۱- دکتر امیرحسین بحرالعلومیان؛
دکتری تخصصی مهندسی پزشکی
۲- مهندس منیره توکلی
(کارشناس ارشدمهندسی پزشکی)

کاربرد هوش مصنوعی در تشخیص های آزمایشگاهی – بخش دوم

هوش مصنوعی چگونه یاد می گیرد؟

در حال حاضر، هوش مصنوعی با فرآیندهای نویی شامل تصمیم‌گیری تحلیلی (تحلیل داده‌ها)، عملکردی (عمل کردن براساس هوش مصنوعی تحلیلی)، تعاملی (ارتباط)، متنی (پردازش زبان طبیعی) و بصری (واقعیت افزوده) کار می کند. هوش مصنوعی بر اساس استراتژی‌های یادگیری مختلف، مانند یادگیری ماشین، یادگیری عمیق، داده‌کاوی، مدل‌سازی مبتنی بر قوانین، منطق فازی، نمایش دانش، استدلال مبتنی بر مورد، داده‌کاوی متنی، تجزیه و تحلیل بصری، و بهینه‌سازی، مدل‌هایی را برای حل مشکلات دنیای واقعی توسعه می دهد.

در ادامه، به‌طور مختصر این استراتژی‌های یادگیری را توضیح می‌دهیم.

یادگیری ماشین (Machine Learning) یک روش شناسایی الگو است که به‌طور خودکار ناهنجاری‌ها را در حجم بالایی از داده‌ها شناسایی می‌کند. این فرآیند تعاملات بین متغیرها را بر اساس روش‌های آماری، ارزیابی می کند و مؤثرترین روش استفاده از آن‌ها را برای دستیابی به هدف از پیش تعیین شده پیدا می‌کند، بدون اینکه نیاز به مداخله انسانی برای تعریف مجموعه‌ای از قوانین سخت‌گیرانه یا فرضیات برنامه‌نویسی داشته باشد. یادگیری ماشین به عنوان یک چارچوب ترجیحی برای پیاده‌سازی برنامه‌های هوش مصنوعی تبدیل شده است که با افزایش دسترسی به داده‌ها (داده‌های کلان) تقویت می‌شود. اگرچه این مفاهیم مشابه و به هم مرتبط هستند، اما از یکدیگر متمایز هستند.

یادگیری ماشین یک روش برای شناسایی الگو می باشد و شناسایی الگو (Recognition Pattern) یک رویکرد ممکن برای هوش مصنوعی است. داده‌ها (Sets Data) برای توسعه مدل ضروری هستند و معمولاً در اشکال مختلفی در دسترس هستند، مانند داده‌های ساختاریافته (بسیار سازمان‌یافته در پایگاه‌های داده (Bases Data)، داده‌های غیرساختاری (بدون فرمت از پیش تعریف شده)، داده‌های نیمه‌ساختاریافته (سازمان‌یافته اما نه در پایگاه‌های داده رابطه‌ای) و متاداده (خصوصیات داده‌ها درباره داده‌ها).

الگوریتم‌های یادگیری ماشین مانند گوسین نایو بیز (GNB)، نزدیک‌ترین همسایگی (KNN)، ماشین‌های بردار پشتیبان (SVM)، درختان تصمیم (DT)، رگرسیون خطی (LR) و ... (برای موارد بیشتر باکس ۱ را ببینید) از رایج‌ترین تکنیک‌هایی هستندکه معمولاً برای استراتژی‌های یادگیری نظارت‌شده به کار می‌روند.

این الگوریتم‌ها از ورودی‌های نمونه برای توسعه مدل و داده‌های بعدی برای پیش‌بینی مدل استفاده می‌کنند. علاوه بر پیش‌بینی بیماری‌های خاص، روش‌های دیگری مانند K-means، تحلیل مؤلفه‌های اصلی (PCA) یا همبستگی پیرسون (ρ) با حداکثر کردن واریانس بین نمونه‌ها، امکان کاوش داده‌ها برای خوشه‌بندی و کاهش ابعاد را فراهم می‌کنند. بنابراین، این روش‌ها امکان کاوش عمیق‌تری از داده‌های بیومدیکال را فراهم می‌آورند که اهمیت زیادی در تشخیص پزشکی دارد.

باکس ۱: واژه‌نامه اصطلاحات کلیدی

AUROC: مساحت زیر منحنی مشخصه عملکرد که با نرخ مثبت واقعی در برابر نرخ مثبت کاذب محاسبه می شود. این معیار اندازه‌گیری ترکیبی از عملکرد در تمام آستانه‌های طبقه‌بندی ممکن را ارائه می‌دهد.

Bootstrapping: تکنیک آماری برای استخراج نمونه با جایگذاری مجدد که امکان آموزش مکرر و آزمون ثابت را فراهم می‌کند.

اعتبارسنجی متقابل (Validation -Cross): روش نمونه‌گیری مجدد برای آزمون و آموزش بخش‌های مختلف داده در چندین تکرار از توسعه مدل.

یادگیری گروهی یا ترکیبی (Learning Ensemble): ترکیبی از پیش‌بینی‌های تخمین پایه برای بهبود مقاومت و تعمیم‌پذیری در برابر تخمین زدن منفرد.

استخراج و انتخاب ویژگی (Extraction Feature) : فرآیندی است که در آن با انجام عملیاتی بر روی داده ها، ویژگی های بارز و تعیین کننده آن مشخص می شود.

ویژگی (Feature): ورودی اطلاعاتی که در طی آموزش و ارزیابی به مدل داده می‌شود.

تابع کرنل (Function Kernel): تابعی که بر روی داده‌های غیرخطی اصلی اعمال می‌شود تا فضاهای با ابعاد بالاتر ایجاد کند که در آن داده‌ها قابل تفکیک خواهند شد.

بیش برازش(Over-Fitting): یکی از خطاهای مدل سازی در علم داده (Science Data) است و عبارت است از فرآیندی که در آن مدل آماری به طور کامل با داده‌های آموزشی تطبیق می‌یابد اما در داده‌های جدید تعمیم‌پذیری خوبی ندارد.

داده های آموزشی (Data Training): زیرمجموعه‌ای از داده‌ها که برای یادگیری و بهینه‌سازی مدل استفاده می‌شود.

داده های اعتبارسنجی (Data Validation): از داده های اعتبارسنجی برای تنظیم پارامترهای مدل و مقایسه مدل های مختلف به کار می روند. این داده ها با داده های آموزشی متفاوت بوده و نباید در مرحله آموزش استفاده شوند در غیر اینصورت مدل دچار بیش برازش خواهد شد.

داده های آزمایشی (Data Test) : برای تست عملکرد مدل نهایی شده از این داده ها استفاده می شود.

الگوریتم‌های یادگیری ماشین نظارت‌شده با الگوریتم‌های بدون نظارت فرق دارند و شامل داده‌های آموزشی هستند که در آن‌ها وضعیت واقعی داده‌ها مشخص است، به عنوان مثال، اینکه کدام افراد دچار کم‌خونی هستند و کدام سالم هستند.

بر اساس داده‌های آموزشی، الگوریتم یک مدل ایجاد می‌کند و برای پیش‌بینی وضعیت یک مجموعه از افراد که وضعیت واقعی آن‌ها ناشناخته است، به کار می‌رود. این پیش‌بینی‌ها به صورت یک مسئله طبقه‌بندی مطرح می‌شوند یعنی وضعیت‌های مجزایی مانند مراحل مختلف کم‌خونی را شناسایی می‌کند. به طور متناوب، این پیش‌بینی‌ها به عنوان یک مسئله رگرسیون تعریف می‌شوند که متغیرهای پیوسته را ارزیابی کرده و به عنوان مثال، مقدار عددی هموگلوبین را پیش‌بینی می‌کند.

اگر وضعیت واقعی داده‌ها ناشناخته باشد، یادگیری به صورت بدون نظارت انجام می شود، یعنی الگوریتم‌ها الگوهای پایه ای را در داده‌های بدون برچسب استنباط می‌کنند تا زیرخوشه‌های داده‌های اصلی را پیدا کنند، نقاط پرت را شناسایی می کنند و نیز نمایش‌های با ابعاد پایین تولید کنند. به این ترتیب، ممکن است بتوان ارتباطات جدیدی را که قابل درک نبوده‌اند، شناسایی کرد. در مثال فوق، الگوریتم می‌تواند دو خوشه جدا کند شامل: بیمارانی که دچار کم‌خونی هستند و بیمارانی که دچار کم‌خونی نیستند، و این کار را بدون اینکه مقدار واقعی هموگلوبین را بداند انجام می دهد.

یادگیری تقویتی (Learning Reinforcement): یک رویکرد مبتنی بر بازخورد است که در آن الگوریتم‌ها از طریق آزمون و خطا یاد می‌گیرند و دانش ورودی را با کاوش داده‌های ناشناخته مدیریت و متعادل می‌کنند. مدل وظیفه را با آگاهی از برخی قوانین پایه‌ای انجام می‌دهد و با وزن‌دهی به متغیرهای خاص، راه‌حل صحیح را پیدا می‌کند و یاد می گیرد. ناظر تنها باید مشخص کند که آیا پاسخ الگوریتم صحیح است؛ این شبیه به یادگیری نظارت‌شده است، اما تصمیم‌گیری (وزن‌دهی) را به آزمون و خطای الگوریتم واگذار می‌کند.

یک کاربرد به‌روز، مدیریت مداوم نرخ جریان اکسیژن برای بیماران مبتلا به COVID-19 در وضعیت بحرانی است، یعنی الگوریتم نرخ جریان مناسب را برای هر بیمار یاد گرفت و باعث کاهش نرخ مرگ و میر و صرفه‌جویی در منابع اکسیژن در دوران پاندمی شد. با این حال، این الگوریتم‌ها به عنوان «داده‌خوار» شناخته می‌شوند زیرا برای آموزش مسیرهای مختلف به مقدار زیادی داده نیاز دارند تا عملکرد پایدار را به دست آورند، که این یک محدودیت در کاربرد آن‌ها برای اطلاعات بالینی غیرساختاریافته است.

در ادامه به بررسی سایر روش های یادگیری ماشین خواهیم پرداخت.