

صلاة الاضلاع



رساله

دوره دکتری تخصصی (PhD) در رشته آمار زیستی

عنوان

طراحی شبکه‌های عصبی مصنوعی برای مدل‌بندی پاسخ‌های دو متغیره آمیخته  
و کاربرد آن در داده‌های پزشکی

نگارش

مرتضی سدهی

استاد راهنما

دکتر یداله محرابی

اساتید مشاور

دکتر انوشیروان کاظم‌نژاد

دکتر وحید جوهری‌مجد

تابستان 1389



بسمه تعالی

### تاییدیه اعضای هیات داوران حاضر در جلسه دفاع از رساله دکتری

آقای مرتضی سدهی رشته آمارزیستی رساله دکتری خود را با عنوان: طراحی شبکه های عصبی مصنوعی برای مدل‌بندی پاسخهای دو متغیره آمیخته و کاربرد آن در داده های پزشکی در تاریخ ۸۹/۴/۲۸ ارائه کردند. اعضای هیات داوران نسخه نهایی این رساله را از نظر فرم و محتوا تایید کرده است و پذیرش آنرا برای تکمیل درجه دکتری پیشنهاد می کنند.

امضاء	نام و نام خانوادگی	اعضای هیات داوران
	دکتر بداله محرابی	استاد راهنما
	دکتر انوشیروان کاظم نژاد	استاد مشاور
	دکتر وحید جوهری مجد	استاد مشاور
	دکتر سقراط فقیه زاده	استاد ناظر
	دکتر فاطمه الحانی	استاد ناظر
	دکتر حجت زراعتی	استاد ناظر
	دکتر مسعود کریملو	استاد ناظر
	دکتر ابراهیم حاجی زاده	نماینده تحصیلات تکمیلی

# آیین نامه حق مالکیت مادی و معنوی در مورد نتایج پژوهشهای علمی

## دانشگاه تربیت مدرس

**مقدمه:** با عنایت به سیاست‌های پژوهشی و فناوری دانشگاه در راستای تحقق عدالت و کرامت انسانها که لازمه شکوفایی علمی و فنی است و رعایت حقوق مادی و معنوی دانشگاه و پژوهشگران، لازم است اعضای هیأت علمی، دانشجویان، دانش‌آموختگان و دیگر همکاران طرح، در مورد نتایج پژوهشهای علمی که تحت عنوان پایان‌نامه، رساله و طرحهای تحقیقاتی با هماهنگی دانشگاه انجام شده است، موارد زیر را رعایت نمایند:

**ماده 1-** حق نشر و تکثیر پایان‌نامه/ رساله و درآمدهای حاصل از آنها متعلق به دانشگاه می باشد ولی حقوق معنوی پدید آورندگان محفوظ خواهد بود.

**ماده 2-** انتشار مقاله یا مقالات مستخرج از پایان‌نامه/ رساله به صورت چاپ در نشریات علمی و یا ارائه در مجامع علمی باید به نام دانشگاه بوده و با تایید استاد راهنمای اصلی، یکی از اساتید راهنما، مشاور و یا دانشجوی مسئول مکاتبات مقاله باشد. ولی مسئولیت علمی مقاله مستخرج از پایان‌نامه و رساله به عهده اساتید راهنما و دانشجو می باشد.

**تبصره:** در مقالاتی که پس از دانش‌آموختگی بصورت ترکیبی از اطلاعات جدید و نتایج حاصل از پایان‌نامه/ رساله نیز منتشر می‌شود نیز باید نام دانشگاه درج شود.

**ماده 3-** انتشار کتاب و یا نرم افزار و یا آثار ویژه (اثری هنری مانند فیلم، عکس، نقاشی و نمایشنامه) حاصل از نتایج پایان‌نامه/ رساله و تمامی طرحهای تحقیقاتی کلیه واحدهای دانشگاه اعم از دانشکده‌ها، مراکز تحقیقاتی، پژوهشکده‌ها، پارک علم و فناوری و دیگر واحدها باید با مجوز کتبی صادره از معاونت پژوهشی دانشگاه و براساس آیین‌نامه-های مصوب انجام شود.

**ماده 4-** ثبت اختراع و تدوین دانش فنی و یا ارائه یافته‌ها در جشنواره‌های ملی، منطقه‌ای و بین‌المللی که حاصل نتایج مستخرج از پایان‌نامه/ رساله و تمامی طرحهای تحقیقاتی دانشگاه باید با هماهنگی استاد راهنما یا مجری طرح از طریق معاونت پژوهشی دانشگاه انجام گیرد.

**ماده 5-** این آیین‌نامه در 5 ماده و یک تبصره در تاریخ 87/4/1 در شورای پژوهشی و در تاریخ 87/4/23 در هیأت رئیسه دانشگاه به تایید رسید و در جلسه مورخ 87/7/15 شورای دانشگاه به تصویب رسیده و از تاریخ تصویب در شورای دانشگاه لازم‌الاجرا است.

«اینجانب مرتضی سدهی دانشجوی رشته آمارزیستی ورودی سال تحصیلی 1384 مقطع دکتری دانشکده علوم پزشکی متعهد می‌شوم کلیه نکات مندرج در آیین‌نامه حق مالکیت مادی و معنوی در مورد نتایج پژوهش‌های علمی دانشگاه تربیت مدرس را در انتشار یافته‌های علمی مستخرج از پایان‌نامه/ رساله تحصیلی خود رعایت نمایم. در صورت تخلف از مفاد آیین‌نامه فوق‌الاشعار به دانشگاه وکالت و نمایندگی می‌دهم که از طرف اینجانب نسبت به لغو امتیاز اختراع بنام بنده و یا هرگونه امتیاز دیگر و تغییر آن به نام دانشگاه اقدام نماید. ضمناً نسبت به جبران فوری ضرر و زیان حاصله براساس برآورد دانشگاه اقدام خواهم نمود و بدینوسیله حق هرگونه اعتراض را از خود سلب نمودم.»

امضا

تاریخ

## آئین نامه پایان نامه (رساله) های دانشجویان دانشگاه تربیت مدرس

نظر به اینکه چاپ و انتشار پایان نامه (رساله) های تحصیلی دانشجویان دانشگاه تربیت مدرس، مبین بخشی از فعالیت های علمی پژوهشی دانشگاه است. بنابراین به منظور آگاهی و رعایت حقوق دانشگاه، دانش آموختگان این دانشگاه نسبت به رعایت موارد ذیل متعهد می شوند:

ماده 1: در صورت اقدام به چاپ پایان نامه (رساله) ی خود، مراتب را قبلاً به طور کتبی به دفتر "دفتر نشر آثار علمی" دانشگاه اطلاع دهد.

ماده 2: در صفحه سوم کتاب (پس از برگ شناسنامه)، عبارت ذیل را چاپ کند:

"کتاب حاضر، حاصل رساله دکتری نگارنده در رشته آمار زیستی است که در سال 1388 در دانشکده علوم پزشکی دانشگاه تربیت مدرس به راهنمایی دکتر یداله محرابی، مشاوره دکتر انوشیروان کاظم نژاد از آن دفاع شده است.

ماده 3: به منظور جبران بخشی از هزینه های انتشارات دانشگاه، تعداد یک درصد شمارگان کتاب (در هر نوبت چاپ) را به "دفتر نشر آثار علمی" دانشگاه اهداء کند. دانشگاه می تواند مازاد نیاز خود را به نفع مرکز نشر در معرض فروش قرار دهد.

ماده 4: در صورت عدم رعایت ماده 3، 50% بهای شمارگان چاپ شده را به عنوان خسارت به دانشگاه تربیت مدرس، تادیه کند.

ماده 5: دانشجو تعهد و قبول می کند در صورت خودداری از پرداخت های بهای خسارت، دانشگاه مذکور را از طریق مراجع قضایی مطالبه و وصول کند، به علاوه به دانشگاه حق می دهد به منظور استیفای حقوق خود، از طریق دادگاه، معادل وجه مذکور در ماده 4 را از محل توقیف کتابهای عرضه شده نگارنده برای فروش، تامین نماید.

ماده 6: اینجانب مرتضی سدهی دانشجوی رشته آمار زیستی مقطع دکتری تعهد فوق و ضمانت اجرایی آن را قبول کرده، به آن ملتزم می شوم.

نام و نام خانوادگی

تاریخ و امضا

تقدیم به :

همسرم، که با صبر و همدلی در طول تحصیل همیشه همراه من بود.

خانواده‌ام، که در طول سالیان تحصیل همواره حامی و مشوق من بودند.

## تشکر و قدردانی

- از استاد راهنما جناب آقای دکتر یداله محرابی به خاطر تمام زحمات و راهنمایی های ایشان و درس های بزرگی که از ایشان آموختم.
- از استاد ارجمند جناب آقای دکتر انوشیروان کاظم نژاد به خاطر تمام زحماتی که در طول سالیان تحصیل به ایشان دادم و اینکه زحمت مشاور پایان نامه را بر عهده داشتند.
- از استاد گرانقدر جناب آقای دکتر وحید جوهری مجد که از راهنمایی ایشان در زمینه شبکه عصبی مصنوعی بهره فراوان بردم.
- از استاد ارجمند جناب آقای دکتر فرزاد حدائق به خاطر راهنمایی های مربوط به زمینه پزشکی .
- از استاد گرانقدر جناب آقای دکتر سقراط فقیه زاده به خاطر تمام زحمات و راهنمایی های ایشان در طول تحصیل در دوره های کارشناسی ارشد و دکترا.
- از استاد ارجمند جناب آقای دکتر ابراهیم حاجی زاده به خاطر مساعدت ها و زحمات ایشان در مدت تحصیل در دوره دکترا.
- از استاد گرانقدر جناب آقای دکتر غلامرضا بابایی که در طول سالیان تحصیل در دوره کارشناسی ارشد و دکترا افتخار شاگردی ایشان را داشتم.
- از همسر عزیزم به خاطر تمام سختی هایی که در طول تحصیل متحمل شدند.
- از کلیه همکلاسی هایم به خاطر همکاری ها و مساعدت های ایشان.

## چکیده

مقدمه: زمانی که در یک مطالعه بیش از یک متغیر پاسخ داشته باشیم که دارای مقیاس اندازه‌گیری متفاوت باشند، این‌گونه پاسخ‌ها را چند متغیره آمیخته می‌گویند که به‌وفور در مطالعات اپیدمیولوژی و پزشکی با آن‌ها مواجه می‌شویم. با توجه به محدودیت‌ها و برقراری نبودن برخی پیش‌فرض‌ها، روش‌های کلاسیک آماری برای مدل‌بندی و پیش‌بینی این نوع پاسخ‌ها کارایی ندارند. هدف این مطالعه، طراحی شبکه عصبی مصنوعی برای مدل‌بندی و پیش‌بینی پاسخ‌های دو متغیره آمیخته می‌باشد.

**روش کار:** این مطالعه شامل سه مرحله طراحی مدل، مطالعه شبیه‌سازی و برازش مدل بر داده‌های واقعی است. پس از طراحی مدل، با در نظر گرفتن پارامترهای مختلف، دو مجموعه داده شبیه‌سازی و مدل‌های یک متغیره و دو متغیره مورد ارزیابی قرار گرفتند. مجموعه اول شبیه‌سازی، حالتی است که همه متغیر(های) کمکی با هر دو متغیر پاسخ ارتباط دارند و در مجموعه دوم برخی متغیرهای کمکی تنها با یکی از متغیرهای پاسخ پیوسته یا دو حالتی ارتباط دارند. برای آموزش شبکه از الگوریتم شیب توام مقیاس شده و برای پایان دادن به هر الگوریتم و نیز تعیین اندازه بهینه تعداد تکرار و ضریب یادگیری از معیار میانگین مجذور خطاها استفاده گردید. برای تعیین مناسب‌ترین مدل در مرحله نهایی از معیار صحت پیش‌بینی استفاده شد. مدل پیشنهادی برای پیش‌بینی توام سندرم متابولیک (کیفی) و شاخص مقاومت به انسولین (کمی) در 347 نفر از افراد شرکت‌کننده در مطالعه قند و لیپید تهران به کار گرفته شد. سن، جنس، وضعیت تاهل، سابقه بیماری قلبی-عروقی، نمایه توده بدن،  $HDL$ ،  $LDL$ ، کلسترول، تری‌گلسیرید، قندخون ناشتا و دو ساعته، مصرف سیگار، فشارخون سیستولیک و دیاستولیک و دور کمر به عنوان متغیرهای مستقل در نظر گرفته شد. برنامه‌های رایانه‌ای در نرم افزارهای  $R 2.9.0$  و  $MATLAB 7.6$  طراحی و اجرا گردید.

**یافته‌ها:** در مجموعه شبیه‌سازی اول، صحت پیش‌بینی در مدل‌های یک متغیره و دو متغیره تقریباً یکسان بود، اما در مجموعه شبیه‌سازی دوم، صحت پیش‌بینی در مدل‌های دو متغیره نسبت به مدل‌های یک متغیره بیشتر بود. در مدل‌های دو متغیره، صحت پیش‌بینی مدل با افزایش همبستگی متغیرهای پاسخ، بیشتر می‌شود. در داده‌های واقعی، مدل با 10 گره در لایه میانی دارای بیشترین صحت پیش‌بینی به ترتیب برابر با 87/37 و 87 درصد برای داده‌های آزمون و اعتبارسنجی به دست آمد. نتایج حاصل از داده‌های اعتبارسنجی در هر دو مجموعه شبیه‌سازی نشان داد که مدل از نظر تعمیم‌پذیری عملکرد مطلوبی دارد.

**نتیجه‌گیری:** تحقیق نشان داد، در حالتی که دو متغیر پاسخ با متغیرهای کمکی مختلف ارتباط دارند مدل دو متغیره نسبت به مدل یک متغیره مناسب‌تر است و با افزایش همبستگی متغیرهای پاسخ صحت پیش‌بینی افزایش می‌یابد.

**کلمات کلیدی:** پاسخ‌های آمیخته - شبکه عصبی مصنوعی - مدل‌های دو متغیره - مطالعه قند و لیپید تهران.



## فهرست مطالب

1	..... فصل اول: مقدمه و مروری بر مطالعات گذشته
2	..... 1-1. مقدمه و بیان مسئله پژوهش
4	..... 2-1. اهداف پژوهش
4	..... 1-2-1. هدف اصلی
4	..... 2-2-1. اهداف فرعی
5	..... 3-1. پاسخ‌های دو متغیره آمیخته
7	..... 4-1. استراتژی‌های موجود برای مدل‌بندی پاسخ‌های آمیخته
7	..... 1-4-1. مدل‌های یک متغیره
8	..... 2-4-1. مدل‌های فاکتورگیری
10	..... 3-4-1. مدل‌های متغیر پنهان
11	..... 4-4-1. معادلات برآورد تعمیم یافته
13	..... 5-1. مبانی شبکه عصبی مصنوعی
13	..... 1-5-1. شبکه عصبی مصنوعی
16	..... 2-5-1. شبکه‌های عصبی زیستی
17	..... 3-5-1. طراحی شبکه عصبی مصنوعی
18	..... 4-5-1. لایه‌ها و ارتباطات بین آن‌ها

19	..... 5-5-1. توابع محرک و انواع آنها.....
21	..... 6-5-1. مدل‌بندی ریاضی شبکه عصبی مصنوعی.....
23	..... 7-5-1. انواع شبکه‌های عصبی مصنوعی.....
25	..... 8-5-1. فرایند یادگیری در شبکه عصبی مصنوعی.....
27	..... 9-5-1. معادلات یادگیری در حالت کلی.....
28	..... 10-5-1. یادگیری شبکه.....
29	..... 11-5-1. امتیازات شبکه‌های عصبی مصنوعی.....
31	..... 12-5-1. کاربرد شبکه‌های عصبی مصنوعی.....
32	..... 13-5-1. شبکه عصبی پرسپترون.....
38	..... 14-5-1. شبکه عصبی پرسپترون چندلایه.....
39	..... 15-5-1. شناسایی الگو و طبقه‌بندی.....
41	..... 16-5-1. مقایسه شبکه عصبی و روش‌های کلاسیک در شناسایی الگو و طبقه‌بندی.....
42	..... 17-5-1. شبکه‌های عصبی و مدل‌های آماری.....
44	..... 18-5-1. شبکه‌های عصبی مصنوعی و مدل‌های رگرسیون.....
47	..... 19-5-1. مدل‌بندی پاسخ‌های دو متغیره به وسیله شبکه عصبی مصنوعی.....
48	..... 20-5-1. مدل‌بندی پاسخ‌های دو متغیره آمیخته به وسیله شبکه عصبی مصنوعی.....

49	..... فصل دوم: روش کار
50	..... 1-2. توپولوژی شبکه عصبی مصنوعی برای مدل‌بندی پاسخ‌های دو متغیره آمیخته
50	..... 1-1-2. انواع توپولوژی شبکه عصبی مصنوعی باری مدل‌بندی پاسخ‌های دو متغیره آمیخته
52	..... 2-2. الگوریتم یادگیری حداقل میانگین مربعات خطا
54	..... 3-2. الگوریتم یادگیری پس‌انتشار
57	..... 1-3-2. تنظیم پارامترهای لایه میانی
60	..... 2-3-2. پس‌انتشار حساسیت‌ها
61	..... 3-3-2. خلاصه الگوریتم BP
64	..... 4-3-2. ملاحظاتی در مورد الگوریتم BP
66	..... 5-3-2. الگوریتم BP برای مدل‌بندی پاسخ‌های دو متغیره آمیخته
69	..... 4-2. مطالعه شبیه‌سازی
69	..... 1-4-2. اهداف مطالعه شبیه‌سازی
70	..... 2-4-2. داده‌های شبیه‌سازی
72	..... 5-2. کاربرد مدل در داده‌های پزشکی
76	..... فصل سوم: نتایج و یافته‌ها
77	..... 1-3. نتایج مطالعه شبیه‌سازی
79	..... 2-3. نتایج داده‌های واقعی
90	..... فصل چهارم: بحث، نتیجه‌گیری و پیشنهادها

91	..... 1-4. بحث
95	..... 2-4. نتیجه گیری و پیشنهادها
96	..... فهرست منابع
100	..... ضمائم
103	..... چکیده انگلیسی

## فهرست جداول

- 45 جدول 1-1. اصطلاحات متداول در شبکه عصبی و معادل آن‌ها در آمار.....
- 77 جدول 1-3. درصد صحت پیش‌بینی مطالعه شبیه‌سازی برای بردار ضرایب  $(b_{b1}, b_{c1}) = (0.2, 2)$ .
- 78 جدول 2-3. درصد صحت پیش‌بینی مطالعه شبیه‌سازی برای بردار ضرایب  $(b_{b1}, b_{c1}) = (1, 2)$ .....
- 79 جدول 3-3. درصد صحت پیش‌بینی مطالعه شبیه‌سازی برای بردار ضرایب  $(b_{b1}, b_{c1}) = (1, 6)$ .....
- جدول 4-3. مقایسه شیوع متغیرهای کیفی مورد بررسی در مبتلایان و غیر مبتلایان به سندرم  
80 متابولیک.....
- جدول 5-3. مقایسه مقادیر میانگین متغیرهای کمی پایه مورد بررسی در مبتلایان و غیر مبتلایان به  
81 سندرم متابولیک.....
- جدول 6-3. طبقه‌بندی افراد مورد بررسی بر اساس پیش‌بینی مدل‌های رگرسیون لجستیک، تحلیل  
82 ممیزی و مدل‌های شبکه عصبی مصنوعی.....
- جدول 7-3. حساسیت، ویژگی، آماره کاپا و سطح زیر منحنی **ROC** برای مدل‌های  
83 مختلف.....
- جدول 8-3. درصد صحت پیش‌بینی سندرم متابولیک و شاخص مقاومت به انسولین (HOMA-  
(IR) به تنهایی و به صورت توأم بر اساس تعداد مختلف گره‌های لایه میانی با استفاده از الگوریتم  
85 آموزش دسته‌ای کاهش شیب.....
- جدول 9-3. درصد صحت پیش‌بینی سندرم متابولیک و شاخص مقاومت به انسولین (HOMA-IR)  
87 به تنهایی و به صورت توأم بر اساس الگوریتم‌های آموزش مختلف.....
- جدول 10-3. درصد صحت پیش‌بینی توأم سندرم متابولیک و شاخص مقاومت به انسولین در  
88

..... داده‌های واقعی بر اساس الگوریتم‌های آموزش مختلف.....

89 جدول 3-11. درصد صحت پیش‌بینی سندرم متابولیک و شاخص مقاومت به انسولین بر اساس تعداد

..... مختلف گره‌های لایه میانی با استفاده از الگوریتم آموزش SCG.....

## فهرست شکل‌ها

- شکل 1-1. سلول عصبی زیستی ..... 16
- شکل 2-1. تابع سیگموئید ..... 19
- شکل 3-1. تابع تانانت هایپربولیک ..... 20
- شکل 4-1. تابع خطی مثبت ..... 21
- شکل 5-1. مدل ریاضی شبکه عصبی تک‌ورودی ..... 21
- شکل 6-1. مدل شبکه عصبی چند ورودی ..... 22
- شکل 7-1. شبکه عصبی پرسپترون چندلایه ..... 39
- شکل 1-3. مدل کامل شبکه عصبی ..... 84

# فصل اول

مقدمه و مروری بر مطالعات

انجام شده



## 1-1. مقدمه و بیان مسئله پژوهش

شناسایی الگو و طبقه‌بندی یکی از مهمترین کاربردهای روش‌های آماری در علوم مختلف است. یکی از اهداف عمده مدل‌سازی و طبقه‌بندی در آمار، پیش‌بینی بر اساس واقعیت‌ها و پارامترهای موجود و اطلاعات در دسترس از یک موضوع خاص می‌باشد. این وظیفه در آمار به‌طور عمده بر دوش روش‌هایی مانند رگرسیون، تحلیل ممیزی، سری‌های زمانی، رده‌بندی و رگرسیون درختی و روش‌های دیگر آماری است. در آمار از رگرسیون برای مدل‌بندی روابط بین متغیرها استفاده می‌شود. رگرسیون چندمتغیره حالتی است که برای هر مشاهده یک بردار پاسخ چندمتغیره وجود دارد. این پاسخ‌ها می‌توانند به صورت کمی، کیفی و یا آمیخته باشند. در حالتی که کلیه متغیرهای پاسخ کمی هستند، روش تحلیل داده‌ها در آمار استفاده از مدل‌های رگرسیون چندمتغیره است. در حالتی که همه پاسخ‌ها کیفی باشند، روش پیشنهادی استفاده از مدل‌های رگرسیون لجستیک چندمتغیره یا مدل‌های رگرسیون لجستیک چندمتغیره تعمیم‌یافته می‌باشد. اما در حالت سوم که متغیرهای پاسخ، برخی کمی و برخی کیفی هستند، با توجه به محدودیت‌هایی که برای در نظر گرفتن توزیع مناسب برای بردار پاسخ در اینگونه موارد وجود دارد، روش‌های کلاسیک در این مورد کارایی لازم را ندارند.

در سال‌های اخیر برای مدل‌سازی و پیش‌بینی پاسخ‌های آمیخته، تحقیقاتی در آمار کلاسیک انجام شده است. استراتژی‌های موجود در این تحقیقات برای مدل‌بندی پاسخ‌های دو متغیره آمیخته را می‌توان به چند دسته تقسیم‌بندی کرد: روش‌های مبتنی بر فاکتورگیری توزیع توام پاسخ‌ها [3-1]،

روش‌های مبتنی بر معرفی یک متغیر پنهان برای مدل‌سازی همبستگی بین متغیرهای پاسخ [10-4] و روش برآورد معادلات تعمیم‌یافته [11-13].

روش‌های کلاسیک رگرسیون برای مدل‌بندی روابط بین متغیرها دارای تعدادی مفروضات می‌باشند. مانند مشخص کردن توزیع متغیرهای پاسخ، خطی بودن رابطه پیشنهادی در برخی مدل‌ها، یکسان بودن واریانس خطاها و دیگر محدودیت‌های روش‌های کلاسیک که هنگام استفاده عملی از این روش‌ها، اگر داده‌های واقعی شرایط مفروض مدل را نداشته باشند، استفاده از این روش‌ها ممکن است موجب خطا و گمراهی محقق گردد. به علاوه اغلب این روش‌ها قابلیت مدل‌بندی روابط پیچیده غیرخطی و همچنین اثرات متقابل درجه بالا را ندارند. حساس بودن بیشتر این مدل‌ها به مشاهدات گم‌شده و داده‌های پرت از دیگر محدودیت‌های این روش‌ها به‌شمار می‌آید [14].

با توجه به محدودیت‌های ذکر شده برای مدل‌های کلاسیک، نیاز به روش‌هایی که با محدودیت‌های کمتری در این زمینه مواجه باشند، احساس می‌شود. در این بین مدل‌های شبکه عصبی مصنوعی می‌توانند یکی از مناسب‌ترین روش‌ها باشند. شبکه عصبی مصنوعی، فرض اولیه‌ای برای توزیع داده‌ها تحمیل نمی‌کند، ضمن اینکه محدودیتی نیز برای شکل تابعی رابطه بین متغیرهای مستقل و وابسته اعمال نمی‌کند بلکه شبکه عصبی، خود، این رابطه تابعی را کشف می‌کند، که لزوماً، یک رابطه خطی نیست. از دیگر مزایای شبکه عصبی این است که در شبکه عصبی اطلاعات به صورت ضمنی<sup>1</sup> پردازش می‌شود، بر این اساس چنانچه بخشی از داده‌های ورودی شبکه دارای موارد گم‌شده یا داده‌های پرت باشند، باز هم احتمال رسیدن به پاسخ صحیح وجود دارد، ضمن اینکه تعمیم‌پذیری شبکه عصبی این امکان را به مدل می‌دهد که مدل در ارتباط با یک مشاهده جدید آموزش داده نشده پاسخ مناسبی ارائه دهد [15].

با توجه به موارد فوق و با در نظر گرفتن اینکه تاکنون مدل‌بندی این‌گونه پاسخ‌ها توسط شبکه‌های عصبی انجام نشده است، در این پایان‌نامه سعی داریم شبکه عصبی مناسب برای مدل‌بندی پاسخ‌های آمیخته شامل یک متغیر کمی و یک متغیر کیفی، با فرض دو حالتی بودن متغیر کیفی را طراحی نماییم.

## 2-1. اهداف پژوهش

### 1-2-1. هدف اصلی

طراحی توپولوژی مناسب شبکه عصبی مصنوعی برای مدل‌بندی پاسخ‌های دو متغیره آمیخته.

### 2-2-1. اهداف فرعی

1-2-2-1. تعیین چگونگی تاثیر میزان همبستگی متغیرهای پاسخ بر روی صحت پیش‌بینی مدل شبکه عصبی مصنوعی

2-2-2-1. تعیین چگونگی تاثیر نوع ارتباط متغیرهای تبیینی با متغیرهای پاسخ بر روی صحت پیش‌بینی مدل شبکه عصبی مصنوعی

3-2-2-1. تعیین چگونگی تاثیر میزان اندازه اثر متغیرهای تبیینی بر متغیرهای پاسخ بر روی صحت پیش‌بینی مدل شبکه عصبی مصنوعی

4-2-2-1. مقایسه اثر توابع مختلف آموزش شبکه عصبی بر روی صحت پیش‌بینی مدل

5-2-2-1. مقایسه صحت پیش‌بینی مدل توام با مدل تک متغیره در حالت‌های مختلف

6-2-2-1. کاربرد مدل برای پیش‌بینی توام سندرم متابولیک و شاخص مقاومت به انسولین در مطالعه قند و لیپید تهران.

## 3-1. پاسخ‌های دو متغیره آمیخته

یکی از مسائلی که در مطالعات کاربردی به‌وفور با آن مواجه می‌شویم، زمانی است که هدف، پیش‌بینی همزمان دو متغیر پاسخ (وابسته) از روی تعدادی متغیر (های) کمکی (پیشگو) باشد. در ادبیات آماری، این‌گونه مدل‌ها را مدل‌های دو متغیره<sup>1</sup> می‌نامند. زمانی که متغیرهای پاسخ، هر دو کمی یا هر دو کیفی باشند، از روش‌های استاندارد آماری مانند رگرسیون دو متغیره<sup>2</sup> یا رگرسیون لجستیک دو متغیره<sup>3</sup> برای مدل‌بندی روابط بین متغیرهای پاسخ و کمکی استفاده می‌گردد. مبنای استنباط در این روش‌ها، در نظر گرفتن توزیع مناسب برای بردار پاسخ می‌باشد. اما زمانی که یکی از متغیرهای پاسخ کمی و دیگری کیفی باشد، اصطلاحاً این‌گونه پاسخ‌ها را پاسخ‌های آمیخته<sup>4</sup> یا پاسخ‌های نامتناسب<sup>5</sup> می‌گویند.

زمانی که با پاسخ‌های آمیخته سروکار داریم، با توجه به محدودیتی که برای در نظر گرفتن توزیع احتمال مناسب برای بردار پاسخ در این‌گونه موارد وجود دارد، روش‌های دو متغیره معمول در آمار کلاسیک، کارایی ندارند. در این حالت ساده‌ترین شیوه تحلیل، استفاده از روش‌های تک متغیره برای هر یک از پاسخ‌های کمی و کیفی به‌صورت جداگانه می‌باشد. از آنجا که این روش، در مدل‌سازی و پیش‌بینی، همبستگی بین پاسخ‌ها را نادیده می‌گیرد، از کارایی لازم برخوردار نیست.

در سال‌های اخیر برای مدل‌سازی و پیش‌بینی پاسخ‌های دو متغیره آمیخته، تحقیقاتی در آمار کلاسیک انجام شده است. استراتژی‌های موجود در این تحقیقات برای مدل‌بندی پاسخ‌های دو متغیره آمیخته را می‌توان به چند دسته تقسیم‌بندی کرد: روش‌های مبتنی بر فاکتورگیری توزیع توأم پاسخ‌ها، روش‌های مبتنی بر معرفی یک متغیر پنهان برای مدل‌سازی همبستگی بین متغیرهای پاسخ و روش برآورد معادلات تعمیم‌یافته.

ایده اصلی روش فاکتورگیری، نوشتن تابع درست‌نمایی به‌صورت حاصلضرب توزیع حاشیه‌ای یکی از متغیرهای پاسخ و توزیع شرطی متغیر پاسخ دوم به شرط متغیر پاسخ اول است. کاکس و ورموت [1]، دو روش فاکتورگیری را برای مدل‌سازی یک پاسخ کمی و یک پاسخ دو حالتی به‌عنوان تابعی از متغیرهای تبیینی شرح دادند. فیتز مائوریس و لارید [2] و کاتالانو و رایان [3] این روش را برای داده‌های خوشه‌ای توسعه دادند.

---

1- Bivariate Models  
2- Bivariate Regression  
3- Bivariate Logistic Regression  
4- Mixed Responses  
5- Non - Commensurate Responses