



حل عددی معادلات دیفرانسیل با مشتقات جزئی با ضرایب نامعین به کمک شبکه عصبی

پایان نامه کارشناسی ارشد ریاضیات کاربردی - آنالیز عددی

ساجد زرین پور نشرودکلی

استاد راهنما: دکتر خدیجه ندائی اصل

استاد مشاور: دکتر پروین رزاقی

شهریور ۱۳۹۹

چکیده

برای مدل سازی پدیده های واقعی با معادلات دیفرانسیل با مشتقات جزئی که شامل عدم قطعیت است، یکی از مشکلات وجود مجموعه ای از پدیده ها است که به عنوان مشکلات ابعاد بالا شناخته می شوند. خوشبختانه، اغلب تغییرات متغیرهای مدل می توانند توسط تعداد کمی خصوصیات دامنه توسط روش های کاهش مدل، ثبت شوند. برای مثال، می توان با استفاده از روش های مبتنی بر شبکه های عصبی متغیرهای مورد نظر را به عنوان تابعی از ضرایب ورودی اندازه گیری کرد. در این صورت، نمایش پذیری متغیرها توسط چنین شبکه ای را می توان با دید شبکه عصبی به عنوان یک تحول زمانی برای پیدا کردن جواب های مدل توجیه کرد. در این پایان نامه، ما یک روش میانبر برای پیدا کردن جواب های مدل روی دو معادله دیفرانسیل با مشتقات جزئی معروف در فیزیک و مهندسی را بازبینی مینمائیم. همچنین، ما به سراغ بررسی یک روش عددی سنتی از نظر تئوری خواهیم رفت و از این طریق، احتمالات جدیدی برای استفاده از شبکه های عصبی در حل معادلات دیفرانسیل را مطرح خواهیم نمود.

واژه های کلیدی: شبکه های عصبی، روش تفاضلات متناهی، روش المان های متناهی، معادلات دیفرانسیل با مشتقات جزئی، عدم قطعیت.

فهرست

دو	چکیده
۱	مقدمه
۲	انگیزه و هدف
۴	تعریف مسئله
۵	شبکه‌های عصبی
۷	روش پیشنهادی
۹	نتایج
۱۰	نتیجه‌گیری و کارهای پیش‌رو
۱۱	واژه‌نامه فارسی به انگلیسی

مقدمه

مقدارسنجی عدم قطعیت (۱۵) در فیزیک و مهندسی اغلب شامل مطالعه معادلات دیفرانسیل با مشتقات جزئی با میدان ضرایب تصادفی است. برای درک رفتار یک سیستم شامل عدم قطعیت، می‌توان کمیت‌های فیزیکی مشتق شده از معادلات دیفرانسیل توصیف کننده آن سیستم را به عنوان توابعی از میدان ضرایب استخراج کرد. اما حتی با گسسته‌سازی مناسب روی دامنه معادله و برد متغیرهای تصادفی، این کار به طور ضمنی به حل عددی معادله دیفرانسیل با مشتق جزئی به تعداد نمایی می‌انجامد.

یکی از روش‌های متداول برای مقدار سنجی عدم قطعیت، روش نمونه برداری مونته کارلو است. گرچه این روش در بسیاری از موارد کاربردی است اما کمیت اندازه‌گیری شده ذاتاً دارای نویز است. به علاوه، این روش قادر به پیدا کردن جواب‌های جدید در صورتی که قبلاً نمونه‌گیری نشده باشند، نیست. ما به دنبال یافت روشی هستیم که نویز داده‌ها در جواب آن تأثیر چندانی نداشته باشند و همچنین، قادر به ارائه جواب برای حالاتی که قبلاً نمونه‌گیری نشده باشند هم باشد.

روش گالرکین تصادفی با استفاده چند جمله‌ای‌های آشوب یک جواب تصادفی را روی فضای متغیرهای تصادفی بسط می‌دهد و به این طریق مسئله با بعد بالا را به تعدادی معادله دیفرانسیل با مشتقات جزئی معین تبدیل می‌کند. این گونه روش‌ها به دقت زیادی درباره تعیین توزیع عدم قطعیت نیازمند هستند و از آن‌جا که پایه‌های استفاده شده مستقل از مسئله هستند، وقتی بعد متغیرهای تصادفی بالا باشد هزینه محاسباتی بسیار زیاد خواهد شد. ما به دنبال مطالعه نقطه ضعف روش‌های عددی سنتی همانند روش گالرکین تصادفی و روش المان‌های متناهی هستیم. برای این منظور، تئوری روش المان‌های متناهی را مورد مطالعه دقیق‌تر قرار می‌دهیم.

هدف کار ما پارامتری کردن جواب یک معادله دیفرانسیل معین به کمک شبکه‌های عصبی (ساخت نمایشی دیگر برای جواب بر پایه ترکیب توابع) و سپس استفاده از روش‌های بهینه‌سازی (مانند روش برای SGD) یافتن جواب معادله است. در این پایان‌نامه تابع مورد نظر برای پارامتری‌سازی روی میدان

ضرایب معادله دیفرانسیل با مشتقات جزئی تعریف شده است. در واقع ما به دنبال کاهش بعد مبتنی بر نمایش شبکه عصبی برای حل معادلات دیفرانسیل با مشتقات جزئی همراه با عدم قطعیت هستیم.

انگیزه و هدف

مدل سازی طبیعت همیشه با پارامترهایی همراه است که مقادیر آنها از کنترل ما خارج است. اما عموماً ما درباره محدوده تغییرات این پارامترها اطلاعاتی داریم. به معادلاتی که شامل این گونه پارامترها هستند، معادلات با ضرایب عدم قطعیت گوییم. به طور مثال، در مورد حرکت مایعات، به طور مثال نفت، در سفره های زیر زمینی؛ برای بیان شیوه حرکت مایعات نیاز به دانستن مکان حفره ها داریم. این امر را می توان به صورت رسانایی مؤثر در حضور ناخالصی نیز در نظر گرفت. به عنوان مثالی دیگر، مسئله ای را مطرح می کنیم که نقطه شروع این رساله بوده است.

برای تشخیص سرطان پستان روش های متعددی موجود است. از جمله آن ها می توان به تصویر برداری پستان با بازتابش اشعه ایکس (XRM)^۱، تصویربرداری با استفاده از ارتعاشات مغناطیسی (MRI)^۲، تصویربرداری فراصوت (US)^۳، توموسنتز دیجیتال (DBT)^۴، ماموگرافی انتشار پوزیترون (PET)^۵ و توموگرافی فراصوت (UST)^۶ اشاره کرد. هر کدام از این روش ها اطلاعات را به طرق مختلفی نمایش می دهند، به این معنا که غده ای که در یکی از این روش ها غیر قابل تشخیص است در روش دیگر قابل تشخیص است؛ غده ای که در یک روش بافت مشکوک معرفی می شود، در روش دیگر می تواند به عنوان غده ای سالم و طبیعی معرفی شود. و این موضوع باعث ایجاد مشکلات بسیاری در روند تشخیص و برنامه ریزی درمان می شود. در این مرحله، راه حلی که به ذهن می رسد، ترکیب نتایج حاصل از این

¹ Projection X-ray mammography

² Magnetic resonance imaging

³ Ultra sound

⁴ Digital breast tomosynthesis

⁵ Positron emission mammography

⁶ Ultra sound tomography

روش‌ها برای بالابردن ضریب دقت است؛ لیکن مشکل دیگری مانع این کار می‌شود. بافت پستان بسیار کشسان است به راحتی تغییر فرم می‌دهد و هرکدام از این روش‌ها نیز به حالت خاصی از قرارگیری بیمار نیاز دارد. به طور مثال، طی MRI بیمار در حالت دمر قرار دارد ولی برای تصویربرداری فراصوت بیمار به پشت می‌خوابد. علاوه بر این، در روش بایوپسی راهنمایی شده توسط MRI¹ بافت پستان توسط صفحه‌های سخت و غیرقابل انعطافی بی حرکت می‌شوند که منجر به فشرده شدن بافت نیز می‌شود. بنابراین، شکل، اندازه و مکان غده در این تصاویر متفاوت خواهد بود. این امر مقایسه تصاویر را با سختی بسیار همراه می‌کند. علاوه بر این، برای برنامه‌ریزی پیش از جراحی، پزشک نیاز به دانستن مکان و اندازه دقیق غده دارد. بنابراین، نیاز به توسعه الگوریتم‌های ثبت غیرسخت² احساس می‌شود.

روش‌هایی مبتنی بر روش المان‌های متناهی³ برای حل این مسئله ارائه شده‌اند. اما مشکل عمده این روش‌ها هزینه محاسباتی بالای آنها است. به طور متوسط اجرای یک شبیه‌سازی صد و بیست دقیقه به طول می‌انجامد که مقرون به صرفه نیست. ما به دنبال ارائه روشی برای کاهش این هزینه محاسباتی با استفاده از شبکه‌های عصبی و ارائه یک مدل مختص به بیمار در زمانی قابل قبول بودیم. این امر مستلزم در نظر گرفتن ضریب کشسانی بدن بیمار، که یک ضریب عدم قطعیت است، می‌باشد. از این رو، برآن شدیم که بدنبال حل عددی معادلات دیفرانسیل (بیضوی) به کمک شبکه‌های عصبی باشیم. در این رساله، بدنبال حل عددی معادلات دیفرانسیل با مشتقات جزئی بیضوی خطی و غیر خطی ناهمگن هستیم. روشی که ما در صدد معرفی آن هستیم، یک روش کاهش بعد برای محاسبه جواب بدون نیاز به حل مستقیم معادله دیفرانسیل است. به عبارت دیگر، در روش‌های عددی متدوال، فرد بدنبال پیدا کردن تقریبی از جواب در یک فضای متناهی یا غیر متناهی است که در حالت متناهی به صورت ترکیب خطی از پایه‌هایی نوشته می‌شوند که در روند روش روی نقاط رأسی بدست می‌آیند. ما به دنبال راهی برای حذف ساخت این پایه‌ها به کمک شبکه عصبی هستیم و به این وسیله در پی کاهش بعد فضای جواب با حفظ خصوصیات اصلی مورد نیاز خود در جواب هستیم. با توجه به اینکه نمایش

¹ MRI-guided biopsy

² Non-rigid registration algorithm

³ Finite element method

جواب بدست آمده از شبکه‌های عصبی به صورت ترکیبی متناهی از توابع بدست می‌آید می‌توانیم با تعویض نمایش جواب از حالت خطی (روش‌های عددی متداول) با حالت غیر خطی (نمایش شبکه‌های عصبی) به این مهم دست یابیم. ایده، استفاده از شبکه‌های عصبی برای یادگیری نگاشتی از دامنه ضرایب عدم قطعیت به فضای جواب بر اساس مجموعه داده‌ای از قبل محاسبه شده است.

تعریف مسئله

در این پایان نامه، هدف ما بررسی یک مدل میانبر برای حل مسائل معادلات دیفرانسیل با مشتقات جزئی به کمک شبکه‌های عصبی است. به عبارت روشن‌تر یافتن نگاشتی از فضای عدم قطعیت مسئله به فضای جواب. مسائلی که در این رساله برای حل انتخاب شده اند از این جهت حائز اهمیت بوده‌اند که هر دو حالت خطی و غیر خطی، معادلات دیفرانسیل غیر همگن شامل عدم قطعیت را پوشش می‌دهند.

یافتن ضریب رسانایی مؤثر در محیط ناهمگون

معادله اول، رسانایی مؤثر در یک جهت انتخاب شده درون یک ماده غیر همگون را توسط ضریب رسانش توصیف می‌کند. ماده غیر همگون، ماده‌ای است که در آن خصوصیات مورد توجه در تمامی نقاط یکسان نیستند. این امر ممکن است به دلایلی همچون جنس‌های گوناگون مواد تشکیل دهنده یا چگالی‌های متفاوت مربوط باشد. فرض ما بر آن است که ضریب رسانایی ماده در جهات متفاوت یکسان نباشد و این ضریب را با $a(x)$ نمایش می‌دهیم. با فرض انتخاب یک جهت دلخواه ثابت $\xi \in \mathbb{R}^d$ ، میزان ضریب رسانش در آن جهت مطلوب است. به عبارت دقیق‌تر، جواب معادله زیر مد

نظر است

$$A_{\text{eff}}(a) = \min_{u(x)} \int_{[0,1]^d} a(x) \|\nabla u(x) + \xi\|^2 dx.$$

یافتن ضریب رسانایی مؤثر در محیط ناهمگون

معادله دوم، معادله غیرخطی شرودینگر دو بعدی است. هدف از این معادله، یافتن میزان انرژی حالت پایه الکترون با پتانسیل اولیه همراه با عدم قطعیت است. حالت پایه سطح انرژی ای است که الکترون مایل به اخذ آن در دمای صفر مطلق می باشد. این معادله به صورت یک مسئله مقدار ویژه به صورت زیر تعریف می شود، که هدف ما در حل این مسئله یافتن کوچکترین مقدار ویژه آن (حالت پایه) است

$$-\Delta u(x) + a(x)u(x) + \sigma u(x)^3 = E_\sigma u(x)$$

$$x \in [0, 1]^d, s.t. \int_{[0,1]^d} u(x)^2 dx = 1.$$

شبکه های عصبی

شبکه عصبی^۱ از تعدادی واحد متصل به هم نام نورون^۲ تشکیل می شود. هر نورون دارای یک وضعیت داخلی است که در ترکیب با داده ورودی تغییر می کند و توسط تابع فعال سازی^۳ خروجی نورون را به حالت روشن یا خاموش تغیر می دهد. به عبارت ریاضی، هر نورون دارای اسکالره های داخلی به نام وزن^۴ و بایاس^۵ و تابع فعال سازی ای است که به ترتیب با W و b و ϕ نمایش داده می شوند. و خروجی نورون در این صورت با فرض اینکه X ورودی نورون باشد به صورت $\phi(WX + b)$ خواهد بود. برای این

¹ Neural network

² Neuron

³ Activation function

⁴ Weight

⁵ Bias

ساختار تابعی به عنوان تابع انحراف^۱ به عنوان تابع هدف تعریف می‌شود که در حالت یادگیری تحت نظارت^۲ به عنوان معیاری برای تعیین انحراف جواب‌های شبکه از جواب‌های واقعی معرفی شده به شبکه است. در این صورت، می‌توان روند یادگیری یک شبکه عصبی را معادل با یک مسئله کمینه‌سازی برای کمینه کردن میزان این تابع هدف در نظر گرفت. ابزار شبکه برای کمینه سازی این تابع هدف، تغییر وزن‌ها و بایاس‌های نورون‌های خود است. قضیه زیر که به قضیه تقریب جهانی مشهور است، شرایط استفاده از شبکه‌های عصبی برای تقریب جواب مسئله را بیان می‌کند.

قضیه. (قضیه تقریب جهانی)

۱. (حالت نامتناهی) فرض کنید $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ یک تابع غیر ثابت پیوسته بی‌کران باشد که آن را تابع فعال‌سازی می‌نامیم. فرض کنید I_m بیان‌گر ابر مکعب m -بعدی $[0, 1]^m$ باشد، و فضای توابع پیوسته حقیقی مقدار روی I_m با $C(I_m)$ نمایش داده شود. در این صورت، به ازای هر $\varepsilon > 0$ دلخواه و هر تابع $f \in C(I_m)$ ، ثوابت حقیقی مانند $v_i, b_i \in \mathbb{R}$ و بردارهای $w_i \in \mathbb{R}^m$ برای $i = 1, \dots, N$ وجود دارند، به طوری که می‌توانیم

$$F(x) = \sum_{i=1}^N v_i \varphi(w_i^T x + b_i)$$

را به عنوان تقریبی از f ارائه دهیم که عبارت است از:

$$|F(x) - f(x)| < \varepsilon$$

که در آن $x \in I_m$ است. به عبارت دیگر، توابع به شکل $F(x)$ در $C(I_m)$ چگال‌اند.

این نتیجه به ازای هر زیر مجموعه فشرده دیگری از \mathbb{R}^m به جای I_m برقرار است.

¹ Loss function

² Supervised learning

۲. (حالت کران‌دار) در شبکه‌های کران‌دار، برای هرتابع انتگرال‌پذیر لبگ مانند $f: \mathbb{R}^n \rightarrow \mathbb{R}$ و

هر $\epsilon > 0$ یک شبکه ReLu^۱ کامل \mathcal{A} با عرض $d_m \leq n + ۴$ ، به گونه ای موجود است که $F_{\mathcal{A}}$ نمایش داده شده با این شبکه در رابطه

$$\int_{\mathbb{R}^n} |f(x) - F_{\mathcal{A}}(x)| dx < \epsilon$$

صدق نماید.

روش پیشنهادی

معادله تعیین ضریب رسانش مؤثر تنها در یک بعد، و معادله شرودینگر در دو بعد حل خواهند شد. هدف این انتخاب تأکید بر این نکته بوده است که با اندک تغییری در ساختار ورودی‌های شبکه، می‌توان از این مدل برای حل معادلات دیفرانسیل در هر بعد دلخواهی استفاده کرد. دلیل این امر نیز استفاده از محاسبات تنسوری در ساختار شبکه عصبی می‌باشد.

توجه خواننده را به این نکته جلب می‌نمائیم که روش عددی‌ای که برای ساخت پایگاه داده اولیه مورد نیاز برای شبکه عصبی استفاده می‌شود، از جهتی حائز اهمیت است؛ زیرا به میزانی که خطای ما روی داده اولیه کمتر باشد اطمینان ما از جواب بر روی شبکه عصبی بیشتر است. از طرفی، بررسی اینکه این خطا در داده‌های اولیه چگونه در شبکه عصبی منتشر می‌شود و شبکه تا چه میزان به خطا در داده‌های ورودی خود حساس است، از حیثه این پایان نامه خارج است. بنابراین، از آنجایی که شبکه عصبی صرفاً از یک جدول داده بهره می‌برد و نسبت به اینکه این جدول داده‌ای از چه راهی بدست آمده اطلاع قبلی ندارد، در این پایان نامه فرض بر این است که روش عددی استفاده شده برای ایجاد پایگاه داده با شبکه عصبی غیر مرتبط است و بنابراین از هر روش عددی‌ای می‌توان استفاده نمود. به همین منظور

¹ Rectified linear unit

از روش تفاضلات متناهی برای تولید پایگاه داده در هر دو مسئله استفاده نموده‌ایم. برای حل معادلات با روش تفاضلات متناهی، معادله تعیین ضریب رسانش مؤثر روی یک شبکه نه نقطه‌ای متساوی‌الفاصله، با مقادیر ضرایب عدم قطعیت با توزیع نرمال $\mathcal{U}[0.3, 1.5]$ ، و معادله شرودینگر غیر خطی روی یک شبکه هشتاد و یک نقطه‌ای متساوی‌الفاصله (گسسته سازی نه نقطه‌ای هر کدام از ابعاد) با مقادیر ضرایب عدم قطعیت با توزیع نرمال $\mathcal{U}[1, 16]$ به تعداد نمونه‌های مورد نیاز حل می‌شوند. سپس درصدی از تکرارها (در اینجا هفتاد و پنج درصد) به عنوان داده برای مرحله آموزش و الباقی برای مرحله آزمون کنار گذاشته می‌شوند.

شبکه عصبی متشکل از سه بخش است. بخش اول و سوم قرینه یکدیگر و متشکل از لایه‌های پیچشی^۱ هستند که به واسطه بخش دوم که یک استخر مجموع^۲ است به هم متصل شده‌اند. ورودی این شبکه برای رسانایی مؤثر و یک ماتریس برای معادله شرودینگر است. دقت شود که در حالت دو بعدی، قبل از لایه‌های پیچشی، ابعاد داده ورودی گسترش می‌یابد. این امر با توجه به اینکه شرط مرزی مسئله دوره‌ای است، به صورت گسترش دوره‌ای انجام می‌شود. خروجی شبکه در هر دو حالت یک اسکالر است. که در مورد ضریب رسانائی مؤثر، این اسکالر برابر ضریب رسانائی مؤثر در جهت ثابت \hat{e} و در مورد معادله شرودینگر، برابر با سطح انرژی پایه است.

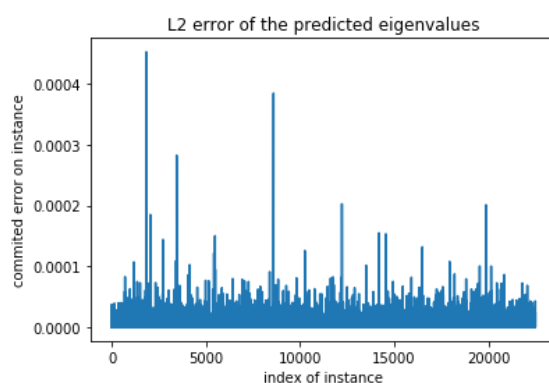
شبکه پس از چندین بار مرور داده‌ها در انتها ضرایب خود را به گونه ای تنظیم می‌کند که تابع هدفی که به آن معرفی کرده‌ایم را کمینه نماید. وقتی تابع مذکور به میزان کمینه خود برسد می‌گوییم آموزش شبکه به اتمام رسیده است. از این پس می‌توانیم با خوراندن ورودی جدید به شبکه از آن برای یافتن جواب استفاده نماییم.

¹ convolutional layers

² sum-pooling

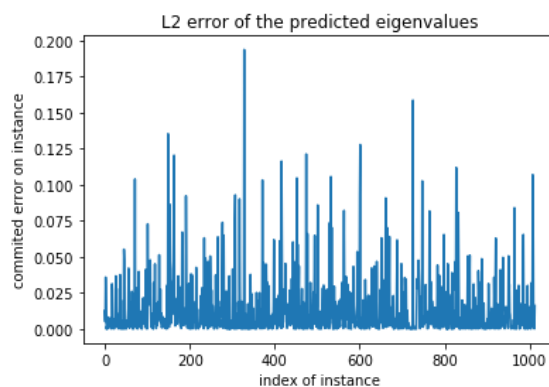
نتایج

مقادیر ضریب رسانایی مؤثر و انرژی حالت پایه به ترتیب 0.76800650 و $10/17474556$ بدست آمده اند که خطای L^2 به ترتیب، عبارت اند از $10^{-3} \times 102100$ و $10^{-5} \times 7235$. نمودار توزیع خطا بر حسب نمونه برای ضریب رسانایی مؤثر به شرح زیر است.



شکل ۱: خطای مرتکب شده روی مجموعه آزمون به تفکیک نمونه

همچنین، نمودار مشابه برای انرژی حالت پایه نیز به شرح زیر است.



شکل ۲: خطای مرتکب شده روی مجموعه آزمون به تفکیک نمونه

نتیجه‌گیری و کارهای پیش‌رو

همان‌گونه که از نتایج مشهود است، شبکه‌های عصبی توانایی بالایی در تقریب روابط پنهان مابین داده‌ها دارند. همچنین سادگی روش، آن را به یک روش در دسترس تبدیل می‌کند. ضمن اینکه پس از طی مرحله آموزش، شبکه عصبی قادر است جواب مسئله را تقریباً به طور آنی ارائه کند. یکی از محدودیت‌های شبکه‌های عصبی در مورد اندازه مقیاس ورودی‌ها است. به این معنی که در صورتی که برچسب‌ها بسیار بزرگ باشند یا با فاصله بسیار از هم، روند یادگیری با مشکل مواجه می‌شود. همان‌گونه که مشاهده می‌شود خطا در معادله شرو دینگر به علت بزرگ بودن برچسب‌ها در مقایسه با ضرایب عدم قطعیت بیشتر است. ما از روش نرمال سازی Z-Score برای نرمال سازی استفاده نمودیم. در ادامه این پایان نامه، سؤالات زیر می‌تواند مورد بررسی بیشتر قرار گیرد

- چه روش‌های دیگری برای حل این مسئله موجود است؟
- آیا این نرمال سازی خود خطایی به مدل تحمیل می‌کند؟ کران این خطای تحمیلی چیست؟
- میزان حساسیت شبکه عصبی به خطا در داده‌های ورودی و همچنین الگوی انتشار خطا در آن به چه صورت است؟
- آیا این انتشار خطا با افزایش عمق شبکه و یا استفاده از ساختارهای متفاوت اعم از توابع متفاوت برای توابع فعال سازی و همچنین نوع آرایش نرون‌ها ارتباطی دارد؟

واژه‌نامه فارسی به انگلیسی

convolutional neural-network	شبکه عصبی پیچشی
deep learning	یادگیری عمیق
finite difference method	روش تفاضلات متناهی
finite element method	روش عناصر متناهی
inhomogeneous media	محیط ناهمگن
neural-network	شبکه عصبی
spatial dimension	بعد فضائی
uncertainty	عدم قطعیت
uncertainty quantification	مقدار سنجی عدم قطعیت
nonlinear	غیر خطی
automatic differentiation	مشتق‌گیری اتوماتیک
graphic processor unit (GPU)	هسته پردازنده گرافیکی
feedforward neural network	شبکه رو به جلو
artificial neural network	شبکه عصبی مصنوعی
bias	بایاس
weight	وزن
activation function	تابع فعال‌سازی
loss function	تابع هزینه
neuron	نورون

پایگاه داده مستخرج از یک منبع dataset
 پایگاه داده مستخرج از منابع متفاوت data set
 بیش همسان سازی overfitting
 کم همسان سازی underfitting
 گشت در داده ها epoch
 دسته batch
 روش کاهش آشوبناک گرادیان stochastic gradient decent (SGD)