

پرسپترون چند لایه

دسته ای از شبکه‌های عصبی مصنوعی پیشخور است. یک MLP شامل حداقل سه لایه گره است: یک لایه ورودی، یک لایه پنهان و یک لایه خروجی. به جز گره‌های ورودی، هر گره یک نرون است که از یک تابع فعال‌سازی غیر خطی استفاده می‌کند. MLP از تکنیک یادگیری نظارت شده به نام بازپرداخت برای آموزش استفاده می‌کند. لایه‌های متعدد آن و فعال‌سازی غیرخطی آن MLP را از یک پرسپترون خطی متمایز می‌کند. در واقع می‌تواند داده‌هایی را متمایز کند که به صورت خطی قابل تفکیک نیستند.

نظریه

تابع فعال‌سازی:

اگر یک پرسپترون چند لایه، تابع فعال‌سازی خطی در تمام نرون‌ها داشته باشد، در واقع با این تابع خطی ورودی‌های وزن دار هر نرون را ترسیم می‌کند. سپس با استفاده از جبر خطی نشان می‌دهد که هر عددی مربوط به لایه‌ها را می‌توان به یک مدل ورودی - خروجی دو لایه کاهش داد. در MLP، برخی از نرون‌ها از یک تابع فعال غیرخطی استفاده می‌کنند که برای مدل‌سازی فرکانس پتانسیل‌های عمل یا شلیک نرون‌های بیولوژیکی توسعه داده شده است.

دو تابع فعال، که هر دو تابع سیگموئید هستند، توسط عبارت زیر توصیف می‌شوند:

$$y(v_i) = \tanh(v_i) \text{ and } y(v_i) = (1 + e^{-v_i})^{-1}$$

اولین جمله، تابع تانژانت هایپربولیک است که بین منفی یک تا یک تغییر می‌کند و بقیه تابع لجستیک هستند که متغیر بین ۰ تا یک است. y_i خروجی i-امین نرون است و v_i مجموع وزنی اتصالات ورودی است. توابع فعال‌سازی جایگزین، از جمله توابع یکسوساز (شبکه عصبی) پیشنهاد شده است. توابع فعال سازی بیشتر تخصصی شامل توابع پایه شعاعی هستند.

لایه‌ها:

MLP شامل سه یا تعداد بیشتری از لایه‌ها است که از گره‌های غیر خطی فعال کننده هستند. از آنجا که MLP ها به‌طور کامل متصل شده‌اند، هر گره در یک لایه با وزن مشخص w_{ij} در هر نود در لایه بعدی متصل می‌شود.

یادگیری:

یادگیری در شبکه عصبی با تغییر وزن اتصال پس از پردازش هر قطعه از داده‌ها، براساس میزان خطا در خروجی در مقایسه با نتیجه مورد انتظار رخ می‌دهد. این نمونه که از یادگیری با نظارت و از طریق بازگشت به عقب و تعمیم الگوریتم حداقل مربعات در پرسپترون خطی انجام می‌شود. خطای موجود در گره خروجی z را در n -امین نقطه داده به صورت $e_j(n) = d_j(n) - y_j(n)$ نشان می‌دهیم که در آن d مقدار هدف و y مقدار تولید شده توسط پرسپترون می‌باشد. مقادیر گره براساس تصحیحات تنظیم می‌شوند که میزان خطا در کل خروجی را به حداقل می‌رساند و به صورت زیر است:

$$\varepsilon(n) = \frac{1}{2} \sum_j e_j^2(n)$$

با استفاده از گرادیان، تغییر در وزن به صورت زیر است:

$$\Delta w_{ij}(n) = -\eta \frac{\partial \varepsilon(n)}{\partial v_j(n)} y_i(n)$$

که در آن y_i نورون پیشین و η «میزان یادگیری» است که انتخاب شده تا اطمینان حاصل شود که وزن‌ها به سرعت به واکنش بدون نوسان همگرا می‌شوند. مشتق محاسبه شده بستگی به میدان محلی القا شده v_j است که خودش تغییر می‌کند. اثبات این که برای گره خروجی این مشتق می‌تواند ساده شود، آسان است.

$$-\frac{\partial \varepsilon(n)}{\partial v_j(n)} = e_j(n) \Phi'(v_i(n))$$

که در آن Φ' مشتق تابع فعال‌سازی است که در بالا توضیح داده شده و خودش هم تغییر نمی‌کند. تجزیه و تحلیل برای تغییر وزن‌ها به یک‌گره پنهان سخت‌تر است، اما می‌توان نشان داد که مشتق مربوطه به صورت زیر است:

$$-\frac{\partial \varepsilon(n)}{\partial v_j(n)} = \Phi'(v_i(n)) \sum_k -\frac{\partial \varepsilon(n)}{\partial v_k(n)} w_{kj}(n)$$

این بستگی به تغییر در وزن گره‌های k دارد که لایه خروجی را نشان می‌دهند؛ بنابراین برای تغییر وزن‌های لایه پنهان، لایه خروجی با توجه به مشتق تابع فعال‌سازی تغییر می‌کند، و بنابراین این الگوریتم تابعی از تابع فعال‌سازی را نشان می‌دهد.

اصطلاحات

دقت کنید که عبارت «پرسپترون چند لایه» به یک پرسپترون منفرد اشاره نمی‌کند که دارای چندین لایه می‌باشد. بلکه شامل بسیاری از پرسپترون‌ها است که در چند لایه سازماندهی شده‌اند. یک جایگزین در واقع «شبکه پرسپترون چند لایه» است. پروپسترون‌ها یک مورد خاص از نورون‌های مصنوعی هستند که از تابع فعال‌سازی آستانه مانند تابع پله‌ای هویساید استفاده می‌کنند. پروپسترون‌های MLP می‌تواند توابع فعال‌سازی دلخواه را بکار گیرد. یک پروپسترون حقیقی طبقه‌بندی باینری را انجام می‌دهد (یا این یا آن)، یک نورون MLP آزاد است که طبق تابع فعال‌سازی، یا طبقه‌بندی یا رگرسیون، آن را انجام دهد. واژه «پرسپترون چند لایه» بعداً بدون توجه به ماهیت گره‌ها و لایه‌ها بکار گرفته شد که می‌تواند متشکل از نورون‌های مصنوعی تعریف شده باشد. این تفسیر از سست شدن تعریف «پرسپترون» در حالت کلی جلوگیری می‌کند.

برنامه‌ها

MLP ها در پژوهش برای توانایی در حل مسائل مفید هستند، که اغلب اجازه راه‌حل‌های تقریبی برای مشکلات بسیار پیچیده مانند تخمین تناسب اندام را می‌دهد. MLP ها توابع جهانی هستند که توسط قضیه سایینکو نشان داده می‌شود، بنابراین می‌توان از آنها برای ایجاد مدل‌های ریاضی از طریق تحلیل رگرسیون استفاده کرد. از آنجا که طبقه‌بندی نمونه خاصی از رگرسیون است، هنگامی که متغیر پاسخ، قطعی است، MLP الگوریتم طبقه‌بندی کننده خوبی را ایجاد می‌کند. MLP یک راه‌حل یادگیری ماشین عمومی در دهه ۱۹۸۰ بود، که کار آن پیدا کردن کاربرد در زمینه‌های مختلف مانند بازشناسی گفتار، بینایی رایانه ای، و نرم‌افزار ترجمه ماشینی بود اما از آن پس از آن با رقابت قوی تری از ماشین‌های بردار پشتیبان بسیار ساده‌تر (و مرتبط) روبرو شدند.