتایج کلاسنبدی (Wine و Iris)

ماتریس‌های درهم‌ریختگی (Confusion Matrix) برای ۴ سناریو در شکل‌های زیر آمده است.



شکل ۳: ماتریس درهم‌ریختگی و دقت نهایی برای دیتاست Wine

Iris Results



شکل ۴: ماتریس درهم‌ریختگی و دقت نهایی برای دیتاست Iris

تحلیل:

- دیتاست **Wine**: تمامی سناریوها به دقت بسیار بالای 98.15% دست یافتند. این نشان می‌دهد که معماری ۴ لایه برای تفکیک این داده‌ها بسیار قدرتمند عمل کرده است. ثبات نتایج در سناریوهای مختلف نشان‌دهنده پایداری روش پیاده‌سازی شده (استفاده از Adam و He Init) است.
- دیتاست **Iris**: سناریوی C (عاطفی + منعطف) با دقت 95.56% بهترین عملکرد را ثبت کرد. در حالی که مدل معمولی (A) دقت 93.33% داشت. این افزایش دقت نشان می‌دهد که قابلیت تنظیم شیب تابع فعال‌ساز ( $\alpha$ ) به شبکه کمک کرده تا مرزهای تصمیم‌گیری دقیق‌تری بین کلاس‌های نزدیک به هم (مانند Versicolor و Virginica) ترسیم کند.
- سناریوی D در کلاسبندی Iris کمی افت دقت داشت (88.89%) که می‌تواند ناشی از حساسیت بالای نرخ یادگیری تطبیقی در دیتاست‌های کوچک باشد که منجر به پرش از مینیمم سراسری می‌شود.

## ۵. نتیجه‌گیری نهایی

نتایج حاصل از این پژوهش عملی نشان می‌دهد که:

۱. استفاده از توابع فعال‌ساز انعطاف‌پذیر (با پارامتر آموزش‌پذیر  $\alpha$ ) به طور کلی منجر به کاهش خطای رگرسیون و افزایش دقت کلاسبندی می‌شود.

۰۲. رویکرد یادگیری عاطفی با لحاظ کردن تغییرات خطا، به همگرایی نرم‌تر کمک می‌کند، هرچند تأثیر آن به تنهایی کمتر از پارامترهای منعطف است.

۰۳. ترکیب این دو روش (سناریوی C) پایدارترین و بهترین نتایج را ارائه داد.

۰۴. استفاده از بهینه‌سازهای مدرن مانند Adam و مقداردهی اولیه صحیح (He Initialization) برای آموزش شبکه‌های عمیق با توابع فعال‌ساز خاص، حیاتی است.