

تکلیف دوم درس شبکه های عصبی

وحید ملکی، پوریا دادستان

۱۴۰۴ آذر ۱۵

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

۱. شرح مسئله و اهداف

در این پروژه، یک شبکه عصبی پرسپترون چندلایه (MLP) با ۴ لایه طراحی و پیاده‌سازی شده است. هدف اصلی بررسی تأثیر پارامترهای انعطاف‌پذیر (تابع فعال‌ساز با پارامتر قابل آموزش α) و مکانیزم یادگیری عاطفی (Emotional Learning) بر عملکرد شبکه در دو وظیفه رگرسیون و کلاسیبندی است. پیاده‌سازی به صورت کاملاً پایه (Scratch) و بدون استفاده از کتابخانه‌های آماده شبکه عصبی انجام شده است.

۲. مدل ریاضی و الگوریتم‌های پیاده‌سازی شده

۱-۱. تابع فعال‌ساز انعطاف‌پذیر

طبق صورت سوال، تابع فعال‌ساز برای لایه‌های پنهان به صورت زیر تعریف شده است:

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

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

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

۱-۲. تابع هزینه و تنظیم کننده‌ها (Regularization)

تابع هزینه کل شامل مجموع مربعات خطأ (MSE) و جریمه‌های وزن برای هر لایه است:

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

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.

۰.۳-۲ روابط پسرو (Backward Pass) و مشتقات

برای آموزش شبکه از الگوریتم پسانشار خطا با بهینه‌ساز Adam استفاده شده است. مشتقات کلیدی عبارتند از:
 الف) مشتق تابع فعال‌ساز نسبت به ورودی (net):

$$\frac{\partial f}{\partial \text{net}} = -\frac{\alpha}{2} \left(1 - \tanh^2 \left(\frac{\alpha \cdot \text{net}}{2} \right) \right) \quad (4)$$

ب) مشتق تابع فعال‌ساز نسبت به پارامتر انعطاف‌پذیر (α): این مشتق برای آموزش پارامتر α در سناریوهای C و D استفاده می‌شود:

$$\frac{\partial f}{\partial \alpha} = -\frac{1}{\alpha^2} - \frac{\text{net}}{2} \left(1 - \tanh^2 \left(\frac{\alpha \cdot \text{net}}{2} \right) \right) \quad (5)$$

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

$$\frac{\partial E}{\partial W} = \delta \cdot X^T + \frac{\partial \text{Reg}}{\partial W} \quad (6)$$

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

۰.۴-۲ یادگیری عاطفی (Emotional Learning)

در سناریوهای C، B و D، سیگال خطای بارگشته طبق فرمول زیر اصلاح می‌شود تا تاریخچه خطا در آن لحاظ گردد:

$$\text{Error}_{\text{emo}} = k_1 \cdot e(t) + k_2 \cdot (e(t) - e(t-1)) \quad (7)$$

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

۰.۳ پیکربندی آزمایش‌ها

داده‌ها به نسبت 70% آموزش و 30% آزمون تقسیم شدند. برای پایداری همگرایی، ورودی‌ها با روش Standard Scaler نرمال‌سازی شدند. همچنین برای کلاس‌بندی، مقادیر هدف (Targets) به بازه [0.1, 0.9] نگاشت شدند تا با دامنه خروجی تابع فعال‌ساز سازگار باشند.

جدول ۱: تنظیمات های پارامترها برای ۴ سناریو

سناریو	نوع شبکه	وضعیت α	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

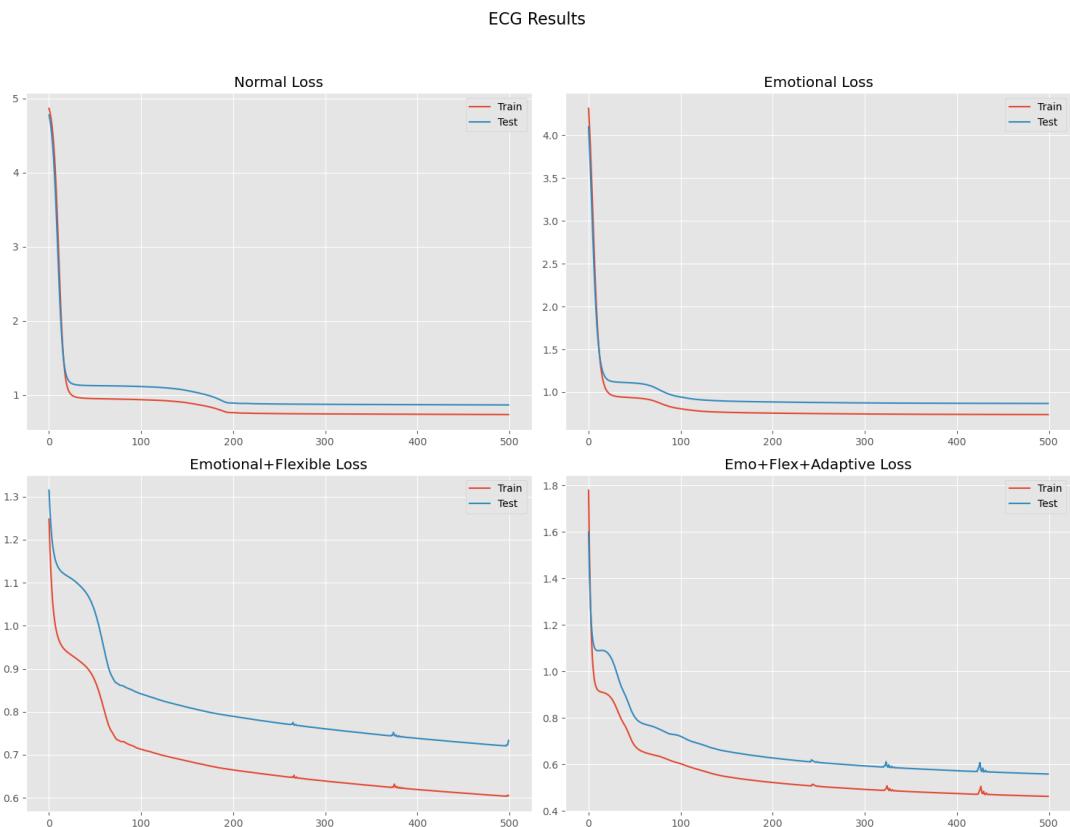
معماری شبکه: برای تمامی آزمایش‌ها از یک شبکه ۴ لایه با تعداد نورون‌های زیر استفاده شد:

[Input, 15, 12, 8, Output]

۴. تحلیل نتایج و نمودارها

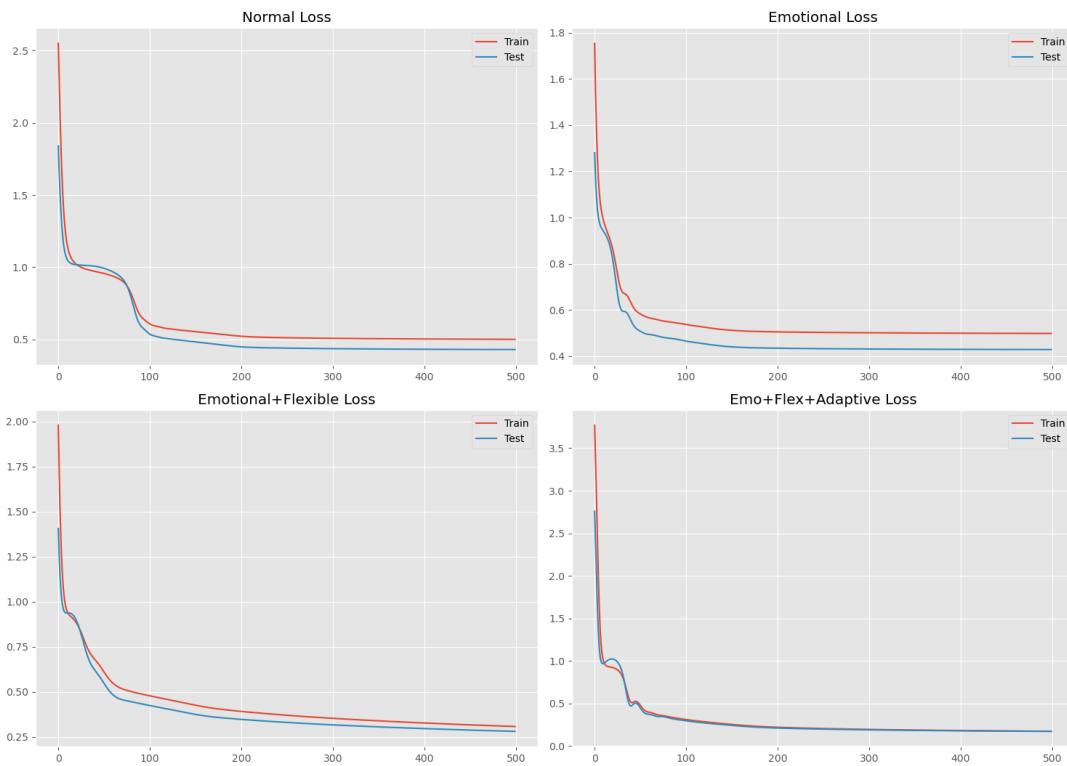
۱-۴. تحلیل نتایج رگرسیون (ECG و Lorenz)

نتایج کاهش خطای (MSE) برای دیتاست های سری زمانی در شکل های زیر نمایش داده شده است.



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

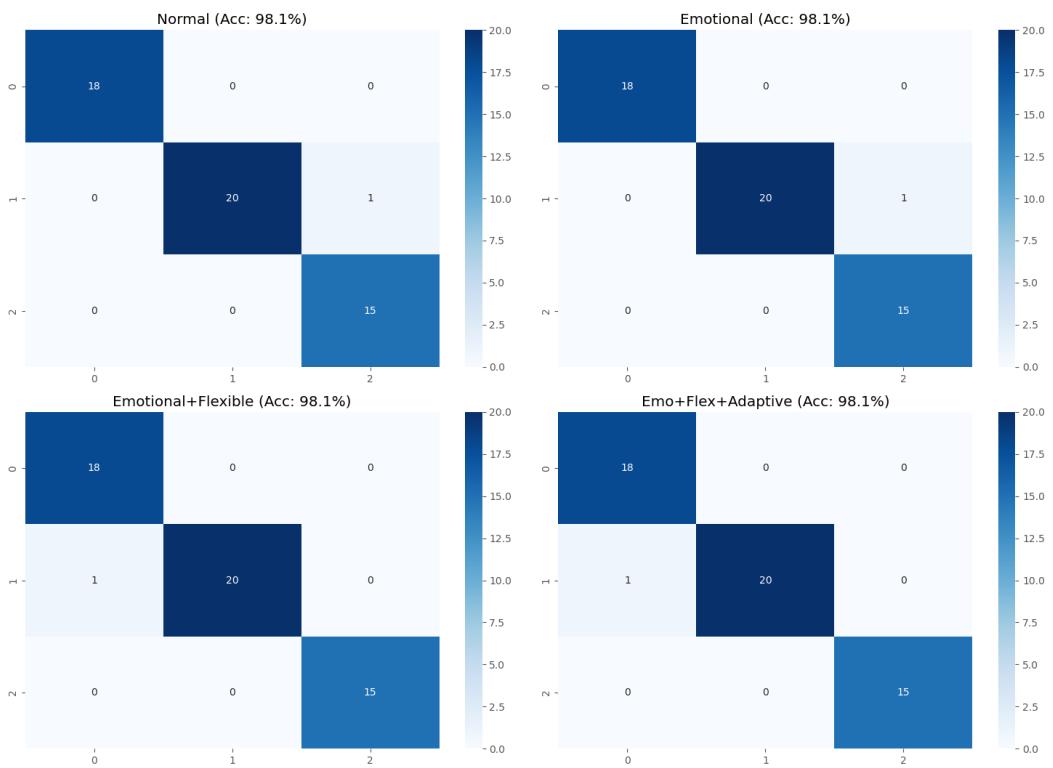
تحلیل: همانطور که در غودارها و جداول خروجی مشاهده می شود، سناریوی C (عاطفی + منعطف) بهترین عملکرد را از نظر کمترین خطای نهایی داشته است.

- 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 (تطیقی) با خطای 0.17 رکورددار کمترین خطاست که نشان دهنده تأثیر مثبت نخ یادگیری تطیقی در داده های آشوب گونه است.
- همگایی در سناریوهای دارای پارامتر منعطف (α) سریع تر و پایدار تر بوده است.

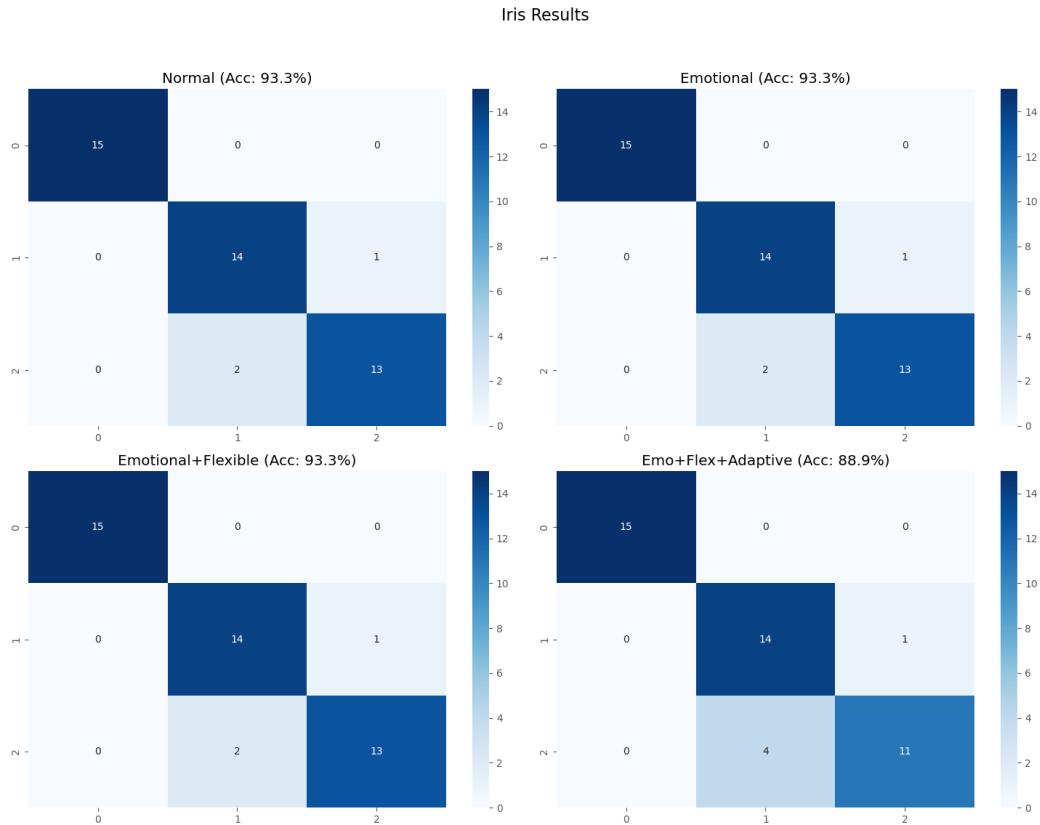
۲-۴. تحلیل نتایج کلاسندی (Wine و Iris)

ماتریس های درهم ریختگی (Confusion Matrix) برای ۴ سناریو در شکل های زیر آمده است.

Wine Results



شکل ۳: ماتریس درهم ریختگی و دقت نهایی برای دیتاست Wine



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 معناری ۴ لایه برای تفکیک این داده‌ها بسیار قدرتمند عمل کرده است. ثبات نتایج در سناریوهای مختلف نشان‌دهنده پایداری روش پیاده‌سازی شده (استفاده از He Init و Adam) 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 شبیه تابع فعال‌ساز (α) به شبکه کمک کرده تا مرزهای تصمیم گیری 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 α) به طور کلی منجر به کاهش خطای رگرسیون و افزایش دقت کلاس‌بندی می‌شود.

- ۰.۲ رویکرد یادگیری عاطفی با لحاظ کردن تغییرات خطای نرم‌تر کمک می‌کند، هرچند تأثیر آن به تنایی کمتر از پارامترهای منعطف است.
- ۰.۳ ترکیب این دو روش (سناریوی C) پایدارترین و بهترین نتایج را ارائه داد.
- ۰.۴ استفاده از بهینه‌سازهای مدرن مانند Adam و مقداردهی اولیه صحیح (He Initialization) برای آموختش شبکه‌های عمیق با توابع فعال‌ساز خاص، حیاتی است.