یادگیری آماری و تعمیم‌پذیری مدل‌های یادگیری ماشین: رویکردی مبتنی بر نظریه VC و کران‌های پیچیدگی

امیرمحمدجلدی [۱]

۱- دانشجوی مقطع ارشد دانشگاه آزاد اسلامی واحد تهران غرب Jaldiamir20@gmail.com

چکیده

در سال‌های اخیر، مفهوم تعمیم‌پذیری به یکی از اصول بنیادین در یادگیری ماشین تبدیل شده است؛ چرا که عملکرد مدل‌ها در مواجهه با داده‌های جدید، به شدت به توانایی آن‌ها در تعمیم از داده‌های آموزشی وابسته است. این مقاله با رویکردی مروری، به بررسی نظریه‌های آماری تعمیم‌پذیری می‌پردازد و مفاهیم کلیدی همچون تابع زیان، ریسک تجربی، ریسک واقعی، و ظرفیت مدل را تشریح می‌کند. تمرکز اصلی بر بُعد VC (VC-dimension) است که معیاری برای سنجش ظرفیت مدل‌ها در یادگیری آماری محسوب می‌شود. همچنین، ارتباط این مفهوم با پدیده‌های بیش‌برازش و کم‌برازش، و نقش آن در کران‌های پیچیدگی نظیر PAC Bound، پیچیدگی نمونه و پیچیدگی رادماخر تحلیل می‌گردد. هدف این مقاله، فراهم آوردن درکی عمیق‌تر از سازوکارهای ریاضیاتی و نظری نهفته در عملکرد مدل‌های یادگیری و کمک به انتخاب مناسب ساختار مدل در مسائل عملی است.

کلمات کلیدی: یادگیری ماشین, بُعد VC , تعمیم‌پذیری, کم‌برازش, بیش‌برازش

۱. مقدمه

در سال‌های اخیر، یادگیری ماشین به یکی از ارکان اصلی توسعه فناوری و بهبود عملکرد سیستم‌ها در حوزه‌های متنوعی نظیر سلامت، مالی و صنایع تبدیل شده است. تصور آنکه تا چه میزان می‌توان از طریق بهره‌گیری صحیح از این فناوری در تسریع و دقت فرآیندهای تصمیم‌گیری در این حوزه‌ها تأثیرگذار بود، اهمیت آن را دوچندان می‌سازد. با این حال، پیش از پرداختن به کاربردها، ضروری است که تعریفی دقیق از این مفهوم ارائه گردد. یادگیری ماشین را می‌توان به عنوان شکلی از برآورد آماری با تمرکز بر استفاده از رایانه‌ها برای تخمین توابع پیچیده و کاهش فواصل اطمینان حول آن‌ها تعریف کرد.[۱] به‌طور کلی، یادگیری ماشین به مجموعه‌ای از روش‌های محاسباتی اطلاق می‌شود که با استفاده از تجربه (یعنی داده‌های پیشین یا آموزشی) (Training Data)، در جهت بهبود عملکرد یا ارائه پیش‌بینی‌های دقیق تلاش می‌کنند. در این زمینه، داده‌های آموزشی به عنوان نمونه‌هایی شناخته می‌شوند که به منظور آموزش الگوریتم‌های یادگیری مورد استفاده قرار می‌گیرند.[۲] هدف از فرآیند یادگیری، ساخت مدلی است که قادر باشد بر اساس داده‌های آموزشی، آموزش ببیند و سپس آموخته‌های خود را به داده‌های جدید تعمیم دهد. در این راستا، «تعمیم‌پذیری» (Generalization) به‌عنوان یکی از مفاهیم بنیادین در یادگیری ماشین، نقش کلیدی ایفا می‌کند. تعمیم‌پذیری به توانایی مدل در انجام پیش‌بینی دقیق بر روی داده‌های نادیده اطلاق می‌شود و به معنای کشف قوانینی از دل داده‌های آموزشی است که بتوان آن‌ها را با دقت بر روی داده‌های جدید نیز اعمال کرد. از این رو، تحلیل و بررسی نظری این مفهوم برای اطمینان از عملکرد مناسب مدل‌ها در

شرایط واقعی ضروری است.[۳, ۴] با این حال، اگر مدلی طراحی شود که عملکرد خوبی تنها روی داده‌های آموزشی داشته و در مواجهه با داده‌های جدید دچار افت دقت گردد، اصطلاحاً گفته می‌شود که مدل دچار بیش‌برازش (Overfitting) شده است. در مقابل، چنانچه مدل حتی در داده‌های آموزشی نیز عملکرد ضعیفی داشته باشد، با پدیده کم‌برازش (Underfitting) مواجه خواهیم بود. هر دو حالت، مانعی در برابر تعمیم‌پذیری مؤثر محسوب می‌شوند و بررسی آن‌ها بخشی از ادبیات نظری یادگیری ماشین است.[۵]

در این مقاله مروری، هدف ما تمرکز بر بنیان نظری تعمیم‌پذیری در یادگیری ماشین است. به‌طور خاص، با نگاهی تحلیلی به مفاهیم نظری از جمله بُعد VC، کران‌های احتمالی PAC و پیچیدگی Rademacher، تلاش می‌شود تا نقش این ابزارهای آماری در تحلیل و درک بهتر عملکرد مدل‌های یادگیری به‌ویژه در زمینه یادگیری عمیق مورد بررسی قرار گیرد.

۲. اهمیت تعمیم‌پذیری و نظریه VC در مدل‌های یادگیری ماشین

در دنیای امروز، هوش مصنوعی به بخش جدایی‌ناپذیری از زندگی انسان‌ها تبدیل شده است. ابزارهای مبتنی بر هوش مصنوعی اکنون در آموزش، بهداشت، صنعت و حتی در تصمیم‌گیری‌های روزمره نقش فعالی ایفا می‌کنند. یکی از چالش‌های اساسی در استفاده مؤثر از این ابزارها، طراحی مدل‌هایی است که نه‌تنها در داده‌های آموزش‌دیده عملکرد مناسبی دارند، بلکه بتوانند در مواجهه با داده‌های جدید و نادیده نیز تصمیم‌های دقیقی اتخاذ کنند. این ویژگی که با عنوان تعمیم‌پذیری شناخته می‌شود، نقشی حیاتی در عملکرد مدل‌های یادگیری ماشین ایفا می‌کند. نظریه VC و سایر چارچوب‌های نظری مرتبط، به ما اجازه می‌دهند که میزان پیچیدگی مدل‌ها را کنترل کرده و تضمین‌هایی درباره عملکرد آن‌ها در شرایط واقعی به‌دست آوریم.

در کاربردهای صنعتی، به‌ویژه در حوزه‌هایی که حساسیت بالایی دارند (مانند پیش‌بینی ریسک یا کنترل فرآیندهای پیچیده)، مدل‌هایی که قابلیت تعمیم‌پذیری بالایی دارند از اهمیت ویژه‌ای برخوردارند. استفاده از مدل‌های عمومی یا آماده در چنین حوزه‌هایی ممکن است منجر به بروز خطاهای غیرقابل جبران شود. بنابراین، درک عمیق از نظریه‌های یادگیری آماری و به‌کارگیری آن‌ها در طراحی مدل‌های خاص و دقیق، پیش‌نیازی مهم برای توسعه سیستم‌های قابل اعتماد در حوزه هوش مصنوعی است. از این‌رو، پرداختن به مفاهیم بنیادین مانند تعمیم‌پذیری و کران‌های پیچیدگی، نه‌تنها ارزش نظری دارد بلکه کاربردهای مستقیمی در طراحی هوش مصنوعی برای حوزه‌های مختلف از جمله آموزش و صنعت دارد.

۳. پیش‌نیازها

قبل از این که به سراغ بحث اصلی یعنی VC-dimension برویم باید یکسری از مفاهیم پایه‌ای و اصلی یادگیری آماری را مرور کنیم. مدل یادگیری ماشینی که طراحی می‌کنیم بر روی داده‌های آموزشی هستند و طبیعتاً مدل ما باید دارای قابلیت تعمیم پذیری باشد، قاعدتاً مدلی که طراحی می‌شود، صد در صد درست نیست و ممکن است دارای مقداری خطا باشد. سنجش میزان و درصد خطا را توسط مفاهیمی به نام ریسک تجربی(خطای تجربی) (Empirical Risk)، ریسک واقعی (خطای واقعی) (True Risk) و تابع زیان (خطا) (Loss Function) اندازه‌گیری می‌کنیم.

ابتدا با تعریف تابع زیان شروع می‌کنیم، اما برای درک بهتر مفهوم تابع زیان باید یک توضیح بسیار کوتاه از مفهوم ساده‌ای به نام برچسب بدهیم. برچسب‌ها مقدار یا دسته‌بندی‌هایی هستند که ما بر روی داده‌های مدل خود انجام می‌دهیم. به زبان ساده این کار برای نشانه‌گذاری یا دسته‌بندی داده‌های مدل است. تابع زیان، تابعی است که تفاوت یا زیان بین برچست پیش بینی شده و برچسب واقعی را اندازه گیری می‌کند. اگر مجموعه تمام برچسب های ممکن را با Y و مجموعه تمام پیش‌بینی‌های ممکن را با Y' نمایش دهیم، آنگاه تابع زیان، به شکل زیر تعریف می‌شود.[۲]

$$L: Y \times Y' \quad (1)$$

که هر جفت (Y, Y') را به یک عدد حقیقی نامنفی، نگاشت می‌کند. در بیشتر موارد داریم $Y = Y'$ و تابع زیان، دارای کران است. اما همیشه این شرایط برقرار نیست.

خطای تجربی میانگین تابع زیان بر روی داده‌ها و نمونه‌های آزمایشی است و نشان‌دهنده میزان دقت در عملکرد داده‌های مشاهده شده است که به شکل زیر تعریف می‌شود:

$$L\left(h(x_i), y_i\right) \sum_{i=1}^{m} \frac{1}{m} = L_s(h) \quad (2)$$

که در این فرمول $h(x_i)$ و y_i به ترتیب ورودی مدل و خروجی مدل و m تعداد نمونه‌های آزمایشی است.

حال مفهوم دیگری وجود دارد به نام حداقل‌سازی خطای تجربی که رویکردی است برای پیدا کردن مدلی که کمترین خطای تجربی را روی داده‌ها دارد. در یادگیری ماشین، هدف اصلی در آموزش یک مدل، عملکرد خوب آن بر روی داده‌های جدید و دیده نشده است. به همین دلیل یکی از معیارهای مهم برای سنجش کیفیت یک مدل، خطای واقعی یا ریسک واقعی است. این معیار بیان می‌کند که پیش بینی های مدل به طور میانگین تا چه اندازه با خروجی‌های واقعی داده‌ها اختلاف دارند. به عبارتی ریسک واقعی برابر است با امید ریاضی تابع زیان نسبت به توزیع داده‌های واقعی.[۶]

$$E_{XY^{P\sim(X,Y)}} \sim [L(f(X), Y)] = R(f) \quad (3)$$

در آن f مدل یادگیرنده، L تابع زیان، مانند خطای صفر و یک یا خطای مربعی، $XY^{P\sim(X,Y)}$ توزیع احتمال ناشناخته داده‌ها، و E امید ریاضی است. در نظریه یادگیری آماری، ظرفیت به میزان پیچیدگی یا انعطاف پذیری یک کلاس یا مدل اشاره دارد و بدین معناست که مدل یادگیری تا چه اندازه قادر است الگوهای مختلف را در داده‌ها تطبیق دهد.[۷]

۴. نظریه VC-dimension

بُعد VC یا Vapnik-Chervoneniks dimension یک معیار بنیادین در نظریه یادگیری آماری است که ظرفیت یا پیچیدگی یک کلاس فرضیه (مدل فرضی) را اندازه‌گیری می‌کند. این مفهوم به صورت رسمی به عنوان بیشترین تعداد نقاطی تعریف می‌شود که مدل و کلاس فرضی ما می‌تواند به همه برچسب‌گذاری‌های دودویی ممکن پاسخ صحیح بدهد. به عبارت دیگر مجموعه‌ای از نقاط را شکسته (shatter) می‌کند. اگر یک کلاس فرضیه قادر باشد تمامی 2^n ترکیب برچسب‌دهی ممکن به n نقطه را مدل‌سازی کند، گفته می‌شود که آن کلاس، آن n نقطه را شکسته است. بنابراین، بعد VC برابر است با بیشترین n برای این که چنین شکستی ممکن باشد. این مفهوم به‌عنوان شاخصی از ظرفیت تعمیم‌پذیری مدل یادگیری عمل می‌کند؛ چرا که هرچه بعد VC بزرگ‌تر باشد، مدل توانایی بیشتری برای انطباق با داده‌ها دارد، اما همچنین ممکن است بیشتر مستعد بیش‌برازش شود. از این رو، بعد VC نقشی کلیدی در درک رابطه بین خطای تجربی، خطای واقعی و

تضمین‌های تعمیم‌پذیری ایفا می‌کند.[۸, ۹]

ظرفیت در مدل بدین معناست که مدل تا چه حدی می‌تواند الگوهای پیچیده را از داده‌ها یاد بگیرد. مدلی با ظرفیت بالا، دارای انعطاف‌پذیری زیادی برای تطبیق با داده‌هاست و می‌تواند مرزهای تصمیم‌گیری پیچیده‌تری بسازد، در مقابل مدل‌های با ظرفیت پایین تنها قادر به یادگیری الگوهای ساده هستند. از منظر بُعد VC هم، هرچه VC بیشتر باشد، مدل توانایی بیشتری در یادگیری دسته‌بندی‌های متنوع دارد. این یعنی مدل ظرفیت بیشتری دارد.

۵. ارتباط بین VC-dimension و Overfitting

در یادگیری ماشین، یکی از اهداف اصلی، یافتن تعادل میان یادگیری و تعمیم است. اگر مدل بسیار انعطاف‌پذیر باشد (ظرفیت بالا)، ممکن است نه‌تنها الگوهای واقعی داده‌ها را بیاموزد، بلکه نویز و اتفاقات تصادفی موجود در داده‌های آموزشی را نیز حفظ کند. این پدیده همان بیش‌برازش است. در این حالت، مدل در داده‌های آموزشی عملکرد عالی دارد، اما در داده‌های جدید و نادیده عملکرد ضعیفی خواهد داشت. در مقابل، مدل‌هایی با VC-dimension پایین، ممکن است دچار کم‌برازش (underfitting) شوند؛ یعنی حتی الگوهای اصلی داده‌ها را هم به‌درستی یاد نمی‌گیرند. بنابراین، مدل با VC بالا دارای ظرفیت بالا بوده و خطر بیش برازش دارد و مدل با VC پایین، دارای ظرفیت محدود و خطر کم‌برازش است.

۶. اهمیت انتخاب مناسب ظرفیت

هدف در طراحی مدل، یافتن نقطه‌ای بین این دو افراط است؛ یعنی ظرفیتی که برای یادگیری مناسب و تعمیم‌پذیر کافی باشد. این تعادل از طریق انتخاب مدل مناسب (مثلاً تعداد لایه‌ها در شبکه عصبی، یا عمق درخت تصمیم) و استفاده از روش‌هایی مانند منظم‌سازی یا اعتبارسنجی متقابل به‌دست می‌آید.

۷. کران‌های پیچیدگی (Complexity Bounds) در یادگیری ماشین

در یادگیری ماشین، یکی از چالش‌های اصلی، درک توانایی تعمیم (generalization) مدل‌ها است؛ یعنی اینکه چقدر یک مدل می‌تواند عملکرد خوبی روی داده‌های جدید و دیده‌نشده داشته باشد. برای تحلیل این موضوع، از مفهومی به نام کران‌های پیچیدگی استفاده می‌شود. این کران‌ها روابطی ریاضی هستند که میزان خطای تعمیم را با توجه به ظرفیت مدل، تعداد داده‌های آموزشی و میزان خطای قابل قبول تخمین می‌زنند. کران‌های پیچیدگی، به‌ویژه PAC Bound و Sample Complexity، ابزارهایی بنیادین برای تحلیل یادگیری ماشین هستند. آن‌ها به ما می‌گویند که برای جلوگیری از بیش‌برازش (Overfitting) یا کم‌برازش (Underfitting)، چقدر داده نیاز داریم و تا چه حد می‌توان به تعمیم یک مدل اعتماد کرد. این کران‌ها بر اساس ظرفیت مدل (مانند VC-dimension یا Rademacher complexity) مشخص می‌شوند و به‌صورت کمی رابطه بین داده، مدل و تعمیم را تبیین می‌کنند.

۸. کران PAC (Probably Approximately Correct)

چارچوب PAC که توسط Leslie Valiant در سال ۱۹۸۴ معرفی شد، از مفاهیم پایه‌ای در تئوری یادگیری آماری است. در این چارچوب، یک مدل یادگیری، موفق تلقی می‌شود اگر با احتمال بالا، عملکردی تقریباً درست داشته باشد. کران PAC کلاسیک برای یک کلاس فرضیه با بعد VC برابر d به صورت زیر بیان می‌شود:

$$\left(\left(\frac{13}{\varepsilon}\right)\log\left(\frac{2}{\delta}\right)+8d\log 4\right)\frac{1}{\epsilon}\leq m \qquad (4)$$

که در آن m حداقل تعداد نمونه مورد نیاز، ε خطای مجاز، δ احتمال شکست و d بعد VC است. این رابطه نشان می‌دهد که با افزایش بعد VC، برای رسیدن به یک خطای پایین و اعتماد بالا، به نمونه‌های بیشتری نیاز داریم. بنابراین، کران PAC پیوندی بین ظرفیت مدل و توانایی تعمیم آن برقرار می‌کند.[۶]

۹. پیچیدگی نمونه (Sample Complexity)

کرانی که تعیین می‌کند برای دستیابی به عملکرد قابل قبول در یادگیری، چه تعداد داده آموزشی مورد نیاز است. این پیچیدگی معمولاً تابعی از بعد VC، میزان خطای قابل قبول و احتمال اطمینان است. در بسیاری از منابع، این کران به صورت تقریبی چنین است:

$$\left(\frac{d+\log\left(1/\delta\right)}{\epsilon^2}\right)O=N \qquad (5)$$

این کران از اصول PAC استخراج شده و برای کاربردهای عملی در تعیین اندازه داده‌های آموزشی بسیار مهم است.[۱۰]

۱۰. پیچیدگی رادماخر (Rademacher Complexity)

این پیچیدگی معیاری داده‌محور برای اندازه‌گیری ظرفیت یک کلاس فرضیه است که مبتنی بر این ایده است که چقدر یک کلاس فرضیه می‌تواند روی داده‌های تصادفی برازش کند. کران رادماخر فرمولی مانند زیر دارد:

$$E_{S,\sigma}\left[SUP_{h\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma_i\,h(x_i)\right]=R_m(H) \qquad (6)$$

که در آن σ_i متغیرهای تصادفی مستقل با مقدارهای ±۱ هستند. کران‌های تعمیم مبتنی بر رادماخر معمولاً دقیق‌تر از کران‌های مبتنی بر VC هستند و می‌توانند در بسیاری از الگوریتم‌های مدرن یادگیری کاربرد داشته باشند.[۲]

۱۱. پیچیدگی محاسباتی (Computational Complexity)

کران‌هایی که زمان یا منابع محاسباتی مورد نیاز برای اجرای یک الگوریتم را مشخص می‌کنند. این کران‌ها معمولاً به صورت زمانی (Time Complexity) یا حافظه (Space Complexity) بیان می‌شوند و با نمادهایی مانند $O(n)$، $O(n^2)$ یا $O(2^n)$ مشخص می‌گردند. این پیچیدگی‌ها در تحلیل عملی الگوریتم‌ها کاربرد دارند ولی تمرکز اصلی در تئوری یادگیری، روی پیچیدگی‌های آماری مانند VC و رادماخر است.

۱۲. چگونه VC-dimension می‌تواند تخمینی از توانایی تعمیم مدل بدهد؟

در یادگیری ماشین، یکی از چالش‌های اساسی آن است که چگونه می‌توان بر اساس عملکرد مدل بر داده‌های آموزشی، توانایی آن را در تعمیم به داده‌های دیده‌نشده ارزیابی کرد. در این زمینه، بعد VC یا VC-dimension به‌عنوان یکی از مفاهیم بنیادین در نظریه یادگیری آماری، ابزار مناسبی برای سنجش توانایی تعمیم مدل فراهم می‌کند. VC-dimension ظرفیت مدل را از منظر توانایی آن در شناسایی الگوها و جداسازی برچسب‌ها در داده‌ها بیان می‌کند. به‌طور خاص، VC-dimension بیشینه تعداد نقاطی است که یک کلاس فرضیه می‌تواند آن‌ها را به تمام حالت‌های ممکن به‌درستی برچسب‌گذاری (shatter) کند. مدلی با VC بالا می‌تواند الگوهای پیچیده‌تری را یاد بگیرد، اما در عین حال، برای اطمینان از عملکرد صحیح آن بر داده‌های جدید، نیاز به داده‌های آموزشی بیشتری دارد.

در چارچوب یادگیری PAC، کران‌های نظری برای میزان تعمیم‌پذیری مدل‌ها ارائه شده‌اند که مستقیماً به VC-dimension مرتبط‌اند. برای هر کلاس فرضیه H با VC برابر با d اگر مدل $h \in H$ بر روی نمونه آزمایشی S با اندازه m آموزش داده شود، آنگاه می‌توان با احتمال بالا تضمین کرد که خطای واقعی L_D (h) با خطای تجربی $L_S(h)$ تفاوت چندانی نخواهد داشت. مشروط بر آنکه m به‌اندازه کافی بزرگ باشد. کران احتمالاتی زیر در حالت realizable بیان‌گر این رابطه است:

$$\varepsilon \geq |L_S(h) - L_D(h)| \Rightarrow \left(\frac{4}{\delta} 4d.\log\frac{2e}{\varepsilon} + \log\right)\frac{1}{\varepsilon} \leq m \quad , \quad (1,0) \ni \delta\forall \qquad (7)$$

در این فرمول ε حداکثر اختلاف قابل‌قبول بین خطای تجربی و خطای واقعی است، δ احتمال شکست یا اطمینان آماری، m تعداد نمونه آزمایشی، d متغیر vc، $L_S(h)$ خطای تجربی و $L_D(h)$ خطای واقعی است.[۲, ۶]

این رابطه نشان می‌دهد که هرچه VC-dimension کلاس فرضیه بیشتر باشد، برای رسیدن به یک تخمین قابل اعتماد از توانایی تعمیم، به تعداد بیشتری نمونه آموزشی نیاز است. بنابراین، VC-dimension به‌صورت غیرمستقیم نقش مهمی در تعیین میزان داده مورد نیاز برای یادگیری مؤثر ایفا می‌کند. به‌عبارت‌دیگر، VC-dimension معیار نظری مهمی برای سنجش پیچیدگی مدل و توانایی آن در تعمیم‌پذیری محسوب می‌شود.

نتیجه‌گیری

در این مقاله، با نگاهی تحلیلی به نظریه‌های آماری یادگیری ماشین، به بررسی مفاهیم کلیدی تعمیم‌پذیری از جمله تابع زیان، ریسک تجربی، ریسک واقعی، و به‌ویژه بُعد VC پرداختیم. نشان داده شد که بُعد VC به‌عنوان معیاری برای سنجش ظرفیت مدل، نقش بنیادینی در درک سازوکار تعمیم‌پذیری ایفا می‌کند و ارتباط مستقیمی با پدیده‌های کم‌برازش و بیش‌برازش دارد. همچنین، با معرفی کران‌های پیچیدگی نظیر PAC Bound و پیچیدگی نمونه، مشخص شد که چگونه می‌توان به‌صورت نظری، عملکرد مدل‌ها را بر روی داده‌های نادیده پیش‌بینی و تحلیل کرد.

درک دقیق این مفاهیم، به‌ویژه در کاربردهای حساس صنعتی و علمی، نقش کلیدی در طراحی مدل‌های یادگیری مؤثر، پایدار و قابل‌اعتماد دارد. مدل‌هایی که بدون درنظر گرفتن ظرفیت مناسب طراحی می‌شوند، یا بیش از حد ساده هستند و توانایی استخراج الگوهای واقعی را ندارند، یا بیش از حد پیچیده‌اند و درگیر جزئیات تصادفی داده‌های آموزشی می‌شوند. از این رو، انتخاب ظرفیت بهینه مدل و استفاده از ابزارهای نظری برای تحلیل تعمیم‌پذیری، پیش‌شرطی حیاتی برای موفقیت در پیاده‌سازی سیستم‌های هوش مصنوعی است.

در نهایت، پیوند میان نظریه‌های آماری و کاربردهای عملی یادگیری ماشین، نه‌تنها به ارتقاء دقت مدل‌ها کمک می‌کند، بلکه مسیر طراحی الگوریتم‌های قابل اتکا برای شرایط واقعی را هموار می‌سازد. این مقاله تلاش کرد تا با نگاهی منسجم و شفاف، زمینه‌ای برای درک عمیق‌تر این پیوند نظری و عملی فراهم آورد.

1. Courville, I.G.a.Y.B.a.A., Deep Learning. 2016: MIT Press.

2. Chen, L.-P., Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar: Foundations of machine learning, second edition. Statistical Papers, 2019. **60**(5): p. 1793-1795.

3. Abbas, A., et al., Effective dimension of machine learning models. arXiv preprint arXiv:2112.04807, 2021.

4. Sonoda, S., et al., Lean Formalization of Generalization Error Bound by Rademacher Complexity. arXiv preprint arXiv:2503.19605, 2025.

5. Zhang, C., et al., Understanding deep learning requires rethinking generalization. arXiv preprint arXiv:1611.03530, 2016.

6. Shalev-Shwartz, S. and S. Ben-David, Understanding machine learning: From theory to algorithms. 2014: Cambridge university press.

7. Zhang, X., Empirical Risk Minimization, in Encyclopedia of Machine Learning, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 312-312.

8. Vapnik, V.L., Esther
Lecun, Yann, Measuring the VC-Dimension of a Learning Machine. Neural Computation, 1994. **6**(9): p. 851-876.

9. Anthony, M. and P.L. Bartlett, Neural network learning: Theoretical foundations. 2009: cambridge university press.

10. Hanneke, S., The optimal sample complexity of PAC learning. Journal of Machine Learning Research, 2016. **17**(38): p. 1-15.