

تکلیف اول درس شبکه های عصبی

وحید ملکی
شماره دانشجویی: ۴۰۳۱۳۰۰۴

۲۶ آبان ۱۴۰۴

سوال ۳

الف: رسم توابع فعال سازی و مشتق آنها

با استفاده از کتابخانه matplotlib، توابع فعال سازی زیر به همراه مشتق آنها در بازه $[-5, 5]$ رسم شده اند:

ReLU •

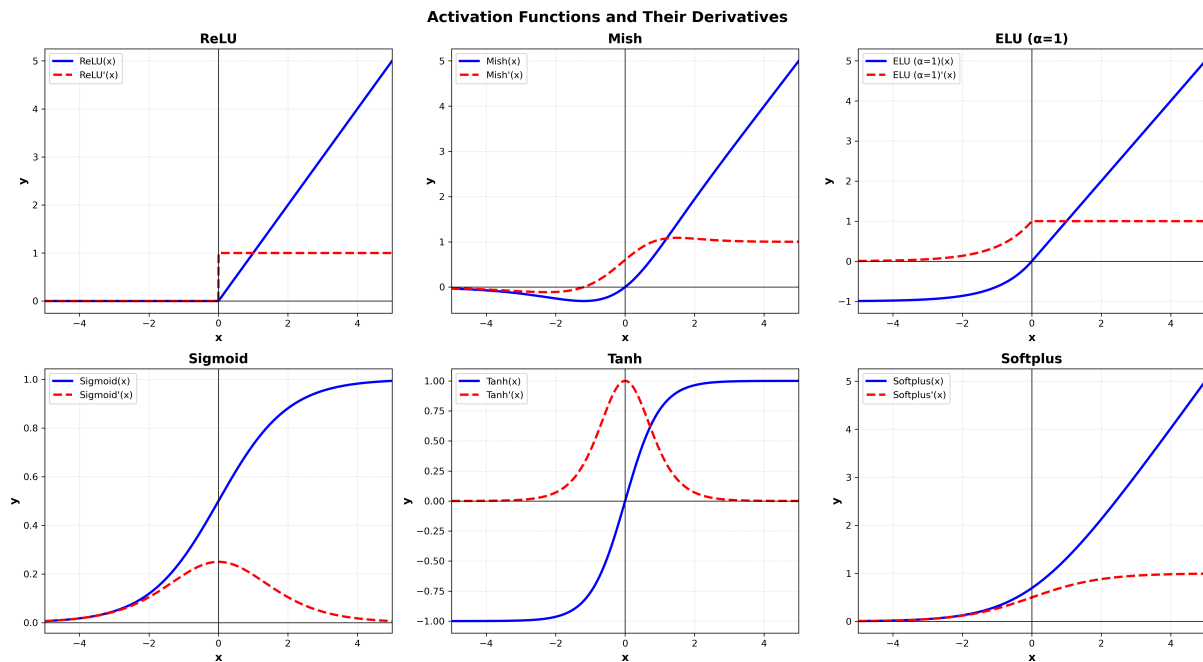
Mish •

ELU ($\alpha = 1$) •

Sigmoid •

Tanh •

Softplus •



شکل ۱: توابع فعال‌سازی (خط آبی پررنگ) و مشتق آن‌ها (خط قرمز نقطه‌چین) برای ReLU، Mish، ELU، Sigmoid، Tanh و Softplus.

مشاهدات از شکل ۱

- **ReLU:** $f(x) = \max(0, x)$ برای $x > 0$ خطی با شیب ۱، برای $x \leq 0$ صفر. مشتق: ۱ در ناحیه مثبت، ۰ در ناحیه منفی.
- **Mish:** $f(x) = x \cdot \tanh(\ln(1 + e^x))$ نرم‌تر از ReLU، در ناحیه منفی کمی منفی می‌شود. مشتق همیشه مثبت و نرم است.
- **ELU:** برای $x > 0$ همانند ReLU، برای $x \leq 0$ به صورت نمایی به α نزدیک می‌شود. مشتق در ناحیه منفی αe^x از مرگ نورون جلوگیری می‌کند.
- **Sigmoid:** $f(x) = \frac{1}{1 + e^{-x}}$ خروجی در $[0, 1]$ ، مشتق حداکثر ۰.۲۵ در $x = 0$ و در $|x|$ بزرگ به صفر نزدیک می‌شود (اشباع).
- **Tanh:** $f(x) = \tanh(x)$ خروجی در $[-1, 1]$ ، متقارن حول صفر، مشتق حداکثر ۱ در $x = 0$.
- **Softplus:** $f(x) = \ln(1 + e^x)$ نسخه نرم ReLU. مشتق آن دقیقاً تابع سیگموئید است.

ب: مشکل اشباع شدن (Vanishing Gradient) در تابع Sigmoid

وقتی $|x|$ بزرگ باشد، خروجی سیگموئید به ۱ یا ۰ نزدیک شده و مشتق آن تقریباً صفر می‌شود:

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \leq 0.25$$

در شبکه‌های عمیق، گرادیان در هر لایه در مشتق تابع فعال‌سازی ضرب می‌شود. اگر چندین لایه مشتق نزدیک به صفر داشته باشند، گرادیان در لایه‌های اولیه ناپدید (vanish) می‌شود و وزن‌ها به‌روزرسانی نمی‌شوند. آیا جایگزینی با Tanh این مشکل را کاملاً برطرف می‌کند؟

خیر، اما به طور قابل توجهی کاهش می‌یابد.
مشتق tanh:

$$\tanh'(x) = 1 - \tanh^2(x) \leq 1$$

حداکثر مشتق آن ۱ است (۴ برابر سیگموئید). در نواحی اشباع همچنان به صفر میل می‌کند، پس مشکل کاملاً حذف نمی‌شود، اما در عمل گرادیان‌ها کمتر ناپدید می‌شوند.

ج: چرا Tanh بهینه‌سازی را نسبت به Sigmoid آسان‌تر می‌کند؟

۱. مشتق بزرگ‌تر: حداکثر مشتق tanh برابر ۱ است (در مقابل ۲۵۰۰ برای سیگموئید) □ گرادیان‌های بزرگ‌تر □ همگرایی سریع‌تر.

۲. خروجی متمرکز حول صفر: (Zero-Centered) خروجی tanh در $[-1, 1]$ و میانگین نزدیک صفر است، در حالی که سیگموئید در $[0, 1]$ و میانگین ≈ 0.5 . خروجی همیشه مثبت در سیگموئید باعث به‌روزرسانی زیگزاگی وزن‌ها می‌شود، اما tanh اجازه جهت‌های مختلف گرادیان را می‌دهد □ بهینه‌سازی روان‌تر.

۳. تقارن حول مبدا: tanh تابع فرد است $\tanh(-x) = -\tanh(x)$ □ یادگیری الگوهای متقارن آسان‌تر.

جدول ۱: مقایسه تابع فعال‌سازی Sigmoid و Tanh

Tanh	Sigmoid	ویژگی
$[-1, 1]$	$[0, 1]$	بازه خروجی
بله	خیر	متمرکز حول صفر؟
۱	۲۵۰۰	حداکثر مشتق
متوسط	بسیار شدید	شدت Gradient Vanishing
سریع‌تر	کند	سرعت همگرایی

در نتیجه، tanh در دهه ۲۰۱۰ جایگزین اصلی sigmoid در لایه‌های مخفی بود، اما امروزه ReLU و خانواده آن به دلیل عدم اشباع در ناحیه مثبت، استاندارد هستند.