



یادگیری ماشین شاخه‌ای از فناوری هوش مصنوعی است که از شیوه‌ها یا الگوریتم‌های خاص برای ساخت خودکار مدل‌هایی از داده استفاده می‌کند. برخلاف یک سیستم معمولی که با پیروی از قوانین مشخص یک وظیفه را انجام می‌دهد، یک سیستم یادگیری ماشین از طریق تجربه کردن وظایف خود را یاد می‌گیرد. در حالی که یک سیستم مبتنی بر قانون هر بار یک وظیفه را به روشی یکسان (چه خوب، چه بد) انجام می‌دهد، وضعیت عملکرد یک سیستم یادگیری ماشین از طریق یادگیری و تمرین از طریق الگوریتم‌های داده کاوی بهبود پیدا می‌کند. الگوریتم‌های یادگیری ماشین با فراگیری از داده‌هایی که دریافت می‌کنند می‌توانند مسائلی را حل کنند که بسیار پیچیده‌تر از آن است که بتوان آن را با برنامه‌نویسی معمولی حل کرد.

الگوریتم‌های یادگیری ماشین اغلب به دو دسته نظارتی یا Supervised (داده‌های آموزشی با پاسخ‌ها برچسب گذاری می‌شوند) بدون نظارت یا Unsupervised (برچسب‌های موجود به الگوریتم آموزش نشان داده نمی‌شود) تقسیم می‌شوند. مسائل **یادگیری ماشین** نظارتی نیز به دو گروه طبقه‌بندی یا Classification (پیش‌بینی پاسخ‌های غیر عددی، مانند احتمال خطا در پرداخت وام مسکن) و رگرسیون یا Regression (پیش‌بینی پاسخ‌های عددی، مانند تعداد چیزهایی که ماه آینده در فروشگاه شما فروخته می‌شود) تقسیم می‌شوند. یادگیری بدون نظارت هم به سه دسته خوشه‌بندی یا Clustering (پیدا کردن گروه‌هایی از اشیاء مشابه مانند کفش‌های ورزشی، کفش‌های طبی و کفش‌های مجلسی)، مشارکتی یا Association (پیدا کردن توالی‌های مشترک از اشیاء، مانند قهوه و خامه) و کاهش ابعاد یا Dimensionality Reduction (طرح‌ریزی، انتخاب ویژگی و استخراج ویژگی) تقسیم می‌شوند.

کاربردهای یادگیری ماشین

ما همه روزه اخبار گوناگونی در مورد کاربردهای مختلف **یادگیری ماشین** می‌شنویم، هر چند همه آن‌ها با موفقیت‌های کامل همراه نیستند. در این زمینه خودروهای خودران مثال خوبی محسوب می‌شوند که در آن وظایف از ساده و موفقیت آمیز (دستیار پارک کردن و دنباله‌روی از خطوط وسط بزرگراه) تا پیچیده و مشروط (کنترل کامل وسیله نقلیه در مکان‌های شهری که به چندین فقره مرگ منجر شده است) گسترش پیدا می‌کند. یادگیری ماشین در انجام بازی در مواردی مانند بازی شطرنج، چکرز و شوگی (شطرنج ژاپنی) کاملاً موفقیت‌آمیز بوده و توانسته قهرمانان انسانی خود را شکست دهد؛ همچنین در ترجمه خودکار زبان تا اندازه زیادی با موفقیت همراه بوده، هر چند در این فناوری بعضی از زبان‌ها از باقی زبان‌ها بهتر عمل می‌کنند و بسیاری از ترجمه‌های ماشینی برای ارتقای سطح کیفی هنوز هم به ویرایش و ترجمه انسانی نیاز دارند.



یک پروژه عملی یادگیری ماشین چگونه می‌توانیم در عمل از یادگیری ماشین استفاده کنیم؟

تبدیل خودکار متن به گفتار برای زبان‌هایی با لهجه اصلی به خوبی کار می‌کند، اما هنوز در برخی لهجه‌های قومی و منطقه‌ای با مشکل مواجه است و وضعیت عملکرد آن به مجموعه آموزش‌های مورد استفاده خدمات‌دهندگان بستگی دارد. تجزیه و تحلیل خودکار احساسات رسانه‌های اجتماعی به سطح موفقیت نسبتاً خوبی دست پیدا کرده و این احتمالاً به دلیل آن است که مجموعه‌های آموزشی (به عنوان مثال، رتبه‌بندی محصولات آمازون، که یک نظر را با امتیازدهی عددی ترکیب می‌کند) گسترده و دسترسی به آن‌ها آسان است. از جمله مباحث بحث‌برانگیز غربالگری خودکار رزومه است. به عنوان مثال، آمازون مجبور شد سیستم داخلی خود را به دلیل بروز اختلالات نمونه‌گیری که باعث پایین آورده شدن رتبه همه سمت‌های شغلی مربوط به زنان شده بود، کنار بگذارد. یکی دیگر از سیستم‌های غربالگری رزومه که در حال حاضر از آن استفاده می‌شود با آموزش‌های متعصبانه‌ای همراه است که باعث می‌شود بدون در نظر گرفتن ملاحظات قانونی امتیازاتی را برای برخی از متقاضیان که خصوصیات مورد نظر کارفرما را دارند در نظر گرفته و آن‌ها را در صدر فهرست قرار دهد. به عنوان نمونه، متقاضیان مرد جوان سفیدپوست که در یک ورزش گروهی مهارت دارند، اولویت بالاتری کسب می‌کنند. تحقیقات زیادی در زمینه حذف تعصبات ضمنی در **یادگیری ماشین** توسط مایکروسافت و سایرین صورت گرفته است. طبقه‌بندی خودکار آسیب‌شناسی و تصاویر رادیولوژی پیشرفت زیادی داشته تا جایی که می‌تواند به آسیب‌شناسان و رادیولوژیست‌ها برای تشخیص نوع خاصی از اختلالات کمک کند. در عین حال، سیستم‌های شناسایی چهره هم به دلایل ملاحظات حفظ حریم خصوصی و کوتاهی کردن در تشخیص چهره زنان و رنگین‌پوستان (باز هم به دلیل تعصبات در رفتارهای آموزشی) بسیار بحث‌برانگیز شده است.

الگوریتم‌های یادگیری ماشین

یادگیری ماشین برای تبدیل مجموعه‌ای از داده‌ها به یک مدل به تعدادی از الگوریتم‌ها وابسته است. این‌که کدام الگوریتم‌ها از همه بهتر کار می‌کند، به نوع مسئله‌ای که شما در حال حل کردن آن هستید، منابع محاسباتی موجود و ماهیت داده‌ها بستگی دارد. مهم نیست، از چه **الگوریتم** یا الگوریتم‌هایی استفاده می‌کنید، در مرحله نخست باید داده‌ها را پاکسازی و مرتب کنید.

الگوریتم‌های طبقه‌بندی

طبقه‌بندی، نوعی مسئله یادگیری نظارتی است که یک انتخاب بین دو یا چند طبقه را درخواست می‌کند. صرف‌نظر از شبکه‌های عصبی و یادگیری عمیق که به سطح بالاتری از منابع محاسباتی نیاز دارند، رایج‌ترین الگوریتم‌ها Naive Support Vector (سرنام SVM) و Bayes, Decision Tree, Logistic Regression, K-Nearest Neighbors (Machine) هستند. شما می‌توانید از روش‌های گروهی (ترکیبی از مدل‌ها) مانند Random Forest, AdaBoost و XGBoost استفاده کنید.

الگوریتم‌های رگرسیون

مسئله رگرسیون، نوعی مسئله یادگیری نظارتی است که این مدل را برای پیش‌بینی یک عدد درخواست می‌کند. ساده‌ترین و سریع‌ترین الگوریتم رگرسیون خطی است، اما به آن بسنده نکنید، زیرا این روش اغلب یک نتیجه متوسط را به شما ارائه می‌دهد. از جمله دیگر الگوریتم‌های رگرسیون یادگیری ماشین رایج می‌توان به Naive Bayes, Decision Tree, K-Nearest Neighbors, LVQ (Learning Vector Quantization), LARS Lasso, Elastic Net, Random Forest, AdaBoost و XGBoost اشاره کرد. لازم به توضیح است که بین **الگوریتم‌های یادگیری ماشین رگرسیون** و طبقه‌بندی اشتراکاتی وجود دارد.



سه حوزه اصلی تخصصی پایتون
پایتون برای انجام چه کارهایی عالی، چه کارهایی خوب و چه کارهایی بد است!؟

الگوریتم‌های خوشه‌بندی

مسئله خوشه‌بندی، نوعی مسئله یادگیری بدون نظارت است که این مدل را برای یافتن گروه‌هایی از نقاط داده مشابه درخواست می‌کند. معروف‌ترین الگوریتم K-Means Clustering است. از جمله الگوریتم‌های دیگر می‌توان به

Mean-Shift Clustering, DBSCAN (Density-Based Gaussian)
HAC (Hierarchical Agglomerative Clustering) و GMM (Gaussian Mixture Models)
Spatial Clustering of Applications with Noise (سرنام)

اشاره کرد.

الگوریتم‌های کاهش ابعاد

مسئله کاهش ابعاد، نوعی مسئله یادگیری بدون نظارت است که این مدل را برای کم کردن یا ترکیب متغیرهایی که تاثیر کمی روی نتیجه دارند یا اصلاً بی‌تاثیر هستند، درخواست می‌کند. از این الگوریتم اغلب در ترکیب با الگوریتم طبقه‌بندی یا رگرسیون استفاده می‌شود. الگوریتم‌های کاهش ابعاد شامل حذف متغیرهایی با مقادیر گمشده زیاد، حذف متغیرهایی با واریانس پایین، Decision Tree, Random Forest، حذف یا ترکیب متغیرهایی با همبستگی بالا، PCA (Principal) و Backward Feature Elimination, Forward Feature Selection, Factor Analysis
Component Analysis هستند.

چگونه از یادگیری ماشین استفاده کنیم؟

ساخت یک مدل یادگیری ماشین با پاکسازی و مرتب کردن داده‌ها شروع می‌شود، با ویژگی‌های مهندسی ادامه پیدا می‌کند و بعد از رسیدن به الگوریتم‌های یادگیری ماشین مناسب استفاده خواهد شد.
در دنیای واقعی هیچ چیز به اندازه پاکسازی و مرتب کردن داده‌های ورودی اهمیت ندارد. برای این‌که داده‌ها در کاربردهای یادگیری ماشین کارآمد باشند باید بلافاصله فیلتر شوند. برای مثال؛ باید داده‌ها را جست‌وجو کرده و هر ستونی را که شامل داده‌های ناقص و از دست رفته است، جداسازی کنید. سپس با نگاه دوباره به داده‌ها ستون‌هایی را که قصد استفاده از آن را برای پیش‌بینی‌های خود دارید، انتخاب کنید.

ویژگی مهندسی برای یادگیری ماشین

مفهوم «ویژگی» به یک متغیر توضیحی مربوط است که در تکنیک‌های آماری مانند رگرسیون خطی استفاده می‌شود. بردارهای ویژگی، تمام ویژگی‌های یک ردیف را با یکدیگر ترکیب کرده و به یک بردار عددی تبدیل می‌کند. بخشی از هنر انتخاب ویژگی‌ها این است که بتوانید مجموعه‌ای از متغیرهای مستقلی که مسئله را توضیح می‌دهند، انتخاب کنید. اگر دو متغیر به میزان زیادی همپوشانی داشته باشند یا باید آن‌ها را به یک ویژگی واحد ترکیب کرد یا باید آن را حذف کرد. گاهی اوقات مردم تجزیه و تحلیل مولفه‌های اصلی را برای تبدیل متغیرهای همبسته به مجموعه‌ای از متغیرهای خطی نامشخص انجام می‌دهند.

تقسیم داده‌ها برای یادگیری ماشین

روش معمول برای یادگیری ماشین نظارتی تقسیم مجموعه داده‌ها به زیر مجموعه‌هایی برای آموزش، اعتبارسنجی و آزمایش است. یک روش برای انجام این کار اختصاص دادن 80% داده‌ها به مجموعه داده‌های آموزشی و 10% برای هر یک از دو بخش مجموعه داده‌های اعتبارسنجی و آزمایش است. البته این را هم باید در نظر داشته باشید که تقسیم‌بندی دقیق به اولویت‌ها بستگی دارد. بخش عمده‌ای از آموزش در مجموعه داده‌های آموزشی انجام می‌شود و

پیش‌بینی در مجموعه داده‌های اعتبارسنجی و در پایان هر دوره انجام می‌گیرد. از خطاهای مجموعه داده اعتبارسنجی می‌توان برای شناسایی معیار توقف یا انجام بهینه‌سازی‌های هایپرپارامتر استفاده کرد. مهم‌تر از همه این‌که، خطاهای موجود در مجموعه داده اعتبارسنجی می‌تواند به شما در شناسایی این‌که آیا این مدل بیش از حد از داده‌های آموزشی برخوردار است یا خیر، کمک کند. پیش‌بینی در مجموعه داده‌های آزمایش در مدل نهایی انجام می‌شود. اگر از مجموعه داده‌های آزمایشی هرگز برای آموزش استفاده نشود، گاهی اوقات آن را مجموعه داده‌های حاوی اطلاعات نام‌گذاری می‌کنند. چندین طرح دیگر برای تقسیم داده وجود دارد. یک تکنیک معروف اعتبارسنجی متقابل نام دارد که شامل تقسیم مکرر تمام مجموعه داده‌ها به زیرمجموعه‌های آموزشی و اعتبارسنجی است. در پایان هر دوره داده‌ها با هم مخلوط شده و دوباره تقسیم می‌شوند.

مطلب پیشنهادی



دنیای امواج نامریی
یادگیری ماشین در شبکه‌های موبایل نسل آینده

بهینه‌سازی هایپرپارامتر و AutoML

بهینه‌سازی هایپرپارامتر و AutoML راهکارهایی را برای آزمایش چندین مدل مختلف برای شناسایی بهترین آن‌ها از طریق کامپیوتر ارائه می‌کند. با AutoML کامپیوتر تمام مدل‌های **یادگیری ماشین** مناسب و همچنین تمام ویژگی‌های مهندسی مناسب و تکنیک‌های گسترش‌پذیری ویژگی را امتحان می‌کند. در بهینه‌سازی هایپرپارامتر شما تعیین می‌کنید که می‌خواهید کدامیک از هایپرپارامترها را از یک مدل خاص (مثل تعداد لایه‌های مخفی)، نرخ یادگیری و نرخ رهاسازی پاکسازی کنید.

گوگل برای Google Cloud AutoML تعریف متفاوتی دارد. به جای امتحان تمام مدل‌های یادگیری ماشین مناسب، این سیستم تلاش می‌کند تا یک مدل یادگیری عمیق مرتبط (بصری، ترجمه یا زبان طبیعی) را با استفاده از یادگیری انتقال عمیق سفارشی‌سازی کند. سرویس Azure Machine Learning خدمات یادگیری انتقال را با اسامی متفاوتی از قبیل بینایی سفارشی، سخنرانی و ترجمه قابل سفارشی‌سازی و جست‌وجوی سفارشی ارائه می‌کند.

یادگیری ماشین در ابر

یادگیری ماشین و یادگیری عمیق را می‌توانید روی سیستم شخصی خود یا روی ابر اجرا کنید. AWS, Azure و Google Cloud همگی خدمات **یادگیری ماشین** را ارائه می‌کنند که بر اساس نیاز خود می‌توانید از آن‌ها استفاده کنید. این خدمات‌دهندگان شتاب‌دهنده‌های سخت‌افزاری را مطابق با نیازتان عرضه می‌کنند. علاوه بر استفاده از خدمات رایگان هر سه شرکت می‌توانید در صورت نیاز به منابع بیشتری از GPU, TPU یا FPGA و از خدمات ماهیانه پولی استفاده کنید. شما باید بین این نوع هزینه عملیاتی با هزینه سرمایه برای خرید کامپیوتر و پردازنده‌های گرافیکی یک توازن برقرار کنید. اگر به طور مداوم به آموزش تعداد بسیار زیادی از مدل‌ها نیاز دارید، باید حداقل یک پردازنده گرافیکی برای موارد استفاده شخصی‌تان خریداری کنید. بزرگ‌ترین مزیت استفاده از ابر برای یادگیری ماشین و یادگیری عمیق این است که شما می‌توانید طی چند دقیقه منابع قابل توجهی را جمع‌آوری کنید، آموزش‌های خود را به سرعت انجام دهید و منابع ابر را آزاد کنید. همچنین هر سه خدمات‌دهنده عمده این حوزه از فناوری خدمات یادگیری ماشین و یادگیری عمیق را به شکلی ارائه می‌کنند که برای راه‌اندازی آن به مهارت تخصصی در زمینه علم داده نیازی نباشد. همچنین می‌توانید از مدل‌های پیش آموزش داده شده آن‌ها استفاده کنید و مدل‌های آن‌ها را بر اساس نیاز خود سفارشی‌سازی کرده یا مدل‌های اختصاصی خود را با استفاده از فریم‌ورک‌های مطرح این شاخه از فناوری مانند Scikit-learn, PyTorch و TensorFlow ایجاد کنید.

نگاهی دقیق‌تر به یادگیری ماشین

فقط با نصب یکی از پکیج‌های یادگیری عمیق، کار با مثال‌های متعدد موجود در آن‌ها و مطالعه منابع آموزشی مرتبط، دانش خود را درباره **یادگیری ماشین** و یادگیری عمیق گسترش دهید.

برخی از این منابع شامل کتاب آنلاین رایگان Neural Networks and Deep Learning اثر مایکل نیلسن، کتاب
آنلاین رایگان A Brief Introduction to Neural Networks، اثر دیوید کریسل و کتاب راهنمای Deep Learning،
اثر هال داوم هستند.

تاریخ انتشار:

30 مرداد 1398

نشانی منبع:

<https://www.shabakeh-mag.com/workshop/15933/%D8%A2%D8%B4%D9%86%D8%A7%DB%8C%DB%8C-%D8%A8%D8%A7-%D8%A7%D9%84%DA%AF%D9%88%D8%B1%DB%8C%D8%AA%D9%85%E2%80%8C%D9%87%D8%A7%DB%8C-%D9%85%D8%B7%D8%B1%D8%AD-%D9%88-%D9%BE%D8%B1%DA%A9%D8%A7%D8%B1%D8%A8%D8%B1%D8%AF-%DB%8C%D8%A7%D8%AF%DA%AF%DB%8C%D8%B1%DB%8C-%D9%85%D8%A7%D8%B4%DB%8C%D9%86>