

یادگیری عمیق چیست؟

فاصله گرفتن از میدان، آغاز نخستین زمستان هوش مصنوعی را رقم زد (که به زمستان هسته‌ای اشاره دارد، زیرا این اتفاق به‌طور کوتاه پس از اوج جنگ سرد رخ داد). این آخرین بار نخواهد بود. در دهه ۱۹۸۰، یک دیدگاه جدید از هوش مصنوعی نمادین، سیستم‌های خبره، شروع به جذب توجه در میان شرکت‌های بزرگ کرد. چند داستان موفقیت اولیه موجی از سرمایه‌گذاری را به راه انداخت، به طوری که شرکت‌ها در سراسر جهان بخش‌های هوش مصنوعی داخلی خود را برای توسعه سیستم‌های خبره شروع کردند. در حدود سال ۱۹۸۵، شرکت‌ها هر سال بیش از یک میلیارد دلار روی این فناوری هزینه می‌کردند؛ اما تا اوایل دهه ۱۹۹۰، این سیستم‌ها ثابت کرد که نگهداری از آن‌ها هزینه‌بر، مقیاس‌پذیری آن‌ها دشوار و دامنه آن‌ها محدود است و به همین دلیل علاقه به آن‌ها کاهش یافت. اینگونه بود که دومین زمستان هوش مصنوعی آغاز شد. ممکن است ما در حال حاضر شاهد چرخه سوم هیاهو و ناامیدی هوش مصنوعی باشیم و هنوز در فاز خوش‌بینی شدید قرار داریم. بهتر است انتظارات خود را برای کوتاهمدت متعادل کنیم و اطمینان حاصل کنیم که افرادی که آشنایی کمتری با جنبه‌های فنی این حوزه دارند، دیدگاه روشنی از آنچه یادگیری عمیق می‌تواند و نمی‌تواند ارائه دهد، داشته باشند.

۱.۱.۸ وعده‌های هوش مصنوعی

اگرچه ممکن است انتظارات کوتاهمدت غیرواقع‌بینانه‌ای از هوش مصنوعی داشته باشیم، اما چشم‌انداز بلندمدت روشن به نظر می‌رسد. ما تازه شروع به استفاده از یادگیری عمیق در بسیاری از مسائل مهم کرده‌ایم که می‌تواند تحول‌آفرین باشد، از تشخیص‌های پزشکی گرفته تا دستیارهای دیجیتال. تحقیقات در حوزه هوش مصنوعی در ده سال گذشته با سرعت شگفت‌آوری پیشرفت کرده است، تا حد زیادی به دلیل سطحی از سرمایه‌گذاری که تاکنون در تاریخ کوتاه هوش مصنوعی بی‌سابقه بوده است. اما تاکنون، بخش نسبتاً کمی از این پیشرفت‌ها به محصولات و فرایندهایی که دنیای ما را شکل می‌دهند راه یافته‌اند. بیشتر یافته‌های پژوهشی در زمینه یادگیری عمیق هنوز به کار گرفته نشده‌اند یا حداقل به تمامی مسائل که می‌توانند در صنایع مختلف حل کنند، اعمال نشده‌اند. پزشک شما هنوز از هوش مصنوعی استفاده نمی‌کند و حسابدار شما نیز همین‌طور. احتمالاً شما هم در زندگی روزمره خود از فناوری‌های هوش مصنوعی خیلی استفاده نمی‌کنید. البته، می‌توانید از گوشی هوشمند خود سوالات ساده‌ای بپرسید و پاسخ‌های معقولی دریافت کنید، می‌توانید توصیه‌های نسبتاً مفیدی برای محصولات در Amazon.com بگیرید، و می‌توانید در Google Photos عبارت «تولد» را جستجو کنید و بلافاصله عکس‌های جشن تولد دخترتان از ماه گذشته را پیدا کنید. این فاصله زیادی با جایی دارد که این فناوری‌ها در گذشته بودند. اما چنین ابزارهایی هنوز تنها لوازم جانبی زندگی روزمره ما هستند. هوش مصنوعی هنوز به مرحله‌ای نرسیده است که به بخش مرکزی نحوه کار، تفکر، و زندگی ما تبدیل شود.

در حال حاضر، ممکن است باور اینکه هوش مصنوعی بتواند تأثیر بزرگی بر دنیای ما داشته باشد دشوار به نظر برسد، زیرا هنوز به طور گسترده به کار گرفته نشده است—همان‌طور که در سال ۱۹۹۵، باور به تأثیر آینده اینترنت سخت بود. در آن زمان، بیشتر مردم نمی‌دیدند که اینترنت چگونه می‌تواند برای آن‌ها مرتبط باشد و چگونه زندگی‌شان را تغییر خواهد داد. همین موضوع امروز در مورد یادگیری عمیق و هوش مصنوعی صدق می‌کند. اما اشتباه نکنید: هوش مصنوعی در راه است. در آینده‌ای نه‌چندان دور، هوش مصنوعی دستیار شما خواهد بود، حتی دوست شما؛ به سوالات شما پاسخ خواهد داد، به آموزش فرزندان کمک خواهد کرد و مراقب سلامت شما خواهد بود. خریدهای شما را در منزلتان تحویل خواهد داد و شما را از نقطه A به نقطه B خواهد رساند. هوش مصنوعی رابط شما به دنیایی خواهد بود که به طور فزاینده‌ای پیچیده و پر از اطلاعات است. و، حتی مهم‌تر، هوش مصنوعی، با کمک به دانشمندان در کشف‌های جدید و انقلابی در تمام زمینه‌های علمی، از ژنومیک تا ریاضیات به کل بشریت کمک خواهد کرد تا به جلو حرکت کند. در این مسیر، ممکن است با چندین چالش مواجه شویم و حتی شاید یک زمستان جدید هوش مصنوعی را تجربه کنیم—به همان شکلی که صنعت اینترنت در سال‌های ۱۹۹۸-۱۹۹۹ بیش از حد تبلیغ شد و سپس با سقوطی مواجه شد که سرمایه‌گذاری‌ها را در اوایل دهه ۲۰۰۰ خشکاند. اما در نهایت به مقصد خواهیم رسید. هوش مصنوعی تقریباً در تمامی فرایندهایی که جامعه و زندگی روزمره ما را شکل می‌دهند به کار گرفته خواهد شد، درست مانند اینترنت امروز. به هیجان‌های کوتاهمدت باور نداشته باشید، اما به چشم‌انداز بلندمدت ایمان داشته باشید. ممکن است مدتی طول بکشد تا هوش مصنوعی به پتانسیل واقعی خود دست یابد—پتانسیلی که هنوز هیچکس جرات نکرده به طور کامل در مورد آن رویاپردازی کند—اما هوش مصنوعی در راه است و دنیای ما را به شکلی فوق‌العاده متحول خواهد کرد.

۱.۲ قبل از یادگیری عمیق: تاریخچه مختصری از یادگیری ماشین

یادگیری عمیق به سطحی از توجه عمومی و سرمایه‌گذاری صنعتی رسیده است که تاکنون در تاریخ هوش مصنوعی بی‌سابقه بوده است، اما این اولین شکل موفق یادگیری ماشین نیست. می‌توان گفت بیشتر الگوریتم‌های یادگیری ماشین که امروزه در صنعت استفاده می‌شوند، الگوریتم‌های یادگیری عمیق نیستند. یادگیری عمیق همیشه ابزار مناسبی برای هر مسئله نیست—گاهی اوقات داده کافی برای به‌کارگیری یادگیری عمیق وجود ندارد و گاهی مسئله با یک الگوریتم دیگر بهتر حل می‌شود. اگر یادگیری عمیق اولین مواجهه شما با یادگیری ماشین باشد، ممکن است خود را در وضعیتی بیابید که تنها ابزارتان چکش یادگیری عمیق باشد و هر مسئله یادگیری ماشینی به نظرتان مانند میخی بیاید. تنها راه برای نیفتادن در این دام، آشنایی با روش‌های دیگر و استفاده از آن‌ها در مواقع مناسب است.

بحث مفصل در مورد رویکردهای کلاسیک یادگیری ماشین خارج از محدوده این کتاب است، اما به‌طور مختصر آن‌ها را مرور می‌کنم و زمینه تاریخی توسعه‌شان را توضیح می‌دهم. این کار به ما کمک می‌کند یادگیری عمیق را در زمینه گسترده‌تر یادگیری ماشین قرار دهیم و بهتر بفهمیم یادگیری عمیق از کجا آمده و چرا اهمیت دارد.

۱.۲.۱ مدل سازی احتمالی

مدل سازی احتمالی به کارگیری اصول آمار در تحلیل داده است. این یکی از اولین اشکال یادگیری ماشین است و تا به امروز نیز به طور گسترده استفاده می شود. یکی از شناخته شده ترین الگوریتم ها در این دسته، الگوریتم «بیز ساده» (Naive Bayes) است.

بیز ساده نوعی الگوریتم طبقه بندی در یادگیری ماشین است که بر اساس اعمال قضیه بیز با فرض مستقل بودن تمام ویژگی ها در داده ورودی عمل می کند (یک فرض قوی یا «ساده لوحانه» که نام آن نیز از همینجا گرفته شده است). این نوع تحلیل داده قبل از اختراع کامپیوترها به صورت دستی انجام می شد و دهه ها پیش از اولین پیاده سازی کامپیوتری آن (احتمالاً به دهه ۱۹۵۰ بازمی گردد) مورد استفاده قرار می گرفت. قضیه بیز و مبانی آمار به قرن هجدهم میلادی بازمی گردند و همین موارد برای شروع استفاده از طبقه بندی بیز ساده کافی است.

مدلی که به بیز ساده نزدیک است، رگرسیون لجستیک (logreg) نام دارد که گاهی اوقات به عنوان «سلام دنیا» در یادگیری ماشین مدرن شناخته می شود. از نام آن گمراه نشوید—رگرسیون لجستیک یک الگوریتم طبقه بندی است و نه یک الگوریتم رگرسیون. مشابه بیز ساده، رگرسیون لجستیک نیز مدت ها پیش از اختراع کامپیوترها توسعه یافته، اما به دلیل سادگی و تطبیق پذیری آن، همچنان مفید است. اغلب اولین چیزی است که یک دانشمند داده روی یک مجموعه داده امتحان می کند تا درک اولیه ای از وظیفه طبقه بندی مورد نظر به دست آورد.

۱.۲.۲ شبکه های عصبی اولیه

نسخه های اولیه شبکه های عصبی به طور کامل توسط نسخه های مدرن که در این صفحات مورد بحث قرار گرفته اند جایگزین شده اند، اما آگاهی از چگونگی پیدایش یادگیری عمیق مفید است. اگرچه ایده های اصلی شبکه های عصبی در قالب های ساده از دهه ۱۹۵۰ مورد بررسی قرار گرفتند، این رویکرد دهه ها طول کشید تا به کار بیفتد. برای مدت طولانی، قطعه گمشده یک روش کارآمد برای آموزش شبکه های عصبی بزرگ بود. این وضعیت در اواسط دهه ۱۹۸۰ تغییر کرد، زمانی که چندین نفر به طور مستقل الگوریتم بازپراکنندگی (Backpropagation) را کشف مجدد کردند—روشی برای آموزش زنجیره ای از عملیات پارامتریک با استفاده از بهینه سازی نزول گرادیان (مفاهیم این روش ها را در ادامه کتاب به طور دقیق تعریف خواهیم کرد)—و شروع به استفاده از آن در شبکه های عصبی کردند.

اولین کاربرد عملی موفق شبکه های عصبی در سال ۱۹۸۹ در آزمایشگاه های بل (Bell Labs) صورت گرفت، زمانی که یان لکون (Yann LeCun) ایده های پیشین شبکه های عصبی کاتولوشنی و بازپراکنندگی را ترکیب کرد و آن ها را برای مسئله طبقه بندی اعداد دست نویس به کار برد. شبکه حاصل، که LeNet نام گذاری شد، در دهه ۱۹۹۰ توسط خدمات پستی ایالات متحده برای خودکار سازی خواندن کدهای پستی روی پاکت های نامه مورد استفاده قرار گرفت.

۱.۲.۳ روش های هسته

وقتی که شبکه های عصبی در دهه ۱۹۹۰، به لطف این اولین موفقیت، شروع به کسب احترام در میان پژوهشگران کردند، یک رویکرد جدید در یادگیری ماشین به شهرت رسید و به سرعت شبکه های عصبی را به فراموشی سپرد: روش های کرنل (Kernel Methods). روش های کرنل گروهی از الگوریتم های طبقه بندی هستند که شناخته شده ترین آن ها ماشین بردار پشتیبان (SVM) است. فرم مدرن SVM توسط ولادیمیر وپنیک (Vladimir Vapnik) و کورینا کورتز (Corinna Cortes) در اوایل دهه ۱۹۹۰ در آزمایشگاه های بل توسعه یافت و در سال ۱۹۹۵ منتشر شد، اگرچه یک فرم خطی قدیمی تر از آن توسط وپنیک و آلکسی چرووننکیس (Alexey Chervonenkis) در سال ۱۹۶۳ منتشر شده بود.

SVM یک الگوریتم طبقه بندی است که با یافتن «مرزهای تصمیم گیری» برای جدا کردن دو کلاس کار می کند (به شکل 1.10 نگاه کنید). SVM ها این مرزها را در دو مرحله پیدا می کنند:

1. داده ها به یک نمایش جدید با ابعاد بالا نگاشت می شوند، جایی که مرز تصمیم گیری می تواند به صورت یک ابرصفحه بیان شود (اگر داده ها دوبعدی باشند، همان طور که در شکل 1.10 نشان داده شده است، ابرصفحه به شکل یک خط مستقیم خواهد بود).
2. یک مرز تصمیم گیری مناسب (یک ابرصفحه جداساز) با تلاش برای بیشینه سازی فاصله بین ابرصفحه و نزدیک ترین نقاط داده از هر کلاس محاسبه می شود، مرحله ای که به آن «بیشینه سازی حاشیه» گفته می شود. این کار به مرز اجازه می دهد تا به خوبی به نمونه های جدید خارج از مجموعه آموزشی تعمیم یابد.

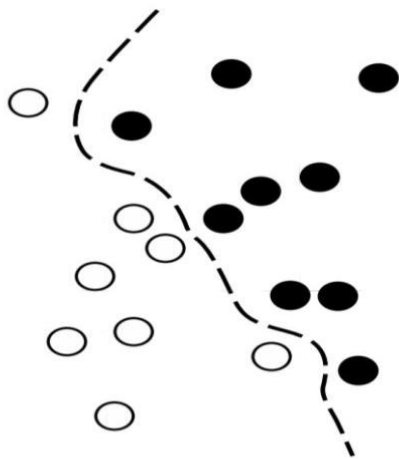


Figure 1.10
A decision boundary

تکنیک نگاشت داده‌ها به یک نمایش با ابعاد بالا که یک مسئله طبقه‌بندی را ساده‌تر می‌کند ممکن است روی کاغذ خوب به نظر برسد، اما در عمل اغلب از نظر محاسباتی غیرقابل اجرا است. اینجاست که ترفند کرنل (Kernel Trick)، ایده اصلی نام‌گذاری این روش‌ها، وارد می‌شود. ایده این است: برای یافتن ابرصفحه‌های تصمیم‌گیری مناسب در فضای نمایش جدید، نیازی به محاسبه صریح مختصات نقاط در فضای جدید ندارید؛ تنها لازم است فاصله بین جفت‌های نقاط در آن فضا را محاسبه کنید، که این کار با استفاده از یک تابع کرنل به طور کارآمد انجام می‌شود. یک تابع کرنل یک عملیات محاسباتی قابل اجرا است که هر دو نقطه در فضای اولیه شما را به فاصله بین این نقاط در فضای نمایش هدف نگاشت می‌کند و کاملاً محاسبه صریح نمایش جدید را دور می‌زند. توابع کرنل معمولاً به صورت دستی طراحی می‌شوند، نه اینکه از داده‌ها آموخته شوند—در مورد SVM، تنها ابرصفحه جداساز آموخته می‌شود.

در زمان توسعه آن‌ها، SVM‌ها عملکرد پیشرفته‌ای در مسائل ساده طبقه‌بندی نشان می‌دادند و یکی از معدود روش‌های یادگیری ماشین بودند که پشتوانه نظری گسترده‌ای داشتند و به تحلیل ریاضی جدی پاسخگو بودند، که باعث می‌شد به خوبی درک شوند و به راحتی تفسیر شوند. به دلیل این ویژگی‌های مفید، SVM‌ها برای مدت طولانی در این حوزه بسیار محبوب شدند.

اما SVM‌ها در مقیاس‌بندی به مجموعه داده‌های بزرگ دشوار بودند و نتایج خوبی برای مسائل ادراکی مانند طبقه‌بندی تصاویر ارائه نمی‌دادند. از آنجا که SVM یک روش کم‌عمق است، استفاده از SVM برای مسائل ادراکی نیازمند استخراج دستی نمایش‌های مفید در ابتدا است (مرحله‌ای که به آن مهندسی ویژگی گفته می‌شود)، که این کار دشوار و شکننده است. برای مثال، اگر بخواهید از SVM برای طبقه‌بندی ارقام دست‌نویس استفاده کنید، نمی‌توانید از پیکسل‌های خام شروع کنید؛ بلکه باید ابتدا به صورت دستی نمایش‌های مفیدی پیدا کنید که مسئله را قابل حل‌تر کنند، مانند هیستوگرام‌های پیکسل که پیش‌تر به آن‌ها اشاره شد.

۱.۲.۴ درختان تصمیم‌گیری، جنگل‌های تصادفی، و ماشین‌های تقویت گرادیان

درخت‌های تصمیم‌گیری ساختارهایی شبیه به نمودار جریان هستند که به شما امکان می‌دهند نقاط داده ورودی را طبقه‌بندی کنید یا مقادیر خروجی را بر اساس ورودی‌ها پیش‌بینی کنید (نگاه کنید به شکل 1.11). این درخت‌ها به راحتی قابل مشاهده و تفسیر هستند. یادگیری درخت‌های تصمیم‌گیری از داده‌ها از دهه ۲۰۰۰ مورد توجه جدی پژوهشگران قرار گرفت و تا سال ۲۰۱۰ اغلب به روش‌های کرنل ترجیح داده می‌شد.

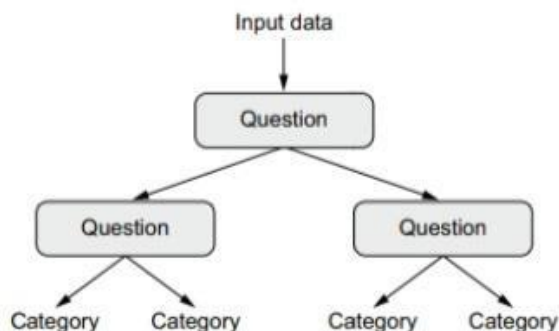


Figure 1.11 A decision tree: the parameters that are learned are the questions about the data. A question could be, for instance, "Is coefficient 2 in the data greater than 3.5?"

به‌ویژه، الگوریتم جنگل تصادفی (Random Forest) یک رویکرد قوی و عملی برای یادگیری درخت تصمیم ارائه داد که شامل ساخت تعداد زیادی درخت تصمیم تخصصی و سپس تجمیع خروجی‌های آن‌ها می‌شود. جنگل‌های تصادفی برای طیف گسترده‌ای از مسائل کاربرد دارند—می‌توان گفت تقریباً همیشه دومین الگوریتم برتر برای هر وظیفه یادگیری ماشین سطحی هستند.

وقتی وبسایت مسابقات یادگیری ماشین محبوب Kaggle (<http://kaggle.com>) در سال ۲۰۱۰ آغاز به کار کرد، جنگل‌های تصادفی به سرعت به یکی از محبوب‌ترین الگوریتم‌ها در این پلتفرم تبدیل شدند—تا سال ۲۰۱۴، که ماشین‌های تقویت گرادیان (Gradient Boosting Machines) جایگزین آن‌ها شدند. ماشین تقویت گرادیان، مشابه جنگل تصادفی، یک تکنیک یادگیری ماشین بر اساس تجمیع مدل‌های پیش‌بینی ضعیف است، معمولاً درخت‌های تصمیم. این تکنیک از تقویت گرادیان استفاده می‌کند، روشی برای بهبود هر مدل یادگیری ماشین با آموزش مکرر مدل‌های جدید که در تخصصی کردن نقاط ضعف مدل‌های قبلی عمل می‌کنند. هنگامی که این تکنیک به درخت‌های تصمیم اعمال می‌شود، مدل‌هایی تولید می‌کند که تقریباً همیشه عملکرد بهتری نسبت به جنگل‌های تصادفی دارند، در حالی که خواص مشابهی دارند.

این تکنیک ممکن است یکی از بهترین، اگر نه بهترین، الگوریتم برای پردازش داده‌های غیرادراکی در حال حاضر باشد. در کنار یادگیری عمیق، یکی از پرکاربردترین تکنیک‌ها در مسابقات Kaggle است.

۱.۲.۵ بازگشت به شبکه‌های عصبی

حدود سال ۲۰۱۰، اگرچه شبکه‌های عصبی تقریباً به طور کامل توسط جامعه علمی کنار گذاشته شده بودند، تعدادی از پژوهشگران که همچنان بر روی شبکه‌های عصبی کار می‌کردند، شروع به دستیابی به پیشرفت‌های مهمی کردند: گروه‌های جفری هینتون در دانشگاه تورنتو، یوشوا بنجیو در دانشگاه مونترال، یان لوکان در دانشگاه نیویورک، و سونیاس در IDSIA.

در سال ۲۰۱۱، دن سیرپسان از IDSIA شروع به برنده شدن در مسابقات دانشگاهی طبقه‌بندی تصویر با شبکه‌های عصبی عمیق آموزش‌دیده بر روی GPU کرد—اولین موفقیت عملی یادگیری عمیق مدرن. اما لحظه تحول‌ساز در سال ۲۰۱۲ رخ داد، با ورود گروه هینتون به مسابقه سالانه طبقه‌بندی تصویر در مقیاس بزرگ ImageNet مسابقه تشخیص بصری در مقیاس بزرگ ImageNet یا به اختصار (ILSVRC) چالش ImageNet در آن زمان به شدت دشوار بود و شامل طبقه‌بندی تصاویر رنگی با وضوح بالا به ۱۰۰۰ دسته مختلف پس از آموزش بر روی ۱.۴ میلیون تصویر بود. در سال ۲۰۱۱، دقت پنجرده برتر مدل برنده، بر اساس رویکردهای کلاسیک بینایی کامپیوتر، تنها ۷۴.۳٪ بود. سپس، در سال ۲۰۱۲، تیمی به رهبری الکس کریژفسکی و تحت نظر جفری هینتون توانست به دقت پنجرده برتر ۸۳.۶٪ دست یابد—یک پیشرفت قابل‌توجه. از آن زمان، مسابقه توسط شبکه‌های عصبی کانولوشنی عمیق هر ساله تحت سلطه قرار گرفته است. تا سال ۲۰۱۵، برنده به دقت ۹۶.۴٪ رسید و مسئله طبقه‌بندی در ImageNet به‌طور کامل حل‌شده در نظر گرفته شد.

از سال ۲۰۱۲، شبکه‌های عصبی کانولوشنی عمیق (convnets) به الگوریتم اصلی برای تمام وظایف بینایی کامپیوتر تبدیل شده‌اند؛ به طور کلی‌تر، آن‌ها بر روی تمام وظایف ادراکی کار می‌کنند. در هر کنفرانس بزرگ بینایی کامپیوتر پس از سال ۲۰۱۵، تقریباً غیرممکن بود که ارائه‌ای پیدا کنید که به شکلی از convnets استفاده نکرده باشد. همزمان، یادگیری عمیق در بسیاری از انواع دیگر مسائل نیز کاربرد پیدا کرده است، مانند پردازش زبان طبیعی. این روش به طور کامل جایگزین SVM ها و درخت‌های تصمیم در طیف وسیعی از کاربردها شده است. برای مثال، برای چندین سال، سازمان اروپایی تحقیقات هسته‌ای (CERN) از روش‌های مبتنی بر درخت تصمیم برای تحلیل داده‌های ذرات از آشکارساز ATLAS در برخورددهنده هادرونی بزرگ (LHC) استفاده می‌کرد، اما در نهایت CERN به شبکه‌های عصبی عمیق مبتنی بر Keras روی آورد، به دلیل عملکرد بالاتر و سهولت آموزش بر روی مجموعه‌های داده بزرگ.

۱.۲.۶ چه چیزی یادگیری عمیق را متفاوت می‌کند

دلیل اصلی موفقیت سریع یادگیری عمیق این است که عملکرد بهتری برای بسیاری از مسائل ارائه می‌دهد. اما این تنها دلیل نیست. یادگیری عمیق همچنین حل مسئله را بسیار آسان‌تر می‌کند، زیرا به طور کامل مرحله‌ای را که قبلاً مهم‌ترین بخش در جریان کاری یادگیری ماشین بود، خودکار می‌کند: مهندسی ویژگی.

تکنیک‌های قبلی یادگیری ماشین—یادگیری کم‌عمق—تنها شامل تبدیل داده‌های ورودی به یک یا دو فضای نمایش پی‌درپی بود، معمولاً از طریق تبدیلات ساده‌ای مانند نگاشت‌های غیرخطی با ابعاد بالا (SVM) ها (یا درخت‌های تصمیم). اما نمایش‌های پالایش‌شده‌ای که برای مسائل پیچیده مورد نیاز هستند، معمولاً با چنین تکنیک‌هایی قابل دستیابی نیستند. بنابراین، انسان‌ها مجبور بودند تلاش زیادی کنند تا داده‌های ورودی اولیه را برای پردازش توسط این روش‌ها مناسب‌تر کنند: آن‌ها مجبور بودند لایه‌های خوبی از نمایش‌ها را برای داده‌های خود به صورت دستی مهندسی کنند. به این فرآیند مهندسی ویژگی گفته می‌شود. در مقابل، یادگیری عمیق این مرحله را کاملاً خودکار می‌کند: با یادگیری عمیق، تمام ویژگی‌ها در یک مرحله آموخته می‌شوند، بدون اینکه نیاز به مهندسی آن‌ها توسط انسان باشد. این کار جریان کاری یادگیری ماشین را به شدت ساده کرده و اغلب جایگزین فرآیندهای پیچیده چنمرحله‌ای با یک مدل یادگیری عمیق ساده و انتها به انتها شده است.

شاید بپرسید، اگر مسئله اصلی داشتن چندین لایه نمایش پی‌درپی است، آیا نمی‌توان روش‌های کم‌عمق را به‌طور مکرر اعمال کرد تا اثرات یادگیری عمیق را شبیه‌سازی کرد؟ در عمل، استفاده مکرر از روش‌های یادگیری کم‌عمق نتایج سریع‌اً کاهشی ایجاد می‌کند، زیرا لایه اول بهینه در یک مدل سه‌لایه، لایه اول بهینه در یک مدل یک‌لایه یا دو لایه نیست. آنچه یادگیری عمیق را متحول کرده است این است که به مدل اجازه می‌دهد تمام لایه‌های نمایش را به‌صورت مشترک و همزمان یاد بگیرد، نه به‌صورت پی‌درپی (که به آن روش حریم‌بانه گفته می‌شود). در یادگیری ویژگی مشترک، هر بار که مدل یکی از ویژگی‌های داخلی خود را تنظیم می‌کند، تمام ویژگی‌های دیگر که به آن وابسته هستند، به طور خودکار با تغییر سازگار می‌شوند، بدون نیاز به دخالت انسان. همه چیز تحت نظارت یک سیگنال بازخوردی واحد است: هر تغییری در مدل به هدف نهایی خدمت می‌کند. این بسیار قدرتمندتر از چیدن حریم‌بانه مدل‌های کم‌عمق است، زیرا اجازه می‌دهد نمایش‌های پیچیده و انتزاعی با شکستن آن‌ها به زنجیره‌ای طولانی از فضا‌های میانی (لایه‌ها) یاد گرفته شوند؛ هر فضا تنها یک تبدیل ساده با فضای قبلی فاصله دارد.

این‌ها دو ویژگی اساسی نحوه یادگیری داده‌ها توسط یادگیری عمیق هستند: روش تدریجی و لایه‌به‌لایه‌ای که در آن نمایش‌های به طور فزاینده پیچیده توسعه می‌یابند، و این واقعیت که این نمایش‌های تدریجی میانی به صورت مشترک یاد گرفته می‌شوند؛ هر لایه برای برآورده کردن نیازهای نمایش لایه بالاتر و نیازهای لایه پایین‌تر به‌روزرسانی می‌شود. این دو ویژگی با هم یادگیری عمیق را بسیار موفق‌تر از روش‌های قبلی یادگیری ماشین کرده‌اند.

۱.۲.۷ چشم انداز یادگیری ماشین مدرن

یک راه عالی برای درک وضعیت فعلی الگوریتم‌ها و ابزارهای یادگیری ماشین، نگاه کردن به مسابقات یادگیری ماشین در سایت Kaggle است. به دلیل محیط رقابتی شدید آن (برخی مسابقات هزاران شرکت‌کننده و جوایز میلیون‌دلاری دارند) و تنوع گسترده مسائل یادگیری ماشین تحت پوشش، Kaggle یک روش واقعی برای ارزیابی این که چه چیزی کار می‌کند و چه چیزی کار نمی‌کند ارائه می‌دهد. بنابراین، چه نوع الگوریتمی به طور قابل اعتمادی مسابقات را برنده می‌شود؟ شرکت‌کنندگان برتر از چه ابزارهایی استفاده می‌کنند؟

اوایل سال ۲۰۱۹، Kaggle یک نظرسنجی انجام داد و از تیم‌هایی که از سال ۲۰۱۷ در پنج رتبه برتر هر مسابقه قرار گرفته بودند پرسید که از کدام ابزار نرم‌افزاری اصلی در مسابقات استفاده کرده‌اند (نگاه کنید به شکل 1.12). مشخص شد که تیم‌های برتر تمایل دارند با از روش‌های یادگیری عمیق (اغلب از طریق کتابخانه Keras) یا از درخت‌های تقویت‌گرایان (اغلب از طریق کتابخانه‌های LightGBM یا XGBoost) استفاده کنند

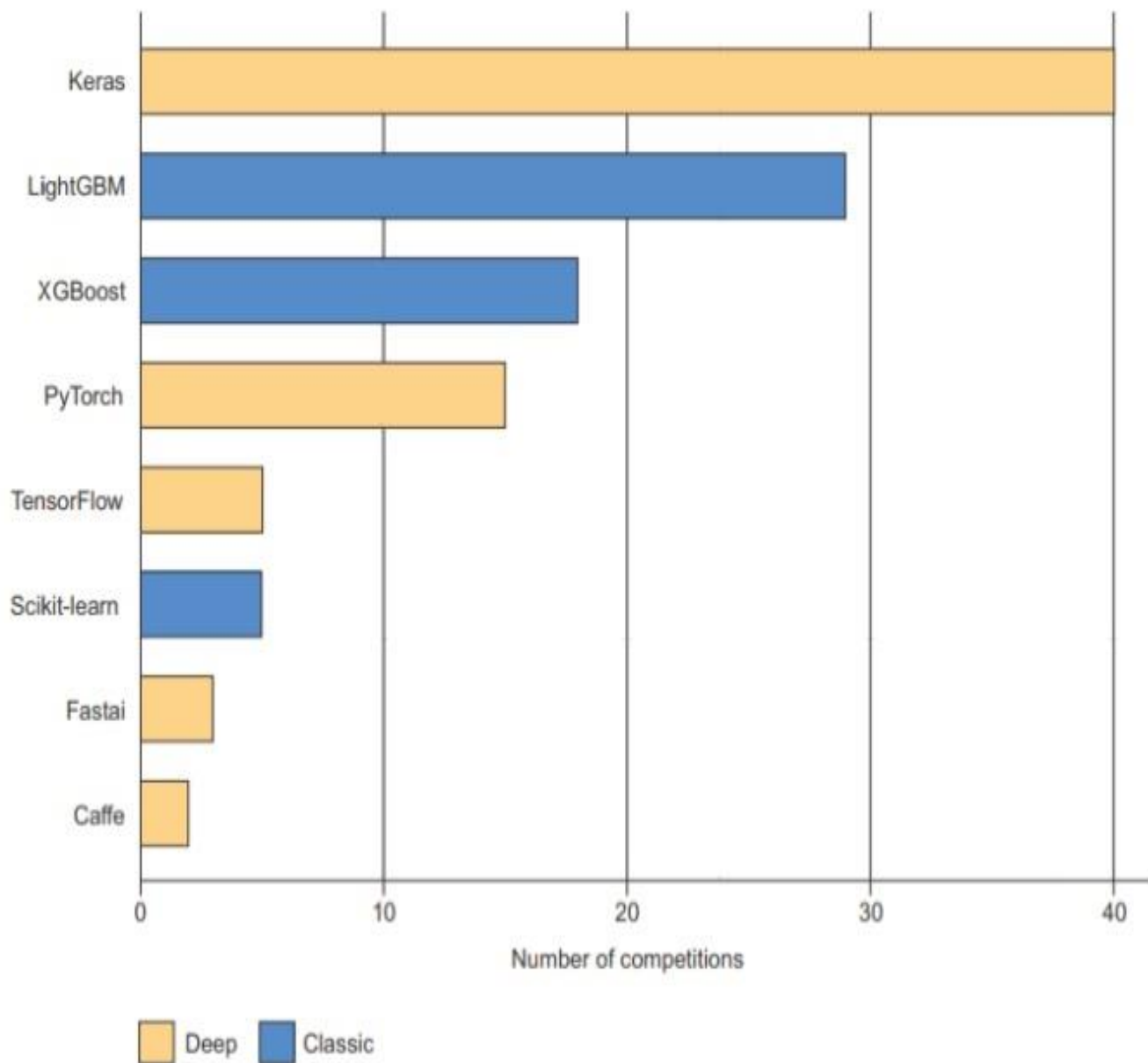


Figure 1.12 Machine learning tools used by top teams on Kaggle

این موضوع فقط مربوط به قهرمانان مسابقات نیست. **Kaggle** همچنین هر سال یک نظرسنجی در میان حرفه‌ای‌های یادگیری ماشین و علوم داده در سراسر جهان انجام می‌دهد. با ده‌ها هزار پاسخ‌دهنده، این نظرسنجی یکی از قابل‌اعتمادترین منابع درباره وضعیت صنعت است. شکل 1.13 درصد استفاده از چارچوب‌های مختلف نرم‌افزاری یادگیری ماشین را نشان می‌دهد.

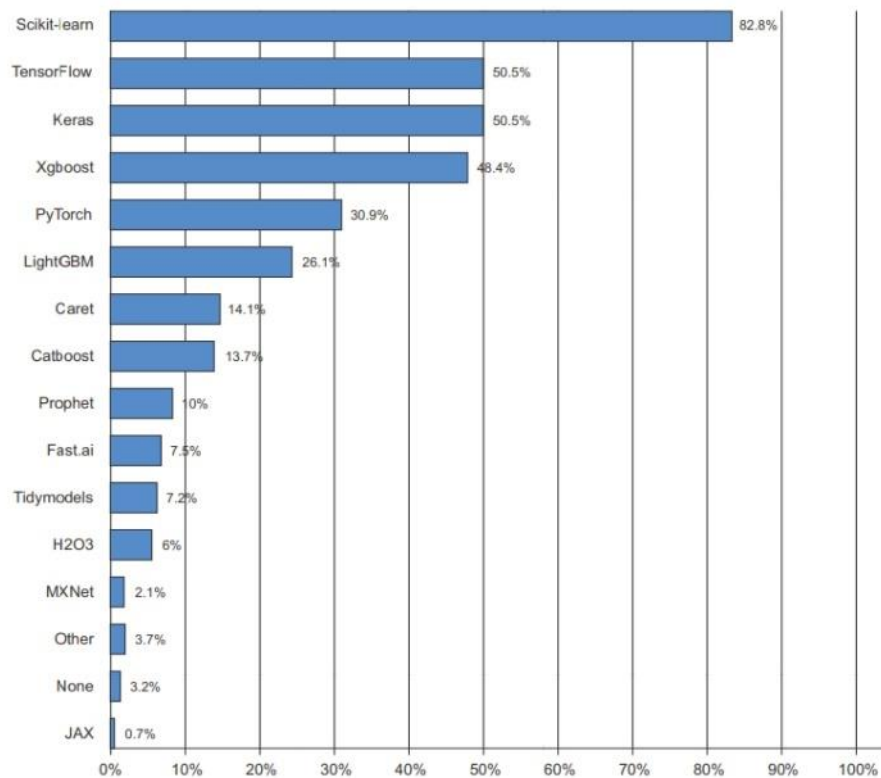


Figure 1.13 Tool usage across the machine learning and data science industry (Source: www.kaggle.com/kaggle-survey-2020)

از سال ۲۰۱۶ تا ۲۰۲۰، کل صنعت یادگیری ماشین و علوم داده تحت سلطه این دو رویکرد بوده است: یادگیری عمیق و درخت‌های تقویت‌گرایان. به طور خاص، درخت‌های تقویت‌گرایان برای مسائلی که داده‌های ساختاریافته در دسترس است استفاده می‌شود، در حالی که یادگیری عمیق برای مسائل ادراکی مانند طبقه‌بندی تصاویر به کار می‌رود.

کاربران درخت‌های تقویت‌گرایان معمولاً از Scikit-learn، XGBoost یا LightGBM استفاده می‌کنند. در همین حال، اکثر افرادی که از یادگیری عمیق استفاده می‌کنند، از Keras استفاده می‌کنند، اغلب در ترکیب با چارچوب اصلی آن، TensorFlow. نکته مشترک این ابزارها این است که همه آن‌ها کتابخانه‌های Python هستند: پایتون با اختلاف، پرکاربردترین زبان برای یادگیری ماشین و علوم داده است.

این دو تکنیک مهم‌ترین تکنیک‌هایی هستند که باید برای موفقیت در یادگیری ماشین کاربردی امروز با آن‌ها آشنا باشید: درخت‌های تقویت‌گرایان، برای مسائل یادگیری کم‌عمق؛ و یادگیری عمیق، برای مسائل ادراکی. از نظر فنی، این بدان معناست که شما باید با Scikit-learn، XGBoost و Keras آشنا باشید—سه کتابخانه‌ای که در حال حاضر در مسابقات Kaggle غالب هستند. با داشتن این کتاب در دست، شما یک قدم بزرگ به جلو برداشته‌اید.

۱.۳ چرا یادگیری عمیق؟ چرا حالا؟

دو ایده کلیدی یادگیری عمیق برای بینایی کامپیوتر—شبکه‌های عصبی کانولوشنی و الگوریتم پس‌انتشار—تا سال ۱۹۹۰ به خوبی درک شده بودند. الگوریتم حافظه بلندمدت کوتاه‌مدت (LSTM)، که برای یادگیری عمیق در سری‌های زمانی اساسی است، در سال ۱۹۹۷ توسعه یافت و از آن زمان تقریباً بدون تغییر باقی مانده است. پس چرا یادگیری عمیق تنها پس از سال ۲۰۱۲ به موفقیت رسید؟ چه چیزی در این دو دهه تغییر کرد؟

به طور کلی، سه نیروی فنی پیشرفت‌های یادگیری ماشین را هدایت می‌کنند:

- سخت‌افزار
- مجموعه داده‌ها و معیارها

• پیشرفت‌های الگوریتمی

از آنجا که این حوزه توسط یافته‌های تجربی هدایت می‌شود و نه تئوری، پیشرفت‌های الگوریتمی تنها زمانی ممکن می‌شوند که داده‌ها و سخت‌افزار مناسب برای آزمایش ایده‌های جدید (یا برای مقیاس‌گذاری ایده‌های قدیمی، که اغلب چنین است) در دسترس باشند. یادگیری ماشین ریاضیات با فیزیک نیست، جایی که پیشرفت‌های بزرگ با یک قلم و یک ورق کاغذ انجام می‌شوند. این یک علم مهندسی است.

تنگنای واقعی در دهه‌های ۱۹۹۰ و ۲۰۰۰ داده‌ها و سخت‌افزار بودند. اما در آن زمان چه اتفاقی افتاد؟ اینترنت گسترش یافت و تراشه‌های گرافیکی با عملکرد بالا برای نیازهای بازار بازی توسعه پیدا کردند.

۱.۳.۱ سخت افزار

بین سال‌های ۱۹۹۰ تا ۲۰۱۰، پردازنده‌های آماده (CPU) حدود ۵,۰۰۰ برابر سریع‌تر شدند. به همین دلیل، امروزه امکان اجرای مدل‌های کوچک یادگیری عمیق روی لپ‌تاپ شما وجود دارد، در حالی که ۲۵ سال پیش این کار غیرممکن بود.

اما مدل‌های معمول یادگیری عمیق که در بینایی کامپیوتر یا تشخیص گفتار استفاده می‌شوند، به قدرت محاسباتی به مراتب بیشتری نسبت به آنچه لپ‌تاپ شما ارائه می‌دهد نیاز دارند. در دهه ۲۰۰۰، شرکت‌هایی مانند NVIDIA و AMD میلیارد‌ها دلار برای توسعه تراشه‌های سریع و به شدت موازی (واحدهای پردازش گرافیکی یا GPU) سرمایه‌گذاری کردند تا گرافیک بازی‌های ویدیویی با کیفیت بالا و واقعی را اجرا کنند—سوپرکامپیوترهای تک‌منظوره ارزان‌قیمتی که برای پردازش صحنه‌های پیچیده سه‌بعدی در زمان واقعی طراحی شده‌اند. این سرمایه‌گذاری در سال ۲۰۰۷ به سود جامعه علمی تمام شد، زمانی که NVIDIA رابط برنامه‌نویسی CUDA را برای خط تولید GPUهای خود معرفی کرد. تعداد کمی از GPUها شروع به جایگزینی خوشه‌های عظیم CPU در برنامه‌های بسیار موازی کردند، ابتدا در مدل‌سازی فیزیکی. شبکه‌های عصبی عمیق که عمدتاً شامل تعداد زیادی ضرب ماتریسی کوچک هستند نیز به شدت موازی هستند، و در حدود سال ۲۰۱۱ برخی از پژوهشگران شروع به نوشتن پیاده‌سازی‌های CUDA برای شبکه‌های عصبی کردند—دن سیریسان و الکس کریژفسکی از اولین‌ها بودند.

آنچه رخ داد این بود که بازار بازی، سوپرکامپیوترهای نسل بعدی برای برنامه‌های هوش مصنوعی را یارانه داد. گاهی اوقات، چیزهای بزرگ با بازی‌ها آغاز می‌شوند. امروزه NVIDIA Titan RTX، یک GPU که در پایان سال ۲۰۱۹ با قیمت ۲,۵۰۰ دلار عرضه شد، می‌تواند تا ۱۶ ترافلاپس عملیات ممیز شناور ۳۲ بیتی در ثانیه انجام دهد (۱۶ تریلیون عملیات float32 در ثانیه). (این تقریباً ۵۰۰ برابر قدرت محاسباتی سریع‌ترین سوپرکامپیوتر جهان در سال ۱۹۹۰، یعنی Intel Touchstone Delta است. روی یک Titan RTX، آموزش یک مدل ImageNet از نوعی که در مسابقات ILSVRC سال ۲۰۱۲ یا ۲۰۱۳ برنده می‌شد، تنها چند ساعت زمان می‌برد. در همین حال، شرکت‌های بزرگ مدل‌های یادگیری عمیق را روی خوشه‌هایی با صدها GPU آموزش می‌دهند.

علاوه بر این، صنعت یادگیری عمیق فراتر از GPUها حرکت کرده و در حال سرمایه‌گذاری روی تراشه‌های تخصصی‌تر و کارآمدتر برای یادگیری عمیق است. در سال ۲۰۱۶، در کنوانسیون سالانه I/O، گوگل پروژه واحد پردازش تنسور (TPU) خود را معرفی کرد: یک طراحی تراشه جدید که از ابتدا برای اجرای شبکه‌های عصبی عمیق با سرعت بسیار بیشتر و بازده انرژی بالاتر نسبت به GPUهای پیشرفته طراحی شده است. امروزه، در سال ۲۰۲۰، سومین نسخه کارت TPU قدرت محاسباتی ۴۲۰ ترافلاپس را ارائه می‌دهد. این ۱۰,۰۰۰ برابر بیشتر از Intel Touchstone Delta در سال ۱۹۹۰ است.

این کارت‌های TPU برای پیکربندی‌های بزرگ‌مقیاس، به نام «پاد» طراحی شده‌اند. یک پاد ۱۰۲۴ کارت TPU می‌تواند به ۱۰۰ پتا FLOPS برسد. برای مقایسه، این حدود ۱۰٪ از قدرت محاسباتی پیک بزرگ‌ترین سوپرکامپیوتر فعلی، IBM Summit، در آزمایشگاه ملی اوک ریج است که از ۲۷,۰۰۰ GPU NVIDIA تشکیل شده و به حدود ۱.۱ اگزا FLOPS می‌رسد.

۱.۳.۲ داده ها

هوش مصنوعی گاهی به‌عنوان انقلاب صنعتی جدید معرفی می‌شود. اگر یادگیری عمیق موتور بخار این انقلاب است، پس داده‌ها زغال‌سنگ آن هستند: ماده خامی که ماشین‌های هوشمند ما را به حرکت درمی‌آورد و بدون آن هیچ چیزی ممکن نخواهد بود. وقتی صحبت از داده‌ها می‌شود، علاوه بر پیشرفت نمایی در سخت‌افزار ذخیره‌سازی طی ۲۰ سال گذشته (بر اساس قانون مور)، تغییر بزرگ ظهور اینترنت بوده است، که جمع‌آوری و توزیع مجموعه‌های داده بسیار بزرگ برای یادگیری ماشین را ممکن کرده است. امروزه، شرکت‌های بزرگ با مجموعه‌های داده تصاویر، ویدیوها، و زبان طبیعی کار می‌کنند که بدون اینترنت امکان جمع‌آوری آن‌ها وجود نداشت. به عنوان مثال، برجسب‌های تولیدشده توسط کاربران بر روی Flickr یک گنجینه ارزشمند از داده برای بینایی کامپیوتر بوده‌اند. ویدیوهای یوتیوب نیز همین‌طور. و ویکی‌پدیا یک مجموعه داده کلیدی برای پردازش زبان طبیعی است.

اگر یک مجموعه داده باشد که به عنوان کاتالیزور برای پیشرفت یادگیری عمیق عمل کرده است، آن مجموعه داده ImageNet است، که شامل ۱.۴ میلیون تصویر است که به صورت دستی با ۱,۰۰۰ دسته تصویر (یک دسته برای هر تصویر) برچسبگذاری شده‌اند. اما چیزی که ImageNet را خاص می‌کند فقط اندازه بزرگ آن نیست، بلکه مسابقه سالانه‌ای است که با آن همراه است.

همان‌طور که Kaggle از سال ۲۰۱۰ نشان داده است، مسابقات عمومی یک راه عالی برای انگیزه دادن به پژوهشگران و مهندسان برای پیشبرد مرزهای دانش هستند. داشتن معیارهای مشترکی که پژوهشگران برای شکست دادن آن‌ها با یکدیگر رقابت می‌کنند، به طور قابل‌توجهی به پیشرفت یادگیری عمیق کمک کرده است، زیرا موفقیت آن در مقایسه با رویکردهای کلاسیک یادگیری ماشین را برجسته می‌کند.

۱.۳.۳ الگوریتم‌ها

علاوه بر سخت‌افزار و داده‌ها، تا اواخر دهه ۲۰۰۰، ما فاقد یک روش قابل‌اعتماد برای آموزش شبکه‌های عصبی بسیار عمیق بودیم. در نتیجه، شبکه‌های عصبی هنوز نسبتاً کم‌عمق بودند و تنها از یک یا دو لایه نمایش استفاده می‌کردند؛ بنابراین، نمی‌توانستند در برابر روش‌های کم‌عمق پیشرفته‌تر مانند SVM ها و جنگل‌های تصادفی خودی نشان دهند. مسئله کلیدی، انتشار گرادیان در میان لایه‌های عمیق بود. سیگنال بازخوردی که برای آموزش شبکه‌های عصبی استفاده می‌شد، با افزایش تعداد لایه‌ها کاهش می‌یافت.

این وضعیت در حدود سال‌های ۲۰۰۹-۲۰۱۰ با ظهور چندین پیشرفت ساده اما مهم الگوریتمی که انتشار بهتر گرادیان را ممکن ساخت، تغییر کرد:

- توابع فعال‌سازی بهتر برای لایه‌های عصبی
- طرح‌های بهینه‌سازی وزن بهتر، که با پیش‌آموزش لایه‌به‌لایه شروع شد، اما به سرعت کنار گذاشته شد
- طرح‌های بهینه‌سازی بهتر، مانند RMSProp و Adam

تنها زمانی که این پیشرفت‌ها امکان آموزش مدل‌هایی با ۱۰ یا بیشتر لایه را فراهم کردند، یادگیری عمیق شروع به درخشش کرد.

در نهایت، در سال‌های ۲۰۱۴، ۲۰۱۵ و ۲۰۱۶، روش‌های پیشرفته‌تری برای بهبود انتشار گرادیان کشف شدند، مانند نرمال‌سازی دسته‌ای (Batch Normalization)، اتصالات باقیمانده (Residual Connections)، و کانولوشن‌های تفکیکی عمق (Depthwise Separable Convolutions).

امروزه می‌توانیم مدل‌هایی با هر عمقی را از ابتدا آموزش دهیم. این قابلیت استفاده از مدل‌های بسیار بزرگ را فراهم کرده است که قدرت نمایشی قابل‌توجهی دارند—یعنی، فضای فرضیه‌های بسیار غنی را رمزگذاری می‌کنند. این مقیاس‌پذیری افراطی یکی از ویژگی‌های متمایز یادگیری عمیق مدرن است. معماری‌های مدل‌های بزرگ‌مقیاس که دارای ده‌ها لایه و ده‌ها میلیون پارامتر هستند، پیشرفت‌های مهمی را هم در بینایی کامپیوتر (مانند معماری‌هایی مثل ResNet، Inception، یا Xception و هم در پردازش زبان طبیعی) مانند معماری‌های بزرگ مبتنی بر Transformer مانند BERT، GPT-3 یا XLNet به ارمغان آورده‌اند.