

## در مورد مسائل مربوط به حقیقت در طبقه بندی نتارت شده

جاناتان کی سو  
آزمایشگاه MIT لیکلن 244  
وود استریت  
*Lexington, MA 02421-6426, USA*

Su@theory.peach.quota

ویراستار: تی شیوانی آگاروال، فرانسیس باخ

### چکیده

طبقه بندی نتارت شده ایده‌آل، برچسب‌های درست ساخته شده را فرصن من کند، اما مسائل مختلف ف صدق در عمل من تواند ایجاد شود: برچسب‌های پر سر و صدا. چندین برچسب متصناد برای یک نم ونه؛ برچسب‌های گمشده؛ و ترکیبات مختلف برچسب‌گذار برای نمونه‌های مختلف. کار قبلی در *noisy-nlabel model* معرف شد که برچسب‌های نویزدار مشاهده شده را به عنوان متغیرها ی تصادفی مشروط به برچسب‌های درست مشاهده نشده مشاهده من کند. عمدتاً بر تخمین توزيع شر ط برچسب‌های پر سر و صدا و کلاس قبلی و همچنین تخمین برچسب‌های صحیح یا آموزش با برج سب‌های پر سر و صدا متمرکز شده است. به روشن مکمل، با توجه به توزيع شرطی و کلاس‌های قبلی، ما تئوری تخمین را برای آزمایش طبقه بندی کننده، آموزش و مقایسه ترکیب‌های مختلف برچسب‌ها اعمال من کنیم. ابتدا، برای طبقه بندی باینزی، یک مدل آزمایشی من سازیم و خلف‌های حاسیه‌ای تقریب را برای دقت، دقت، یادآوری، احتمال هشدار استیبه و امتیاز F، و خلف‌های مشترک برای تحلیل R OC و فراخوان دقیق به دست من آوریم. ما تست حداقل میانگین مربع خطای MMSE(t) ر ا پیشنهاد من کنیم که از الگوریتم‌های تجربی بیز برای تخمین پارامترهای مدل آزمایشی استفاده من کن د و سپس تخمین‌های نقطه بینه و مناطق معتبر را برای معیارها محاسبه من کند. ما رویکرد طبقه بندی چند طبقه را گسترش من دهیم تا تخمین‌های بینه دقت و عناصر ماتریس سردرگمی را به دست آوریم. دوم، یک دیدگاه واحد از آموزش ارائه من کنیم که مدل‌های احتمالی (یعنی افتراقی با مولد) و غیر احتمالی را پوشش من دهد. برای اولی، ما آموزش‌داداگر احتمال با حداقل پیشین را برای مسائل حقیقت تنظیم من کنیم. برای دومی، ما آموزش MMSE را پیشنهاد من کنیم که تخمین MMSE ریسک تجریب را به حداقل من رساند. ما همچنین آموزش‌های کمتر از حد مطلوب را که با زبرساخته های موجود سارگار است، توصیف من کنیم. سوم، مشاهده من کنیم که اطلاعات متقابل اجازه من دهد هر ترکیب برچسب‌زنی را به عنوان یک برچسب‌گذار تک خبره آموزنده یا آموزنده تر باشند. آزمایش‌ها اثربخش روشن‌ها را نشان من دهنده و کاربرد آن را تأیید من کنند.

طبقه بندی نتارت شده، خطاهای حقیقت، برچسب‌های پر سر و صدا، تخمین بیزی، کلمات کلیدی: بیز تجربی، اطلاعات متقابل، جمع سپاری

تی بیانیه توزيع A. برای انتشار عمومی تایید شده است. توزيع نامحدود است.  
این مطالب بر اساس کار پشتیبانی شده توسط معاون وزیر دفاع در امور تحقیقات و مهندسی تحت قرارداد نیروی هوا  
ی شماره 0001-D-15-702-FA8702 است. هر گونه نظرات، یافته‌ها، نتیجه‌گیری‌ها یا توصیه‌های بیان شده در این مطالب  
متعلق به نویسنده است و لزوماً منعکس کننده نظرات معاون تحقیقات و مهندسی وزیر دفاع نیست.

تی 2023 ctt موسسه فناوری ماساچوست.

مشمول حقوق ثبت اختراع 11-227 FAR52.227 - مالکیت پیمانکار (ماه 2014) به دولت ایالات متحده  
با حقوق نامحدود، همانطور که در DFARS قسمت 7013-252.227 یا 7014 (فوریه 2014) تعریف شده است، تحویل داد  
ه شد. علی رغم هرگونه اطلاعیه حق چاپ، حقوق دولت ایالات متحده در این اثر توسط 7013 DFARS 252.227-7014 یا 2  
52.227-7014 همانطور که در بالا توضیح داده شده است، تعریف شده است. استفاده از این اثر به غیر از مواردی که به طور  
خاص توسط دولت ایالات متحده مجاز است، ممکن است هرگونه حق نسخه برداری موجود در این اثر را نقصان کند.

## 1. معرفی

کلاس بندی تحت نظرارت از داده های برچسب گذاری شده برای آموزش و تست یک مدل پیش بینی با طبقه بندی کننده استفاده من کند، که برای پیش بینی برچسب های داده های برچسب نسده استفاده خواهد شد. مدل پیش بینی یک نقشه گذاری است:  $Y \rightarrow X : g$ ، که توسط  $\theta$  پارامتر بندی شده است، از فضای ویرگی ها  $X$  به مجموعه ای از  $C-1 = \{0, 1, \dots, C\}$  کلاس های یا برچسب های منحصر به فرد و کامل استفاده من کند. ورودی مدل یک درجه ویرگی  $X \in \mathbb{X}$  است و خروجی مدل برچسب پیش بینی شده  $Y \in \mathbb{Y} = g(\mathbb{X}; \theta)$  است که ممکن است با برچسب صحیح  $Y \in \mathbb{Y}$  موافق یا ناموافق باشد. یک نمونه برچسب گذاری شده  $y \in \mathbb{Y}$  شامل یک درجه ویرگی  $x \in \mathbb{X}$  و برچسب صحیح مربوط به آن  $y$  است، در حالی که یک نمونه برچسب نسده فقط یک درجه ویرگی  $x \in \mathbb{X}$  است. مجموعه ای از  $N$  نمونه برچسب گذاری شده به عنوان  $\{x_i, y_i\}_{i=1}^N$  نشان داده من شود، جایی که  $x_i \in \mathbb{X}$  و  $y_i \in \mathbb{Y}$  است. نتیجه اعمال یک مدل پادگیری به هر ویرگی در  $X$ ، لیست برچسب های پیش بینی شده  $\{y_i\}_{i=1}^N$  است.

تمرین مدل را از مجموعه ای از نمونه های برچسب گذاری شده یاد من گیرد. تست از مجموعه ای جدا گانه از نمونه های برچسب گذاری شده استفاده من کند، مدل یاد گرفته شده را به ویرگی های ویکتورهای خود اعمال من کند و معیارهایی را محاسبه من کند که توافق بین هر برچسب پیش بینی شده و برچسب صحیح مربوطه را کمیت من کنند. نشانه گذاری ما بین مجموعه های آموزش و تست تقاضوت نمی گذارد، زیرا مجموعه مورد نظر از متن واضح خواهد بود.

### 1.1 مسائل حقیقت آمیز

ممدوحاً فرص من شود که برچسب های صحیح در طول آموزش و تست سناخته من شوند؛ ما به این و صنعتی به عنوان موقف ایده آل اشاره من کنیم. اما این فرص من مهم اغلب در عمل نقص من شود (متلای Northcutt et al. (2006, p. 172), Fr'enay and Verleysen (2014) و (2021b)، که من تواند مدل آموزش دیده را کاهش دهد و ساخته های تست گمراه کننده تولید کند. انحراف از موقف ایده آل من تواند ساعت های بسیاری از کار سخت و پردازش کامپیوترا را بی فایده کند و همچنین باعث شود که کاربران و مشتریان نسبت به کارایی و عملکرد یک طبقه بندی من کاک شوند.

در نتیجه، منابع قابل توجهی اغلب به تعیین درستی اختصاص داده من شود، که فرآیندی است برای برچسب گذاری data به بهترین شکل ممکن. ما از اصطلاح برچسب دهنده برای اشاره به موجودیتی که برچسب ها را به نمونه ها اختصاص من دهد، استفاده من کنیم؛ برخی نویسندها از عباراتی مانند "مع لم" یا "برچسب گذار" استفاده من کنند. برچسب دهنده من تواند یک انسان، یک حسگر، یک آزمایشگاه، یا حتی یک طبقه بندی کننده دیگر باشد. با وجود تلاش های بهترین برچسب دهنده، خطاهای من توانند و همچنان رخ من دهنند. انسان ها اشتباه من کنند، خسته من شوند و میران تخصص، توجه و انتیزه شان متفاوت است؛ حسگرهای تحت تأثیر نویز، انسداد و دیگر زوالات قرار من گیرند؛ تاریخ آزمایشگاه هم من شه قطعی نیستند؛ و طبقه بندی کننده ها معمولاً پیش بینی کنندها کامل نیستند. علاوه بر این، برچسب ب گذاری از طریق منابع جمعی به این معناست که بیش از یک برچسب دهنده من تواند برچسبی به ویرگی های درجه بندی سده<sup>۲</sup>؛ اختصاص ده د و اگر برچسب ها با هم در تصاد باشند، حداقل یکی از آن ها باید نادرست باشد.

در کوتاه، تعداد زیادی مسئله درستی سنجی من تواند به وجود آید: خطاهای درستی، چند برچسب دهنده که برای یک نمو نه های مسابه برچسب های متعارض ارائه من دهنند، برچسب های گم شده و ترکیبات مختلف برچسب دهندها برای نمونه های مختلف. ما در اینجا برخی نشانه گذاری ها را معرفی من کنیم و یک مثال از آن را در جدول 1 آورده ایم. ما فرص من کنیم که تعداد  $T$  برچسب دهنده وجود دارد، که از 1 به  $T$  شمارش من شوند، و ما  $\{1, 2, \dots, T\} = T$  را من گذاریم. بگذارید  $Y \subseteq \{\emptyset\}$  اگر  $t \in T$ ، نشان دهنده برچسب نویز ده شده به  $x_t$  توسط  $tth$  برچسب دهنده باشد، جایی که  $z_{i,t} = \emptyset$  اگر هیچ برچسب اختصاص داده نشده باشد. ما نیاز داریم که حداقل یک برچسب دهنده به هر نمونه برچسب اختصاص دهد؛ یعنی  $z_{i,t} \in T$  برای برخی  $i$ . بنابراین، یک برچسب دهنده تنها من تواند یک برچسب اختصاص دهد

## در مسائل درستی ساری در طبقه بندی تحت نطارت

سر و صدا $i$	نمونه ویرگلی صحیح ویکتور $x_i$	نمونه برچسب برچسب $y_i$	نشان دهنده برچسب				
			1	2	3	4	5
1		0	0	1	0	0	حقیقی
2	$\sqrt{x}$	1	1	حقیقی	حقیقی	1	حقیقی
3	$x^3$	0	0	0	0	0	0
4	$x/4$	1	1	1	0	0	1
5	$x \cdot 5$	0	0	حقیقی	حقیقی	0	0
...							
N	نمودار TN	0	0	0	حقیقی	0	حقیقی

جدول 1: نمونه ای از نشانه گذاری برای طبقه بندی باینری و پنح برچسب. این مثال به مثال آموزشی در بخش 5.4 پاسخ می دهد.

به یک نمونه الیته  $t$  ممکن است نادرست باشد و با متفاوت باشد. مجموعه ای از برچسب های noisyn را نشان دهید  $xti$  fort  $zti, tt$  ; برچسب های پر سر و صدا نیازی به توافق ندارند. در نهایت.  $.lett zt = (ztiti) tNit = 1t$   $t$   $byt ztit = (tzti, tt) ttttTt$  یک رویکرد طبیعی، که توسط داوید و اسکن (1979) معرفی شد، این است که برچسب های صحیح و برچسب های پر سر و صدا را به عنوان متغیرهای تصادفی  $(RV)$  (در نظر بگیریم. ما این دیدگاه را اتخاذ می کنیم و از حروف بزرگ  $(Mtl)$  برای  $RV$  و حروف کوچک  $(m)$  مثلاً پیرای همتای غیر تصادفی آن استفاده می کنیم. از این رو،  $t$   $Ytit RVt$  با برچسب صحیح برای آن نمونه نیست،  $sot Yt = (tYtit) tNit = 1t$   $t$   $Ytit RVt$  های دارای برچسب صحیح است، و  $yt$   $nandt$  لیستی از برچسب های صحیح است. به طور متسابه،  $RV$  های بردار ویرگل  $Xt$   $inandt Xt = (tXt it) tNit = 1t$  هستند، و تحقق آنها است. همچنین  $t$   $Zti, tt, t Zt, t Ztit = (tZti, tt) tttTt$ ,  $nandt Zt = (tZtit) tNit = 1t$   $RVst$  با برچسب پر سر و صد  $t$   $zti, tt, t zt, t ztit, andt zt$  تحقق مربوطه خود را نشان می دهد.

ما فرض می کنیم که  $RV$  های مرتبط با نمونه های مختلف مستقل و به طور یکسان توزیع شده اند و  $RV$  های دارای برچسب صحیح دارای کلاس prior  $\pi t = t$   $(t\pi t(tyt)t) tCtt1tyt = 0t$  هستند، جایی که  $\pi t(tyt) = t$   $\pi t(tyt) = Pr(tYt = t yt)$  توجه به  $RV$  با برچسب صحیح به صورت شرطی وابسته هستند، و توزیع شرطی با برچسب نویز را نشان می دهیم  $pt(tz | tyt; t \psi t)$ ، که بایت  $\psi$  پارامتر می شود، و جایی که یک نیمه ویرگول  $RV$  ها را از پارامترهای غیر تصادفی جدا می کند. ما به  $tot pt(tz | tyt; t \psi t) t\pi t(tyt)$  مدل نویز برچسب اشاره می کنیم. برای این نمونه، پارامترها  $\psi$  هستند. فرض می کنیم که  $Ztit$  و  $Ztjt$  از همان توزیع شرطی تبعیت می کنند. در این مرحله اجازه می دهیم پارامترها کاملاً عمومی بمانند.  $npt(tz | tyt; t \psi t) ift \psi tit = t \psi tjt = t \psi t$  همچنین، اجازه دهید  $\psi$  لیستی از پارامترها برای همه نمونه ها باشد.

بسیاری از کارهای مرتبط، که در بخش 1.2 توصیح داده شده است، بر ساخت مدل  $\psi$  nfort  $pt(tz | tyt; t \psi)$ ، تخمین پارامتر  $\psi$  و کلاس  $prior t$   $\pi t$ ، و تخمین  $Yt$  با برچسب صحیح متمرکز  $\pi t$  ده است. این مقاله و کار مرتبط به طور صمنی فرض می کند که  $n$  به  $t pt(tz | tyt; t \psi) t6t = t pt(tz | tyt; t \psi) t$  طوری که  $RV$  های دارای برچسب نویز اطلاعاتی در مورد  $RV$  با برچسب صحیح ارائه می دهند. اگر  $zt$  آمورنده نیست، هیچ یک از این روش ها موثر نخواهد بود. برچسب صحیح تخمین زده برای نمونه را به عنوان  $tyt = (tytit) tNit = 1t$  نشان دهید و اجازه دهید  $\psi$  در این مقاله، ما به دنبال پاسخ به سوالات زیر هستیم:

1. چگونه می‌توان یک طبقه بندی کننده را در حضور مشکلات درستی یابی تست کرد؟
2. چگونه می‌توان یک طبقه بندی کننده را در حضور مشکلات درستی یابی آموزش داد؟
3. چگونه می‌توان ترکیبات مختلف برچسب گذاران با توانایی های مختلف را مقایسه کرد؟

## 1.2 کارهای مرتبط

ما کارهای مرتبط را به مدل سازی و یادگیری برچسب های نویز، آموزش با مشکلات درستی یابی، تست با مشکلات درستی یابی و مقایسه ترکیبات برچسب گذاران تقسیم بندی کرده ایم.

### 1.2.1 مدل سازی و یادگیری برچسب های نویز

چندین نویسنده مدل های برچسب های نویز  $p(z|y)$  را پیشنهاد داده اند و روش هایی برای یادگیری آنها توسعه داده اند.<sup>2</sup> یعنی تخمین زدن  $\pi$  و  $\alpha$ . اغلب، آنها همچنین RV هایی برچسب های صحیح را تخمین زده اند. جداول 2 یک خلاصه ارائه می‌دهد. همه مجموعه های نویسندگان به جز یک مجموعه، چندین برچسب دهنده را مجاز می‌دانند. جداول نشان می‌دهد که بیشتر آنها فرض کرده اند که RV برچسب های نویز دار conditionally independent given the correct-label RV است.

و Skene (1979) یک تحقیق اولیه و تأثیرگذار در این زمینه انجام دادند. آنها به مسائل طبقه بندی نظارت شده پرداختند، اما به مشکلی که شامل جمع آوری مشاهدات بیماران از چندین کلینیکان که ممکن است با هم اختلاف نظر داشته باشند یا اشتباه کنند، بود، پرداختند. در این زمانه، آنها فرمول RV را معرفی کردند که اکنون رایج است. وقتی برچسب های صحیح موجود هستند، آنها نشان دادند که برآورد احتمالاتی (ML)  $\pi$  و  $\alpha$  ساده است. با توجه به نگاه به گذشته، این وضعيت به یادگیری یک طبقه بندی از یک مجموعه داده های جانشینی  $tz$  منجر می‌شود که برچسب های نویز دار  $z$  به عنوان بردار ویژگی عمل می‌کنند. به عنوان مثال، ممکن است یک مجموعه داده های جانشینی کوچک داشته باشد که شامل برچسب های نویز دار از کلینیکان و برچسب های صحیح از یک آزمایشگاه استاندارد جداگانه باشد و مقدار زیادی داده با فقط برچسب های نویز دار.

مهم تر از همه، وقتی برچسب های صحیح موجود نیستند، Skene و Dawid کاربرد الگوریتم expectation-maximization (EM) برآورد  $\pi$  و  $\alpha$  را نشان دادند. بسیاری از نویسندگان دیگر بر این کار بنا نهادند.

Donmez et al. (2010) به مسائل طبقه بندی و پیش بینی پرداختند و همچنین پیشنهاد کردند که از الگوریتم EM برای یادگیری مدل برچسب های نویز دار استفاده شود. آنها شرایطی را که برآورد های ML پارامترهای مدل برچسب های نویز دار سارگار هستند، مطالعه کردند. برای طبقه بندی، آنها نشان دادند که سارگاری تحت شرایطی خاص، مانند زمانی که برچسب دهنده گذاران یادگیرندگان صنعتیف یا بهتر هستند و پیش بینی های کلاس  $\pi$  نیستند، وجود دارد.

Whitehill et al. (2009) مدلی برای  $p(y|z)$  پیشنهاد داد که دشواری نمونه و تخصص برچسب دهنده را در نظر می‌گیرد. یک گسترش بیانی از مدل به این پارامترها پیش بینی ها را اجازه می‌دهد.

آنها از الگوریتم EM و نسخه ای بیانی از آن برای برآورد  $\pi$  و  $\alpha$  استفاده کردند. Welinder et al. (2010) و Perona (2010) مدل را گسترش دادند تا شامل انحراف برچسب دهنده و پشتیبانی از برچسب های چندکلاس و توصیحات ارزش دار پیوسته شود.

1. ما به دلیل اینکه روش آزمایش ما بی نظیر از نحوه آموزش طبقه بندی است، به آزمایش قبل از آموزش  $\pi$  می‌پردازیم. همچنان ن، اگر مشکلات حقیقت یابی وجود داشته باشد، اگر نمی‌توان عملکرد طبقه بندی را ارزیابی کرد، هیچ نکته ای در آموزش طبقه بندی وجود ندارد.

2. شاید بتوان مدل هایی از فرم  $y \neq z(p), y \neq z(q), y \neq z(r)$  (از یک مجموعه جانشینی) بنا کراین اگر چنین مجموعه ای در دسترس باشد، من توان مدل پیش بینی کننده را از  $\{x\}, \{y\}$  یاد گرفت.

## در مسائل درستی ساری در طبقه بندی تحت نطارت

مرجع	توصیحات	تعداد کلاس‌ها	برجسب‌گذارواب ستگیرچسب‌گذارواب
Skene و Dawid (1979)	<ul style="list-style-type: none"> <li>کار مهم</li> <li>ML تخمین مدل برچسب‌های نویز، اگر صحیح باشد برچسب‌های موجود</li> <li>برآورد EM مدل برچسب‌های نویز آلود و</li> </ul> <small>برچسب‌های صحیح اگر برچسب‌های صحیح، اگر برچسب‌های صحیح در دسترس نباشد</small>	چندگانه	مستقل
Donmez و همکاران(2010)	<ul style="list-style-type: none"> <li>همچنین رگرسیون</li> <li>ارزیابی مدل‌های بر اساس برچسب‌های نویز</li> <li>شرایط یکپارچگی برای برآورد ML</li> </ul>	چندگانه	مستقل
Whitehill و همکاران(۲۰۰۹)	<ul style="list-style-type: none"> <li>labeler-expertise و Sample-difficulty</li> <li>برآورد نویز برچسب‌ها با روش EM یا بیانی مدل</li> </ul>	دوگانه	مستقل
Perona(2010) و Welinder و Welinder(؛10 مکاران)(2010)		دو یا چند گانه	مستقل
برانسن و همکاران(۲۰۱۰)	<ul style="list-style-type: none"> <li>توسعه Welinder و همکاران(2010) به بخشی از نقاط کلیدی و برچسب‌بندی‌های محدوده</li> <li>درایافت پیوسته برچسب‌های نویزدار</li> </ul>	دوگانه	مستقل
Van Horn و همکاران(2018)	توسعه Branson و همکاران(2017) به چندین کلاس و برچسب‌دهندگان وابسته به ترتیب	چندگانه	وابسته به ترتیب
Karger و همکاران(2014)	<ul style="list-style-type: none"> <li>تخصیص برچسب‌دهندگان به نمونه‌های برچسب‌نشده</li> <li>تبادل پیام‌های مکرر</li> </ul>	دوگانه	مستقل
Zhou و همکاران(2015)	اصل ناگهانی سازی حداقل-بیشینه‌ای انرژی	چندگانه	مدل‌های هیرا
	رشیک	دوگانه	مستقل یا وابسته
	نمونه گیری کهans		
(2021a)	<ul style="list-style-type: none"> <li>فرض کنید کلاس بندی کشته‌ای که قبلاً آموزش دیده است</li> <li>مقداردهی فاصله مستقر، <math>\psi</math>، (بر روی مجموعه داده‌های جدید</li> <li>مستندات اشتباه یا حذف شود</li> </ul>	چندگانه	نه مربوط

جدول 2: کارهای مرتبط در مدل سازی و یادگیری برچسب‌های نویز. متابعی که مدل برچسب‌نویز و مدل پیس<sup>۷</sup> بینی کشته را به طور مشترک یادگیری و آموزش می‌دهند در جدول 3 آورده شده‌اند.

برانسن و همکاران(۲۰۱۰) مدل‌های ویلیندر و پروندا و ویلیندر و همکاران را گسترش دادند. برای افزودن دیگر اشکال توصیحات. همچنین یک الگوریتم معرفی کردند که به ترتیب برچسب‌های پر سر و صدا را به دست می‌آورد تا خطر برچسب صحیح برآورد شده به زیر یک حد<sup>۸</sup> بین کاهش یابد. Van Horn و همکاران(2018) کار Branson و همکاران(2017) را به طبقه بندی چندگانه و برچسب‌دهندگان وابسته به ترتیب گسترش دادند.

Karger و همکاران(2014) مساله تخصیص برچسب‌دهندگان به نمونه‌های برچسب‌نشده را بررسی کرده‌اند و یک الگوریتم انتقال پیام برای تخمین برچسب‌های صحیح را پیشنهاد داده‌اند. Zhou و همکاران(2015)

در مدل خود نمونه‌های دشواری را شامل شدند و  $\psi$ ،  $\pi$ ،  $\omega$ ،  $\psi$ -estimates،  $\pi$ ،  $\omega$  and  $\pi$ ،  $\psi$  با بهینه سازی معیار minimax  $Y$  با روی هندسه شرطی  $Z$  با توجه به  $Y$  بازیگرانه (Platanios et al. 2016)

مدل‌های تولیدی مختلفی برای  $\psi$ ،  $\pi$ ،  $\omega$ ،  $\psi$ -estimates،  $\pi$ ،  $\omega$  و  $\psi$  پیشنهاد دادند و آنها Gibbs sampling را برای استنتاج  $\psi$  و  $\pi$  و  $\omega$  استفاده کردند.

هیچ یک از این نویسنده‌گان آموزش<sup>۹</sup> یا تست را در نظر نگرفته‌اند و از برچسب‌های معتبر برای محا سبه خطاهای تخمین در آزمایشات خود استفاده کرده‌اند. برای اهداف ما، این آثار مدل‌های برچسب نویز را ارائه می‌دهند که می‌توانند یاد گرفته شوند—حتی بدون برچسب‌های صحیح—و با روش‌های آموزش و تست ما استفاده شوند. مدلی که توسط Whitehill et al. (2009) معرفی شده است

با ایتالیک نوشته شده است زیرا، در بخنس<sup>2</sup> 5.1.2، ما یک مدل مسایه برای ( $\psi | ty; p(tz)$ ) ارائه می دهیم که شامل دشواری نمونه و خطاهای برچسب گذار است؛ اما هدف ما برآورد  $\psi$  نیست بلکه نشان دادن آموزش و تست زمانی که  $\psi$  موجود است.

طبقه بندی کننده موجود موجود است، که قبلاً بر روی یک مجموعه داده جانبه با تعداد کافی نمونه های برچسب گذاری شده صحیح آموزش دیده است تا از وجود برخی نمونه های برچسب گذاری شده نویزدار جلوگیری کند. با توجه به یک ویژگی کننده، طبقه بندی کننده موجود احتمالات طبقه برای همه طبقات پیش بینی می کند، برخلاف برچسب گذار انسانی که برچسب نویزدار را نشان می دهد که یک طبقه را نشان می دهد. برای یک مجموعه داده جدید با برچسب  $\psi(z)$  و  $p(z)$  استفاده کردند. آن ها از این تخمین ها برای شناسایی نمونه هایی در مجموعه داده جدید که احتمالاً برچسب گذاری شده نادرست هستند و برای تصحیح یا حذف آن ها استفاده کردند. سپسی من توان یک طبقه بندی کننده جدید را با استفاده از برچسب های تصحیح شده آموزش داد یا آزمایش کرد. این روش روش متفاوتی برای دستیابی به مدل نویزدار آرائه می دهد و روش های ما می توانند با حل نمونه تند، آن را تکمیل کنند.

برخی از نویسندهای یادگیری مستترک یک مدل برچسب نویز و آموزش یک مدل پیش بینی کننده را انجام می دهند. این منابع در ادامه بحث می شوند.

### 1.2.2 آموزش (و یادگیری مستترک) (با مسائل درستیابی)

کارهای انجام شده در آموزش با مشکلات دروغ گویی در جدول 3 لیست شده اند و در اینجا بررسی شده اند.

(cid-sueiro و همکاران) 2012 (از دیدگاه بیانی استفاده کردند و به بررسی تلفیق با برچسب های ناقص پرداختند که به طور متفاوتی از برچسب های نویزیدیر مورد بررسی در اینجا مدل سازی شده اند. نویسندهای تلفیق با برچسب های ناقص را به یک تلفیق معادل برای برچسب های صحیح مرتبط کردند و جنبه های نظری ساخت یک تلفیق از یک تلفیق معادل داده شده را بررسی کردند. رویکرد آن ها با دیدگاه یکپارچه ای که در بخنس<sup>3</sup> 3.2 ارائه می دهیم، همخوانی ندارد؛ من توان آن را معکوس روش آموزش با حداقل خطای مربعات میانگین ارائه شده در بخنس<sup>3</sup> 3.4.1 تفسیر کرد. دو رویکرد در بخنس<sup>3</sup> 3.6.1 بحث شده اند.

(Natarajan و همکاران) 2013, 2018 (دیدگاه کلاسیک (یعنی فرکانس گرا) را اتخاذ کردند) د و بر روی برچسب های نویزآلود از یک برچسب گذار، با تشکیل یکتابع زیان جایگزین برای  $Z$  که یک برآوردهای طرف (در معنای کلاسیک) از تابع زیان ایده آل برای  $\psi$  است، محدوده بندی کنندهای دودویی آموزش دادند. روش آن ها در دیدگاه یکپارچه ما جا نمی گیرد؛ بحث بیشتری در بخنس<sup>3</sup> 3.6.2 آورده شده است. van Rooyen و Williamson (2018) یک چارچوب کلی برای یادگیری با برچسب های نویزآلود ارائه دادند که بخنس های از آن کارهای Cid-Sueiro (2012)، Cid-Sueiro (2014)، و همکاران (2013, 2018) را تعمیم می دهد.

Ratner و همکاران (2016, 2017) روش برای تولید برنامه ای چند برچسب نویز ب رای بسیاری از نمونه های برچسب نشده و پس از آن آموزش محدوده های تشخیص کننده با برچسب  $\psi$  ای نویز پیشنهاد کردند. برای مدل سازی ( $\psi | ty; p(tz)$ ), مدل های تولیدی با برچسب دهندهای مستقل یا وابسته معرفی کردند و  $\psi$  را با فرض پیش بینی های برابر برای کلاس ها برآورد کردند. آموزش به حداقل رساندن خطر تجربی انتظاری در برابر ( $\psi | ty; p(tz)$ ) بود.

نویسندهای باقی مانده به یادگیری و آموزش مستترک با برچسب های نویز توجه کردند. al. Raykaret (2010). از الگوریتم EM برای تخمین مستترک  $\psi$ ,  $\pi$  و  $\theta$  و آموزش یک کلاس بندی دوگانه لجستیک اس تقاده کردند. همچنین، فرم بایه ای این الگوریتم با پیش فرض های بر روی احتمالات خطای برچسب دهندگان و ارائه احتمالات پس تریم تقریبی از احتمالات خطای ارائه دادند.

توانایی ها و روش را به طبقه بندی چندگانه، رگرسیون مرتبه ای و رگرسیون گسترش دادند.

Khetan و همکاران(2018) نسخه ای از الگوریتم EM را ارائه دادند که بین آموزشی ک طبقه بندی دوگانه با استفاده از  $Z$  و تخمین های فعلی  $\hat{y}$  و  $\hat{\pi}$  جایجا من شود و سپس تخمین های  $\hat{y}$  و  $\hat{\pi}$  را با استفاده از پیش بینی های فعلی به روزرسانی می کند. آنها تابع ضرر آموزشی را با استفاده از پس زمینه برچسب صحیح برای هر یک از مقادیر ممکن برچسب صحیح وزن من دهند و به این ترتیب R ۷ برچسب صحیح را حذف می کنند.

برای آموزش یک شبکه عصبی قویون پیکربندی شده(CNN) بر روی برچسب های نویزآلود از یک برج سب دهنده تنها و تخمین زدن (2015) p(t|y)، Sukhbaatar et al. (2016) softmax یک شبکه پایه اضافه کردند. در طول آموزش، شبکه پایه یاد می گیرد که برج صافی به خروجی شبکه p(t|y) را پیش بینی کند در حالی که لایه اضافی (y|t) را تخمین می زند و برچسب نویز از خروجی شبکه پایه پیش بینی می کند. پس از آموزش، لایه اضافی می تواند بربده شود و شبکه پایه برای پیش بینی استفاده شود. این روش آموزش خارج از دیدگاه یکپارچه ما از آموزش قرار دارد و در بخش 3.6.3 بحث شده است.

Tanno و همکاران(2019) (یک شبکه عصبی عمیق)confusion matrices را به طور مشترک آموزش دادند و ماتریس های گیج کننده برای چند برچسب دهنده مستقل برآورد کردند. آن ها از ماتریس های گیج کننده برای تبدیل خروجی های شبکه عصبی عمیق به پیش بینی های برچسب های نویزی(noisy labels) استفاده کردند که مشابه کار Sukhbaatar و همکاران است. (2015). به ویژه، تانو و همکاران یک اصطلاح تقویت کننده را برای تقویت تقویت کننده را تضمین می کنند. شباه را به حداقل می رساند و تحت شرایط خاص، ارزیابی مناسب را تضمین می کند. نویسندها روش خود را با چندین روش دیگر مقایسه کرده اند، از جمله روش های Raykar et al. (2018)، Sukhbaatar et al. (2015) و Khetan et al. (2018) این رویکرد با دیدگاه یکپارچه ما را آموزش همخوانی ندارد و بحث بیشتری در بخش 3.6.4 آورده شده است.

تمام این نویسندها در طول تست یا ارزیابی های آزمایش دیگر، برچسب های صحیح را استفاده کرده اند؛ آنها به تست با مشکلات درستی ساری توجه نکردند.如表3 نشان می دهد، بسیاری از آنها چندین برچسب دهنده ممکنه وابسته را در نظر نمی گیرند که کار ما این امکان را فراهم می کند. برخی از آنها یک مدل برچسب دهنده نویز و یک مدل پیش بینی کننده را به طور مشترک یاد می گیرند که ما این را در نظر نمی گیریم. بسیاری از این آثار در چشم انداز یکپارچه ما از آموزش (بخش 3) فزار دارند. جدول بخشندهایی از Ratner et al. (2016، 2017) و Khetan et al. (2018) را که با استفاده ما از برآورد حداقل خطای مرتعات میانگین تجربی (بخش 3.4.1) سارگار هستند، برجسته ساری می کند.

در کار Raykar و همکاران(2010)، رگرسیون ۱۰۵ لستیک با استفاده از آنکاه italicized (شده است زیرا ما از آن به عنوان یک مثال توصیحی (بخش 5.4) نیز استفاده می کنیم.

### 1.2.3 تست با مشکلات حقیقت آمیز

همانطور که در بالا ذکر شد، تمامی کارهای مرتبط در زمینه آموزش از مجموعه ای از برچسب های صحیح اختصاصی برای تست استفاده کرده اند. کارهای مرتبط با تست و مسائل حقیقت یابی در جدول ۴ آورده شده است.

در سال 1994، Smyth و همکاران بررسی کردند که چگونه می توان از برچسب های نویزآلود که یک دانشمند یا طبقه بندهی کننده به تصاویر ماهواره ای ناحیه های سطح سیاره و نوساخته ای از مجموعه ای از برچسب های نویزآلود استفاده کرد. به وضوح، در این مورد، حقیقت مطلق در دسترس نیست. برای آزمایش برچسب های پیش بینی شده توسط یک دانشمند، نویسندها از الگوریتم EM پیشنهادی توسط Dawid Skene (1979) با استفاده از برچسب های نویزآلود از تمام دانشمندان دیگر برای تخمین برچسب های صحیح استفاده کردند که سپس به عنوان برچسب های صحیح در نظر گرفته شدند. همچنین، یک طبقه بندی کننده با مقایسه پیش بینی های خود با برچسب های صحیح تخمین زده همکاران، 1994) آموزش داده شد. این طبقه بندی کننده با مقایسه پیش بینی های خود با برچسب های صحیح تخمین زده شده توسط EM با استفاده از برچسب های نویزآلود تمام دانشمندان آزمایش شد.

مرجع	توصیحات	تعداد کلاس ها	برچسب گذارواب ستگیرچسب گذاروایستکی	با مدل یاد گیری برچسب ها ی نویزدار؟	Fits into our
Cid-Sueiro (۲۰۱۲) (2014)	<ul style="list-style-type: none"> <li>• نظری نظریه بیانی</li> <li>• طراحی شده صنعتی از دست دادن برای برچسب های جزئی</li> <li>• بازده معادل برای برچسب های صحیح</li> </ul>	دو یا چند گانه	نه )	نه	نه
Natarajan و همکاران (2013, 2018)	<ul style="list-style-type: none"> <li>• نظریه کلاسیک یا نظریه فراوانیست</li> <li>• تابع از دسترسی به جای که در معنای کلاسیک بی طرف است</li> </ul>	دوگانه	نه قابل اعمال(برح سب دهنده یک نوعی)	نه	نه
W van Rooyen illiamson (2018)	<ul style="list-style-type: none"> <li>• چارچوب کلی توسعه یافته از Cid-Sueiro (۲۰۱۲) و Natarajan (2014)؛ (2018) همکاران (2013، 2016)</li> </ul>	دوگانه	نه قابل اعمال(برح سب دهنده یک نوعی)	نه	نه
Ratner و همکاران (2017)	<ul style="list-style-type: none"> <li>• رگرسیون لجستیک</li> <li>• همچین بلند مدت کوتاه مدت شبکه حافظه</li> <li>• نویز برنامه ای برچسب زدن</li> <li>• به حداقل رساندن خطر تجربی مورد انتظار</li> </ul>	دوگانه	مستقل و وا بسته	نه	بله
Raikar و همکاران (2010)	<ul style="list-style-type: none"> <li>• بازبینی لوگاریتم تقریب به عنوان مثال اصلی</li> <li>• همچنین تقریب، تقریب ترتیبی</li> <li>• EM یا بیانی تقریب</li> </ul>	دو یا چند گانه	مستقل	بله	بله
ketan و همکاران (2018)	<ul style="list-style-type: none"> <li>• RNN تجربی کاهش متنی سازی</li> </ul>	دو یا چند گانه	مستقل	بله	بله
سوخباتر و همکاران (2015)؛ جیندال و همکاران (2016)	<ul style="list-style-type: none"> <li>• CNN</li> <li>• شبکه پایه یاد می گیرد pt(tyt txt)</li> <li>• لایه اضافی یاد می گیرد pt(tzt tyt)</li> </ul>	چندگانه	نه قابل اعمال(برح سب دهنده یک نوعی)	بله	نه
تano و همکاران (2019)	<ul style="list-style-type: none"> <li>• CNN</li> <li>• منظم کردن سیستم های برچسب دهنده دگان و ماتریس های گیج کننده</li> </ul>	چندگانه	مستقل	بله	نه

جدول 3: کارهای مرتبط در آموزش با مشکلات حقیقت آمیز.

## در مسائل درستی ساری در طبقه بندی تحت نطارت

مرجع	توضیحات	تعداد کلاس‌ها	برچسب گذارواب ستگیرچسب گذارواب
Smyth و همکاران (1994)	مقایسه برچسب‌های نویز دار یک برچسب دهنده با • برچسب‌های صحیح برآورده شده با اعمال EM به برچسب های نویز دار دیگر برچسب دهنگان • برچسب‌های چندگانه (نوع آتش‌سنجان یا غیر آتش‌سنجان) به برچسب‌های دوگانه (آتش‌سنجان یا غیر آتش سنج) (کاهش یافته)	چندگانه، کاهش یافته به دوگانه	مستقل
Lam و Stork (2003)	تأثیر برچسب‌های نویز بر احتمال خطا و اریانس احتمال خطای تخمین زده شده به صورت الف تابع احتمال خطای برچسب‌گذار، تعداد نمونه‌ها	چندگانه، کاهش یافته به دوگانه	نه قابل اعمال(بر جسم دهنده یک نوعی)
کارلوتو (2009)	بررسی تأثیر برچسب‌های پر سر و صدا بر دقت رابطه بین دقت‌های محاسبه شده بر روی برچسب‌های صحیح و روی برچسب‌های پر سر و صدا • تخمین تقریبی دقت ایده آل	چندگانه	نه قابل اعمال(بر جسم دهنده یک نوعی)
هولودنک و همکاران (2018)	مطالعه تجربی روش‌های تخمین دقت از برچسب های پر سر و صدا	چندگانه	مستقل و با مسته

جدول 4: کار مرتبط در مورد آزمایش با مسائل مربوط به صدق.

برای طبقه بندی دودویی، Stork و Lam (2003) احتمال خطای ایده آل ( $Pr(y^*) = z$ ) – دقت را به احتمال خطای برچسب دهنده های  $y = z = Pr(y^*)$  (و احتمال خطای طاهری طبقه بندی کنندگان) مرتبط دانستند. آن‌ها یک برآورد از  $y = z = Pr(y^*)$  (با توجه به یک‌قدر فرضی از ارائه دادند و تغییرات این برآور د را به عنوان یک تابع از  $z$  و تعداد نمونه ها  $N$  بررسی کردند.

تأثیر خطاهای حقیقی بر دقت اندازه‌گیری شده برای یک‌قدر برچسب دهنده Carlotto (2009) را تحلیل کرد. Carlotto یک معادله ای به دست آورد که دقت ایده آل را به دقت محاسبه شده در برابر برچسب‌های نویز دار مرتبط می‌کند و یک برآورد تقریبی از دقت را پیشنهاد کرد زمانی که احتمال خطای برچسب دهنده شناخته شده باشد.

یک مطالعه تجربی با داده‌های واقعی و شیوه سازی شده Holodnak et al. (2018) برای مقایسه گوناگون تکنیک‌ها برای تخمین دقت یک طبقه بندی کننده از برچسب‌های نویز دار انجام دادند. آن‌ها دو مدل معرفی کردند که وابستگی‌ها بین برچسب دهنگان یا RV برچسب‌های نویز دار را در نظر می‌گیرند و نشان دادند که تکنیک‌های تخمین که فرضی از کنند وابستگی شرطی مستقل هستند، با افزایش وابستگی برچسب دهنده ها برآوردهای کم اعتمادتری ارائه می‌دهند.

در یک مقاله بررسی بر طبقه بندی با نویز برچسب (Frénay and Verleysen, 2014)، ص 862، نویسنده‌گان یادداشت می‌کنند: "مشکلی که در ادبیات کمتر دکر می‌شود این است که تأیید مدل می‌تواند در حضور نویز برچسب دشوار باشد." واقعًا، تعداد ارجاعات به آزمایش بسیار کمتر از تعداد ارجاعات به یادگیری یا آموزش است و این کارها عمدهاً دقت را بررسی کرده‌اند. با این حال، کارهای بالا روش‌ها ای اصلی را منعکس می‌کنند:

محاسبه برچسب‌های درست تخمین زده شده  $y$  و استفاده از آن‌ها به عنوان مرجع؛ تخمین دقت به برخی از روش‌ها؛ و مطالعات مقایسه‌ای. کار ما شامل بسیاری از معیارهای آزمایش معمولی، از جمله دقت، دقت مثبت، بارگشت یا احتمال تشخیص، و احتمال خطای نادرست است. ما بر تخمین معیارها آزمایش تمرکز داریم تا برچسب‌های درست، و الگوریتم‌هایی برای محاسبه برآوردهای بھیه بیانی از معیارهای سکولار و مشترک توسعه داده ایم.

آزمایش‌های ما از برچسب دهنگان مستقل شرطی برای راحتی اجرایی استفاده می‌کنند، اما روش‌های مدل‌های برچسب‌های نویز دار با برچسب دهنگان وابسته شرطی را پذیرا هستند.

### 1.2.4 مقایسه ترکیبات برچسب دهنگان

در مورد مقایسه ترکیبات مختلف برچسب دهنگان، ما یک مقاله نسبتاً دور از موضوع پیدا کرده ایم.

برای طبقه بندی دودویی و یک برچسب دهنده، RV (1992) Lugosi برچسب های درست و RV برحسب های نویزدار را به عنوان ورودی و خروجی جداگانه یک کانال ارتباطی در نظر گرفت. لوزی جنبه های نظری تأثیر برچسب های نویز بر دقت را بررسی کرده است اگر یک طبقه بندی کنند ه از قاعده تصمیم گیری پس زمینه ای حداکثر یا نزدیک ترین همسایه استفاده کنند. در بخش 4، ما تشییه کانال را برای یک یا چند برچسب گذار انجام می دهیم، اما در مسیر متفاوتی ادامه می دهیم. ما پیشنهاد می دهیم که اطلاعات مشترک من تواند برای مقایسه ترکیبات برچسب گذارها استفاده شود، که به این معناست که اطلاعات انتقال دهنده توسط برچسب گذارهای متوجه چندگانه من تواند برابر با بیشتر از آنچه که توسط یک برچسب گذار کارشناس ارائه می شود، باشد.

### 1.3 نظریه طبقه بندی و ارزیابی تحت نطارت

این کار نظریه‌ی تخمین زنی را در زمینه‌ی طبقه بندی نطارت شده با مشکلات حقیقت یابی اعمال می کند و تصویر شماره 1 نمودارهای مفهومی برای این دو زمینه را ارائه می دهد. این نمودارها شبیه به نظر من رسند اما ان در روش اساسی ای متفاوت هستند: در طبقه بندی نطارت شده، فرآیند واقعی ناشناخته است. ان در حالی که در نظریه‌ی تخمین زنی، متغیر واقعی ناشناخته است. طبقه بندی نطارت شده به دنبال یافتن مدل پیش‌بینی کننده ای خوب است که به واقعیت های آینده، خارج از نمونه های ناشناخته از یک فرآیند ناشناخته، تعمیم می یابد، با توجه به مجموعه ای از نمونه های برچسب گذاری شده از فرآیند و خانواده ای از مدل ها. در این کار، نظریه‌ی تخمین زنی به تخمین دقیق ارزش فعلی، در نمونه، از یک متغیر غیر مشاهده شده، با توجه به اندازه گیری های نویزآلود و مدل فرآیند اندازه گیری، می پردازد.

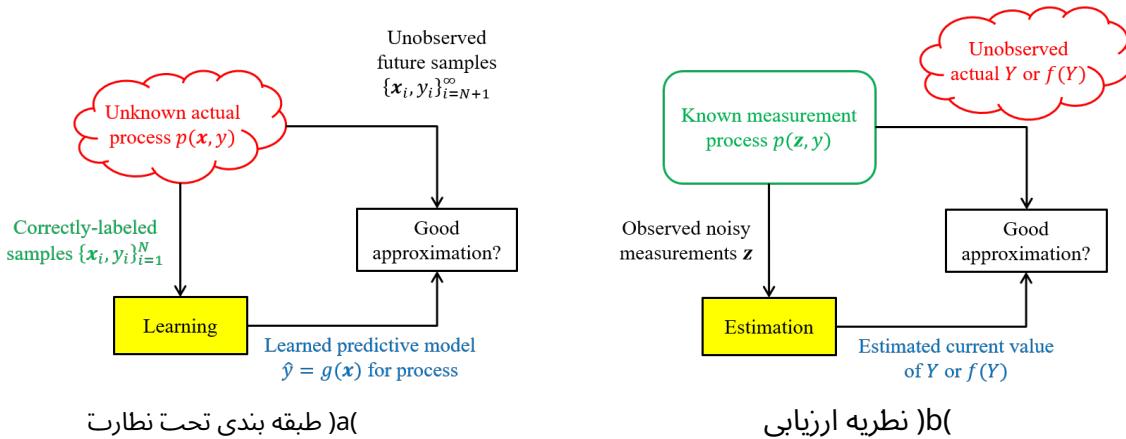
این زمینه ها نیز از عبارات و اهداف مشابه استفاده می کنند. طبقه بندی تحت نطارت پارامترهای مدل  $\theta$  را یاد می گیرد، نظریه ارزیابی یک برآورد یا محاسبه  $\hat{y}$  یا  $\hat{y}^*$  پیدا می کند و هر دو زمینه به دن بال پاسخی هستند که بر اساس برخی معیارها بهترین باشد. ما چند مثال ارائه می دهیم. در این مثال ها، معیار اساساً مشابه است؛ تفاوت ها در اجزایی که فرصت می شود شناخته شده باشند، فضای راه حل و روش های بهینه سازی است.

ابتدا، هر دو زمینه از تخمین ML استفاده می کنند: طبقه بندی پارامترهای غیر تصادفی را انتخاب می کند  $\theta$  به حداکثر رساندن  $p(y|tx; \theta)$  یا  $p(y|tx; \theta)$ ، و برآورد انتخاب می کند  $\hat{y}$  به حداکثر رساندن  $(y|z)^*$ . دوم، وقتی در طبقه بندی مدل پارامترها به عنوان  $\theta$  با پیش‌بینی  $p(\theta)$  استفاده می شود، از معیار داکثر احتمال (MAP) استفاده می کند و  $\theta$  را انتخاب می کند تا  $(x|y, p)$  را بیشینه کند. وقتی ارزیابی از معیار MAP استفاده می کند،  $y = \arg \max y p(y|z)^*$

سوم، طبقه بندی می تواند کاهش میانگین ضرر مربع را به حداقل برساند  $E[\sum_{i=1}^N (y_i - \hat{y}_i)^2]$  در حالی که تخمین گیری ممکن است خطای مربع های میانگین (MSE)  $E[(Y - Z(h))^2]$  را به حداقل برساند. ما از تخمین گیری میانگین مربع های کمینه (MMSE)  $E[(Y - Z(h))^2]$  را به حداقل میانگین صفر-یک را کمینه کند. نهایت، طبقه بلندی می تواند صفر میانگین صفر-یک را کمینه کند  $E[Y|Z]$ .

$$h^{\text{MMSE}}(Z) = \arg \min_h E[(Y - h(Z))^2] = E[Y|Z]. \quad (1)$$

در نهایت، طبقه بلندی می تواند صفر میانگین صفر-یک را کمینه کند  $E[Y|Z]$ .



تصویر 1: مقایسه طبقه بندی تحت نطارت و نظریه ارزیابی. در طبقه بندی تحت نطارت، هدف استفاده از مجموعه ای از نمونه های برچسب گذاری شده صحیح از یک فرآیند واقعی ناشناخته برای یادگیری یک مدل پیش بینی کننده است که به واقعیت های آینده و خارج از نمونه های فرآیند تعمیم می یابد. در نظریه ارزیابی، هدف ارزیابی ارسان نمونه ای یک متغیر غیر مشاهده شده از اندازه گیری های نویرآلد تولید شده توسط یک فرآیند اندازه گیری شناخته شده است. برای هر مسئله، عنصر ناشناخته کلیدی به رنگ قرمز، عنصر شناخته شده کلیدی به رنگ سبز و عنصر مورد نظر به رنگ آبی نشان داده می شود.

نظریه (در اصنافه نامه D) این است که محاسبه کننده MPE با محاسبه کننده MAP همخوانی دارد:

$$h^{\text{MPE}}(Z) = \arg \min_h \mathbb{E}[\mathbb{1}(Y \neq h(Z))] = \arg \max_y p(y|Z) = h^{\text{MAP}}(Z). \quad (2)$$

استفاده ما از نظریه ارزیابی با فرص این آغاز می شود که یک مدل برچسب گذاری نویرآلد خوب ( $t$ )  $\psi(z|ty)$  در دسترس است. اگر مجموعه داده های جانبی  $\{y', t'\}$  در دسترس باشد، این مدل من تواند از طریق یادگیری تحت نطارت به دست آید. بسیاری از کارهای مرتبط به مسئله یادگیری مدل هنگامی که برچسب های صحیح در دسترس نیستند اما برخی از دانش در مورد رفتار خطای برچسب گذاران و/یا پیش بینی های کلاس وجود دارد. بنابراین، این مقاله به بسیاری از کارهای مرتب و به طور کلی با آن ها سارگار است.

#### 1.4 وکلاد نوین و سازمان دهنده

ما از خط تیره برای برجسته سازی وکلاد نوین استفاده می کنیم. ما برخی از وکلاد مفهومی را بحث می کنیم و سپس از سه سوال مطرح شده در بخش 1.1 برای پوشش سازمان دهنده باقی مانده مقاوله و ذکر دیگر وکلاد استفاده می کنیم. به عنوان مرجع، جداول های 5، 6 و 7 کوتاه ترین ها، نمادها و توزیع های مهم استفاده شده در این مقاله را لیست می کنند.

طبقه بندی تحت نطارت با مشکلات حقیقی شامل سه عنصر مختلف است که هر کدام نیاز به مدل دارند. فرآیند برچسب گذاری یا برچسب گذاری نیاز به مدلی برای مدل نویرآلد دارد، آمورنیش مدل پیش بینی کننده یاد می گیرد و تست نیاز به مدلی برای تست دارد. بسیاری از کارهای مرتبط به بررسی روش هایی برای یادگیری مدل نویرآلد و/یا استفاده از مدل نویرآلد برای آموزش مدل پیش بینی کننده می پردازند.

بزرگترین مفهومی این مقاله، استفاده از نظریه ارزیابی بیانی در آموزش و تست است. به باور ما، این مقاله اولین مقاله ای است که این کار را انجام می دهد.

این دیدگاه را در نظر بگیریم و امکانات آن را تا این حد گستردۀ دنبال کنیم. ما روی آموزش و آزمایش تمرکز می‌کنیم و با این فرصت شروع می‌کنیم که یک مدل برچسب نویزدار  $t\pi(t|ty, t)$  است، بنابراین این کار بیشتر کارهای مرتبط را تکمیل می‌کند. رویکرد بیزی به طور طبیعی اجراه می‌دهد تا چندین برچسب، ترکیب متفاوتی از برچسب‌ها را برای هر نمونه، و برچسب‌های نویزدار ناپدید و/یا متناقض داشته باشد. همچنین از مدل‌های دارای برچسب نویز با برچسب‌های وابسته به شرط  $t_3$  پشتیبانی می‌کند.

در این کار، من گوییم که یک برآورده تنها در صورتی بهینه است که سه شرط را برآورده کند: از یک تخمین مناسب استفاده کند، از تمام اطلاعات موجود به طور کامل بهره برداری کند، و معیار جرمیه به خوبی تعریف شده را به حداقل برساند یا یک معیار مطلوبیت تعریف شده را به حداقل برساند. برآورده کننده ای که هر یک را این الزامات را برآورده نمی‌کند، زیر بهینه در نظر گرفته می‌شود.

مشارکت‌های فنی ما شامل روش‌های آزمایش و آموزش است که در این مفهوم دقیق بهینه هستند. رویکرد تست بهینه ما بر تخمین معیارها تمرکز می‌کند، مدل آزمایشی را معرفی می‌کند که بهره برداری کامل از اطلاعات موجود را امکان پذیر می‌سازد، و از معیار MMSE استفاده می‌کند. در مقابل، برحی از روش‌های غیربهینه نظریه تخمین را به طور کامل حذف می‌کند، برعکس دیگر به جای معیارها، برچسب‌های صحیح را تخمین می‌زنند، و برعکس دیگر یک مدل آزمایش را حذف می‌کند و نمی‌توانند از اطلاعات موجود استفاده کامل کنند.

روش‌های تمرینی بهینه ما یک تابع احتمال مناسب، پسین، خطرا یا خطر را انتخاب می‌کنند که به دفعه اصلی از تمرین ایده آل وفادار باشد. آنها از مدل noisy-labels برای بهره برداری کامل از اطلاعات موجود استفاده می‌کنند. و معیار  $ML$ ,  $MAP$ ,  $MMSE$  را اعمال می‌کنند. برعکس از روش‌های غیربهینه، نظریه تخمین را نادیده می‌گیرند، در حالی که برعکس دیگر به جای رسیک، برچسب‌های صحیح را تخمین می‌زنند.

روش‌های پیشنهادی ما هرگز برچسب‌های صحیح را تخمین نمی‌زنند، زیرا از نظر ما، آنها تخمین درست نیستند. این دیدگاه سهم مفهومی جدید دیگری را نشان می‌دهد: انتخاب خودآگاه ما از تخمین برحسب صحیح خودداری می‌کند. ما عمدها از تصمیم‌گیری سخت در مورد برچسب‌های صحیح مشاهده نشده اجتناب می‌کنیم، زیرا انجام این کار باعث ایجاد خطاهایی می‌شود که در آموزش و آزمایش منتشر می‌شوند. به این ترتیب، کار ما با رویکردهای موجود و غیربهینه که برچسب‌های صحیح را تخمین می‌زنند و سپس به گونه‌ای پیش می‌روند که گویی برچسب‌های تخمین زده شده درست هستند، متفاوت است.

چگونه می‌توان یک طبقه بندی کننده را در حضور مسائل مربوط به صدق آزمایش کرد

این سوال عمیقاً در بخش 2.2 برای طبقه بندی بازیزی بررسی شده است، ما مدل آزمایش را برابر برچسب‌های نویزدار و پیش‌بینی شده پیشنهاد می‌کنیم (بخش 2.1، شکل 2). سپس روش های آزمون تخمین-نظری را به صورت زیر استخراج می‌کنیم:

ما معیارهای مختلف را بر حسب دو RV معمولی که مستقل و تقریباً به طور معمول توزیع  $\text{Beta}$  شده اند بیان می‌کنیم (بخش 2.2).

ما پیشنهادی حاسیه ای تقریبی را برای دقت، دقت، یادآوری یا احتمال تشخیص، احتمال  $\text{F}$ -هشدار کاذب، و  $F$ -اسکورن و پیشنهادی تقریبی مفصل برای احتمال تشخیص و عدم احتمال هشدار کاذب و همچنین دقت قلعه و فراخوانی بدست می‌آوریم (بخش 2.3).

ما آزمایش MMSE را پیشنهاد می‌کنیم و الگوریتم‌های بیز تجربی را برای تخمین پارامتر مد (ح) ل آزمایش از طریق تخمین تکراری MMSE توسعه می‌دهیم.

3. برای سهولت پیاده سازی، شبیه سازی ها و آزمایش های ما از برچسب های مستقل مستروط استفاده می‌کنند، اما مستقایت و الگوریتم ها این کار را نمی‌کنند.

(الگوریتم های 1 و 2، شکل 3)، و ما رابطه آنها را با الگوریتم EM و همگرایی مورد بحث قرار من دهیم (بخش 2.4).

ما نحوه محاسبه تخمین های یهیه بیزی (MAP، MMSE، و مناطق باورنکردنی) معیارها را (د) ز پارامترهای مدل آزمایشی تخمین و پسین توضیح من دهیم (بخش 2.5).

ما همچنین برخی از روش های آزمایش جایگزین (بخش 2.6.1، الگوریتم 3 و تخمین کامل بیزی (بخش 2.6.2) را توصیف من کنیم.

برای طبقه بندی چند کلاسه (بخش 2.7)، مدل آزمایشی را گسترش دهید (بخش 2.7.1)، الگوریتم بیز تجربی دیگری را برای آزمایش MMSE ارائه کنید (الگوریتم 4، شکل 4)، و پسین های تقریبی را برای دقت و 2.7 اجرای مجرای مجزا استخراج کنید.

- چگونه من توان یک طبقه بندی کننده را در حضور مسئکلات درستی یاب آموزش داد؟  
ما این سوال را در بخش 3 بررسی من کنیم. فرص مستقلیت نمونه ها (بخش 3.1) را بارگو من کنیم و دیدگاه یکپارچه ای از آموزش ارائه من دهیم که شامل و سازماندهی برخی از کارهای مرتبط است (بخش 3.2).

برای مدل های احتمالی (مثلًا تبعیض آمیز یا مولد)، ما تابع احتمال یا پسین پارامترهای مدل پیش بینی کننده را برای حقیقت یاب مسائل استخراج من کنیم، به طوری که اصل آموزش ایده آل (MAP) (حفظ منجذب) شود (بخش 3.3).

برای مدل های غیر احتمالی (بخش 3.4)، که با به حداقل رساندن ریسک تجربی آموزش داده من شوند، ما آموزش MMSE را پیشنهاد من کنیم که تخمین MMSE ریسک تجربی را به حداقل من رساند، و نشان من دهیم که این رویکرد به همان هدف آموزش پیشنهاد شده توسعه برخی از کارها مرتبط منجر منجذب شود (بخش 3.4.1). ما ویژگی های مرتبط با تخمین MMSE را بررسی من کنیم (بخش 3.4.2)، شکل مناسب آن را در محاسبه گرادیان ذکر من کنیم (بخش 3.4.3)، و برخی موارد خاص را در نظر من گیریم (بخش 3.4.4). ما یک شرط اساسی را برای سارگاری برآورده اگر MMSE پیشنهاد من کنیم (بخش 3.4.5).

در مرحله بعد، به برخی از جنبه های آموزش MMSE اشاره من کنیم که آن را حذف، تر از آموزش nML و MAP من کنند (بخش 3.5). ما همچنین برخی از رویکردهای آموزش جایگزین را که با دیدگاه یکپارچه مطابقت ندارند مورد بحث قرار من دهیم (بخش 3.6). در نهایت، روش هایی را برای آموزش توندو با زیرساخت هایی که برای مسائل مربوط به حقیقت طراحی نشده اند توصیف من کنیم (بخش 3.7).

- چگونه من توان ترکیبات مختلف برچسب گذاران با توانایی های مختلف را مقایسه کرد؟  
این سوال به طور خلاصه در بخش 4 مورد مطالعه قرار من گیرد. ما یک قیاس ساده بین مدل نویز-برچسب و یک کانال ارتباطی ایجاد من کنیم که اطلاعات متقابل را به عنوان مبنای برای مقایسه ترکیبی از برچسب ها پیشنهاد من کنند. ما بر روی کانال پخش متقابله دودویی (بخش 4.1) تمرکز من کنیم، و پیشنهاد من کنیم اطلاعات متقابله مجموعه ای از برچسب ها را بر حسب اطلاعات در برچسب معادل واحد بیان کنیم (بخش 4.2).  
ما مشاهده من کنیم که، در تئوری، چندین برچسب گذار متوسط من توانند به عنوان یک برچسب ب شناس خبره، اطلاعاتی یا آموزنده تر باشند (بخش 4.3).

نتایج تجربی در بخش 5 طاهر من شوند. ما برای تولید بسیاری از برچسب های درست، پر سر و صد ا و پیش بینی شده به شبیه سازی تکیه کردیم (بخش 5.1). درست شبیه سازی کردیم و پیش بینی کردیم

گسترش	محفظ
کanal پخس متقارن باينزى	بى بى سى
کanal متقارن باينزى	BSC
قصصيە حد مرکزى	CLT
شبکه عصبي کانولوشنال	CNN
حداکثر كردن انتطار	در
کاهش خطر تجربى	ERM
شرایط تقریب لحظه	MAC
حداکثر تفاوت مطلق	دیوانه
حداکثر از عقب	نقشه
حداکثر احتمال	ML
حداقل میانگین مربعات خطأ	MMSE
حداقل احتمال خطأ	9
خطاي ميانگين مربع	MSE
نقطه عملیاتی	روشن
فراخوان دقیق	P-R
ويژگی عملکرد گیرنده	ROC
منغير تصادفي	RV

## جدول 5: لیست اختصارات.

برچسب‌ها به صورت ساده (بخش 5.1.1) و ما از یک مدل برچسب‌های نویزدار خاص (بخش 5.1.2) که متشابه مدل Whitehill et al. (2009) است، استفاده کردیم. برای تست، نتایج چندین آزمایش بر روی طبقه بندی دوگانه (بخش 5.2) را بررسی می‌کنیم و در یک آزمایش بر روی طبقه بندی چندگانه (بخش 5.3) گزارسش می‌دهیم. برای آموزش، مثالی در مورد رگرسیون لوژیستیک دوگانه ارائه می‌دهیم (بخش 5.4). برای مقایسه برچسب دهنده‌گان، آزمایش‌هایی را نشان می‌دهیم که از روش‌های آموزشی و تستی پی شنهادی برای اثبات امکان اطلاعات مشترک معادل استفاده می‌کنند (بخش 5.5).

بخش 6 شامل یک خلاصه و نتیجه گیری ها (بخش 6.1)، یک روند کاری برای طبقه بندی تحت نطارت با مسائل حقیقت یابی (بخش 6.2)، و پیشنهادات آینده (بخش 6.3) من شود. همچنین چندین پیوست نیز شامل می شود. پیوست A معیارهای تست را بر اساس RV های معمول از بخش 2.2 استخراج می کند. پیوست B نتایج مربوط به نسبت های RV های نرم افزار مسترک را خلاصه می کند؛ که برای محسنه احتمالات پس زمینه معیارهای سکولار مفید هستند. پیوست های C و D MMSE، MPE و برآورد MAP را بررسی می کنند. پیوست E استنتاج های روش های آموزش در بخش های 3 و 3.4 را ارائه می دهد. پیوست F جزئیات مربوط به مثال آمورنژ رگرسیون LOGRارتومیک در بخش 5.4 را می دهد.

## 2. تست های با مشکلات واقعی: ارزیابی با کلیدهای پاسخ کثیف

ما تست ها را قبل از آموزش انجام می دهیم زیرا نتایج و روش های ارائه شده در اینجا، بدون توجه به اینکه یک طبقه بندی کننده چگونه آموزش دیده است، اعمال می شوند. آن ها حتی اگر آموزش با توجه به مسائل حقیقت یابی انعام نشده باشد، می توانند استفاده شوند. علاوه بر این، اگر کسی نیاز به آموزش مدل با مسائل حقیقت یابی دارد، احتمالاً نیاز خواهد داشت که مدل آموزش دیده با مسائل حقیقت یابی نیز تست شود. ممکن است کسی تمایل به شروع آموزش نداشته باشد اگر شک کند که مسائل حقیقت یابی معیارهای ای تست را باطل می کند. این بخش اطمینان می دهد که تست قابل اعتماد ممکن است.

مطالعات سنتی، شاخص‌ها را بر روی یک مجموعه آزمایش‌ایده‌آل  $\{y^*, y\}$  محاسبه می‌کند تا اختلاف بین برچسب‌های پیش‌بینی شده  $y^*$  و برچسب‌های صحیح  $y$  را اندازه‌گیری کند. در یادگیری ماشین،

## در مسائل درستی ساری در طبقه بندی تحت نطارت

نماد	توضیحات	نماد	وصف
B	توزيع برنولی	tg	پیس آستانه پیس. مدینک
بـتا	توزيع بتا	htMMSE	برآورده دلخواه
سـس	تعداد کلاس ها	j	MMSE برآورده
Ctempt. [tCt]	ماتریس سردرگمی	ptDt. [tPtDt]	نمونه indexn
Cat	توزع طبقه بندی شده	FA, [P]	تکرار یا ساختار عمومی FItem
ptBt. [tFtBt]	F-score	~pD ~pFA   P_D pFA	احتمال تشخیص احتمال خطای اعلام
من	اطلاعات متقابل	سـ	پارامترهای OP
J ideal	هدف آموزش ایده آل	ـ	و صنعتی پیس از پلانک
J pri	عيارت اصلی	ـ	ساختار برچسب
Jtreg	مدت تنظیم	ـ	ویرگی
K	. ماتریس شرایط . پارامتر ماتریس conf	xt, t xt, t xt, [tXt, t Xt, t Xt]	برچسب های صحیح
L	عملکرد از دست دادن	ـ	برچسب های پیس بین شده
n^tLtMMSE	MMSE تخمین تابع صفر	ـ	برچسب های صحیح تخمین زده است
ـ	نه در الگوهای نمونه گیری ترسیم می کند.	ـ	برچسب های پرس و صدا
ـ	تعداد نمونه	ـ	clipping valuen^t
Nt1	تعداد دفعات 1^tytit =	ـ	پارامتر F-scoreRte
mpt. [tRt]	ـ	ـ	tRtMMSE نمونه دشواری
ـ	تخمین زده از RV با ریسک	ـ	مسئل خطا (57) (یا 60)
ـ	تعداد لیبل کننده ها	ـ	بروب های معمول پرس و صدا Ut, t Vt [ ]
ـ	ـ	ـ	پارامترهای pred.-model
ـ	ـ	ـ	ـ تنظیم وزن
ـ	ـ	ـ	ـ کلاس قبل
ـ	ـ	ـ	ـ آستانه تصمیم گیری
ـ	ـ	ـ	ـ خطاطیدنی برچسب گذار
ـ	ـ	ـ	ـ پارامترهای برچسب نویزدار
ـ	ـ	ـ	ـ بردار صفر
ـ	ـ	ـ	ـ عملکرد نشانگر

جدول 6: لیست از نمادهای اصلی. گوشه ها نشان دهنده RV ها هستند.

توزيع	توضیحات
p(tz   ty; ψ)	توزيع شرطی برچسب های پرس و صدا
p(tz   ty; ψ)π(y)	مدل برچسب نویز
p(~yt   ti; ~pD, ~pFA)	توزيع شرطی برچسب پیس بین شده (3)، (4)، (5)، (6)
(iy; iz(p) FAp ~Dp ; iy) (p)	مدل تست (7)
FAp ~Dp ~iy; iz, iy	کلاس پس زمینه تست (10)
p (y_i   z_i; ψ_i)	پس زمینه کلاس آموزشی (41)

جدول 7: لیست توزیع های کلیدی.

$y^*$  با اعمال مدل یادگیری و به هر ویرگی در  $\text{eKTOP}$  نیست، بنابراین نتایج این بخش به مشکلات خارج از یادگیری ماشین نیز اعمال می‌شود، مانند هماهنگ مشاهدات توسط چندین پژوهش برای مثال، بینید: Dawid and Skene 1979; Raykar et al., 2010).

با مشکلات درستی سنجی، تست باید با مجموعه تستی  $\{y^*, z, \psi, \pi\}$  کار کند در عوض  $\{y, y^*\}$ . شاخص‌ها توابعی از RV‌های برچسب صحیح هستند که غیر مشاهده اند، بنابراین شاخص‌ها RV‌هایی هستند که مقادیر آن‌ها غیر مطمئن باقی مانند. بنابراین ما تست را به عنوان یک مسئله برآورد در نظر می‌گیریم: ما می‌خواهیم مقادیر در نمونه شاخص‌ها بر روی مجموعه تست را بدانیم. ما به احتمالات R $V$ ‌های شاخص می‌پردازیم، تحت شرط  $y^*$  و  $y$ . سپس برآورد بھینه یا مناطق قابل اعتماد برای RV‌های شاخص می‌توانند از احتمالات محاسبه شوند.

نزدیک ترین روش ما استفاده از تمام اطلاعات موجود است، از جمله برچسب‌های پیش‌بینی شده  $y^*$ . مشکل مشابه با ارزیابی یک آزمون چندگزینه ای با استفاده از کلید پاسخ‌ها از یک یا چند دستیار آموزش با دستخط ضعیف است. کلید پاسخ‌ها اطلاعات مربوط به پاسخ‌های صحیح را ارائه می‌دهند، اما پاسخ‌های دانش آموز نیز همینطور. ما می‌توانیم بهترین برآورد نمره را با مشاوره پاسخ‌های دانش آموز و کلید پاسخ‌ها به دست آوریم، نه اینکه تنها به کلید پاسخ‌ها اعتماد کنیم.

بیشتر این بخش به کلاس بندی دوگانه اختصاص دارد؛ بخش ۷.۲ به گسترش‌های کلاس بندی چند گانه می‌پردازد. برای کلاس بندی دوگانه، چندین معیار سکالر رایح را در نظر می‌گیریم: دقت، دقت، بارگشت یا احتمال تشخیص، احتمال خطأ، و F-score. هر یک از این شاخص‌ها مقادیر دار محدوده  $[0, 1]$  دارد. برای احتمال خطأ، مقادیر کوچکتر نشان دهنده عملکرد بهتر هستند؛ برای سایر شاخص‌ها، مقادیر بزرگتر نشان دهنده عملکرد بهتر هستند. جدول ۸ فرم‌های تجربی این شاخص‌ها را رائه می‌دهد که به آنها acc، prec<sub>emp</sub>، F<sub>β</sub> و  $F_{\beta} = \frac{F_{\beta} \cdot rec}{D_p + F_{\beta}}$  به ترتیب. همچنین دو معیار مشترک مشتراك را در نظر می‌گیریم: خط ویره گیرند  $ROC(P-R)$  و نقاط عملکرد دقت-بارگشت  $(P-R)_c$ ، به ترتیب  $(P-R)_c = \frac{rec}{F_{\beta} + D_p}$ . معیارهای تجربی در حالت ایده‌آل استفاده می‌شوند وقتی  $y^*, y$  در دسترس است. فرم‌های RV مربوطه  $D_p$ ،  $PRECACC$ ،  $REC$ ،  $PREC$ ،  $FAPDP$ ،  $REC$ ،  $FAP$ ،  $REC$ ،  $F_FAP$ ،  $REC$ ،  $F_FAP$ ،  $REC$  همچنین  $\beta$  معنای یا معیار تجربی یا واقعیت یک RV استفاده می‌کنیم.

## 2.1 آزمایش فرضیات

در آزمایش ایده‌آل، معیارها شامل ویرگی‌های توبولوژی نمی‌شود، بنابراین  $X$  و  $X$  را از بررسی خود می‌کنیم. اگر ما یک مدل آزمایش خوب برای  $p$ ،  $z$ ،  $y$  (داده‌ی  $p$ ) داشتیم، می‌توانستیم از آن برای تخمین معاشرهای RV استفاده کنیم. ممکن است پیشنهاد شود که آن را از مجموعه ای جانبی  $\{y^*, y, z\}$  یاد بگیریم، اما اگر چنین مجموعه‌ای موجود باشد، می‌توان فقط آزمایش ایده‌آل  $\{y^*, y, z\}$  انجام داد.<sup>6</sup>

در عوض، ما باید مدلی برای  $p$ ،  $z$ ،  $y$  (طراحی کنیم و پارامترهای آن را از  $\{y^*, y, z\}$  تخمین بزنیم، بدون این که امیدی به یادگیری آنها از داده‌های جانبی باشد. ما این مشکل را با اعمال تکنیک‌های نظریه برآورد به جای یادگیری ماشین حل می‌کنیم.

فرضیات خود را در اینجا بیان می‌کنیم و شکل 2 مدل گرافیکی نشان دهنده آنها را نشان می‌دهد.

— —

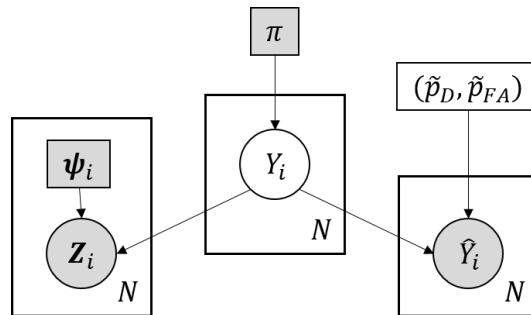
4. به یاد بیاورید که احتمال تشخیص، حساسیت و نرخ مثبت واقعی معانی مشابهی دارند.

5. احتمال خطأ در اعلام مشابه است با  $1 - specificity$  (و نرخ مثبت تاریخت).

6. ما انتخاب کرده ایم که هیچ وابستگی به ویرگی‌های توبولوژی را حذف کنیم. اگر کسی پیشنهاد کند که یک مدل آزمایش ما نند  $p$ ،  $x$ ،  $y$ ،  $z$ ،  $y^*$  یاد بگیرد، سپس یک تناقض مشابه به وجود خواهد آمد.

مقیاس	فرم تجربی		فرم متغیر تصادفی	
	نماد	بیان	نماد	بیان
دقت	$acc$	$\frac{tytit = t ytit^N}{N}$	$ACC$	$\frac{tNt1}{tVt + Nt1}$ به عنوان
دقت	مقدمه	$\frac{tytit = 1 \text{ and } ytit = 1^N}{tits = 1^N}$ خوب بارها	$PREC$	$\frac{\sum_{t=1}^N}{\sum_{t=1}^N}$ در
ofn Detection.n Recall	$ptDt, t rec$	$\frac{tytit = 1 \text{ and } ytit = 1^N}{ytit = 1^N}$ نه بارها	$PtDt, t REC$	$\frac{\text{در}}{tVt + \text{صورت}}$
پرور هشدار کاذب	$ptFA$	$\frac{tytit = 1 \text{ and } ytit = 0^N}{ytit = 0^N}$ نه بارها	$PtFA$	$\frac{tNt1t / Nt t U}{1 t (tUt + t Vt)}$
F ( $t\beta > t 0$ ) امتیاز	$F_{temp, \beta}$	$\frac{(1 + t \beta t2t) t prect t rex}{\beta t2t prect + t rex}$	$Ft\beta$	$\frac{(1 + \beta^2)U}{\beta * 2 * (U + V) + ^N1 / N}$

جدول 8: شاخص های طبقه بندی دوگانه: فرم های تجربی و فرم های RV در مفهوم RV های مشترک U و V از (8) و (9).



تصویر 2: مدل گرافیکی برای روش آزمایش. مستطیل های کوچک نشان دهنده متغیرهای غیرتصادفی هستند؛ دایره ها نشان دهنده RV ها هستند. سایه دهنده نشان دهنده متغیری است که به طور کامل مشاهده شده است. مستطیل های بزرگ نشان دهنده  $N$  نمونه های مستقل متغیرهای محصور شده هستند که توسط شمارش می شوند.

دارای توزیع شرطی  $p(z_i | \psi_i)$  و وابسته به برچسب پیش بینی نشده نیستند. به عنوان یک عمل متعارف، همچنین فرض من کنیم نمونه ها مستقل هستند، بنابراین  $p(z_i | \psi_i) = \prod_{i=1}^N p(z_i | \psi_i)$ .

در مرحله بعد، مشاهده من کنیم که دقت، دقت، F-score من توانند هر کدام تنها به صورت کلاس  $acc = \pi(0)$  پیش بینی شده، احتمال تشخیص و احتمال خطأ نادرست نوشته شوند. به عنوان مثال،  $1 - \pi(1)$  بینی شده،  $pD$  و  $pFA$  احتمال تشخیص و احتمال خطأ نادرست نوشته شوند. ما پارامترهای نقطه کار (OP) را معرفی من کنیم  $\tilde{pFA}$ ، که برای تعیین سایر شاخص ها کافی هستند زیرا کلاس  $\tilde{pFA}$  بینی شده فرض من شود که شناخته شده است. پارامترهای OP نمایانگر عملکرد انتظاری بر روی مجموعه آزمایشی قبل از مشاهده  $\hat{Y}$  و  $Z$  هستند.

سپس هر برچسب پیش بینی شده  $\hat{Y}$  را به عنوان یک واقعیت از یک RV برچسب پیش بینی شده  $\hat{Y}$  در نظر می گیریم که تحت شرط RV برچسب صحیح  $\hat{Y}$  و  $\tilde{pFA}$  قرار دارد. توزیع شرطی برچسب پیش بینی شده

به سادگی است

$$p_{\hat{Y}_i|Y_i}(0|0; \tilde{p}_D, \tilde{p}_{FA}) = 1 - \tilde{p}_{FA}, \quad (3)$$

$$p_{\hat{Y}_i|Y_i}(1|0; \tilde{p}_D, \tilde{p}_{FA}) = \tilde{p}_{FA}, \quad (4)$$

$$p_{\hat{Y}_i|Y_i}(0|1; \tilde{p}_D, \tilde{p}_{FA}) = 1 - \tilde{p}_D, \quad (5)$$

$$p_{\hat{Y}_i|Y_i}(1|1; \tilde{p}_D, \tilde{p}_{FA}) = \tilde{p}_D. \quad (6)$$

فرص مستقل بودن نمونه ها به معنای  $(\hat{y} \perp \!\!\! \perp y)_{(\tilde{p}=1)}$  )  $\text{FAP} \sim \text{DP}$  ;  $\psi$

فرص نهایی ما این است که، با توجه به  $y^i, Y^i, Z^i$  و  $X^i$  به صورت conditionally independent گذارند، به ترتیب، این فرص ممکن است قوی باشد، اما این ام که در آن  $(\text{FAP} \sim \text{DP})$  و  $\psi$  تنها بر  $Y^i$  و  $Z^i$  تأثیر می گذارند، به ترتیب. این فرص ممکن است قوی باشد، اما این ام ر منعکس کننده مورد معمولی است که در آن  $Y^i$  بدون دسترسی به  $Z^i$  تولید می شود. به عنوان مثال،  $Y^i$  می تواند خروجی یک مدل پیش بینی کننده باشد که تنها ورودی آن  $X^i$  است، یا  $Y^i$  می تواند تشخیص باشد که تو سط یک پزشک انجام سده است که تشخیص های دیگر پزشکان را ندیده است.

بنابراین، مدل آزمایش ما

$$p(\hat{y}_i, z_i | y_i; \psi_i, \tilde{p}_D, \tilde{p}_{FA}) = p(\hat{y}_i | y_i; \tilde{p}_D, \tilde{p}_{FA}) p(z_i | y_i; \psi_i). \quad (7)$$

ما  $X^i$  را حذف کردیم، بنابراین  $Y^i$  تنها RV است که می تواند  $Z^i$  و  $\hat{Y}^i$  را به هم متصل کند و مدل شامل چنین اتصالی است. مدل با داشتن تنها دو پارامتر، ساده است. همانطور که در ابتدای این بخش توضیح داده شده، آنها از طریق یادگیری نطارت دار تعیین نمی شوند. بخش 2.4 روش های تکراری برای تخمین آنها از  $\{\pi, \psi, z, y\}$  را ارائه می دهد.

## 2.2 RV های متريک بر اساس RV های معمولی

در اينجا نشان می دهيم که هر شاخص RV می تواند به صورت RV مستقل و تقريباً نرمال نوشته شود.

### 2.2.1 مقیاس های RV به صورت مقیاس های معمولی

ما با در نظر گرفتن یک معیار تنها آغاز می کنیم، به ویژه احتمال تشخیص. با توجه به مشکلات حقیقت یابی، این معیار به یک RV شرطی تبدیل می شود که بر  $y^i, z^i, \psi^i, \text{FAP} \sim \text{DP}$  بستگی دارد، که ما آن را به صورت زیر می نویсим:

$$P_D = \frac{\sum_{i=1}^N \mathbb{1}(\hat{y}_i = 1 \text{ and } Y_i = 1)}{\frac{1}{N} \sum_{i=1}^N \mathbb{1}(Y_i = 1)}.$$

در اين عبارت، برچسب پيش بینی شده به صورت حروف کوچک نمایش داده می شود زира مساهده شده است و برچسب صحیح به صورت حروف بزرگ نمایش داده می شود زира یک RV غیرمساهده شده است.

ما  $V$  و  $U$  را تحت شرط  $\{\text{FAP} \sim \text{DP}, \psi^i, z^i, y^i\}$  به صورت زیر تعریف می کنیم:

$$\begin{aligned} U &= \frac{1}{N} \sum_{i=1}^N \mathbb{1}(\hat{y}_i = 1 \text{ and } Y_i = 1) \\ &= \frac{1}{N} \sum_{i:\hat{y}_i=1} \mathbb{1}(Y_i = 1), \end{aligned} \quad (8)$$

$$\begin{aligned} V &= \frac{1}{N} \sum_{i=1}^N \mathbb{1}(\hat{y}_i = 0 \text{ and } Y_i = 1) \\ &= \frac{1}{N} \sum_{i:\hat{y}_i=0} \mathbb{1}(Y_i = 1). \end{aligned} \quad (9)$$

نمایانگر  $P_D$  دقیقاً برابر است. نرمال تقسیم کننده

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \mathbb{1}(Y_i = 1) &= \frac{1}{N} \sum_{i:\hat{y}_i=1} \mathbb{1}(Y_i = 1) + \frac{1}{N} \sum_{i:\hat{y}_i=0} \mathbb{1}(Y_i = 1) \\ &= U + V. \end{aligned}$$

در نتیجه،

$$P_D = \frac{U}{U + V}.$$

با دستکاری های متسابه (مطابق صنیمه A)، می توانیم هر یک از RV های دیگر جغرافیایی را به صورت RV های مشترک U و V بنویسیم. جدول 8 روابط را خلاصه می کند.

### 2.2.2 استقلال و نزدیکی به نرمال

ما بلافاصله نتیجه می گیریم که U و V مستقل هستند زیرا آن ها مجموعه های جداگانه از  $\hat{y}$  را در بر می گیرند. اکنون توزیع آن ها را در نظر می گیریم.

در (8) و (9)، هر  $i$  یک RV برنولی با احتمال موفقیت برابر با  $p_{\text{D}} = \pi(\hat{y}_i | \mathbf{z}_i, \psi_i)$  است. با استفاده از قانون بیس، ما این احتمال را از احتمال پس مورد آزمون کننده می گیریم:

$$\begin{aligned} p(y_i | \hat{y}_i, \mathbf{z}_i; \psi_i, \tilde{p}_{\text{D}}, \tilde{p}_{\text{FA}}) &= \frac{\pi(y_i) p(\hat{y}_i, \mathbf{z}_i | y_i; \psi_i, \tilde{p}_{\text{D}}, \tilde{p}_{\text{FA}})}{\sum_{y'_i \in \mathcal{Y}} \pi(y'_i) p(\hat{y}_i, \mathbf{z}_i | y'_i; \psi_i, \tilde{p}_{\text{D}}, \tilde{p}_{\text{FA}})} \\ &= \frac{\pi(y_i) p(\hat{y}_i | y_i; \tilde{p}_{\text{D}}, \tilde{p}_{\text{FA}}) p(\mathbf{z}_i | y_i; \psi_i)}{\sum_{y'_i \in \mathcal{Y}} \pi(y'_i) p(\hat{y}_i | y'_i; \tilde{p}_{\text{D}}, \tilde{p}_{\text{FA}}) p(\mathbf{z}_i | y'_i; \psi_i)}, \end{aligned} \quad (10)$$

که مقدار مناسب برای  $\tilde{p}_{\text{D}}$  از (3) تا (6) تعیین شود.

با  $p(B)$  توزیع in (8)

$$p_i = p_{Y_i | \hat{Y}_i, \mathbf{Z}_i}(1 | 1, \mathbf{z}_i; \psi_i, \tilde{p}_{\text{D}}, \tilde{p}_{\text{FA}}), \quad i = 1, \dots, N. \quad (11)$$

Let  $N(\mu, \sigma^2)$  denote a normal RV with mean  $\mu$  and variance  $\sigma^2$ . By the central limit theorem (CLT),  $U$  is approximately distributed  $N(\mu_U, \sigma^2_U | \hat{y}, z; \psi, \tilde{p}_D, \tilde{p}_{FA})$  with

$$\begin{aligned}\mu_U(\hat{y}, z; \psi, \tilde{p}_D, \tilde{p}_{FA}) &= \frac{1}{N} \sum_{i:\hat{y}_i=1} p_i \\ &= \frac{1}{N} \sum_{i:\hat{y}_i=1} p_{Y_i|\hat{Y}_i, Z_i}(1|1, z_i; \psi_i, \tilde{p}_D, \tilde{p}_{FA}),\end{aligned}\quad (12)$$

$$\begin{aligned}\sigma_U^2(\hat{y}, z; \psi, \tilde{p}_D, \tilde{p}_{FA}) &= \frac{1}{N^2} \sum_{i:\hat{y}_i=1} p_i(1 - p_i) \\ &= \frac{1}{N^2} \sum_{i:\hat{y}_i=1} p_{Y_i|\hat{Y}_i, Z_i}(1|1, z_i; \psi_i, \tilde{p}_D, \tilde{p}_{FA}) \\ &\quad \cdot (1 - p_{Y_i|\hat{Y}_i, Z_i}(1|1, z_i; \psi_i, \tilde{p}_D, \tilde{p}_{FA})),\end{aligned}\quad (13)$$

جایی که  $\hat{y}$  وابستگی را به  $y$  نشان داده ایم.  
همچنین، هر  $i$  در  $(9)$  توزیع  $B$  دارد

$$q_i = p_{Y_i|\hat{Y}_i, Z_i}(1|0, z_i; \psi_i, \tilde{p}_D, \tilde{p}_{FA}), \quad i = 1, \dots, N, \quad (14)$$

بنابراین  $U$  است تقریباً توزیع شده  $\mathcal{N}(\mu_U, \sigma_U^2)$

$$\begin{aligned}\mu_V(\hat{y}, z; \psi, \tilde{p}_D, \tilde{p}_{FA}) &= \frac{1}{N} \sum_{i:\hat{y}_i=0} q_i \\ &= \frac{1}{N} \sum_{i:\hat{y}_i=0} p_{Y_i|\hat{Y}_i, Z_i}(1|0, z_i; \psi_i, \tilde{p}_D, \tilde{p}_{FA}),\end{aligned}\quad (15)$$

$$\begin{aligned}\sigma_V^2(\hat{y}, z; \psi, \tilde{p}_D, \tilde{p}_{FA}) &= \frac{1}{N^2} \sum_{i:\hat{y}_i=0} q_i(1 - q_i) \\ &= \frac{1}{N^2} \sum_{i:\hat{y}_i=0} p_{Y_i|\hat{Y}_i, Z_i}(1|0, z_i; \psi_i, \tilde{p}_D, \tilde{p}_{FA}) \\ &\quad \cdot (1 - p_{Y_i|\hat{Y}_i, Z_i}(1|0, z_i; \psi_i, \tilde{p}_D, \tilde{p}_{FA})).\end{aligned}\quad (16)$$

این عبارات به راحتی نمونه های برچسب‌گذاری شده درست را پذیرا می‌شوند. برای نمونه  $i$  از  $U$ ، اگر  $\hat{y}_i = 1$  باشد یا بلووان آن را دقیقاً از  $z$  بازیابی کرد، پس  $p_i = 1$  اگر  $y_i = 1$  و  $0$  اگر  $y_i = 0$ . نمونه های برچسب‌گذاری شده به جمع های  $U$  و  $V$  یک عدد ثابت اضافه می‌کند و به جمع های  $U$  و  $V$  صفر اضافه می‌کند.

### 2.3 احتمالات بعدی مقیاس های RV

ما اکنون می‌توانیم احتمالات بعدی مقیاس های RV را به دست آوریم.

#### 2.3.1 مقیاس سکولار RVs

ما بلافاصله می‌فهمیم  $ACC$  تقریباً نرمال است با میانگین  $\mu$  و واریانس  $N^{\frac{2}{3}} + U^{\frac{2}{3}}$ . به طور مشابه  $PREC$  تقریباً نرمال است با میانگین  $\mu$  و واریانس  $N^{\frac{2}{3}} / U^{\frac{2}{3}}$ .

از بخش "فرم متغیر تصادفی" جدول 8، مشاهده می کنیم که هر RV متريک مقاييس باقيمانده نس بتي از RV های مشترک تقربياً معمولی است. عبارات شكل بسته برای اين پسيين ها در دسترس نیست، اما کار مارساليا (1965, 2006) توضيح می دهد که چگونه می توان پسيين ها را محاسبه کرد. صنيمه B اين روش را بررسی می کند و جدول 16 در پيوست، پaramترهای لازم برای هر يك از اين معيارها را پوس س می دهد.

علاوه بر اين، Marsaglia (2006, t 4) تقريب های بسته اى را برای ميانگين و واريانس نسبت RV های مشترک معمولی ارائه می کند. اگر شرط تقريب لحظه اى ساده (MAC) برآورده شود، تقريب ها معتبر هستند.<sup>7</sup>

In summary, for any scalar metric  $RV_{Qt}$   $\{tACt, t PRECt, t PtDt, t PtFAt, Ft\beta t\}$ , we can computethe momentst  $\mu tUt, t \sigma t^2, \mu t(tqt|t^tyt, t zt; t \psi t, t tptDt)$  لازم برای به دست آوردن تقريب های  $t \mu tUt, t \sigma t^2tVt$  fromt  $\{t^tyt, t zt, t \psi t, t tptDt, t tptFAt\}$  and استفاده کنيد.  $var(tQt|t^tyt, t zt; t FAt, t zt; t FAt) = E[tQt|t^tyt, t zt; t \psi t, t tptDt, t tptFAt]$

### های متريک مشترک 2.3.2

برای تجزيه و تحليل ROC، مشاهده می کنیم که  $PtDt$  و  $PtFAt$  مستقل و تقرير  $RV_{st}$   $Ut$  and  $Vt$  تابع دو  $PtDt$  اعمال می کنند. ما تبدیل استاندارد را برای دو  $RV$  به دست آوردن قسمت خلفي مشترک ( $tPtDt, t PtFAt$ ) برای به دست آوردن قسمت خلفي مشترک ( $t6-3, 991$ )

$$p(p_D, p_{FA} | \hat{y}, \underline{z}; \underline{\psi}, \tilde{p}_D, \tilde{p}_{FA}) \approx \left| \frac{(\hat{N}_1/N - p_{FA})(\hat{N}_1/N - p_D)}{(p_D - p_{FA})^3} \right| \frac{1}{2\pi\sigma_U\sigma_V} \cdot \exp \left[ -\frac{1}{2} \left( \frac{\left( \frac{\hat{N}_1/N - p_{FA}}{p_D - p_{FA}} p_D - \mu_U \right)^2}{\sigma_U^2} + \frac{\left( \frac{\hat{N}_1/N - p_{FA}}{p_D - p_{FA}} (1 - p_D) - \mu_V \right)^2}{\sigma_V^2} \right) \right], \\ 0 \leq p_D \leq 1, 0 \leq p_{FA} \leq 1, p_D \neq p_{FA}. \quad (17)$$

که در آن  $Ut$  و  $Vt$  از (11)(16) محاسبه می شود.

واضح است که قسمت خلفي غير گاوسي است. اين عبارت به خوبی تعریف شده است، به جز وق تى اگرچه ما موفق به یافتن يك عبارت بسته برای اين حد را مشکل ساز نمی دانيم. ابتدا، بخش 2.3 نشده ايم)، فقدان اين حد را استفاده از  $tPtDt, t PtFAt$  به دست آوردن قسمت خلفي ( $t6-3$ ) توصیح می دهد که چگونه می توانيم از نمونه برداری برای تقریب قسمت خلفي (دوام) استفاده از ما استفاده کنیم (17). دوم، يك طبقه بندی کننده باينزی تنها زمانی مفید است که دور از خ ط شناس عمل کند،  $sonone$  احتمالاً فقط به طبقه بندی کننده هایي با چگالی احتمال ناچير نزدیک به خ ط شناس علاقه مند است. در نهايیت، در بخش 2.4.1، ما يك الگوريتم تخمين ارائه می کنیم که فقط از پسيين های حاشیه ای  $PtDt$  و  $PtFAt$  استفاده می کند و نیازی به (17) ندارد.

<sup>7</sup>. به طور کل، توزيع نسبت دو متغیر تصادفی نرمال مشترک نمی تواند به خوبی با توزيع نرمال تقریب شود. اما Marsaglia (2006) و همچنین بخش پيوست نيز يك آزمون برای زمانی که تقریب نرمال معقول است ارائه می دهد.

برای تحلیل PREC و REC نیز تابعی از  $U$  و  $V$  هستند. ما تبدیل توابع  $RV$  را برای به دست آوردن پستربور جوینت  $REC \cdot PREC$  (اعمال من کنیم:

$$p(prec, rec | \hat{\mathbf{y}}, \underline{z}; \underline{\psi}, \tilde{p}_D, \tilde{p}_{FA}) \approx \left( \frac{\hat{N}_1}{N} \right)^2 \frac{prec}{rec^2} \frac{1}{2\pi\sigma_U\sigma_V} \\ \cdot \exp \left[ -\frac{1}{2} \left( \frac{\left( \frac{\hat{N}_1}{N} prec - \mu_U \right)^2}{\sigma_U^2} + \frac{\left( \frac{\hat{N}_1}{N} prec \left( \frac{1-rec}{rec} \right) - \mu_V \right)^2}{\sigma_V^2} \right) \right], \\ 0 \leq rec \leq 1, 0 \leq prec \leq 1 \quad (18)$$

با لحاطات  $U$  و  $V$  محاسبه شده از  $(11)-(16)$ . این توزیع نیز غیر-کارائی gaussian است. خط شناس در فضای  $P$ - $R$   $\pi(1)$  است (بیانید Rehmsmeier, 2015) و همچنین مسئلکی ایجاد نمی کند. قانون  $L'H^{\infty}$ opit  $FAP = 0$ , اعمال شود، اما تواند برای نشان دادن این که محدوده  $rec$   $prec$   $(p_0 - DP, \psi; z, y) \cap rec$   $prec$   $= \emptyset$ , بنا برای این عبارت در کل مرربع یک جا به خوبی تعریف شده است.

### 2.3.3 پستیوری از طریق نمونه گیری

ما همچنین می‌توانیم از نمونه‌گیری برای تقریب احتمالات استفاده کنیم. ما  $M$  نمونه‌های طول  $N$  از  $R$  با برچسب صحیح  $Y$  را تولید می‌کنیم. برای هر نمونه  $y = y_1, y_2, \dots, y_m$  با استفاده از  $F_{\text{AP}}(y)$  (انتخاب می‌شود،  $y_i$ )، ما مقیاس مورد نظر تجربی (مقیاً س یا مشترک) را محاسبه می‌کنیم که  $M$  نمونه از  $R$  مقیاس را تولید می‌کند. سپس احتمال تقریبی  $m$  را تواند به عنوان یک نمودار یک بعدی یا دو بعدی محاسبه شود.

اعتبار احتمالات وابسته به این است که چقدر خوب می توانیم لحطاب  $T$  را تخمین زد و  $V$ ، که به نوبه خود به پارامترهای  $OP$  ( $\sim D^p$ )  $FAP$  بستگی دارند. ما دو الگوریتم تکراری برای یافتن برآور MMSE پارامترهای  $OP$  ارائه می دهیم، بنابراین این روش را به عنوان آزمایش MMSE معرفی می کنیم. الگوریتم ها به دسته ی برآورد کنندگان بیایس تجربی تعلق دارند. برآورد بیایس استاندار دیبا مشخص کردن پیش بینی برای متغیرهای مراحم ناساخته، آن ها را به صورت مرگشایی (marginalizing) در نظر می گیرد تا برآورد پسین از متغیر مورد برآورد (estimand) به دست آمده در مقابل، برآورد بیایس تجربی پیش بینی را نمی پذیرد.<sup>8</sup> به جای آن، داده های موجود را برای برآورد م تغیرهای مراحم مشاوره می کند (Casella, 1992). اغلب از برآورد MAP یا ML استفاده می کند. این الگوریتم بین برآورد متغیرهای مراحم و متغیر مورد برآورد جابجا می شود. این فرآیند تکراری به طور پیوس ته هر برآورد را بهبود می بخشد و از نظر روحیه مشابه الگوریتم EM و Dempster and Mekaran (1977) است. در الگوریتم های پیشنهادی، پارامترهای  $OP$  متغیرهای مراحم هستند، متغیر مورد برآورد از احتمال تشخیص و احتمال خطا در اعلام خطر (probability-of-false alarm RVs) تشکیل شده است، و برآور MMSE استفاده می شود. بخش 2.4.4 رابطه ی این الگوریتم ها با الگوریتم EM را بررسی می کند.

انگیزہ 2.4.1

فرص کنید در تکرار ح، پارامترهای قبلی OP را داشته باشیم  
 با همراهی  $\{\pi, \psi, z, y\}$ . نتایج قبلی ما را قادر می سازد تا لحطات  $U$  و  $V$  را محاسبه کنیم.

به عنوان نتیجه، گاهی اوقات می‌گویند روش‌های بیايس تجربی "کاملاً بیايس" نیستند.

. )FAP<sub>D</sub>P( شرطی کردن بر روی  $\hat{y}, \psi, z, y$  و سپس به منظور دریافت احتمال تقریبی  $\hat{P}_{\hat{y}, \psi, z, y}^{(j-1)}$  ما می‌توانیم پارامترهای OP را بهبود بخشیم با به روزرسانی  $\hat{\pi}_{\hat{y}}^{(j-1)}$  با ارزیابی بهینه )FAP<sub>D</sub>P( با توجه به  $\hat{P}_{\hat{y}, \psi, z, y}^{(j-1)}$  . ما می‌توانیم این فرآیند را تکرار کنیم تا پارامترهای OP به هم نزدیک شوند.

ما معیار MMSE را از نظریه برآورد بیایی پذیرفته ایم. بیایید  $\hat{Y} = \hat{Z}$ ,  $\hat{Y} = \hat{Y}_{FA}$  را یک برآورد از  $Z$  (با توجه به  $\hat{Y}_{FA}$ ) در نظر بگیریم. MSE چنین برآورده است و برآورده کننده MMSE تعریف شده به صورت  $E[\hat{Y}_{FA}^2 - Z^2]$  است و برآورده کننده  $\hat{Y}_{FA}$   $E[\hat{Y}_{FA}^2 - (\hat{Y}_{FA} - Z)^2]$  است. نتیجه استاندارد این است که  $\hat{Y}_{FA} = \arg \min_{\hat{Y}} E[\hat{Y}^2 - (\hat{Y} - Z)^2]$  باشد.

$$h^{MMSE}(\hat{Y}, \underline{Z}) = E_{p(p_D, p_{FA} | \hat{y}, z; \underline{\psi}, \tilde{p}_D^{(j-1)}, \tilde{p}_{FA}^{(j-1)})} [P_D, P_{FA} | \hat{Y}, \underline{Z}; \underline{\psi}, \tilde{p}_D^{(j-1)}, \tilde{p}_{FA}^{(j-1)}],$$

جایی که توزیع مرتبط به عنوان زیرنویس نشان داده می‌شود ( مقایسه کنید: Thomas Cover (1991) و 2.2.2). بنابراین، با توجه به  $y = \hat{Y}$  و  $z = \underline{Z}$ ، ما تنظیم می‌کنیم

$$\begin{aligned} \tilde{p}_D^{(j)} &= E[P_D | \hat{y}, z; \underline{\psi}, \tilde{p}_D^{(j-1)}, \tilde{p}_{FA}^{(j-1)}], \\ \tilde{p}_{FA}^{(j)} &= E[P_{FA} | \hat{y}, z; \underline{\psi}, \tilde{p}_D^{(j-1)}, \tilde{p}_{FA}^{(j-1)}]. \end{aligned}$$

هر معیار میانگین شرطی شامل یک پسوند مرزی است. به عنوان مثال،

$$E[P_D | \hat{y}, z; \underline{\psi}, \tilde{p}_D^{(j-1)}, \tilde{p}_{FA}^{(j-1)}] = \int_0^1 p_D p(p_D | \hat{y}, z; \underline{\psi}, \tilde{p}_D^{(j-1)}, \tilde{p}_{FA}^{(j-1)}) dp_D. \quad (19)$$

بنابراین، همانطور که قبلاً ذکر شد، فقدان محدودیت در خط شناس در (17) مانع از محاسبه معیار میانگین شرطی FAP<sub>D</sub>P( نمی‌شود.

من توانیم معیار میانگین شرطی D<sub>FAP</sub> یا  $\hat{P}_{\hat{y}, \psi, z, y}^{(j-1)}$  به سه روش محاسبه کنیم. اولاً، اگر شرط MAC برآورده شود، من توانیم از فرمول بسته ای در (2006) Marsaglia یا بخش B استفاده کنیم. دوماً، من توانیم یک عددی یک بعدی به عمل آوریم. سومین روش، استفاده از نمونه گیری است که در بخش 3.3.2 تو صیح داده شده است.

#### 2.4.2 انتخاب پارامترهای نقطه عملی اولیه

ما از دیدگاه بیایی استفاده می‌کنیم ( $\hat{P}_{\hat{y}, \psi, z, y}^{(0)}$ ) این بخش در مورد پارامترهای نقطه عملی اولیه بحث می‌کند ( $\hat{P}_{\hat{y}, \psi, z, y}^{(0)}$ ) و یک پیش‌فرص غیر اطلاعاتی - به عنوان مثال، توزیع یکنوا با RVهای نقطه عملی اولیه مقداردهی‌ها توسط  $\hat{P}_{\hat{y}, \psi, z, y}^{(0)}$  ( ۷۰٪ بر روی مربع یک).

بانگین شرطی، که دقیق‌تر  $\hat{P}_{\hat{y}, \psi, z, y}^{(0)}$  است، داده می‌شوند.

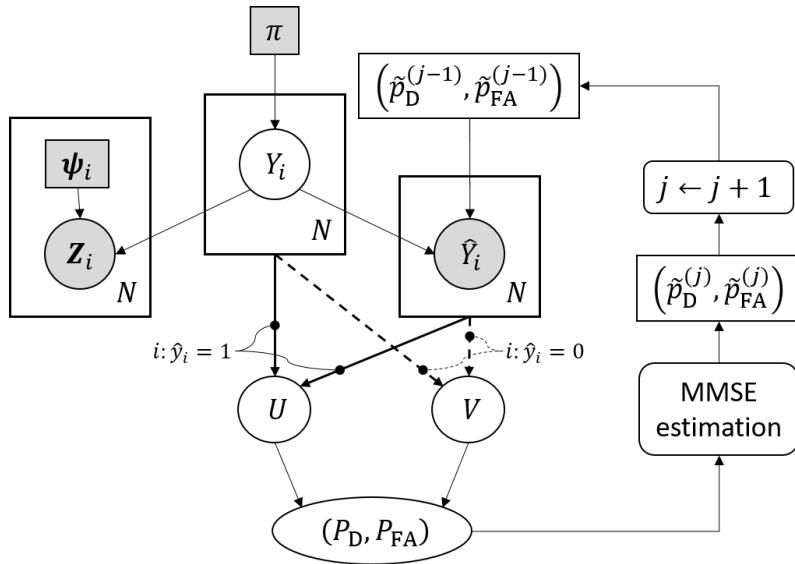
در نتیجه،  $\hat{P}_{\hat{y}, \psi, z, y}^{(0)} = \hat{P}_{\hat{y}, \psi, z, y}^{(1)}$  برابر با  $\hat{P}_{\hat{y}, \psi, z, y}^{(2)}$  برابر با  $\hat{P}_{\hat{y}, \psi, z, y}^{(3)}$  برابر با  $\hat{P}_{\hat{y}, \psi, z, y}^{(4)}$  برابر است، بنابراین پس بینی کلاس آزمایش (10) به

$$p(y_i | \hat{y}_i, z_i; \underline{\psi}, \tilde{p}_D = 1/2, \tilde{p}_{FA} = 1/2) = \frac{\pi(y_i) p(z_i | y_i; \underline{\psi})}{\sum_{y'_i \in \mathcal{Y}} \pi(y'_i) p(z_i | y'_i; \underline{\psi})}. \quad (20)$$

طرف راست با پس بینی کلاس عادی همخوانی دارد، که شرطی بر  $\hat{y}, \psi, z, y$  و  $\hat{P}_{\hat{y}, \psi, z, y}^{(0)}$  قرار نمی‌دهد:

$$p(y_i | z_i; \underline{\psi}) = \frac{\pi(y_i) p(z_i | y_i; \underline{\psi})}{\sum_{y'_i \in \mathcal{Y}} \pi(y'_i) p(z_i | y'_i; \underline{\psi})}. \quad (21)$$

( ۷۰٪ دارای همان تأثیر است که اگر ما برچسب پیش‌بینی شده را ( بنابراین، انتخاب اصلًا شامل نشده بودیم.



شکل 3: مدل گرافیکی تخمین‌گیری مکرر برای آزمایش. RV های مشترک  $U$  و  $V$  بر اساس بخش‌های جدا گانه  $Y_i$  و  $\hat{Y}_i$  بستگی دارند.

#### 2.4.3 الگوریتم‌های تخمین

رابطه مرتبط بین پارامترهای OP پیاپی منجر به دو تخمین دهنده تجربی بیاییس برای MMSE آزمایش می‌شود، به ترتیب الگوریتم‌های 1 و 2. شکل 3 مدل گرافیکی و روش را نشان می‌دهد. در هر تکرار، هر اگوریتم پارامترهای OP خود را با محاسبه تخمین  $P_D$  و  $P_{FA}$  (MMSE) داده شده از  $z$ ،  $y$ ،  $\hat{y}$  و پارامترهای OP قبلی به روزرسانی می‌کند. در تکرار بعدی، تخمین MMSE پارامترهای OP جدید را فراهم می‌کند.

الگوریتم 1 از این واقعیت بهره می‌برد که  $P_D$  و  $P_{FA}$  هر دو نسبت‌های RV های تقریباً نرمال مشترک هستند، بنابراین میانگین‌های شرطی آن‌ها می‌توانند بدون نمونه گیری محاسبه شوند.

با استفاده از  $tpt(tjt1)FAt$  computet ptit andt qtit fornit،  $tpt(tjt1)Dt$  و  $tpt(tjt1)Nt$  را محاسبه می‌کند. برای بدست  $t$ ، تکرار  $tjt1$  را آغاز می‌کند. ...  $t$  و سپس لحظه‌های  $U$  andt  $V$  را محاسبه می‌کند.

بت‌آوردن  $t$  OPnparameterst ( $tpt(tjt)Dt$ ،  $t$   $tpt(tjt)FAt$ ) بهبود یافته، از تقریب فرم بسته برای میانگین شرطی  $ofnPtDt$  یا  $PtFAt$  استفاده می‌کند اگر MAC برقار باشد، و در غیر این صورت به یکیارچکی عددی بازمی‌گردد. برای جلوگیری از انحطاط عددی، پارامترهای OP جدید به  $[ta,t 1 t at]$  بریده می‌شوند. این رویه تا زمانی تکرار می‌شود که حداقل اختلاف مطلق بین تخمین‌های متوالی کمتر از مقدار  $t$  ر تحمل شود یا به حداقل تعداد تکرار  $jtmaxt$  برسد. سپس پارامترهای نهایی  $(OP(tptDt,t tptFAt)=t)$

را بر می‌گرداند.

الگوریتم 2 از تحقق نمونه‌های  $Ytto$  با برچسب صحیح استفاده می‌کند. در تکرار  $t$ ،  $Mt$ ،  $tpt(tjt)FAt$  را بررسی می‌کند.  $t$ ،  $t$   $tpt(tjt)Dt$ ،  $t$   $tpt(tjt)FAt$ ،  $t$   $tpt(tjt)Nt$  را محاسبه می‌کند. میانگین شرطی  $tPtDt,t PtFAt(t)$  با میانگین گیری نقاط عملیاتی  $Mt$  بدست می‌آید که پارامتر OP جدید  $(tpt(tjt)Dt,t tpt(tjt)FAt)$  را به دست می‌دهد.

الگوریتم ۱ آزمایش MMSE با تخمین تجربی بیز (tptFAt, tptDt,t) از طریق نسبت RV های معمولی م شترک.

:EmpiricalBayesViaRatiost(^tyt,t zt,t ψt,t πt)n2 ﻊ t:1

مقداردهی اولیه  $tpt(0)Dtt$  در حالی که  $j$  کمتر از  $j_{max}$  است و  $MAD$  حد مجاز را برآورده می‌کند

$$j \leftarrow j + 1$$

برای  $i = 1$  تا  $N$  انجام دهید

معادلات (10)، (11)، (12) محسوبه

معادله (13) برای محاسبه استفاده از  $\hat{p}_{t,y,z}$

معادله (14) برای محاسبه استفاده از  $\hat{p}_{t,y,z}$

معادله (15) برای محاسبه استفاده از  $\hat{p}_{t,y,z}$

معادله (16) برای محاسبه استفاده از  $\hat{p}_{t,y,z}$

متغیر ift mean approximation fort PDT format  $\mu tUt$ ,  $\sigma t2tU$ ,  $\mu tVt$ ,  $\sigma t2tV$  ب.

App. Marsaglia (2006), tenn.t

10:	$tpt(tjt)Dtt E \left[ \frac{PtD^{t^*tyt,tzt;t\psi,t} tpt(tjt1)D}{- \quad , t tpt(tjt1)FA} \right]$	از تقریب میانگین
11:	دیگر	
12:	$pn \left( \frac{ptD^{t^*tyt,tzt;t\psi,t} tpt(tjt1)D}{- \quad , t tpt(tjt1)FA} \right)$ fromt $\mu tUt$ , $\sigma t2tUt$ , $\mu tVt$ , $\sigma t2tV$	. t برنامه. Marsaglia (1965, 2006)
13:	$tpt(tjt)Dtt E \left[ \frac{PtD^{t^*tyt,tzt;t\psi,t} tpt(tjt1)D}{- \quad , t tpt(tjt1)FA} \right]$	توسط ادعام عددی 1 بعدی
14:	میانگین تقریبی ift fort PtFAt fromt $\mu tUt$ , $\sigma t2tUt$ , $\mu tVt$ , $\sigma t2tVt$ معتبر است پس	
15:	$tpt(tjt)FAt t E \left[ \frac{PtFA^{t^*tyt,tzt;t\psi,t} tpt(tjt1)D}{- \quad , t tpt(tjt1)FA} \right]$	از تقریب میانگین
16:	دیگر	
17:	$pn \left( \frac{ptFA^{t^*tyt,tzt;t\psi,t} tpt(tjt1)D}{- \quad , t tpt(tjt1)FA} \right)$ fromt $\mu tUt$ , $\sigma t2tUt$ , $\mu tVt$ , $\sigma t2tV$	
18:	$tpt(tjt)FAt t E \left[ \frac{PtFA^{t^*tyt,tzt;t\psi,t} tpt(tjt1)D}{- \quad , t tpt(tjt1)FA} \right]$	توسط ادعام عددی 1 بعدی
19:	$]ta,t 1 t at[ 4 clip tpt(tjt)Dtand tpt(tjt)FAt$	
20:	MADt t max $\left( tpt(tjt)Dtt tpt(tjt1)Dtt, t tpt(tjt)FAt t tpt(tjt1)FAt \right)$	پارامترهای نهایی OP. t.

هر دو الگوریتم از  $t_{at} = 10tt3t$ ,  $t_{jtmaxt} = 30$ ,  $t_{oltt} = 10tt3t$  و الگوریتم 2 از  $Mt = 5000$  استفاده کردند. آنها همچنین اجازه می‌دهند تا برعی از نمونه‌هایی که به درستی برچسب گذاری شده‌اند، Ift  $ytit$  شناخته شده است یا دقیقاً از قابل پارسیابی است. سپس  $FAt$  تابعی از  $t_{ptit}$ ,  $t_{qtit}$ ,  $t_{ort}$ ,  $t_{ptYt}$  |  $t_{tytit}$ ,  $t_{tztit}$ ,  $t_{tpf(tjt1)}$ ,  $t_{tpf(tjt1)Df}$ ,  $t_{tpf(tjt1)Rf}$  اضافه مناسب را به لحظه‌های  $Ut$  و  $Vt$  اضافه می‌کند و الگوریتم 2 همیشه تحقق مناسب  $RV$  است ( $= 1$ ). الگوریتم 1 تابعی از  $t_{tytit}$  است که بازگشتی است.

#### 2.4.4 ارتباط يا الگوريتم EM و نکاتي در مورد همگرایی

مانند بسیاری از روش‌های تجربی بیز، الگوریتم‌های تکراری ما شبیه به الگوریتم EM هستند، اما از این نظر تفاوت دارند که متغیرهای پنهان یا پنهان - یعنی برچسب‌های صحیح - به جای مقادیر غیرتصادفی  $PtFA$ .  $PtDt$  و  $PtFA$  روابط بالا به ما اجازه می‌دهد که میانگین شرطی تقریبی اغلب  $RVsnin$  را مستقیماً بدست آوریم، بنابراین «گام E» ما از یک تابع کمکی مانند الگوری EM استفاده نمی‌کند. همچنین، سایر متغیرهای پنهان - یعنی پارامترهای  $(tptDt, t, tptFAt)$  -  $OPn$  با استفاده از MMSE به جای تخمین ML تعیین می‌شوند، بنابراین "M-nstep" ما به جای حداکثر کردن یک تابع کمکی، حداقل کردن MSE را انجام می‌دهد.

الگوریتم  $t$  آزمایش MMSE با تخمین تجربی بیز ( $tptDt, t$   $tptFAt$ ) از طریق نمونه گیری.

مقداردهی اولیه  $tpt(0)Dtt$   $0t.t5t,t$   $tpt(0)FAt$   $t$   $0t.t5t,jt$   $t$   $0t,t$  MADt می باشد  
در حالی که زکمتر از  $jmax$  است و MAD حد مجاز را برآورده می کند

- 3:  $j \leftarrow j + 1$
- 4: **mt t 1 :t Mt do**
- 5:     برای  $i = 1$  تا  $N$  انجام دهید
- 6:          $B$  تا  $m$  در  $y$  ترسیم (  $\psi_{p(j-1)FA}(\dots, y, y^*, \dots, y_i, y_{i+1}, \dots, y_{i+Z}, \dots, y_{i+Y}, \dots, y_{i+Z+1})$  )
- 7:          $\{y, y^*\}^m$   $\leftarrow$  معیارهای تجربی برای  $\{y, y^*\}$
- 8:          $(\frac{D}{p}(j,m), \frac{F}{p}A(j,m))$   $\leftarrow$  میانگین  $(\frac{\sum_{t=1}^{An(j)} t M}{tpt(tj, mt)D, tpt(tj, mt)F})$
- 9:          $tpt(tj)Dt, tpt(tj)FA \leftarrow$  میانگین شرطی تجربی  $t.$
- 10:          $|ta, t 1t t at| \leftarrow clip(tpt(tj)Dt and tpt(tj)FA$
- 11:          $MADt t max \leftarrow \max\{tpt(tj)Dt, tpt(tj)FA, tpt(tj)Dtt, tpt(tj)FAt, tpt(tj)FAt\}$
- 12:          $(tptDt, tptFAt)t n \leftarrow \frac{1}{tpt(tj)Dt, tpt(tj)FA}$
- 13:          $)tpt Dt, tpt FAt$  پارامترهای نهایی  $t.$

ما حدس می‌زنیم که، صرف نظر از پارامترهای اولیه OP، الگوریتم‌ها نباید به یهینه جهانی ( $tpttDt,t$ ) بودند. همگرا شوند.

(PtFAt t tptFA The : استدلال بر اساس استدلال .t2tt^tyt,t zt:t ψt,t tptDt,t tptFAt]t)

$tPtDt, t PtFAt$  محدود است. به طور کلی، میانگین شرطی  $(tPtDt, t PtFAt)$  nestimand (برای

هر پارامتر OP یکتا است ( $FAt(t) = tpt(t) - tpt(t-1)$ ) و MSE در هر تکرار کاهش می‌یابد، مگر اینکه در حال حاضر در بهینه‌جوانی است. بنابراین، الگوریتم ها باید بدون توجه به اعلب او  $tpt(t+1) = tpt(t) + tpt(t-1) \cdot EA$

این ویژگی با الگوریتم thenEM در تصادف (قات به بھینه جهانی همگرا شوند)  $tpt(0)Dt, t tpt(0)$  است، که فقط می‌توان گفت که به یک بھینه محلی همگرا می‌شود، و این opti-nnum ممکن است بسته به مقداردهی اولیه به طور قابل توجهی متفاوت باشد (Koller and Friedman, 2009, t 19.2.). اینطور که  $(p^*_D, p^*_FA)$  ما همچنین می‌توانیم الگوریتم را را  $\tilde{z}$ -بنال نقطه ثابت منحصر به فرد در نظر بگیریم

$$E[P D, P FA | \gamma y, z; \Psi, \sim p^* D, \sim p^* FA] = (\sim p^*_D, \sim p^*_FA).$$

یک جزئیات مانع از رأئه یک ادعا قوی تر در مورد نزدیکی من شود.  
هر دو  $FAP_{DP}$  و  $FAP_{RV}$  به عنوان نسبت های RV های نرم افزار مشترک تقریب داده می شوند، که میانگین و واریانس آن ها وجود ندارد (بیانیه Marsaglia و صمیمه B، 1965، 2006). به عنوان یک نتیجه، منحصر به فرد بودن محتاطی MMSE برای  $FAP_{DP}$  (تصمیم نمی شود، بنابراین ممکن است که اثبات دقیق نزدیکی این دو مقدار را نشود).

این موانع ممکن است بیشتر یک نگرانی نظری باشد، و به استثنای چندین موقعیت بیمارگونه، الگوریتم ها ممکن است همیشه در عمل به نتیجه برسند. بخس 5.2.3 شواهد تجربی برخی از نتیجه گیری ها یعنی پی و بیو پست مانند با خود سنت جلسه

## 2.5 ارزیابی یهودی مقیاس‌های RV

الگوریتم های 1 و 2 هر کدام پارامترهای نهایی OP را تولید می کنند ( $\tilde{D_p}$ ,  $\tilde{FAp}$ ). هنوز باقی مانده است که برآورد نهایی مقیاس های RV داده شده ( $\tilde{D_p}$ ,  $\tilde{\pi}$ ,  $\tilde{\psi}$ ,  $\tilde{z}$ ,  $\tilde{y}$ )  $\tilde{FAp}$  را محاسبه کنیم. ما این کار را با محا سبه لحطات RV های مشترک  $U$  (بخنس 2.2) و سپس با استفاده از برآورد های پس زمینه ای مقیا س های RV (بخنس 2.3) انجام می دهیم. با توجه به پس زمینه های تقریبی، برآورد بهینه یک موضوع سا ده است؛ ما آن را به طور خلاصه برای کامل بودن بحث من کنیم.

ما دو تخمین نقطه ای و یک تخمین دامنه ای را در نظر می گیریم. اولین تخمین نقطه ای میانگین شرطی است که در معنای MMSE بھینه است. RV های دقت و دقیق تقریباً نزمال هستند، بنابراین فقط میانگین است. برای REC یا  $P_{\text{D}}$ , یا  $P_{\text{FA}}$ , یا  $\beta$ ,  $F$ , می توانیم از تقریب میانگین استفاده کنیم اگر وجود دارد یا به عددی سازی پا نموده گیری متوجه شویم.

برای تحلیل P-R یا ROC، ما می‌توانیم فقط از میانگین‌های شرطی عناصر فردی استفاده کنیم. نقطه برآورد دوم برآورد MAP است، ارزش احتمالی بیشترین برای معیار RV. این می‌تواند با محاسبه احتمال پس زمینه در یک رزولوشن دقیق و یافتن اوچ بدست آید.

محدوده بهینه تخمین، منطقه‌ی قابل اعتماد  $\%95$  است که منطقه‌ای را مسخن می‌کند که با اختصار  $L/100p$ ، متغیر RV مأخذ در داخل این منطقه قرار دارد. منطقه‌ی قابل اعتماد لزوماً منحصر به فرد نیست، اما یک راه برای دستیابی به منطقه‌ی قابل اعتماد معقول، اعمال جستجوی دودویی برای پیدا کردن حد  $c$  است که در آن، جمع عددی مأخذ در نقاطی که مأخذ از  $c$  بیشتر است، برابر با  $100p/100$  در برخی است.

در نهایت، ما تکرار می کنیم که الگوریتم های 1 و 2 و محاسبات ارزیابی بهینه‌ی بعدی هرگز تلاش نمی کنند تا RVs با برچسب صحیح<sup>۷</sup> را تخمین بزنند. گرفتن تصمیم قاطع در مورد<sup>۸</sup> در هر نقطه من تواند خطاهایی را در فرآیندهای بعدی معرفی کند، همانطور که در بخش ۱.۴ ذکر شده است. رویکرد ما از این کار اجتناب می کند در حالی که از تمام اطلاعات موجود برای تولید تخمین های خود استفاده می کند.<sup>۹</sup>

## 2.6 روش‌های حاگزین؛ ارزنایی

برای مقایسه، چندین روس<sup>۳</sup> جایگزین برای تست را دکر می‌کنیم. ابتدا چهار روس<sup>۳</sup> ناکارآمد را ارائه می‌دهیم و سپس یک روس<sup>۳</sup> کاملاً بیایی را توصیح می‌دهیم که بهینه است اما پارامترهای OP را به گونه‌ای متفاوت از روس<sup>۳</sup> بیاییس تحریب مدبیریت می‌کند.

#### 2.6.1 روش‌های آزمایشی غیر اندیز

اولین روش ناکافی این است که برچسب های صحیح را تخمین بزیریم و سپس تخمین ها را به عنوان اگر آنها درست بودند، برای یهیوDp استفاده کنیم. ما یک روش مکرر در الگوریتم ۳ ارائه می دهیم و تخمین خود از RV برچسب های صحیح  $Z_{iY}$  اتسمیه می کنیم. در حین تکرار، الگوریتم از پارامترهای OP قبلی ( $Dp$ ) استفاده می کند و تخمین های  $Y$  را بر اساس معیار MPE براورد می کند. بیایید  $\hat{y}(Z_{iY})$  را یک برآوردگر  $Y$  بدار  $(\hat{y}(Z_{iY}))_{Dp,t}$  با توجه به  $Z_{iY}$  و پارامترهای  $Dp$  و  $D1$  که به دلیل اینکه  $RV$  ها نیستند، پنهان شده اند. احتمال خطای  $iY$  است  $p_{iY} = \text{Pr}(Z_{iY} \neq \hat{y}(Z_{iY})) = \sum_{j \neq i} \sum_{y_j} p_{ijy_j}$ . این روش به دنبال  $(\hat{y}(Z_{iY}))_{Dp,h}$  است. این روش  $\text{arg min}_{h=1}^N \text{MPE}_h$  است.

راه حل شامل پیدا کردن مطلوب ترین برآوردگر (MPE) (برای هر فرد $\gamma$ ) است و نتیجه استاندارد این است که برآورده MAP احتمال خطا را به حداقل می‌رساند (بیانیه 2) (یا صنیمه D). بنابراین، الگوریتم از:

$$\check{y}_i^{(j)} = \arg \max_{y \in \mathcal{Y}} p(y | \hat{y}_i, \mathbf{z}_i; \boldsymbol{\psi}_i, \tilde{p}_{\text{D}}^{(j-1)}, \tilde{p}_{\text{FA}}^{(j-1)}), \quad i = 1, \dots, N. \quad (22)$$

سیس الگوریتم از این تخمین‌ها برای محاسبه احتمالات تجربی تشخیص و هشدار غلط  $tpt(tjt)Dta$  استفاده می‌کند. این رویکرد مشابه الگوریتم‌های 1 و 2 است به جزء  $tpt(tjt)FAt$

9. اگر کسی بخواهد  $Y_t$  را تخمین بزند، پس از تعیین  $tptDt, t$   $tptFAt$  (می توان این کار را با استفاده از 10) انجام داد. با این حال، متفاوتاً استفاده از این تخمین برای محاسبه معیارهای آزمایش با رویکرد ما مغایرت دارد.

---

### الگوریتم 3 تخمین زیر بهینه ( $tptDt, t tptFAt$ ) با تخمین $Yt$ با برچسب صحیح.

---

1:t function EstimateOPParametersViaEstimatedCorrectLabelst(`tyt,t zt,t  $\psi_t$ ,t  $\pi_t$ )n2:

```

3:   tpt(0)Dtt 0t.t5t,t tpt(0)FAt t 0t.t5t, jt t 0t,t MADt
4:   در حالی که  $j \leq j_{max}$  است و MAD حد مجاز را برآورده من کند
5:    $j \leftarrow j + 1$ 
6:   برای  $i = 1$  تا  $N$  انجام دهد
7:    $y_t | t^* y_{tit}, t z_{tit} \sim \mathcal{N}(y_t | t^* y_{tit}, t z_{tit}; \psi_t, t tptDt = tpt(t_{tit})D, t tptFAt = tpt(t_{tit})FA)$  معادله (22).  $t$ .
8:    $y_t | t^* y_{tit}, t z_{tit} \sim \mathcal{N}(y_t | t^* y_{tit}, t z_{tit}; \psi_t, t tptDt = tpt(t_{tit})D, t tptFAt = tpt(t_{tit})FA)$  محاسبه تجربی
9:    $MADt = \max_{i=1}^{Nt} |y_t - \hat{y}_t|$ 
10:   $(tptDt, t tptFAt) = \frac{\partial \log p(y_t | t^* y_{tit}, t z_{tit}; \psi_t, t tptDt, t tptFAt)}{\partial \psi_t, t tptDt, t tptFAt}$  پارامترهای نهایی  $t$ .
11:  بازگشت( $y_t | t^* y_{tit}, t z_{tit}; \psi_t, t tptDt, t tptFAt$ )

```

---

این روش تصمیم سخت گیری در مورد برچسب های صحیح را در هر آنیترو (iteration) اتخاذ می کند. این رویکرد از مدل تست بهره می برد و معیار مجازات معین شده را به حداقل می رساند، اما بر اساس بخش 1.4، آن ناکارآمد است زیرا زمانی که باید ( $FA, P, DP$ ) را تخمین می زند،  $Y$  را تخمین می زند. ما هیچ ادعایی در مورد ویژگی های نزدیک شدن آن نداریم. ما دوباره از  $tol = 10^{-3}$ ،  $jmax = 30$ ،  $\alpha = 10^{-3}$  استفاده کردیم.

مانند الگوریتم های 1 و 2، الگوریتم 3 تنها پارامترهای نهایی  $OP$  را تخمین می زند ( $\tilde{FAP}, \tilde{DP}, \tilde{P}$ ). پس از این مرحله، این روش (22) را برای دریافت برآورده نهایی MPE یا MAP اعمال می کند.  $Y$  داده شده  $\{\tilde{pFA}, \tilde{P}, \tilde{D}, \tilde{z}, \tilde{\psi}\}$ . برآورده نهایی معیارها با در نظر گرفتن  $\tilde{y}$  به عنوان درست و محاسبه معیارهای تجربی برای  $\{\tilde{y}, \tilde{y}\}$  به دست می آید.

راه دوم ناکارآمد کاملاً نظریه ارزیابی را نادیده می گیرد؛ فقط ساخته های محاسبه شده برای هر برج سب دهنده ی فردی را با هم ترکیب می کند. برای هر  $T \in t$ ، برچسب های  $t$  labeler را به عنوان درست در نظر می گیرد و ساخته تجربی برای این نمونه ها را محاسبه می کند. این کار باعث تولید  $T$  نمونه از یک ساخته می شود. سپس با استفاده از آمار مرکزی مانند میانگین یا مدیان از این  $T$  نمونه ها، برآورده نهایی به دست می آید.

راه های سوم و چهارم به جای مدل آزمایشی (7)، توزیع  $p(tz | ty)$  conditionally noisy-label برای  $t$  استفاده می کنند، بنابراین  $p(tz | ty)$  استفاده می کنند. آنها پهینه نیستند زیرا رابطه بین  $\tilde{Y}$  و  $\tilde{Y}$  را نادیده می گیرند و بنابراین تتوانسته اند برچسب های پیش بینی  $\tilde{y}$  را به طور کامل بهره برداری کنند. سومین روش از برآورده MMSE استفاده می کند اما در (11)-(16) (جایگزین  $\tilde{FAP}, \tilde{DP}, \tilde{P}, \tilde{z}, \tilde{\psi}$ ) را با  $\tilde{pFA}, \tilde{P}, \tilde{D}, \tilde{z}, \tilde{\psi}$  می کند. چهارمین روش برچسب های صحیح را برآورده می کند و همان جایگزینی را در (22) انجام می دهد. (معادله آن این است که  $\tilde{y} = \tilde{y}_{max} - \tilde{y}$  را در الگوریتم های 1، 2 یا 3 تنظیم کنیم، بنابراین  $\tilde{FAP}, \tilde{DP}, \tilde{P}, \tilde{z}, \tilde{\psi}$  به (21) کاهش می یابد.) 2.6.2 برآورده بی درنگ بیایی

یک راه جایگزین دیگر یک روش بی ایستگی کامل است که ( $\tilde{FAP}, \tilde{DP}, \tilde{P}, \tilde{z}, \tilde{\psi}$ ) را به عنوان یک واقعیت غیر مشاهده شده از  $RV$  های پارامتر مراحم ( $\tilde{FAP}, \tilde{DP}, \tilde{P}, \tilde{z}, \tilde{\psi}$ ) با پیش بینی  $p(\tilde{y} | t, z, \psi)$  در نظر می گیرد. توزیع یکنواخت بر وی مربع یکم به عنوان پیش بینی غیر اطلاعاتی عمل می کند.

این روش هر یک از ساخته های  $RV$  را با محور کردن ( $\tilde{FAP}, \tilde{DP}, \tilde{P}, \tilde{z}, \tilde{\psi}$ ) تخمین می زند، بنابراین تخمین تحت شرط یک نمونه خاصی از ( $\tilde{FAP}, \tilde{DP}, \tilde{P}, \tilde{z}, \tilde{\psi}$ ) نیست و نیاز به تکرار ندارد. به عنوان مثال، تخمین MMSE از ساخته  $RV$  دقیق را در نظر نگیرید. برای یک تخمین

تقریباً  $ACC_{\text{MMSE}} = \min_{h \in \mathcal{H}} E[\text{E}[Y(h) | X] - Y]^2$  است، و برآورده MMSE است  $h_{\text{MMSE}} = \arg \min_h E[\text{E}[Y(h) | X] - Y]^2$ . نتیجه استاندارد (بینید 1) (یا پیوست) این است که راه حل میانگین شرطی است، بنابراین برآورد بی طرفانه کامل  $ACC_{\text{MMSE}}$  است.

$$\mathrm{E}_{p(\tilde{p}_D, \tilde{p}_{\text{FA}})}[ACC \mid \hat{y}, \underline{z}; \underline{\psi}] = \int_0^1 \int_0^1 \mathrm{E}[ACC \mid \hat{y}, \underline{z}, \tilde{p}_D, \tilde{p}_{\text{FA}}; \underline{\psi}] p(\tilde{p}_D, \tilde{p}_{\text{FA}}) d\tilde{p}_D d\tilde{p}_{\text{FA}}.$$

این تخمین MSE را تحت فرض اینکه پارامترهای  $OP$ ها هستند ( $FAP \sim DP$ ) به حداقل می‌رساند تا غیر از مقادیر تصادفی.

هر دو روش بر اساس بخش ۱.۴ بهینه هستند، اما در برخورد با پارامترهای OP متفاوت هستند. محاسبه میانگین گیری است، بنابراین روش  $FAP$  به طور کامل بیزیان یک میانگین نمره م و کردن)  $DP$  (نوعی میانگین گیری است. روش بیزی تجربی بیشتر به تشییه نمره دهندر ابتدای این بخش اعتماد دارد؛ امتحانی ک نمره خاص دارد، که با ( $\sim DP$ ) نشان داده می شود، نه یک میانگین نمره. با این حال، نیاز به آزمایش برای مقایسه عملکرد این روش‌ها وجود دارد که در بخش ۵.۲.۱ آورده شده است.

۲.۷ آزمایش MMSE برای طبقه بندی چندگانه

این بخش به بحث در مورد گسترش تست MMSE به طبقه‌بندی چندگانه نیز 2)

مدل ساری  $\pi$  یا تخمین زدن  $\pi$  و دشوار باشد، اما کارهای مرتبط توصیح داده شده در

ما با معرفی  $C \times C$  ماتریس شفافیت شرطی شروع می‌کنیم. ما ادیم برای نشان دادن فرم تجربی آن استفاده می‌کنیم و یک عنصر از آن را با  $K_{\text{emp}}$  نشان می‌دهیم تا ماهیت شرطی آن را برجسته کنیم. سپس  $K_{\text{emp}}$  تعریف می‌شود به صورت زیر:

$$\mathbf{K}_{n|\ell}^{\text{emp}} = \frac{\text{no. of times } \hat{y}_i = n \text{ and } y_i = \ell}{\text{no. of times } y_i = \ell}, \quad \ell, n \in \mathcal{Y}. \quad (23)$$

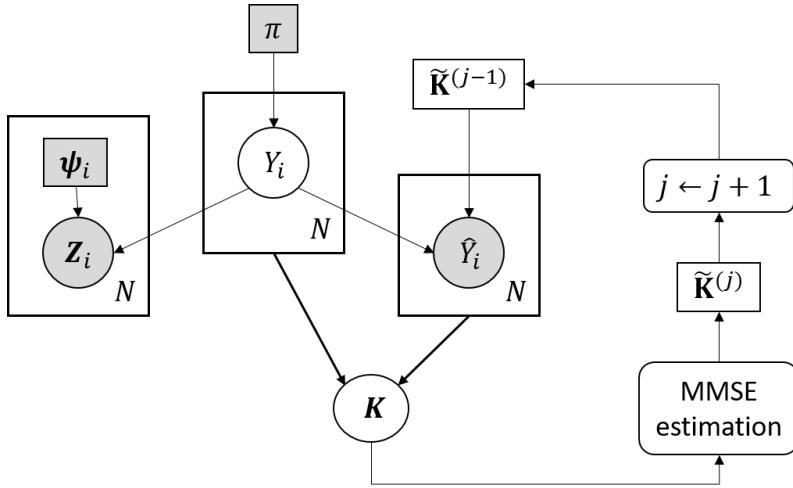
## فرم RV ماتریس $K$ ، با

$$K_{n|\ell} = \frac{\frac{1}{N} \sum_{i=1}^N \mathbb{1}(\hat{y}_i = n \text{ and } Y_i = \ell)}{\frac{1}{N} \sum_{i=1}^N \mathbb{1}(Y_i = \ell)}, \quad \ell, n \in \mathcal{Y}.$$

2.7.1 روش سایس تحریر از طریق نمونه گیری (کلاس بندی چندگانه)

ما همان فرصتیات را مانند بخش 2.1 برای طبقه بندی دودویی در نظر می‌گیریم. ما داده‌ای داریم مجموعه  $\{z_t, y_t, \psi_t, \pi_t\}$ , هر  $i$   $Z$  دارای توزیع شرطی  $(\psi_i | y_i; p(z_i | y_i))$  است و نمونه‌ها مس تقل هستند. مدل گرافیکی در شکل 4 نشان داده شده است. هر برچسب پیش‌بین شده  $y_i$  یک وافع است از یک  $A_{RV}$  است که برای طبقه بندی حندگانه توزیع شرطی،  $(K_i | y_i; p(y_i | A_{RV}))$  دارد، حار، که

ما از شمارش ماتری بر اساس ستون و با شمارش از صفر استفاده می کنیم.  $K_{n+1, n+1}^{\text{emp}}$  خوانده می شود که در نماد ماتری، اسناندارد.



تصویر 4: طبقه بندی چندگانه: مدل نموداری تخمین مکرر برای تست.

$\tilde{\mathbf{K}}$  یک ماتریس از پارامترهای ماتریس مخلوط conditionally است که مشابه پارامترهای OP است و بنابراین

$$p(\hat{y}_i|y_i; \tilde{\mathbf{K}}) = \tilde{\mathbf{K}}_{\hat{y}_i|y_i}. \quad (24)$$

هر ردیف از  $\tilde{\mathbf{K}}$  باید یک  $\mathcal{C}$ -سیمپلکس احتمالی تشکیل دهد: برای هر  $K \in \mathcal{C}$ ،  $\sum_{n=0}^N K_{jn} = 1$ . وقتی  $C = 1$ ، کافی بود تنها یک عنصر در هر ردیف از  $\tilde{\mathbf{K}}$  در نظر گرفت، و ما از  $K = DP_0$  استفاده کردیم؛ این ویژگی به ما امکان داد تا تقریب‌های میانگین یا تعاملات پس تریم‌بندی FAP و DP را در الگوریتم 1 استفاده کنیم. این تکنیک برای 2 قابل اجرا نیست زیرا، برای هر ردیف از  $K$ ، باید احتمال مشترک  $P_{\text{RVs}}$  را تعیین کنیم و باید تعاملات عددی یک بعدی  $\mathcal{C}$ -بعدی را انجام دهیم.

در این حالت، برای مقادیر متوسطی از  $\mathcal{C}$  و یک پیش‌بینی اولیه که به شدت متمایل به یک سمت نیست، من توانم به راحتی روش نمونه‌گیری بیاییس تجربی الگوریتم 2 را به طبقه بندی چندطبقه گسترش داد. کد نمونه در الگوریتم 4 آورده شده است. با استفاده از ماتریس پارامتر  $\tilde{\mathbf{K}}$ ، ما  $M$  واقعی سازی‌های  $\mathbf{Y}$  بر اساس احتمال پس زمینه‌ای آزمون چندکلاس، که به همان روشی به دست می‌آید که (10):

$$\begin{aligned} p(y_i|\hat{y}_i, z_i; \psi_i, \tilde{\mathbf{K}}) &= \frac{\pi(y_i)p(\hat{y}_i, z_i|y_i; \psi_i, \tilde{\mathbf{K}}_{\text{PD}}, \tilde{\mathbf{K}}_{\text{FA}})}{\sum_{y'_i \in \mathcal{Y}} \pi(y'_i)p(\hat{y}_i, z_i|y'_i; \psi_i, \tilde{\mathbf{K}})} \\ &= \frac{\pi(y_i)p(\hat{y}_i|y_i; \tilde{\mathbf{K}})p(z_i|y_i; \psi_i)}{\sum_{y'_i \in \mathcal{Y}} \pi(y'_i)p(\hat{y}_i|y'_i; \tilde{\mathbf{K}})p(z_i|y'_i; \psi_i)} \\ &\stackrel{(a)}{=} \frac{\pi(y_i)\tilde{\mathbf{K}}_{\hat{y}_i|y_i}p(z_i|y_i; \psi_i)}{\sum_{y'_i \in \mathcal{Y}} \pi(y'_i)\tilde{\mathbf{K}}_{\hat{y}_i|y'_i}p(z_i|y'_i; \psi_i)}, \end{aligned} \quad (25)$$

جایی که (a) از (24) است. ما ماتریس شتاب conditional برای هرنمایندگی طول- $N$  محاسبه می‌کنیم و ماتریس‌های  $M$  را میانگین می‌گیریم تا تخمین میانگین  $\hat{\mathbf{K}}$  داده  $\{\mathbf{K}, \psi, z, y\}$  را بدست آوریم. برای الگوریتم 4،  $C = 50$ ،  $M = 2500$ ،  $\max_j j^3 = 10^3$  را تنظیم می‌کنیم.

الگوریتم 4 آزمایش MMSE برای طبقه بندی چندگانه با تخمین تجربی بی ایز (empirical Bayes) از K از طریق نمونه گیری.

---

```

 $) , M\pi, \psi, z, \hat{y} ($  ^ MULTICLASSEMPIRICALBAYESVIASAMPING
2:  $0 \leftarrow \in \mathcal{Y} \times \mathcal{M}, \ell / C = \mathbf{K}_n^{(0)}$  ابتدا  $\mathbf{K}$  تراویح
3: تکرار
4:  $j \leftarrow j + 1$ 
5:  $mt t 1 : t Mt do$ 
6: برای  $i = 1$  تا N انجام دهید
7:  $i \sim p(Y | \hat{y}_i, Z_i; \psi_i, \tilde{K}(j-1))$  نفاش کنید (y را به صورت
8:  $\tilde{\mathbf{K}}^{(j,m)} \leftarrow \text{empirical matrix from } \{\hat{y}_i, \hat{y}^{(m)}\}$  داشته باشیم
9:  $1 = m_M \dots j, m_{\ell} \leftarrow \dots$  میانگین  $(\hat{y}_i, \hat{y}^{(m)})$ 
10: هر ردیف t را طوری تنظیم کنید که هر عنصر در  $[ta, t \dots t]_t$  قرار گیرد و مجموع ردیف ها به یک برسد.
11: تا  $tjt1t1tt1tt1ttt1t|t tKt(tjt)t$  برآورد نهایی  $K$ .
12: طاقت طاقت (اتجاه)
13: بارگشت به O

```

---

این روش اجازه من دهد تا برخی از نمونه ها به درستی برچسب گذاری شده باشند. اگر برچسب صحیح برای  $\hat{y}_i$  ن نمونه برابر با  $t^i$  شناخته شود، سپس  $\hat{y}_i = t^i$  و نمونه برداری همیشه برای این نمونه،  $t^i$  تولید خواهد کرد. پارامتر اولیه ماتریس  $tKt(0)$  نتیجه استفاده از یک ردیف قبلی غیر اطلاعاتی برای هر ردیف از ماتریس است - یعنی توزیع دیریکله از مرتبه  $C_t$  با تمام پارامترهای غلطت برابر با واحد. الگوریتم  $tKt$ ، میانگین شرطی تقریبی را برمی گرداند. عناصر اعلی  $Kt$  محدود می شوند و تخمین MMSE استفاده می شود، بنابراین اطهارات بخش 2.4.4 در مورد همگرای نیز ممکن است قابل اجرا باشد.

## 2.7.2 پشتی RV های متريک (طبقه بندی چند کلاسه)

پس از اينکه برآورد  $K$  در دسترس باشد، می توانیم احتمالات مختلف RVs را به دست آوریم. دقت در روش مسایه برای طبقه بندی دوگانه و چندگانه تعریف می شود. دقت تجربی  $= acc$  (تعداد دفعات که  $\hat{y}_i = y_i$ ) /  $N$  و فرم RV آن  $ACC = N - \sum_{i=1}^N \mathbb{1}(Y_i = \hat{y}_i)$ . در حالت چندگانه، هر RV برآورده با احتمال موقوفیت  $p_{Y_i}(\hat{y}_i | Y_i, Z_i; \psi_i, \tilde{K})$  است.

RV برآورده ها مستقل هستند، بنابراین  $ACC$  با استفاده از CLT تقریباً نرمال است.

علاوه بر این، ماتریس خلط معمول  $C \times C$  فرم تجربی  $C_{temp}$  با  $11C_{temp}$  دارد.

$C_{temp}$  دفعات که  $\hat{y}_i = n$  و  $y_i = n$  (تعداد دفعات که  $\hat{y}_i = n$  و  $y_i = n$ )

$$C_{n,\ell} = \sum_{i=1}^N \mathbb{1}(\hat{y}_i = n \text{ and } Y_i = \ell) \\ = \sum_{i:\hat{y}_i=n} \mathbb{1}(Y_i = \ell).$$

11.  $C_{temp}$  در نماد ماتریس استاندارد مطابقت دارد.

بنابراین، عنصر ماتریس  $\sum C_{tn} \cdot \hat{y}_i$  تقریباً توزیع نرمال دارد با  $\sim K$ .  
 $= \text{var}(C_{tn} \cdot \hat{y}_i) = p_{Yi} \cdot \text{var}(\hat{y}_i)$   
 این نتیجه برای قسمت خلفی هر عنصر جداگانه از  $Ct$  اعمال می شود.  
 آن قسمت خلفی مشترک عناصر متعدد اغلب  $Ct$  را توصیف نمی کند.  
 در نهایت، اشاره می کنیم که می توان روش نسبت نرمال را برای محاسبه پسین تقریبی  $Ktnt | t^t$  اعمال کرد.

۳. آموزش با مسائل مربوط به حقیقت: یادگیری از برجسب‌های پرس و صدا

این بخش به آموزش با مسائل مربوط به حقیقت می پردازد. به یاد بیاورید که ما از `gt(txt;t θt)` برای نشان دادن طبقه بنده کننده یا مدل پیش بینی کننده با پارامتر مدل  $\theta_t$  استفاده می کنیم. با توجه به `xt`, طبقه بنده کننده `anstatistic st = tgt(txt;t θt)` را محاسبه می کند، که شامل محاسبه مدل های شناسی است که  $t$  به هر کلاس تعلق دارد، و سپس یک قانون تصمیم `st tot` را برای انتخاب  $^t\text{yt}$  اعمال می کند. در طبقه بنده باینی،  $t$  به طور معمول یک اسکالار است، و  $^t\text{yt}$  با مقایسه `st` با آستانه `tt` انتخاب می شود. یعنی  $(ts > tt) \cdot yt$ . طبقه بنده `Innmul`

آموزش فرآیند یادگیری  $\theta_t$  از مجموعه آموزش است. در حالت ایده‌آل، مجموعه آموزشی  $\{x_t, t \in T\}$  است، و آموزش به دنبال پارامتر  $\theta_{tt}$  است که در زمانی که در مدل آموزش دیده برای نمونه های دیده نشده اعمال می‌شود، دقیق ترین پیش‌بینی ها را ایجاد می‌کند. عموماً آموزش به عنوان یک مسئله بهینه سازی مطرح می‌شود:

$$\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta}} J_{\text{ideal}}(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{y}) = \arg \min_{\boldsymbol{\theta}} [J_{\text{pri}}(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{y}) + \lambda J_{\text{reg}}(\boldsymbol{\theta})], \quad (26)$$

در جایی که عبارت اولیه  $Jtprit(t\theta t; t xt, t yt)$  هزینه یا جریمه ای را برای تفاوت بین برجسون و  $Jtgt(txtit; t \theta t)$  می‌نماید، عبارت اولیه به صورت  $Jtgt(txtit; t xt) + anyt(gt(txtit; t \theta t), t xt)it t xt; t anyt gt(txtit; t \theta t)$  تحریم می‌گردد. وزن  $\lambda$  سطح منظم شدن را کنترل می‌کند. عبارت اولیه به مدل پیش‌بینی  $Lt2t:t$  کننده آن زمان بستگی دارد، و عبارت منظم سازی به انتخاب منظم سازی، چنین تنظیم سازی  $Jtregt(t\theta t)t tjt | t\theta tjt | t\theta tjt$  دارد.

3.1 مفروضات آموزشی

در صورت وجود مسائل مربوط به صدوق، برچسب های صحیح در دسترس نیستند و مجموعه آموزشی  $t_{\text{txt}}, t_{\text{zt}}, t_{\text{t}}, t_{\text{pt}}, t_{\text{psi}}, t_{\text{xt}}, t_{\text{tyt}}$  تبدیل می شود. با توجه به  $\pi t$  and  $\pi t$ , مدل  $\pi t$   $\text{t}_{\text{pt}}(\text{t}_{\text{zt}} \mid \text{tyt}; t \cdot \text{psi})$  نویزدار برای هر نمونه شناخته شده است. ما فرض معمول نمونه های مستقل را انجام می دهیم. (نمایه  $sot$   $\text{pt}(\text{t}_{\text{zt}} \mid \text{tyt}, t \cdot \text{xt}; t \cdot \text{psi})$ )

$$p(\underline{z}|\mathbf{y}, \underline{x}; \underline{\psi}) = \prod_{i=1}^N p(z_i|y_i; \psi_i). \quad (27)$$

توجه ما را به آموزش‌هایی که من توانند به صورت(26) بیان شوند محدود من کنیم. سپس مسئله بهینه سازی برای آموزش داده شده با  $\{x, z, \psi, \pi\}$  تبدیل من شود

$$\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}; \underline{\boldsymbol{x}}, \underline{\boldsymbol{z}}, \underline{\boldsymbol{\psi}}, \boldsymbol{\pi}) = \arg \min_{\boldsymbol{\theta}} [J_{\text{pri}}(\boldsymbol{\theta}; \underline{\boldsymbol{x}}, \underline{\boldsymbol{z}}, \underline{\boldsymbol{\psi}}, \boldsymbol{\pi}) + \lambda J_{\text{reg}}(\boldsymbol{\theta})], \quad (28)$$

جایی که تنها واژه اصلی تغییر کرده و به  $\theta(\rho, \psi, z, x)$  تبدیل شده است. چون شرط تنظیم شده تغییر تکرده است، ما در ادامه بر روی واژه اصلی تمرکز می‌کنیم.

### 3.2 دیدگاه یکپارچه

ما یک دیدگاه یکپارچه از آموزش با مسائل حقیقت یابی ارائه می‌دهیم که روش‌های کلی برای آموزش مدل‌های پیش‌بینی احتمالی یا غیراحتمالی را توضیح می‌دهد.

#### نظرة یکپارچه به آموزش با مسائل حقیقت‌آمیز

برای مدل‌های احتمالی، اصل بھینه سازی از آموزش ایده‌آل را حفظ کنید و اصطلاح اصل را تغییر دهید تا حساب شود  $\{\pi, \psi, z, x\}$  به جای  $\{\psi, \pi, x, z\}$ .

برای آموزش Machine Learning، که بھینه سازی عملکرد تابع احتمال باشد (الف) به جای استفاده از  $p(t|y \perp tx; \theta)$  با  $p(t|y; \theta) p(tx; \theta)$  به ترتیب استفاده کنید (الف)  $p(tz|tx; \psi, \theta)$ .

برای آموزش MAP، به جای به حداقل رساندن توزیع پسین  $(x, y, z)$  از  $p(\theta|y, x, z)$  به استفاده کنید  $, \psi | x, z$ .

برای مدل‌های غیر احتمالی که از یک تابع زیان استفاده می‌کنند و بھینه سازی ریسک تجربی  $R_{exp}(\theta; x, y)$  را بھینه می‌کنند، آموزش MMSE را انجام دهید: تابع زیان را حفظ کنید و  $R_{MMSE}(\theta; x, z)$  را بھینه کنید، برآورد RV از  $RV$  ری تجربی داده شده با  $\{\pi, \psi, z, x\}$ .

نظرة یکپارچه ساده، زیبا و جذاب به نظر من رسد. برای مدل‌های احتمالاتی، آموزش به اصل بھینه سازی اصلی از آموزش ایده‌آل وفادار می‌ماند. برای مدل‌های غیر احتمالاتی، آموزش به تابع زیان اصلی از آموزش ایده‌آل وفادار می‌ماند و برآورد MMSE (مقدار متوسط مربعات خطأ) از خطر تجربی را بھینه سازی می‌کند.

هر روش بر اساس بخشن 1.4 بھینه است. مقیاس‌های تخمین مناسب هستند: آموزش برای مدل‌های احتمالاتی هدف گیری از تابع احتمال یا پس ترتیب دارد و آموزش برای مدل‌های غیر احتمالاتی هدف گیری از خطر تجربی است. استفاده از تابع احتمال، پس ترتیب یا محاسبه کننده M MSE به این معناست که تمام اطلاعات موجود در  $\{tx, z, y, \psi, \pi\}$  به طور کامل استفاده می‌شود. معیار تنبیه یا بهره وری به وضوح تعریف شده است. هیچ یک از روش‌ها سعی نمی‌کنند برچسب‌های صحیح را تخمین بزنند.

نظرة یکپارچه همچنین برخی از کارهای مرتبط را سازماندهی می‌کند. روش آموزشی که توسط Raykar و همکاران (2010) پیشنهاد شده با بند 1a مطابقت دارد. Ratner و همکاران (2016)، Khetan (2017) و Khetan و همکاران (2018) روش‌های آموزشی ای پیشنهاد کرده اند که معادل بند 2 هستند، اگرچه به آنها دست نیافتدند با استفاده از تخمین MMSE. ما فرض می‌کنیم که  $p(y|\psi, \pi, z)$  شناخته شده یا قبلًا یاد گرفته شده است. اما Raykar و همکاران (2010) و Khetan و همکاران (2018) نشان داده اند که می‌توان از این روش‌های آموزشی در حالی که مدل برچسب‌های نویز را به طور مسترک یاد می‌گیرند، استفاده کرد.

دو بخشن بعدی توابع احتمال، احتمالات پس از توزیع و محتاطی MMSE را استخراج می‌کنند که اصلاح شده ترین شرایط اصلی را ارائه می‌دهند. استنتاج‌ها بسیار ساده‌تر از آن‌هایی برای ترسیم هستند دیگر هیچ برچسب پیش‌بینی شده ای  $\hat{y}$  برای در نظر گرفتن وجود ندارد. در واقع، استنتاج‌ها به معنای محاسبه حاشیه ای بر روی برچسب‌های صحیح است. محاسبه حاشیه ای باعث معرفی برخی تنظیمات صمنی می‌شود که به ابهام برچسب‌های صحیح توجه دارد.<sup>12</sup>

نویسنده از یکی از نویسندگان ناشناس برای این مشاهده تشکر می‌کند.

## مدل های پیش بینی احتمالی آموزش ML یا MAP

در اینجا، ما مدل های پیش بینی بر اساس دیدگاه احتمالاتی را بررسی می کنیم. دو جنبه باید در نظر گرفته شود. یک جنبه شکل مدل پیش بینی است: جداکننده یا تولیدی. یک مدل جداکننده فرم پیشین  $p(t|tx; \theta)$  را فرصت من کند و مستقیماً از  $p(tx; \theta) = p(tx|tx; \theta)$  استفاده می کند. یک مدل تولیدی فرم توزیع مشترک  $p(tx, ty; \theta) = p(tx; \theta)p(ty; \theta)$  را فرصت من کند و از  $p(tx; \theta) \propto p(ty, tx; \theta) / p(ty, tx; \theta)$  استفاده می کند. بسیاری از مدل های تولیدی از فاکتوریزاسیون  $p(tx, ty; \theta) = p(tx|ty; \theta)p(ty; \theta)$  استفاده می کنند و از  $p(tx|ty; \theta)p(ty; \theta) \propto p(tx|ty; \theta)p(ty; \theta) / p(tx; \theta)$  استفاده می کنند. انتخاب  $\theta$  تعیین کننده  $J_{\text{prior}}(\theta)$  است. رگرسیون LOGISTICK و شبکه های عصبی مثال های معمولی مدل های جداکننده هستند و بیانس ساده یک مثال کلاسیک از مدل تولیدی است (بینید Jordan, 2001 و Ng).

نکته دیگر، درمان پارامترهای مدل پیش بینی کننده  $\theta$ : غیر تصادفی یا تصادفی است. وقتی پارامترها غیر تصادفی هستند، آموزش ML استفاده می شود. برای یک مدل تشخیص، این به معنای پیدا کردن  $\theta$  برای حداقل کردنتابع احتمال  $p(y|x; \theta)$  است؛ برای یک مدل تولیدی، این به معنای پیدا کردن  $\theta$  برای حداقل کردنتابع احتمال  $p(y; x; \theta)$  است. وقتی پارامترها  $\theta$  هستند و پیش بینی های اولیه  $p(\theta)$ ، آموزش MAP استفاده می شود. این به معنای پیدا کردن  $\theta$  است که پست ترتیب  $p(y|x; \theta)$  را بیشینه می کند؛ پست ترتیب به گونه ای متفاوت گسترش می یابد که پسته به اینکه مدل متماز یا تولیدی باشد.

با در نظر گرفتن هر دو این جنبه ها، من توانیم واژه اصلی—تابع احتمال یا پس ترتیب—در تابع هدف آموزشی استخراج کنیم. تابع اصلاحی بدون تغییر باقی می ماند زیرا شامل برچسب های صحیح نیست.

### مثال 3.3.1: مدل دسیمیناتیو، پارامترهای غیر تصادفی

برای توضیح این روش، ما یک مدل جداکننده با پارامترهای غیر تصادفی را در نظر می گیریم. در حالت ایده آل، ما به  $\underline{x}$ ،  $y$  دسترسی داریم و از آموزش Machine Learning برای پیدا کردن  $\theta$  برای حداقل کردن تابع احتمال  $p(y|\underline{x}; \theta)$  که به شرح زیر است، استفاده می کنیم:

$$p(y|\underline{x}; \theta) = \prod_{i=1}^N p(y_i|\underline{x}_i; \theta). \quad (29)$$

به طور مشابه، من توانیم تابع نرمال سازی شده می منفی لیگودود hood likelihood را به حداقل برسانیم، که منجر به

$$\begin{aligned} J_{\text{prior}}(\theta; \underline{x}, y) &= -\frac{1}{N} \log p(y|\underline{x}; \theta) \\ &= -\frac{1}{N} \sum_{i=1}^N \log p(y_i|\underline{x}_i; \theta). \end{aligned} \quad (30)$$

## در مسائل درستی ساری در طبقه بندی تحت نطارت

با مسائل مربوط به صدقه، ما به دنبال  $\text{pt}(tzt | txt; t \psi_t, t \theta_t)$  هستیم که توسط

$$\begin{aligned}
p(\underline{z} | \underline{x}; \underline{\psi}, \boldsymbol{\theta}) &= \prod_{i=1}^N p(z_i | x_i; \psi_i, \theta) \\
&\stackrel{(a)}{=} \prod_{i=1}^N \sum_{y_i \in \mathcal{Y}} p(y_i, z_i | x_i; \psi_i, \theta) \\
&= \prod_{i=1}^N \sum_{y_i \in \mathcal{Y}} p(z_i | y_i, x_i; \psi_i, \theta) p(y_i | x_i; \psi_i, \theta) \\
&\stackrel{(b)}{=} \prod_{i=1}^N \sum_{y_i \in \mathcal{Y}} p(z_i | y_i; \psi_i) p(y_i | x_i; \theta),
\end{aligned} \tag{31}$$

جایی که (tat) پیس از برچسب های درست حاشه می شود، و (bt) به این دلیل است که RV های دارای برچسب نویز به بردار و  $y_i$  یا پارامتر  $\theta_t$  وابسته نیستند و به این دلیل که طبقه بندی کننده فقط بر اساس  $x_{tit}$  and  $t$  پیس بین می کند. ما می توانیم به طور معادل آن را به حداقل برسانیم

$$\begin{aligned}
J_{\text{pri}}(\boldsymbol{\theta}; \underline{x}, \underline{z}, \underline{\psi}, \boldsymbol{\pi}) &= -\frac{1}{N} \log p(\underline{z} | \underline{x}; \underline{\psi}, \boldsymbol{\theta}) \\
&= -\frac{1}{N} \sum_{i=1}^N \log \sum_{y_i \in \mathcal{Y}} p(z_i | y_i; \psi_i) p(y_i | x_i; \theta).
\end{aligned} \tag{32}$$

رایکار و همکاران (2010) (31) را به عنوان بخشی از تخمین مشترک  $\text{pt}(tzt | tyt)$  و رگرسیون لجستیک در آموزش پیشنهاد کردند.

### 3.3.2 همه موارد

چهار ردیف بالای جدول 9 نتایج را برای همه موارد ممکن خلاصه می کند. مشتقات در صنایعه های E.1، E.2، E.3 و طاهر می شوند. تنها تفاوت بین موارد ایده آل و مسائل مربوط به حقیقت این است که مورد دوم پیس از مقادیر احتمالی  $RV_t$  با برچسب صحیح به حاشه رانده می شود. سپس تفاوت های دیگر مانند حالت ایده آل است. تغییر از غیرتصادفی به پارامتر تصادفی  $\theta_t$  عامل دیگری را برای پارامتر  $\text{prior}_t$  معرفی می کند و هدف op-ntimization MAP را از  $\text{ML}$  به  $\text{MAP}$  تغییر می دهد. جابجایی از یک مدل متمایز به یک مدل مولد، تابع حداقل شده را از یک  $t$  به یک تابع شامل  $npt(txtit | tytit; t \theta_t) = t \text{pt}(txtit, tytit; t \theta_t) t \pi_t(tytit)$  تغییر می دهد.

### 3.4 مدل های پیس بین غیر احتمالی (به حداقل رساندن ریسک تجربی)

برخی از طبقه بندی کننده ها، مانند ماشین های بردار پشتیبان، فرمول احتمالی ندارند، و گاهی اوقات مسئله نظری یک مدل احتمالی تمرکز اصلی نیست. در چنین مواردی، آموزش ایده آل از اصل کمینه ساری ریسک تجربی (ERM) استفاده می کند (Vapnik, 1991) و به دنبال آن است که خطر تجربی یا ضرر متوسط را به حداقل برساند.

$$J_{\text{pri}}(\boldsymbol{\theta}; \underline{x}, \underline{y}) = R^{\text{emp}}(\boldsymbol{\theta}; \underline{x}, \underline{y}) = \frac{1}{N} \sum_{i=1}^N L(\tilde{g}(x_i; \boldsymbol{\theta}), y_i), \tag{33}$$

قبل از مدل	پار./ متقد.*	برچسب‌ها و مجموعه آموزشی	
		ایده‌آل: استفاده از $\{txt, t, yt\}$	Uset $\{txt, t, zt, t, \psi t, t, \pi t\}$ نویز
بی‌پیش‌آمده	خیر/ ML	$\begin{aligned} & \text{احتمال} (t \mid txt; t \theta t) \\ &= \sum_{it=1}^N pt(ytit \mid txtit; t \theta t) \end{aligned}$	$\begin{aligned} & \text{احتمال} \sum_{i=1}^N \psi it \mid txtit; t \theta t \\ &= \sum_{i=1}^N ytit \cdot pt(ytit \mid txtit; t \theta t) \end{aligned}$
	RV/ نقشه	$\begin{aligned} & \text{قسمت پسین} (t \mid txt, t \theta t) \\ &= \sum_{it=1}^N \frac{pt(ytit \mid txtit, t \theta t)}{pt(t \theta t)} \end{aligned}$	$\begin{aligned} & \text{قسمت پسین} \sum_{i=1}^N \frac{pt(ytit \mid txtit, t \theta t)}{pt(t \theta t)} \\ &= \sum_{i=1}^N ytit \cdot pt(ytit \mid txtit, t \theta t) \end{aligned}$
نویز پذیرکننده	خیر/ ML	$\begin{aligned} & \text{احتمال} (t \mid txt, t \theta t) \\ &= \text{Prepared} \text{Prepared} \end{aligned}$	$\begin{aligned} & \text{احتمال} \sum_{i=1}^N \psi it \mid txtit, t \theta t \\ &= \sum_{i=1}^N ytit \cdot pt(ytit \mid txtit, t \theta t) \end{aligned}$
	RV/ نقشه	$\begin{aligned} & \text{قسمت پسین} (t \mid txt, t \theta t) \\ &= \sum_{it=1}^N \frac{pt(ytit \mid txtit, t \theta t)}{pt(t \theta t)} \end{aligned}$	$\begin{aligned} & \text{قسمت پسین} \sum_{i=1}^N \frac{pt(ytit \mid txtit, t \theta t)}{pt(t \theta t)} \\ &= \sum_{i=1}^N ytit \cdot pt(ytit \mid txtit, t \theta t) \end{aligned}$
احتمال پذیر	خیر	$\begin{aligned} & \text{رسک تجربی} (Remp) = \sum_{i=1}^N \psi it \\ &= N - 1 \end{aligned}$	$\begin{aligned} & \text{تخمین MMSE RV NERM} \\ &= \sum_{i=1}^N \bar{y}_{tit} \cdot \bar{L}(t \mid txtit; t \theta t) \\ &= \sum_{i=1}^N ytit \cdot pt(ytit \mid txtit, t \theta t) \end{aligned}$

تصادفی (و معیار بهینه سازی) ML: حداقل احتمال، MAP: حداقل پسین، ERM: تغییرهای تصادفی (Par./Crit)\*. "Nشان دهنده درمان پارامتر مدل پیش‌بینی کننده (NR)  $\theta t$ : غیرتصادفی، RV: تغییرهای

به حداقل رساندن رسک تجربی).

برای مدل‌های افتراقی یا مولد، عبارت اولیه برابر است با log-likelihood زمانی شده منفی یا log-posterior. به عنوان مثال، ورودی بالا سمت چپ با  $Jtprit(t \theta t; t \mid xt, t \mid yt) = t \mid Nt t \mid 1 \log t \mid \log pt(ytit \mid txtit; t \theta t)$  مطابقت دارد. Fornnon-probabilistic models, the primary term appears in the bottom row.

جدول 9: مقایسه توابع هدف آموزشی برای مدل‌های پیش‌بینی.

جایی که  $L_t(ts, yt)$  تابعی در از دست دادن است که انحرافات بین  $t = st$  و علامت صحیح  $yt$  را جرمیمه می کند. مثال هایی برای طبقه بندی دودویی شامل ماشین های بردار پشتیبان، که ممکن است با از دست دادن لولا خم شوند، و رگرسیون لجستیک، که مربوط به یک مدل خطی آموزش دیده با تلفات لجستیک است. به عنوان مثال از طبقه بندی چند کلاسه، آموزش شبکه های عصبی اغلب از بردار خروجی از لایه های کاملاً متصل نهایی شبکه استفاده می کند و تلفات آنتروپی متقابل را اعمال می کند.

در این بخش، فرصت من کنیم که  $t = \theta$  تصادفی نیست، زیرا پارامترهای تصادفی نیاز به مدل سازی احتمالی دارند. مانند قبل، اصطلاح قاعده سازی تحت تأثیر مسائل صدق نمی باشد.

### 3.4.1 تخمین MMSE برای RV خطر تجربی آموزش

با مشکلات درست گوین، RV های برچسب صحیح مشاهده نمی شوند و تابع های صخر و خطر تجربی تابعی از آن ها هستند، بنابراین ارزش های آن ها غیرقابل تعیین است. بگذارید  $L(\tilde{g}(\mathbf{x}_i; \boldsymbol{\theta}), Y_i)$  است و از  $(33)$ ، RV خطر تجربی را بنویسیم به RV تابع صخر  $\psi$  باشد، که تابعی از RV برچسب صحیح  $\tilde{Y}$  است و از  $(33)$ ، RV خطر تجربی را بنویسیم

$$J_{\text{pri}}(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{Y}) = R(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{Y}) = \frac{1}{N} \sum_{i=1}^N L(\tilde{g}(\mathbf{x}_i; \boldsymbol{\theta}), Y_i). \quad (34)$$

ما استدلال من کنیم که اگر یک برآورد خوب از RV خطر تجربی برای مجموعه آموزشی  $\{\mathbf{x}, \psi, z\}$  به دست آید، سپس آموزش که این برآورد را به حداقل می رساند باید عملکرد خوبی داشته باشد. بنابراین، ما بر روی برآورد RV خطر تجربی تمرکز من کنیم.

ناکامی در شناخت خطر تجربی RV به دلیل عدم مشاهده RV های برچسب صحیح  $\tilde{Y}$  به وجود من آید. ویرگی های توپولوژیکی  $x$  به دلیل شناخته بودن به ناطمینانی کمک نمی کند. با فرمول مدل برچسب های نویز  $p(y|\psi)$ ، قانون بیز می تواند  $p(\psi|y)$  را ارائه دهد اما نه  $p(x|\psi)$ . به این دلیل، هنگام برآورد خطر تجربی RV،  $\mathbf{Y}$  و  $\mathbf{Z}$  را به عنوان RV ها در نظر می گیریم، اما نه  $x$ . با این حال، در رمینه گسترده تر یادگیری، ویرگی های توپولوژیکی من توانند به عنوان RV ها در نظر گرفته شوند.

ما معیار MMSE را در تخمین  $R(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{Y})$  استفاده می کنیم. یک تخمین گر برای  $R(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{Y})$  را با  $\hat{R}(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{Z}, \psi, \pi)$  نشان می دهیم، جایی که برای اختصار  $\psi$  و  $\pi$  را از  $\hat{R}$  حذف می کنیم. ما اصطلاح اصلی را برابر با تخمین گر MMSE برای  $R(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{Y})$  می گذاریم:

$$\hat{R}_{\text{pri}}(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{Z}, \psi, \pi) = \hat{R}^{\text{MMSE}}(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{Z}) = \arg \min_{\hat{R}} \mathbb{E}[(\hat{R}(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{Z}) - R(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{Y}))^2].$$

نتیجه استاندارد این است که مطلوب ترین تخمین دهنده MMSE معادل میانگین شرطی  $R(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{Y})$  است ( ایست ز (بینید ) ( یا پیوست های C و E.4)، بنابراین

$$\hat{R}^{\text{MMSE}}(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{Z}) = \mathbb{E}_{p(y|\mathbf{z}; \boldsymbol{\psi})}[R(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{Y}) | \mathbf{Z}; \boldsymbol{\psi}] \quad (35)$$

$$= \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{p(y_i|\mathbf{z}_i; \boldsymbol{\psi}_i)}[L(\tilde{g}(\mathbf{x}_i; \boldsymbol{\theta}), Y_i) | \mathbf{Z}_i; \boldsymbol{\psi}_i] \quad (36)$$

$$= \frac{1}{N} \sum_{i=1}^N \sum_{y_i \in \mathcal{Y}} L(\tilde{g}(\mathbf{x}_i; \boldsymbol{\theta}), y_i) p(y_i | \mathbf{Z}_i; \boldsymbol{\psi}_i). \quad (37)$$

نمونه ها مستقل هستند، بنابراین  $R(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{Y})$  به صورت نرمال توزیع شده توسط CLT است. برآورده  $\hat{R}^{\text{MMSE}}$  معادله میانگین شرطی این RV نرمال را با توجه به  $Z$  حدس می زند.

با توجه به  $Z_t = t$  از RV با MMSE ریسک تجربی است

$$\hat{R}^{\text{MMSE}}(\boldsymbol{\theta}; \underline{x}, \underline{z}) = E_{p(y|\underline{z}; \boldsymbol{\psi})}[R(\boldsymbol{\theta}; \underline{x}, Y)|\underline{z}; \boldsymbol{\psi}] \quad (38)$$

$$= \frac{1}{N} \sum_{i=1}^N E_{p(y_i|\underline{z}_i; \boldsymbol{\psi}_i)}[L(\tilde{g}(\underline{x}_i; \boldsymbol{\theta}), Y_i)|\underline{z}_i; \boldsymbol{\psi}_i] \quad (39)$$

$$= \frac{1}{N} \sum_{i=1}^N \sum_{y_i \in \mathcal{Y}} L(\tilde{g}(\underline{x}_i; \boldsymbol{\theta}), y_i) p(y_i|\underline{z}_i; \boldsymbol{\psi}_i), \quad (40)$$

جایی که  $\text{pt}(tyt|tztit;t \Psi tit)$  13t پسین است:

$$p(y_i|\underline{z}_i; \boldsymbol{\psi}_i) = \frac{\pi(y_i)p(\underline{z}_i|y_i; \boldsymbol{\psi}_i)}{\sum_{y'_i \in \mathcal{Y}} \pi(y'_i)p(\underline{z}_i|y'_i; \boldsymbol{\psi}_i)}. \quad (41)$$

پس از یافتن تخمین گر RV با ریسک تجربی، اکنون می‌توانیم اغلب از MMSE trainingt صلح بست کنیم، که به دنبال  $\theta_{\text{tt}}$  است که (40)، برآورد RV MMSE از خطر تجربی را به حداقل برساند. معادله (28) من شود

$$\begin{aligned} \boldsymbol{\theta}^* &= \arg \min_{\boldsymbol{\theta}} [\hat{R}^{\text{MMSE}}(\boldsymbol{\theta}; \underline{x}, \underline{z}) + \lambda J_{\text{reg}}(\boldsymbol{\theta})] \\ &= \arg \min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^N \sum_{y_i \in \mathcal{Y}} L(\tilde{g}(\underline{x}_i; \boldsymbol{\theta}), y_i) p(y_i|\underline{z}_i; \boldsymbol{\psi}_i) + \lambda J_{\text{reg}}(\boldsymbol{\theta}). \end{aligned}$$

آموزش MMSE نمونه هایی را که به درستی برچسب گذاری شده اند به روشنی طبیعی مدیریت می‌کند. برای این نمونه، اگر برچسب صحیح به شرط  $t = \text{pt}(tytit|tztit;t \Psi tit)$  باشد، ساخته می‌شود، و نمو نه ارزش تابع صفر معمول را به عبارت ریسک تجربی نمی‌دهد.

ردیف پایین جدول 9 آموزش مدل های غیر احتمالی را خلاصه می‌کند. Rat-nner و همکاران (2017) (از 39) برای مدل های تمایزی از جمله رگرسیون لجستیک باینزی و شبکه های عصبی بارگش تی حافظه کوتاه مدت استفاده کرد.ختان و همکاران (2018).

(40) را برای توابع از دست دادن دلخواه پیشنهاد کرد، از آن برای تخمین مشترک اع لب (tzt|tyt) و آموزش استفاده کرد و صنمانت های عملکرد را برای طبقه بندی باینزی ارائه کرد. هیچ ی ک از گروه های نویسنده پیوند به تخمین MMSE را نشان ندادند، بنابراین این بخش راه دیگری برای ایجاد ان گیزه و رسیدن به این نتیجه ارائه می‌دهد.

ما این بخش را با در نظر گرفتن تابع صفر برای یک نمونه می‌بندیم. در (36)، هر ترم در جمع با برآورد گر MMSE یک تابع تلفات منفرد  $nRVt Lt(tgt(txtit;t \theta)t, Ytit)$  مطابقت دارد.

$$\hat{L}^{\text{MMSE}}(\tilde{g}(\underline{x}_i; \boldsymbol{\theta}), Z_i) = E_{p(y_i|\underline{z}_i; \boldsymbol{\psi}_i)}[L(\tilde{g}(\underline{x}_i; \boldsymbol{\theta}), Y_i)|Z_i; \boldsymbol{\psi}_i], \quad i = 1, \dots, N, \quad (42)$$

$$= \sum_{y_i \in \mathcal{Y}} L(\tilde{g}(\underline{x}_i; \boldsymbol{\theta}), y_i) p(y_i|Z_i; \boldsymbol{\psi}_i), \quad i = 1, \dots, N. \quad (43)$$

اگر تابع صفر دارای ویژگی هایی مانند غیر منفی بودن یا تحدب باشد، آنگاه (43) چنین ویژگی هایی را ح فطر می‌کند زیرا  $\text{pt}(tytit|tztit;t \Psi tit)$  یک توزیع احتمال است. جای تعجب نیست که تخمین گر RV با ریسک تجربی فقط میانگین تخمین گرهای MMSE در RV های منفرد با تابع صفر را سنت.

13. پس زمینه کلاس آموزشی با پس زمینه کلاس تستی (10) که شامل برچسب های پیش بینی شده و پارامترهای OP بود، متفاوت است. این همان چیزی است که در بخش 2.4.2 به عنوان (21) ذکر شده است.

### 3.4.2 ویرگی های استاندارد

ویرگی های استاندارد برآورده کنندگان  $Z_{\underline{x}; \theta}^{MMSE}$  به  $R(\underline{x}; \theta)$  اعمال می شود؛ بینید در پیوست C.1 آن<sup>14</sup> بی طرف است از نظر بیایی:

$$E[\hat{R}^{MMSE}(\theta; \underline{x}, \underline{Z})] = E[R(\theta; \underline{x}, \underline{Y})]. \quad (44)$$

همچنین، برآورده کنندگان MMSE هر تابع زیان نیز بی طرف است:

$$E[\hat{L}^{MMSE}(\tilde{g}(\underline{x}_i; \theta), Z_i)] = E[L(\tilde{g}(\underline{x}_i; \theta), Y_i)], \quad i = 1, \dots, N. \quad (45)$$

از (44)، میانگین خطای برآورد صفر است:

$$E[\hat{R}^{MMSE}(\theta; \underline{x}, \underline{Z}) - R(\theta; \underline{x}, \underline{Y})] = 0,$$

و واریانس خطای برآورد برابر MSE است:

$$\text{var}(\hat{R}^{MMSE}(\theta; \underline{x}, \underline{Z}) - R(\theta; \underline{x}, \underline{Y})) = E[(\hat{R}^{MMSE}(\theta; \underline{x}, \underline{Z}) - R(\theta; \underline{x}, \underline{Y}))^2] \quad (46)$$

$$= \frac{1}{N^2} \sum_{i=1}^N E[\text{var}(L(\tilde{g}(\underline{x}_i; \theta), Y_i) | Z_i)]. \quad (47)$$

پیوست E.4 را استخراج می کند و نشان می دهد که خطای برآورد به یک RV عادی با این لحاظات نزدیک می شود. توجه داشته باشید که  $Z_{\underline{x}; \theta}^{MMSE}$  برابر  $R^{MSE}$  است. ویرگی های دیگر، مانند داصل راست کلی، نیز صادق هستند اما در اینجا مرتب نیستند.

برآورده کنندگان MMSE یک تابع از  $ZRV$  است، بنابراین آن نیز یک RV است و (46) و (47) را با توجه به توزیع  $Z$  توضیح می دهند. در مقابل، برآورده کنندگان MMSE برآورده کنندگان ای است که در  $z = Z$  ارزیابی می شود و تصادفی نیست. MSE که توسط (39) یا (40) به دست آمده است برابر واریانس  $\text{condition}_{\text{all}}$   $z = Z$  است:  $z = Z$   $\text{RV empirical-risk}$

$$\begin{aligned} E[(\hat{R}^{MMSE}(\theta; \underline{x}, \underline{Z}) - R(\theta; \underline{x}, \underline{Y}))^2 | Z = z] &= \text{var}(R(\theta; \underline{x}, \underline{Y}) | Z = z) \\ &= \frac{1}{N^2} \sum_{i=1}^N \text{var}(L(\tilde{g}(\underline{x}_i; \theta), Y_i) | Z_i = z_i). \end{aligned}$$

### 3.4.3 محاسبه مشتق

آموزن اغلب از نوعی از نزولگردی استفاده می کند تا  $\theta^*$  را پیدا کند. به عنوان مثال، آموزن شبکه های عصبی عمیق از دیفرانسیل خودکار برای محاسبه مشتق خطای تجربی استفاده می کند. در حالت ایده ای،

$$\frac{\partial J_{\text{ideal}}}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} [R^{\text{emp}}(\theta; \underline{x}, \underline{y})] + \lambda \frac{\partial J_{\text{reg}}}{\partial \theta_j}, \quad (48)$$

و از (33)

$$\frac{\partial}{\partial \theta_j} [R^{\text{emp}}(\theta; \underline{x}, \underline{y})] = \frac{1}{N} \sum_{i=1}^N \frac{\partial}{\partial \theta_j} [L(\tilde{g}(\underline{x}_i; \theta), y_i)] = \frac{1}{N} \sum_{i=1}^N \frac{\partial L}{\partial \tilde{g}}(\tilde{g}(\underline{x}_i; \theta), y_i) \frac{\partial \tilde{g}}{\partial \theta_j}(\underline{x}_i; \theta). \quad (49)$$

<sup>14</sup>).  $E[\underline{x}; \theta](R[\underline{y}]) = E[\underline{y}; \theta](R[\underline{x}])$   $E[\underline{y}; \underline{z}] = E[\underline{z}; \underline{y}]$ . به ویژه،

تدریب بر اساس تخمین MMSE جایگزین مشتق جزئی  $\partial$  می شود

با استفاده از  $R_{\text{temp}}(t_{\text{theta}}; x, y)$

$$\begin{aligned} \frac{\partial}{\partial \theta_j} [\hat{R}^{\text{MMSE}}(\boldsymbol{\theta}; \underline{x}, \underline{z})] &= \frac{1}{N} \sum_{i=1}^N \sum_{y_i \in \mathcal{Y}} p(y_i | z_i; \psi_i) \frac{\partial}{\partial \theta_j} [L(\tilde{g}(\underline{x}_i; \boldsymbol{\theta}), y_i)] \\ &= \frac{1}{N} \sum_{i=1}^N \sum_{y_i \in \mathcal{Y}} p(y_i | z_i; \psi_i) \frac{\partial L}{\partial \tilde{g}}(\tilde{g}(\underline{x}_i; \boldsymbol{\theta}), y_i) \frac{\partial \tilde{g}}{\partial \theta_j}(\underline{x}_i; \boldsymbol{\theta}). \end{aligned} \quad (50)$$

این یک ترکیب محاسبه از مشتقات جزئی تابع صفر است. همچنین با روش های خودکار مشتق گیری سا رگار است و محاسبه آن ارزان است. برای یک شبکه عصبی عمیق، محاسبه  $\frac{\partial}{\partial \theta_j} \tilde{g}(\underline{x}_i; \boldsymbol{\theta})$  بیشتر باز محاسباتی را تشکیل می دهد و این مقادیر باید فقط یک بار محاسبه شوند—همان بار محاسباتی که در حالت ایده آل وجود دارد. محاسبه  $\frac{\partial}{\partial \theta_j} \tilde{g}$  برای یک تابع صفر معمولی مقدار قابل توجهی محاسبه نیاز دارد، بنابراین محاسبه آن برای هر یک از مقادیر ممکن  $\underline{z}$  بار محاسباتی را به طور قابل توجهی افزایش نمی دهد.

یک مثال از مشتق های جزئی برای رگرسیون لوجستیک دودویی در بخش 5.4 و اپنیکس F.2 آمده است.

#### 3.4.4 موارد خاص

ما دو مورد خاص در دو انتهای مخالف را به طور خلاصه در نظر می گیریم. اولاً، فرض کنید که برچسب صحیح می تواند به طور کامل از برچسب های نویز دار بازیابی شود؛ یعنی،  $(i; \psi_i) = (y'_i | z_i; \underline{y})$ . پس تخمین MMSE مقادیر صحیح ریسک تجربی و تابع صفر را بارمی گرداند. به عنوان مثال، (40) به دست می دهد  $\hat{R}_{\text{MMSE}}^{\text{mp}} = y_i \underline{x}_i \underline{\theta}^T \underline{R}^{-1} \underline{x}_i \underline{\theta}$  (43) و (49) (کاهش می یابد)  $= z_i \underline{x}_i \underline{\theta}^T \underline{R}^{-1} \underline{x}_i \underline{\theta}$  (50) (کاهش می یابد). همچنین، مشتق جزئی در (49) (به 50) (به 43) (به 40) (به 49) (کاهش می یابد)  $=$   $- \frac{\partial}{\partial \theta_j} \underline{R}^{-1} \underline{x}_i \underline{\theta}$ .

دوم، فرض کنید برچسب های نویز دار هیچ اطلاعاتی در مورد RVs برچسب های صحیح ارائه نمی دهند؛ یعنی،  $(\psi_i) = p(z_i | y_i; \underline{\theta})$ . سپس  $p(y_i | z_i; \underline{\theta}) = p(z_i; \underline{\theta})$  (37) به

$$\hat{R}^{\text{MMSE}}(\boldsymbol{\theta}; \underline{x}, \underline{z}) = \frac{1}{N} \sum_{i=1}^N \sum_{y_i \in \mathcal{Y}} L(\tilde{g}(\underline{x}_i; \boldsymbol{\theta}), y_i) \pi(y_i) = \mathbb{E}_{\pi(y)} [R(\boldsymbol{\theta}; \underline{x}, \underline{Y})].$$

برای هر مقدار  $\underline{z}$ ، محاسبه کننده MMSE همیشه میانگین RV خطر تجربی را توجه به پیش بینی های کلا س  $\pi(y)$  (برمی گرداند. مشتق جزئی در (50) به طور مشابه رفتار می کند. محاسبه کننده MMSE بی طرف باقی می ماند، بنابراین (44) همچنان صادق است. اما محاسبه کننده یک مقدار ثابت است، بنابراین واریانس آن صفر است. بر اساس قانون واریانس کل، MSE آن به بزرگترین مقدار ممکن می رسد و برابر با واریانس  $R(\underline{Y}; \underline{x}; \boldsymbol{\theta})$  (است) ببینید در پیوست C.1.

#### 3.4.5 یکپارچگی برآورده

برای آموزش ایده آل، یک دلیل مهم برای کمینه سازی ریسک تجربی (33) برای یک دسته از مدل های پیش بینی کننده، تطابق اصل ERM است: ریسک تجربی ایده آل به احتمال نزدیک به کمترین ریسک قابل دستیابی می رسد وقتی که  $N$  به نهایت نزدیک می شود، حتی اگر توزیع واقعی  $X_i, Y_i$  ناشناخته باشد (Vapnik, 1991). رابطه بین برآورد MMSE از RV ریسک تجربی و تطابق اصل ERM نیاز به مطالعه بیشتری دارد. در عین حال، ما به تطابق برآورده کننده MMSE خود توجه می کیم.

مقداردهنده  $\hat{R}_{MMSE}(\theta; \underline{x}, \underline{Z})$  هموار است اگر به احتمال به  $R(\theta; \underline{x}, \underline{Z})$  نزدیک شود به عنوان  $N \rightarrow \infty$ . به این ترتیب، مقدار واقعی  $RV$  خطر تجربی را به دست می‌آورد. همچنین، یک مقداردهنده اگر  $MSE$  آن به صفر برسد به عنوان  $N \rightarrow \infty$ , هموار است و هموار بودن  $MSE$  به معنای هموار بودن است. بنابراین، اگر  $\hat{R}_{MMSE}(\theta; \underline{x}, \underline{Z})$  هموار باشد، آن هموار است.

از (46) و (47) به شرح زیر است:

$$E[(\hat{R}_{MMSE}(\theta; \underline{x}, \underline{Z}) - R(\theta; \underline{x}, \underline{Y}))^2] = \frac{1}{N^2} \sum_{i=1}^N E[\text{var}(L(\tilde{g}(\underline{x}_i; \theta), Y_i) | \underline{Z}_i)].$$

شرط کافی برای وقتی که برآورده  $MMSE$  همگن است به شرح زیر است:

$$\lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{i=1}^N E[\text{var}(L(\tilde{g}(\underline{x}_i; \theta), Y_i) | \underline{Z}_i)] = 0. \quad (51)$$

اگر یک ثابت  $b$  وجود داشته باشد که

$$E[\text{var}(L(\tilde{g}(\underline{x}_i; \theta), Y_i) | \underline{Z}_i)] < b, \quad i = 1, \dots, N, \quad (52)$$

بنابراین، اگر تابع زیان محدود باشد، واضح است که (52) (برآورده می‌شود)  $b/N \rightarrow 0$  و (51) (برآورده می‌شود)  $\lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{i=1}^N E[\text{var}(L(\tilde{g}(\underline{x}_i; \theta), Y_i) | \underline{Z}_i)] = 0$ . بنابراین، اگر تابع زیان محدود باشد، واصح است که (52) (برآورده می‌شود). اگر تابع زیان اصلی نامحدود باشد، ممکن است بتوان آن را تغییر داد تا محدود شود؛ به عنوان مثال، وقتی تابع زیان اصلی از یک حد مشخص فراتر می‌رود، می‌تواند به یک محدوده بالایی به صورت نزدیک به محدوده نهایی تبدیل شود.

### 3.5 مزایای آموزش MMSE

در دیدگاه یکپارچه آموزش، آموزش MMSE برای مدل‌های غیر احتمالی دارای مزایای جالبی در مقایسه با آموزش ML یا MAP برای مدل‌های احتمالی دارد. آموزش MMSE می‌تواند همچنان از تابع صرر اصلی استفاده کند و شامل تغییرات ساده‌ای در محاسبه خطر تجربی است. برآورده  $MMSE$  ویرگی‌های استانداردی مانند بی طرفی بیایی را به دست می‌آورد. می‌توان از نزول کردن gradient و خودتولیدکنندگی با تغییرات کمی استفاده کرد. اگر تابع صرر محدود باشد، برآورده  $MMSE$  یک برآورده ثابت از  $RV$  خطر تجربی است.

در مقابل، روش آموزش ML یا MAP ممکن است نیاز به استنتاج‌های جدید داشته باشد و ممکن است نتواند از تابع‌های صرر موجود و مستقایات آن‌ها استفاده کند. این تغییرات ممکن است تحلیل نظری را پیچیده تر کند.

این تفاوت‌ها در مثال برای رگرسیون لوجستیک دودویی، که در بخش 5.4 ارائه شده، قابل مشاهده است. معادلات برای شرایط اصلی و گرادیان‌ها در پیوست F آورده شده‌اند.

### 3.6 روش‌های آموزش جایگزین

نظره یکپارچه تمام روش‌های ممکن برای آموزش را پوشش نمی‌دهد و در اینجا ما درباره برخی جایگزین‌ها صحبت می‌کنیم.

### 3.6.1 ایجاد تلفات ضعیف برای برچسب‌های جزئی

یک رویکرد بیزی جایگزین توسط Cid-Sueiro (2012) و همکاران (2014) خواص نظری تلفات ضعیف را برای برچسب‌های جزئی مورد مطالعه قرار داد. برای classt، برچسب صحیح یک داغ به صورت  $t_{\text{tyt}} = t_{\text{tetyt}}$  کدگذاری شده بود، که در آن بردار  $t_{\text{tetjt}}$  برابر یک در عنصر  $j,t$  و صفر در جای دیگر است. بردار برچسب جزئی  $t_{\text{tzt}}$  به طور مشابه رمزگذاری شده بود، اما چندین عنصر را می‌توان روی یک تنظیم کرد، بنابراین برچسب دار نویزدار می‌تواند بیش از یک کلاس را نشان دهد. نویسنده‌گان محدودیت‌های مختلفی را بر روی بردارهای مجاز جزئی برچسب در نظر گرفتند، بنابراین تعداد بردارهای برچسب ب جزئی ممکن،  $t$ ، می‌تواند یک عدد صحیح بین  $C_{\text{nand}} 2tC_t$  باشد.

تابع زیان تعریف کردند که برای کار بر روی برچسب‌های جزئی به جای برچسب‌های صحیح طراحی شده است. آنها سپس معادل صفر را برای برچسب‌های صحیح معرفی کردند:

$$L^{eq}(s, \bar{y}) = E_{p(\bar{z}|\bar{y})}[L^{wk}(s, \bar{Z})|\bar{Y} = \bar{y}], \quad (53)$$

و این را نشان دادند

$$\begin{aligned} E_{p(\bar{z})}[L^{wk}(s, \bar{Z})] &= \sum_{\bar{z}} p(\bar{z}) L^{wk}(s, \bar{z}) \\ &= \sum_{\bar{z}} \sum_{\bar{y}} p(\bar{y}) p(\bar{z}|\bar{y}) L^{wk}(s, \bar{z}) \\ &= \sum_{\bar{y}} p(\bar{y}) \sum_{\bar{z}} p(\bar{z}|\bar{y}) L^{wk}(s, \bar{z}) \\ &= E_{p(\bar{y})}[E_{p(\bar{z}|\bar{y})}[L^{wk}(s, \bar{Z})|\bar{Y}]] \quad (54) \\ &= E_{p(\bar{y})}[L^{eq}(s, \bar{Y})]. \quad (55) \end{aligned}$$

از این رابطه آنها استدلال کردند که آموزش بر روی برچسب‌های جزئی با  $L_{\text{twkt}}(ts, t \ tz)$  مانند آموزش روی برچسب‌های صحیح با  $L_{\text{teqt}}(ts, t \ tyt)$  رفتار نخواهد کرد. با توجه به تابع صفر اصلی  $L_{\text{t}}(ts, yt)$  که برای برچسب‌های صحیح در نظر گرفته شده است، صفر معادل است  $L_{\text{teqt}}(ts, t \ tetyt) = t L_{\text{t}}(ts, yt)$  و شخص می‌خواهد یک صفر ضعیف مربوطه را پیدا کند. برای این منظور، نویسنده‌گان (53) را به یک معادله ماتریسی بسط دادند:

$$\begin{bmatrix} L^{eq}(s, \bar{e}_0) \\ L^{eq}(s, \bar{e}_1) \\ \vdots \\ L^{eq}(s, \bar{e}_{C-1}) \end{bmatrix} = \begin{bmatrix} p_{\bar{Z}|\bar{Y}}(\bar{b}_0|\bar{e}_0) & p_{\bar{Z}|\bar{Y}}(\bar{b}_1|\bar{e}_0) & \cdots & p_{\bar{Z}|\bar{Y}}(\bar{b}_{B-1}|\bar{e}_0) \\ p_{\bar{Z}|\bar{Y}}(\bar{b}_0|\bar{e}_1) & p_{\bar{Z}|\bar{Y}}(\bar{b}_1|\bar{e}_1) & \cdots & p_{\bar{Z}|\bar{Y}}(\bar{b}_{B-1}|\bar{e}_1) \\ \vdots & \vdots & \ddots & \vdots \\ p_{\bar{Z}|\bar{Y}}(\bar{b}_0|\bar{e}_{C-1}) & p_{\bar{Z}|\bar{Y}}(\bar{b}_1|\bar{e}_{C-1}) & \cdots & p_{\bar{Z}|\bar{Y}}(\bar{b}_{B-1}|\bar{e}_{C-1}) \end{bmatrix} \begin{bmatrix} L^{wk}(s, \bar{b}_0) \\ L^{wk}(s, \bar{b}_1) \\ \vdots \\ L^{wk}(s, \bar{b}_{B-1}) \end{bmatrix}, \quad (56)$$

که در آن  $t_{\text{bt}} = 0, t_1, t_2, \dots, B$ . ماتریس  $t \ it = 0, t_1, t_2, \dots, B$  دهنده این است که بردار برچسب جزئی ممکن است. اشاره Cid-Sueiro اس اشاره س توزیع شرطی  $t \ pt(tz | tyyt)$  را جدول بندی می‌کند، بنابراین دارای ابعاد  $Ct \ tBt$  است.  $L_{\text{twkt}}(ts, t \ tz)$  وجود دارد.

راه کار آموزش MMSE که در بخش 3.4.1 پیشنهاد شده نیز به روس بیزینس است، اما به جای طراحی یک تابع صفر جدید برای برچسب‌های نویز، از محاسبه گر  $\hat{LMMSE}(s, Z)$  تابع صفر اصلی  $L(s, Y)$  که برای برچسب‌های صحیح طراحی شده استفاده می‌کند. آموزش MMSE فرص من کند

بخش‌هایی از کار van Rooyen و (2018) به این روش تردید کی است.

برچسب های نویز به ویرگی های ویکتوریایی وابسته نیستند زیرا مدل برچسب نویز فرم  $p(z|y)\pi(y)$  را دارد.

از (42) و (43)، محاسبه گر MMSE معیار میانگین condition تابع صرر اصلی با توجه به  $Z$  است و تنها ترکیب محدب تابع صرر اصلی برای هر ممکن  $y$ ، با وزن  $p(y|Z)$  است. برای روش که توسط Cid-Sueiro et al (53) پیشنهاد می کند که تابع صرر معادل می تواند ب  $\hat{L}^{MMSE}(s, Z)$  نویز  $y$  را تفسیر شود. برای پیدا کردن تابع صرر صنعیف یعنوان برآورد MMSE تابع صرر صنعیف با توجه به  $\bar{Y}$  (56) را حل کند.

از (45)، محاسبه گر MMSE در معنای بیزینس بی طرف است، نتیجه استانداردی که توسط انتظارا ت تکراری به دست می آید:

$$E_{p(z)}[\hat{L}^{MMSE}(s, Z)] = E_{p(z)}[E_{p(y|z)}[L(s, Y)|Z]] = E_{p(y)}[L(s, Y)].$$

Cid-Sueiro et al (55) می دهند. یک رابطه مشابه به دست می آورند، اما در ترتیب معکوس: (54) و (55) می دهند

$$E_{p(\bar{z})}[L^{wk}(s, \bar{Z})] = E_{p(\bar{y})}[E_{p(\bar{z}|\bar{y})}[L^{wk}(s, \bar{Z})|\bar{Y}]] = E_{p(\bar{y})}[L^{eq}(s, \bar{Y})].$$

این رابطه به بی طرف بیزینس محاسبه گر MMSE تابع صرر صنعیف با توجه به  $\bar{Y}$  مربوط است.

آموزش MMSE شرط می گذارد که برچسب های نویز  $Z$  باشد. پس از اینکه  $Z = z$  مشاهده می شود، از  $p(y|z)$  برای برآورد تابع صرر اصلی استفاده می کند. کار Cid-Sueiro et al (56) شرط می گذارد که برآورد  $\bar{Y}$  با  $RV$  باشد. قبل از اینکه  $Z$  مشاهده شود، آنها  $p(\bar{z}|y)$  برای تمام ترکیب های  $\bar{z}$  و  $\bar{y}$  را جدول بندی می کنند و معادله realization of  $Z$  که ممکن است رخ دهد، محاسبه کنند.

در خلاصه، این دو روش بیزینس برخی شباهت ها دارند اما مسائل حقیقت یابی را به روش های مختلف حل می کنند. هیچ یک از آنها برچسب های صحیح را برآورد نمی کنند. می توان یکی را به عنوان عم ل در جهت معکوس دیگری تفسیر کرد. تابع صرر معادل برای برچسب های صحیح را می گیرند و تابع صرر صنعیف برای برچسب های نویز می سازند. آموزش MMSE برچسب های نویز را می گیرد و تابع صرر اصلی برای برچسب های صحیح را برآورد می کند.

### استفاده از تابع های صرر پراکنس

یک راه جایگزین دیگر، که توسط Natarajan و همکاران (2013) (پیشنهاد شده، دیدگاه کلاسیک) یعنی فراوانی  $\hat{g}(x)$  را هنگام جایگزینی تابع صرر اصلی  $L(s, y)$  برای  $s$  و  $y$  با یک تابع صرر واسطه  $\tilde{g}(x)$  می تواند تابع صرر صنعیف برای نویز  $y$  را تخمین زند. آنها به طبقه بندی دودویی با یک برجسته دهنده تنها توجه کردند و دیدگاه کلاسیک را اتخاذ کردند، بنابراین  $y$  ناشناخته اما غیرتصادفی است.

— — —

16. واین راین و ویلیامسون (2018) این روش را به شرایط دیگری تعمیم دادند که در آن تابع از دست دادن برای برچسب های صحیح می تواند تغییر کند تا از برچسب های نویز دار حساب شود.

تابع صفر پراکسی به عنوان راه حل یک سیستم ارج C معادلات خطی در  $C$  معیار ناشناخته ها است:<sup>17</sup>

$$\begin{bmatrix} p_{Z;y}(0;0) & p_{Z;y}(1;0) & \cdots & p_{Z;y}(C-1;0) \\ p_{Z;y}(0;1) & p_{Z;y}(1;1) & \cdots & p_{Z;y}(C-1;1) \\ \vdots & \vdots & \ddots & \vdots \\ p_{Z;y}(0;C-1) & p_{Z;y}(1;C-1) & \cdots & p_{Z;y}(C-1;C-1) \end{bmatrix} \begin{bmatrix} L^{pr}(s,0) \\ L^{pr}(s,1) \\ \vdots \\ L^{pr}(s,C-1) \end{bmatrix} = \begin{bmatrix} L(s,0) \\ L(s,1) \\ \vdots \\ L(s,C-1) \end{bmatrix},$$

که بسیار مشابه (56) در روس (2012) Cid-Sueiro et al (2012) و Cid-Sueiro (2014) است؛ تا زمانی که رتبه کامل داشته باشد، یک راه حل منحصر به فرد برای  $s, Z^{pr}$  وجود دارد.

This approach utilizes knowledge of  $p(z; y)$  and does not estimate the correct labels. As a classical approach, it does not involve a class prior  $\pi$ . Extending it to multiple labelers and varying combinations of labelers might be difficult. For  $T$  labelers, if every labeler provides a label for every sample, then the proxy loss function must satisfy  $E[p(z; y)][L(pr(s, Z))] = L(s, y)$ , برای هر  $y$  در  $Z$  ...  $Z^T$ . جایی که  $Z^T$  م معادلات خطی در  $C$  معلوم نا شناخته ایجاد می کند که برای  $T$  ناکافی است. اگر برچسب گذاران بتوانند از ارائه برچسب خودداری کنند اما حداقل یک گذار باید برای هر نمونه برچسب ارائه دهد، تعداد معلوم ناشناخته  $+1$  - 1 می شود.

در مقابل، آموزش MMSE در بخش 3.4.1 یک روش بیانی است که از پیش بینی های کلاس  $\pi$  و هم چنین  $p(tz | ty)$  استفاده می کند. این روش از برآورده MMSE برای  $L(ts, Y)$  استفاده می کند که از معادله (42) به دست می آید:  $E[p(ty | tz)][L(ts, Y)] = E[p(ty)]E[L(ts, Y)]$ . این برآورده در معنای بیانی، بی طرف است، یعنی  $[Z]^\wedge$ : اولی  $y$  را به عنوان یک مقدار ناشناخته و غیرتصادفی در نظر می گیرد و دومی  $y$  را به عنوان یک واقعیت ناشناخته از  $Y$  RV در نظر می گیرد. بنابراین، بی طرف کلاسیک یک هدف اصلی در آمار بیانی نیست (Bibinid 2001 Kay, 1993, §10.3). دیدگاه های کلاسیک و بیانی به طور بینایی متفاوت هستند (Breiman 2001 Gelman et al. 2013, §4). روش آموزش ما از تابع از دست دادن اصلی استفاده می کند، بنابراین نیازی به تابع از دست دادن و اسطه نیست. همچنین، از معادله (43)، برآورده MMSE یک ترکیب محدب از مقادیر تابع از دست دادن چندین برچسب دهنده و ترکیب های مختلف برچسب گذاران برای نمونه های مختلف را پذیرا شود.

### 3.6.3 پیش بینی برچسب های نویز

در یک روش آموزش جایگزین، Jindal et al. (2015) Sukhbaatar و (2016) در یک شبکه پایه دنبال شده توسط یک لایه اصناف تشکیل شده است تا برچسب های نویز  $Z$  از یک برچسب دهنده تنها پیش بینی کند. تابع از دست دادن بدون تغییر باقی مانده است، اما  $y$  جایگزین شده توسط  $Z$  می شود. در طول آموزش، نویسندگان هر دو بخش شبکه ترکیبی را با تنظیمات استاندارد کرده اند تا شبکه پایه  $(x|y)p(y|x)$  را یاد بگیرد و لایه اصنافی  $p(z|y)$  را یاد بگیرد. پس از آن موزش، شبکه پایه می تواند استخراج شده و برای پیش بینی برچسب های صحیح استفاده شود.

اجازه دهید  $b(x; \theta)$  شبکه پایه را نشان دهیم و  $a(\cdot)$  لایه اصنافی را نشان دهیم، جایی که  $\Psi$  یک ماتریس  $C \times n$  است که تخمین لایه از  $(y|z)p(y|z)$  را نشان می دهد. سپس شرط اصلی این روش

$$S = \sum_{i=1}^n L(a(nn b(x; \theta); \Psi), z_i)$$

این روش برچسب های صحیح را تخمین نمی زند و به طور مشترک  $p(y|x)$  و  $p(z|y)$  را تخمین می زند. پیش بینی های کلاس  $\pi$  را در نظر نمی گیرد. برای یک برچسب دهنده تنها، هیچ تغییر در تابع از دست دادن لازم نیست، اما تابع از دست دادن باید تغییر کند تا برای چندین برچسب دهنده یا ترکیب های مختلف برچسب دهنده ها حساب شود.

17. Natarajan et al. 2017. تنها به طبقه بندی دوگانه توجه کرده اند؛ ما فرمولاسیون برای  $C$  متغیر را ارائه می دهیم.

لازم است تغییرات در تابع از دست دادن انجام نشود، اما تابع از دست دادن باید تغییر کند تا برای چندین برجس ب دهنده یا ترکیب های مختلف برجس ب دهنده ها حساب شود.

### 3.6.4 3 پیش بینی برجس ب های نویز با تنظیمات ریدیابی

برای برجس ب دهنده های مستقل چندگانه، Tanno et al. (2019) پیشنهاد کرد که همزمان یک شبکه CNN را آموزش دهد و ماتریس های گیج کننده برجس ب دهنده ها را تخمین بزند. آن ها از ماتریس های گیج کننده برای تنظیم خروجی softmax شبکه CNN استفاده کردند تا امتیازهای پیش بینی شده برای برجس ب های نویز  $z$  را به دست آورند. هدف آموزش یک شرط از دست دادن خطرات و یک شرط تنظیم ریدیابی را ترکیب کرد. شرط از دست دادن خطرات به خروجی تنظیم شده شبکه CNN و برجس ب های نویز  $z$  اعمال شد که به معنای پیش بینی برجس ب های نویز مانند (2015) Sukhbaatar et al. است. تنظیم ریدیابی به دلیل اینکه، همانطور که نویسندهای شان دادند، این کار ماتریس های گیج کننده تخمین زده شده را به ارزش های واقعی شان هدایت می کند، معروف شد.<sup>18</sup> اگر نتیجه، به محض اینکه ماتریس های گیج کننده برجس ب دهنده ها تخمین زده شوند، شبکه CNN یاد می گیرد که برجس ب های صحیح را پیش بینی کند.

این روش برجس ب های صحیح را تخمین نمی زند. ماتریس های گیج کننده تخمین زده می توانند برای تخمین  $y$  و  $p(z_{tt}, y)$  برای هر برجس ب دهنده و همچنین  $\pi$  استفاده شوند. Tanno et al. (2010) EM Raykar et al. (2018) و Khetan et al. (2010) توصیح داده است.

### 3.7 آموزش ناکافی، سارگار با زیرساخت

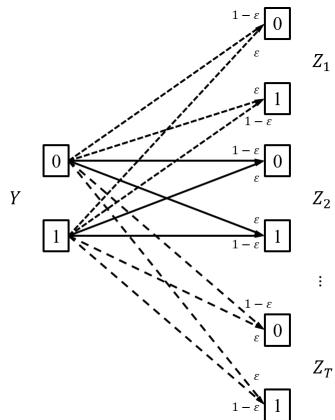
وجود زیرساخت های موجود مهمی وجود دارد، مانند بسته های نرم افزاری یا چارچوب های یادگیری ماشین، که انتظار برجس ب های صحیح یا فرصنی صحیح دارند و تغییر آن ها برای یکی از روش های پیشین ممکن است غیرعملی یا گران قیمت باشد. در ادامه، برخی از روش های آموزشی که مطلوب نیستند امتحان زیرساخت هایی سارگار هستند را توضیح می دهیم.

ابتدا، برآورد برجس ب یک برآورد  $y$  از  $RV$  های برجس ب صحیح  $Y$  تولید می کند و بر روی  $\{x, y\}$  آموزش می بیند به جای  $\{x, y\}$ . برای برآورده  $Y$ ، من توان از کارهای مرتبط استفاده کرد  $MPE$  را مانند  $MPE = \text{کمینه کننده} \min_{y \in Y} \arg \max_{y \in Y} p(y | x)$  داده شده است. براورده  $MPE$  با  $y$  اعمال کرد. براورده  $MPE$  با  $y$  داده است که راه حل MAP استیماتور است (بینید 2) (یا ضمیمه D)، بنابراین  $y = \arg \max_{y \in Y} p(y | x; \theta)$ . تخمین گذاری برجس ب ها نا یهینه است (مقایسه با بخش 3.4.1).<sup>19</sup> زیرا آموزش باید شرط اصلی را تخمین بزند—مثلًا،  $RV$  خطر تجربی، مانند بخش 3.4.1—به جای  $Y$ .

دوم، انتخاب تابع های طبقه بندی  $T$ ، که طبقه بندی  $g-t$  به صورت نمونه ها و برجس ب های نویزدار—که به عنوان ن برجس ب های صحیح در نظر گرفته می شوند—از فقط برجس ب دهنده  $t-th$  t-th، یعنی  $\{x_i, y_i\}_{i=1}^N \neq \emptyset$ ، آموزش دیده است. با توجه به نمونه بدون برجس ب  $x$ ، برجس ب پیش بینی شده  $y$  با رای اکثریت بین  $\{x_i\}_{i=1}^N$  انتخاب می شود. این تکنیک نظریه ای ارزیابی را نادیده می گیرد و به همین دلیل زیر یهینه است و پیچیدگی آموزش و مستقر کردن را به صورت  $T$  افزایش می دهد.

سوم، تکرار نمونه ها، که توسط Raykar و همکاران (2010) پیشنهاد شده است، هر  $x$  را چندین بار کپی می کند، بر اساس  $\{x_i | z_i, y_i\}$  بر روی کپی های آن برجس ب ها را اختصاص می دهد و بر روی کپی ها و برجس ب های آن ها آموزش می دهد. به عنوان مثال، در طبقه بندی دوگانه، اگر  $\{x_i | z_i, y_i\}$  باشد، پس ۵  $x$  بار کپی می شود، با یک کپی که برجس ب ۰ اختصاص داده شده و چهار کپی که برجس ب ۱ اختصاص داده شده است. اجرای مستقیم می تواند نیاز به ذخیره سازی و محاسبات برای آموزش را چندین برابر کند.

Sukhbaatar و همکاران (2015) این نتیجه را دکر کرده اند اما به دلیل دلایل اجرایی، تنظیم متفاوتی برای آن اعمال کرده اند.



تصویر 5: کanal پخنس دوگانه متقارن.

این روش‌ها بین حسابداری مشکلات حقیقت یابی و تغییر زیرساخت‌های موجود، تنازل‌های محتل فی ارائه می‌دهند. تخمین برچسب‌گذاری یک تصمیم سخت در مورد  $Y$  قبل از شروع آموزش می‌گیرد، که می‌تواند برچسب‌های استباهی را به آموزش معرفی کند اما نیازی به تغییرات دیگری ندارد. رای کیری در نظر من گیرد تصمیمات سخت مختلفی در مورد  $Y$  از برچسب‌گذاران فردی، پیچیدگی پیاده سازی را توسط  $T$  چندین برابر می‌کند و نیاز به مکانیزم رای گیری ساده‌ای دارد.

تکرار نمونه‌ها هرگز تصمیم سخت در مورد  $RV$ ‌های برچسب صحیح  $Y$  نمی‌گیرد، تقریب قاتی برای روشن‌های ارائه شده در بخش‌های 3.2، 3.3 و 3.4 ارائه می‌دهد و نیاز به منابع آموزشی را توسط تعداد کمپانی‌ها چندین برابر می‌کند، اما پیچیدگی پیاده سازی را تحت تأثیر قرار نمی‌دهد.

#### 4. مقایسه ترکیبات برچسب‌گذاران: دیدگاه اطلاعاتی

مشکلات حقیقت یابی یک تفسیر اطلاعاتی ساده دارند. برای یک نمونه خاص،  $RV$  برچسب صحیح  $Y$  می‌تواند به عنوان ورودی یک کanal در نظر گرفته شود و  $RV$ ‌های برچسب نویز  $Z = Z_1, \dots, Z_T$  می‌توانند به عنوان خروجی‌های کanal در نظر گرفته شوند. سپس اطلاعات مشترک  $Z^{(I)}$  ( $\sum_{y \in \mathcal{Y}} p_y \log(p_y)$ ) میزان اطلاعاتی که  $Z$  در مورد  $Y$  و بر عکس انتقال می‌دهد را تعیین می‌کند (Bibinić et al., 1991). Thomas Over در اینجا، نشان می‌دهیم که این تفسیر به ما اجازه می‌دهد ترکیبات مختلف برچسب‌گذاران را از نظر اطلاعات مشترک مقایسه کیم، حداقل در تئوری. در بخش 5.5، نشان می‌دهیم که تکنیک‌های آموزش و تست بخش‌های 2 و 3 به ما اجازه می‌دهد این اطلاعات را در عمل استفاده کنیم.

##### 4.1 کanal پخنس دوگانه مشابه

ما این دیدگاه را برای طبقه‌بندی دوگانه توصیح می‌دهیم. برای سادگی، فرض می‌کنیم که برچسب‌گذاران یا  $RV$ ‌های برچسب نویز مستقل شرطی هستند، اگرچه این تفسیر برای برچسب‌گذاران مستقل شرطی نیز اعمال می‌شود. همچنین فرض می‌کنیم که تمام برچسب‌گذاران توزیع شرطی مشابهی دارند و هر برچسب‌گذار به هر نمونه برچسب نویزی اختصاص می‌دهد، بنابراین کافی است به یک نمونه توجه کنیم.

19. در یک زمینه نظری، Lugosi (1992) تفسیر کanal برای یک برچسب‌گذار خاص ارائه داد.

حالا می توانیم یک کانال پخن<sup>ش</sup> دوگانه مشابه (Thomas, 1991, § 14.6 Cover و BSBC) (بینید: عرفی کنیم که در شکل ۵ نشان داده شده و توسط احتمال خطای برچسب گذاری  $\varepsilon \in [0, 1]$  پارامترسازی شده است. برای  $y, z \in t$  و  $T \in \{0, 1\}$  ،

$$p(z_t|y; \varepsilon) = \begin{cases} 1 - \varepsilon, & \text{if } z_t = y; \\ \varepsilon, & \text{if } z_t \neq y. \end{cases} \quad (57)$$

برای  $\varepsilon \in [0, 1/2]$  in (59) مدل معادل تنظیم  $i$  و  $C = 2\varepsilon$ ,  $\delta_i \equiv 0$ ,  $\forall i$  است.

قانون بی نمایی می تواند برای بدست آوردن  $I(Y | T, \varepsilon)$ , اطلاعات مشترک برای  $T$  برچسب دهنده ه، هر کدام با احتمال خطای استفاده شود:

$$\begin{aligned} I(Z; Y | T, \varepsilon) &= \pi(0) \sum_{m=0}^T \binom{T}{m} \varepsilon^m (1-\varepsilon)^{T-m} \log_2 (\varepsilon^m (1-\varepsilon)^{T-m}) \\ &\quad - \sum_{m=0}^T \binom{T}{m} (\pi(0)\varepsilon^m (1-\varepsilon)^{T-m} + \pi(1)(1-\varepsilon)^m \varepsilon^{T-m}) \\ &\quad \cdot \log_2 (\pi(0)\varepsilon^m (1-\varepsilon)^{T-m} + \pi(1)(1-\varepsilon)^m \varepsilon^{T-m}). \end{aligned}$$

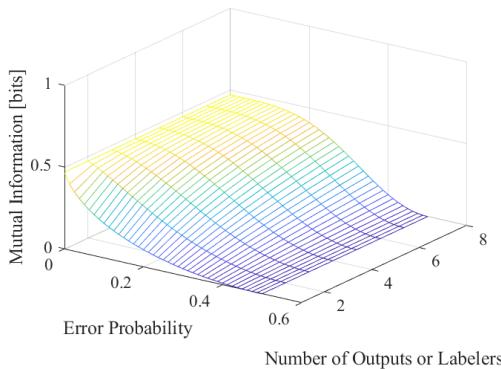
برای یک مقدار ثابت  $\varepsilon \in \{0, 1\}$ ,  $I(Y | T, \varepsilon) = \pi(1) \log_2 \pi(1) - \pi(0) \log_2 \pi(0)$  است. اگر  $\varepsilon = 1/2$  باشد، آنتروپی  $I(Y | T, \varepsilon) = H(Y)$  است. اگر  $\varepsilon = 0$  باشد، آنتروپی  $I(Y | T, \varepsilon) = 0$  است. اگر  $\varepsilon = 1$  باشد، آنتروپی  $I(Y | T, \varepsilon) = H(Y)$  است. اگر  $\varepsilon \in (0, 1/2)$  باشد، محدودیت  $I(Y | T, \varepsilon) = H(Y)$  است. اگر  $\varepsilon \in (1/2, 1)$  باشد، محدودیت  $I(Y | T, \varepsilon) = 0$  است.

شکل 6 برای مقادیر مختلف  $\varepsilon$  و  $T$  وقتی  $\varepsilon = 0.1$  نشان می دهد. نمودار فقط برای  $T \leq 1/2$  نشان داده می شود. به عنوان  $\varepsilon = 1/2$  به صفر کاهش می یابد. برای  $\varepsilon = 0$ ,  $I(Y | T, \varepsilon) = 0$  باشد. از صفر به حداقل افزایش می یابد که انتظار می رود. برای  $\varepsilon = 1$ ,  $I(Y | T, \varepsilon) = H(Y)$  است. اگر  $\varepsilon \in (0, 1)$  باشد، از صفر به  $H(Y)$  افزایش می یابد زیرا تعداد بیشتری از مشاهده های برچسب دهنده نویزدار برای تخمین  $Y$  در دسترس است.

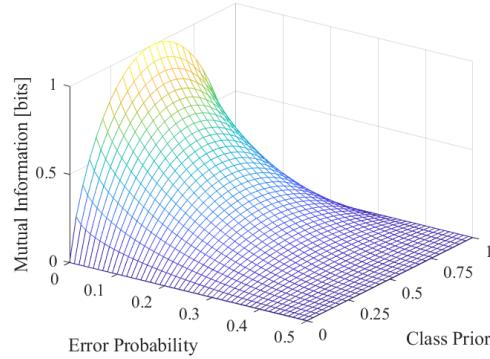
#### 4.2 برچسب دهنده یکسان

ما می توانیم از  $I(Y | T_1, \varepsilon_1)$  و  $I(Y | T_2, \varepsilon_2)$  برای مقایسه اطلاعات ارائه شده توسط دو گروه مختلف برچسب دهنده استفاده کنیم، اما ممکن است به فکر یک برچسب دهنده واحد بیفتد. گذشته از  $I(Y | Z; \varepsilon)$  معرف اطلاعات مشترک برای یک برچسب دهنده با احتمال خطای  $\varepsilon$ ، که با کانال دوگانه استاندارد (وروپی یک، خروجی یک) (BSBC) (بینید: Thomas, 1991, §8.1 Cover و BSBC) مطابقت دارد. داده های  $T, \pi$  و  $\varepsilon$ ، می توانیم از جستجوی دودویی برای پیدا کردن  $\varepsilon$  استفاده کنیم که  $I(Y | Z; \varepsilon) = I(Y | T, \varepsilon)$ . داده های  $T, \pi$  و  $\varepsilon$  برچسب دهنده ها را به عنوان یک برچسب دهنده واحد معادل با احتمال خطای  $\varepsilon$  بیان کنیم. بنابراین، می توانیم تعداد مختلف برچسب دهنده ها با احتمال خطای مختلف را از نظر اطلاعات مشترک برچسب دهنده واحد مقایسه کنیم.

شکل 7 رابطه بین  $\varepsilon$ ,  $\pi$ ,  $I(Y | Z; \varepsilon)$  برای کانال دوگانه استاندارد نشان می دهد. اطلاعات مشترک در  $\varepsilon = 1/2$  متقاضی است. برای یک مقدار ثابت  $\varepsilon$ , با کاهش  $\pi$  از  $1/2$  کاهش می یابد، که نشان می دهد که حل مشکلات حقیقت یابی بیشتر چالش برانگیز می شود تا کلاس ها بیس تر نابرابر شوند.



تصویر 6: نمودار اطلاعات مشترک  $I(Y; Z | T, \epsilon)$  برای BSBC با  $\pi(1) = 0.1$  به عنوان یک تابع از  $T$  و  $\epsilon$ .



تصویر 7: اطلاعات مشترک  $I(Y; Z | \epsilon)$  استاندارد BSC به عنوان یک تابع از پیش‌بینی‌های کلاس  $(1)\pi$  و احتمال خطای  $\epsilon$ .

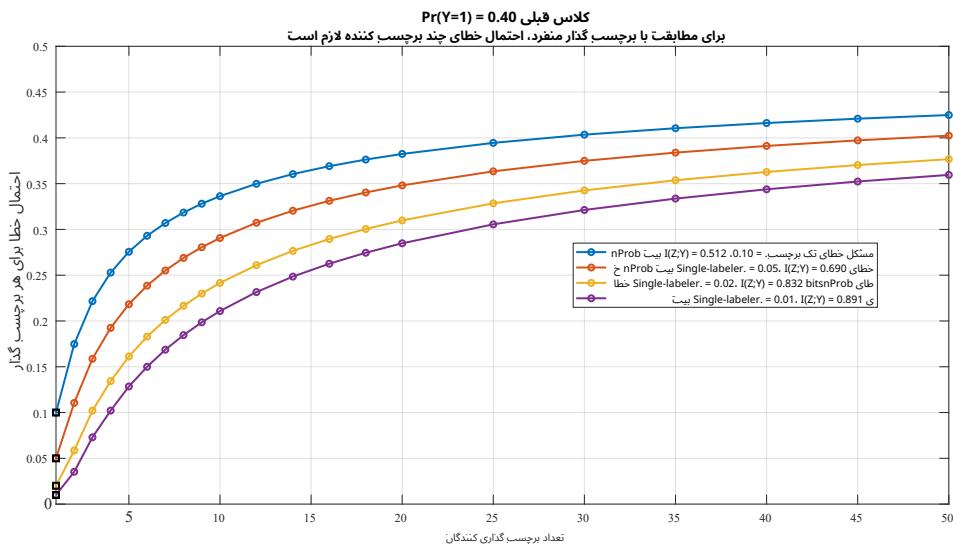
#### 4.3 برجسب‌گذاری کنندگان متوسط چندگانه یا برجسب‌گذاری کننده کارشناسانه یک نفر

تصویر 6 همچنین نشان می‌دهد که برای  $\epsilon < 1/2$ ,  $I(Y; Z | T, \epsilon) > I(Y; Z | \epsilon)$  باشد. بنابراین، برجسب‌گذاری کنندگان متوسط چندگانه می‌توانند به اندازه‌یا بیشتر از یک برجسب‌گذاری کننده کارشناسانه اطلاعاتی داشته باشند. این نتیجه به توجیه و توصیح موفقیت‌های منبع جمعی کمک می‌کند. ما می‌توانیم دوباره از جستجوی دودویی برای پیدا کردن کمترین مقدار  $T$  مورد نیاز برای رضایت  $I(Y; Z | T, \epsilon) \geq I(Y; Z | \epsilon)$  استفاده کنیم.

تصویر 8 نمونه منحنی‌ها برای  $\epsilon = 0.01, 0.02, 0.05, 0.10$  و  $0.4 = \{0.01, 0.02, 0.05, 0.10\}\pi$  را نشان می‌دهد. منحنی‌ها برای مقادیر کوچک  $T$  به طور شدید افزایش می‌یابند و سپس به دلیل کاهش بازدهی کاهش می‌یابند؛ به عنوان  $\epsilon \rightarrow \infty$ ,  $T \rightarrow \infty$ . یک محدوده نزدیک به  $1/2$  می‌رسد. بنابراین، چند برجسب‌گذاری کننده متوسط ممکن است کافی باشد تا یک احتمال خطای معادل کوچک  $\epsilon$  را به دست آوریم. یا اینکه ممکن است فردی بخواهد یک مقدار بسیار کوچک  $\epsilon$  داشته باشد که هیچ کارشناس نمی‌تواند به تنهایی به آن برسد، اما چند برجسب‌گذاری کننده انسانی—not superhuman—همه با هم ممکن است به آن برسند.

این امکان یادآور روش تقویت (boosting) است (بینید Schapire, 1990; Freund و Schapire, 1997)، در آن چند یادگیرنده صنعتی برای دستیابی به عملکرد یک یادگیرنده قوی استفاده می‌شوند. این رابطه را بیشتر بررسی نکرده ایم اما چند نکته کوتاه می‌زنیم. در تقویت، یادگیرنده‌های صنعتی بهتر از حد تصادفی عمل می‌کنند، اما برجسب‌های صحیح شناخته می‌شوند و این اطلاعات برای بهبود عمل کرد استفاده می‌شود. با مشکلات حقیقت یابی، برجسب‌های نویز ممکن است بسیار نادرست باشند و برجسب‌های صحیح مشاهده نشوند، اما توزیع شرطی  $p(tz | ty)$  شناخته می‌شود و این اطلاعات درباره برجسب‌های صحیح برای آموزش و تست ارائه می‌دهد.

معنای نظریه اطلاعات از اطلاعات معادل جالب است، اما این نمی‌تواند توصیح دهد که چگونه از اطلاعات که  $Z$  یا  $Z$  انتقال می‌دهد درباره  $Y$  استفاده کنیم. خوشبختانه، روش‌هایی که در بخش‌های 2 و 3 توسعه یافته اند، راهی برای این کار ارائه می‌دهند. آنها باید با افزایش  $T$  بهتر تخمین‌ها را تولید کنند زیرا واریانس یک برآورد بهینه با افزایش تعداد کاهش می‌یابد.



شکل 8: تعداد برچسب دهنده های مورد نیاز برای دستیابی به اطلاعات متقابل یکسانی مانند singlenlabeler fort  $\pi t(1) = 0t.t4$ .

مشاهدات پر سر و صدا (مستقل) در دسترس هستند.  $T20t$  به اندازه کافی بزرگ است، تخمین ها برای برچسب زنان متوسط باید با تخمین های برچسب گذار متخصص منفرد قابل مقایسه باشند. بخش 5.5 چند آزمایش را ارائه می کند که مفهوم را تأیید می کند.

## 5. آزمایشات

ما تعدادی آزمایش انجام دادیم تا بینیم روش های مختلف آزمایش و آموزش چگونه عمل می کنند و مفهوم اطلاعات متقابل معادل را برای ترکیب های مختلف برچسب ها بررسی کنیم.

### 5.1 شبیه سازی

برای اعمال روش های تست، به برچسب های صحیح، پر سر و صدا و پیس بینی شده نیاز داریم. برای مطالعه روش های آموزش فقط به برچسب های صحیح و پر سر و صدا نیاز داریم. همه اینها می توانند از طریق شبیه سازی تولید شوند. البته، روش ها از برچسب های درستی استفاده نمی کنند، که در برنامه های مورد نظر قابل دسترس نیستند، اما شبیه سازی به ما امکان می دهد نتایج را با حالت ایده آل مقایسه کنیم.

#### 5.1.1 شبیه سازی برچسب های صحیح و پیس بینی شده

مدل سازی برای پیس بینی کننده مورد نظر کلاس π ساده است. برای طبقه بندی دوگانه، نمونه های برجسب گذاری صحیح  $\pi^{(0)}$  مستقل از توزیع  $B(\pi)$  انتخاب می شوند. برای طبقه بندی چندگانه، آنها

2. این رفتار به 740 مریوط است، تعداد برچسب های نویز برای هر نمونه، به  $\mathcal{N}$ ، تعداد نمونه ها. افزایش  $T^*$  به معنای اینکه تعداد بیشتری از مشاهده های نویز برای  $\gamma$  در دسترس قرار می گیرد تا عدم قطعیت در مورد برجسب صحیح یک نمونه کاهش یابد. صرفاً افزایش  $N$  نمی تواند مشاهده بیشتری برای نمونه های قبلی فراهم کند.

مستقل از توزیع دسته‌ای با پیش‌بینی کننده  $\pi$  و دسته‌ها  $0, 1, \dots, C-1$ . جنین توزیع را  $Cat(\pi, 0, 1, \dots, C-1)$  می‌نامند.

آزمایی‌ها بر روی روش‌های تست بخنس 2 نیاز به برچسب‌های پیش‌بینی شده نمونه سازی شده دارند. پیش‌بینی شده - (ptdesDr. t ptdesFA) برای نمونه سازی یک طبقه بندی دوگانه با نقطه عملکرد مورد نظر

برچسب  $\hat{tyt}$  به طور مستقل ترسیم می شود، با  $\hat{tytit}$  ترسیم شده  $B(tptdes)$  و  $FAt(0)tit = 0$  ترسیم شده  $1 = tit(0)tft$  به همین ترتیب، عملکرد یک طبقه بندی کننده چند کلاسه را می توان با یک ماتریس سردرگمی شرطی مورد نظر  $Ktdest$ ، که در آن عنصر ماتریس  $|t' t|$   $Ktdestnt$  حاولی  $t' t | nt$  نیست، مشخص کرد. بنابراین، هر  $\hat{tytit}$  با برچسب پیش بینی شده به طور مستقل ترسیم می شود.

#### 5.1.2 مدل سازی و شبیه سازی پرچسب های نویزدار

از آنجایی که نمونه‌ها مستقل هستند، هنگام توصیف مدل ساری و شبیه ساری RVs دارای برچسب نویز، نمونه را به طور موقع ایندکس می‌کنیم. تا این مرحله،  $t$  pt(tzt; t ψt) کاملاً عمومی باقی مانده است، بنابراین برچسب‌های واپسی به شرطی را مجاز می‌کند. یعنی واپسی بین Ztt و RVst Zttt با برچسب نویزدار مختلف برای یک نمونه. ما همچنین ψunspecified را گذاشت tt هایم. برای شبیه ساری، به یک مدل ملموس تر نیاز داریم.

اول، ما فرصت می‌کنیم که RV های دارای برچسب نویز برای یک نمونه به طور مشروط مستقل هستند:

$$p(\mathbf{z}|y; \boldsymbol{\psi}) = \prod_{t: z_t \neq \emptyset} p(z_t|y; \psi_t), \quad (58)$$

جایی که  $\psi_{tttTt}$  حاوی پارامترهایی است که تعیین می‌کنند برچسب گذار چگونه  $ttth$  را در نمونه نشان می‌دهد. این فرصت در درجه اول برای سهولت اجرایی است. ما می‌توانیم از مدل با برچسب‌های وابسته مانند یکی از مدل‌های Holodnak و همکاران استفاده کنیم. (2018).

دوام، ما تعریف می‌کنیم که  $\phi = \psi_{\delta},$  جایی که  $\delta = [0, 1]$  نمونه دشواری است و  $\phi \in [0, 1]$  خطا را برچسب دهنده است. نمونه دشواری، ایهام موجود در برچسب صحیح نمونه خود را نشان می‌دهد. به عنوان مثال، در تشخیص تصویر، می‌تواند میزان تار یا نویز را نشان دهد. خطا برچسب دهنده، ویرگی برچسب دهنده است؛ می‌تواند تجربه یک تحلیل گر یا اعتماد یک شرکت کننده در منابع جمعی را نشان دهد.

مدل پس از آن است

$$p(z|y; \delta, \phi_t) = \begin{cases} 1 - \varepsilon(\delta, \phi_t), & \text{if } z = y \in \mathcal{Y}; \\ \frac{\varepsilon(\delta, \phi_t)}{C-1}, & \text{if } z \neq y \in \mathcal{Y}, z \in \mathcal{Y}; \\ 0, & \text{if } z \notin \mathcal{Y}; \end{cases} \quad (59)$$

جایی که احتمال خطای برچسب گذاری  $t(t\delta, \phi_{ttt})$  یکتابع دوخطی است:

$$\varepsilon(\delta, \phi_t) = (\delta - \delta\phi_t + \phi_t) \frac{C-1}{C}, \quad 0 \leq \delta \leq 1, 0 \leq \phi_t \leq 1. \quad (60)$$

برچسب گذار Zttt را به عنوان برچسب صحیح با احتمال  $\text{et}(t\delta, \phi_{ttt})$  در نظر می‌گیریم. اگر استیاهی رخ دهد،  $Z_{ttt}$  احتمالاً به طور مساوی بر روی برچسب  $t [0, t]$  است. اگر  $\delta$  توزیع  $C_t$  را نادست احتمالی توزیع می‌شود، آنگاه  $\text{et}(t\delta, \phi_{ttt}) = 0$ . اما اگر هریک از پارامترها غیر صفر باشد،  $\text{et}(t\delta, \phi_{ttt})$  با هر دو  $t$  و  $\delta$  افزایش می‌یابد. اگر یکی از  $t$  و  $\delta$  صفر باشد،  $\text{et}(t\delta, \phi_{ttt}) = 1$  است. بنابراین  $Z_{ttt}$  به طور مساوی بر روی  $t$  توزیع می‌شود.

مدل ما شبیه مدل Whitehill و همکاران است. (2009)، اما فرمول ما کمی متفاوت است، و هدف ما به طور قابل توجهی متفاوت است. مدل آنها برای طبقه بندی باینزی است، در حالی که مدل ما برای Ct دلخواه اعمال می شود. مدل آن ها به برچسب زن های متخصص اجازه می دهد تا بر عکس برج سب دودویی درست را انتخاب کنند، اما وقتی  $t > C$ ، «بر عکس» برچسب the incorrect مشخص نیست. مهمتر از آن، Whitehill و همکاران، بر تخمین پارامترهای مدل متمرکز است، در حالی که ما بر روی یهود برداری از مدل زمانی که پارامترهای آن شناخته شده است تمکز می کنیم. در مورد دوم، یک برج سب زن متخصص که به عنوان متضاد برچسب صحیح شناخته می شود، در واقع آموزندگی تر از برچسب زنی نامرتب است که برچسب ها را احتمالاً به طور تصادفی اختصاص می دهد.<sup>21</sup>

در نهایت، دوباره همه نمونه ها را در نظر می گیریم و سپس  $\text{lett}$  را در  $\text{t} \Delta t$  و  $\text{t Nit} = 1$  اندت  $\phi_t = (\text{t} \phi \text{tttt}) \text{t} \text{Ttt} = 1$  اندت  $\psi_t = (\text{t} \phi \text{tit}, \text{t} \phi \text{t})$  بنا نماییم.

$$p(\underline{\mathbf{z}}|\mathbf{y}; \underline{\boldsymbol{\psi}}) = \prod_{i=1}^N p(\mathbf{z}_i|y_i; \delta_i, \boldsymbol{\phi}) \\ \stackrel{(a)}{=} \prod_{i=1}^N \prod_{t: z_{i,t} \neq \emptyset} p(z_{i,t}|y; \delta_i, \phi_t), \quad (61)$$

حاب، که (tat) از (58) است.

در طول شبیه سازی،  $t_{Nt}$ ,  $\text{and } \phi_{ttt,t} = 1, \dots, t_{Tt}, \dots$  را ترسیم کردیم.  $\delta t_{it,t}$  را مساقط از توزیع بتای یکنواخت. ما همچنین یک بردار  $t_{ttt} = (t_{tttt})t_{Ttt} = 1t$  رسم می کنیم، که در آن  $\eta_{ttt}$  احتمال تقریبی است که برچسب  $tttt$  یک برچسب برای یک نمونه ارائه می دهد. هر  $ttt$  مساقط از  $U(t_{ta}, bt)$  ترسیم می شود که در آن  $U(t_{ta}, bt) = U(t_{ta}, bt)$  نشان دهنده توزیع یکنواخت را دارد. برای اولین نمونه، ما به طور مکرر  $\eta_{tit,t} = (tB(t_{tttt}))t_{Ttt} = 1t$  را با عناصر توزیع  $t_{ta}, bt$  برونوی مساقط ترسیم کردیم تا زمانی که حداقل یک عنصر اغلب  $\eta_{tit,t}$  برابر با واحد باشد. سپس  $z_{ti,tt}$  را مطابق با  $(59)$  و  $(60)$  شبیه سازی می کنیم.

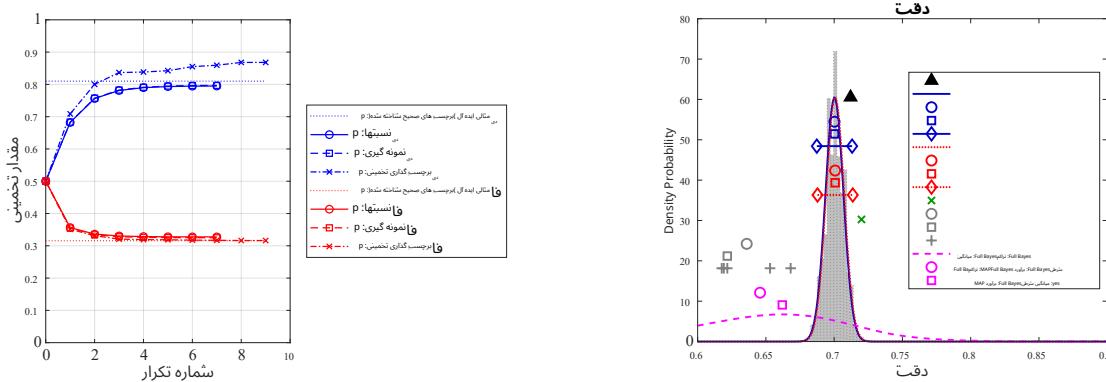
### 5.2 نمونه های از آزمایش: طبقه بندی، یافته

نمونه های مختلفی که جنبه های روش های آزمایش برای طبقه بندی باینزی را نشان می دهند در اینجا طا  
هر می شوند. ما تأکید می کنیم که به جز حالت ایده آل، در طول آزمایش از هیچ برچسب صحیح استفاده  
نشده است.

مثا ، اصل ، تست

تصویر 9 پیشرفت تخمین ها  $\hat{y}_p$  و  $\hat{y}_{p_{\text{des}}}$  برای روش های تکراری را نشان می دهد. خط های نقطه دار مقادیر ایده آل  $y^*$  و  $F_k$  را نشان می دهند اگر برچسب های صحیح شناخته می شدند؛ چون آنها نتیجه نمونه کیری و شبیه سازی هستند. (هفقاری ایده آل کم از  $p$ ، بسیار تسبیح MMSE، روش های بیاسی تحریری) الگوریتم های 1 و 2.

Ipeirotis و همکاران (2010). 35 (مسایه این مشاهده را در زمینه ارزیابی هزینه یک پرچسب گذار انجام دادند.



تصویر 9: مثال اصلی تست: تخمین دقت تو  
سط همه روش‌ها.  
 $FAP_p$  در روش‌های تخمین مکر  
ری.

تصویر 10: مثال اصلی تست: تخمین دقت تو  
سط همه روش‌ها.

نیازمند است که صورت عمودی فاصله داده شده اند  
تا خوانایی راحت‌تری داشته باشد.

تخمین‌ها در هر آنتیراسیون بهبود یافتد و پس از هفت آنتیراسیون، هر دو به تقریباً همان پارامترهای نزدیک شدند. روش ناقص‌ال-economic تخمین برچسب‌های صحیح (الگوریتم 3) پس از نه آنتیراسیون به نتیجه رسید؛ پارامتر نهایی  $OP_{\text{آن}}(\sim DP, \sim FAP)$  کمی دقیق‌تر از روش‌های Bayes تجربی بود، اما پارامتر نهایی  $OP_{\text{آن}}(\sim DP)$  به طور قابل توجهی کم دقیق‌تر از آن‌ها بود.

جدول 10 تخمین‌های همه ساختنیک‌های ارائه شده در بخش 2 را جمع‌بندی می‌کند و تصویر 10 نتایج دقت را نشان می‌دهد. نتایج روش‌های ناقص‌ال-economic که برچسب‌های پیش‌بینی شده را به طور کامل بهره برداری نمی‌کنند در بخش 5.2.2 آورده شده است. برای تست Bayes با روش تجربی ("نسبتی‌ها" و "نمونه گیری")، تصویر تخمین‌های احتمال پس زمینه، میانگین شرطی، تخمین MAP و محدوده‌های 95%-قابل اعتماد برای ACC را نشان می‌دهد. تصویر همچنین یک هیستوگرام حاکستری را نشان می‌دهد که با گرفتن پارامترهای نهایی  $OP_{\text{آن}}(\sim DP, \sim FAP)$  از الگوریتم 2، تولید 5000 نمونه گیری از ۷ از 10)، و محاسبه دقت تجربی برای هر نمونه ایجاد شده است. تخمین‌های بهینه تقریباً به دقت ایده‌آل نزدیک هستند و محدوده‌های قابل اعتماد شامل دقت ایده‌آل هستند. نتایج دو روش بسیار مشابه هستند؛ این رفتار برای همه مثال‌ها مشاهده شده است، بنابراین تصاویر بعدی نتایج الگوریتم 2 را فراتر از هیستوگرام ها شامل نمی‌کنند.

تصویر همچنین تخمین دقت روش ناقص‌ال-economic در الگوریتم 3 ("تخمین برچسب‌های تجربی") را نشان می‌دهد. این تخمین نسبتاً خوب است، اما مانند محدوده‌های قابل اعتماد هیچ احساسی از عدم قطعیت ارائه نمی‌دهد. همچنین، این الگوریتم تخمین‌های نادرستی از REC یا DP در تصاویر بعدی 11 و 12 نشان داده من شود (تولید من کند).

همچنین، تصویر تخمین‌های T برای هر برچسب دهنده‌ی (单独) "برچسب دهنده‌ی" (را نشان می‌دهد. ما همچنین میانگین و مدیان این تخمین‌ها را محاسبه کرده‌ایم؛ ما الگوریتم وردی و زانک (2000) را برای محاسبه مدیان چند بعدی اعمال کرده‌ایم. این تخمین‌ها نسبتاً کم دقت هستند، بنابراین میانگین و مدیان آن‌ها نیز تخمین‌های بدی ارائه می‌دهند.

آخرین مجموعه نتایج در تصویر برای روش به طور کامل Bayes است. تخمین‌های نزدیک آن به دلیل محدودیت بر  $FAP, DP$  (بسیار پراکنده هستند و میانگین شرطی آن

	مقیاس های ساده							
	ACC	PREC	$\frac{D}{REC}$	PtFA	F1	$(PREC, REC) \cap (PRE, REC)$	$(PDPFA)$	
مطلوب(برچسب های صحیح شناخته شده)								
مثالی	0.712	0.421	0.810	0.316	0.554	(0.421, 0.810)	(0.810, 0.316)	
تست MMSE: تخمین تجربی بیز از طریق نسبت RV های معمولی مشترک (Alg. 1)								
میانگین مشروط 0.795, 0.328) معتبر پایانی (تر) (بالایی) منطقه	0.700 tat 0.asahta 0.687 0.373 0.713 0.420	0.795 0.795 0.764	0.328 0.325 0.317	0.534 0.529 0.506	- - -	(TbnMAP (0.397, 0.795)tbt (0.397, 0.795)tct(0.796, 0.325)tcn	برآورد	
آزمایش MMSE: تخمین تجربی بیز از طریق نمونه گیری (Alg. 2)								
میانگین مشروط P تخمین ن ترین (قابل اعتماد (بالایی) منطقه	asahta 0.701 0.688 0.374 0.714 0.421	0.796 0.796 0.765	0.327 0.325 0.316	0.535 0.530 0.506	- - -	(0.796, 0.327)MA((0.397, 0.796) (0.796, 0.325)((0.397, 0.796))پایی	برآورد	
تخمین برچسب های صحیح (الگوریتم 3)								
تخمین برچسب ها	0.720	0.402	0.868	0.316	0.550	(0.402, 0.868)	(0.868, 0.316)	
ادغام معیارها از برچسب گذاری های فردی								
میانگین میانگین	0.636 0.431 0.622 0.421	0.615 0.605	0.354 0.354	0.506 0.504	- -	(0.431, 0.615) (0.421, 0.605)	(0.615, 0.354) (0.605, 0.354)	
برچسب گذار تعداد برچسب ها	1 727 2 879 3 272 4 667 5 229	0.622 0.653 0.618 0.619 0.668	0.450 0.414 0.405 0.421 0.465	0.574 0.643 0.605 0.587 0.667	0.354 0.343 0.377 0.366 0.331	0.505 0.504 0.485 0.490 0.548	(0.450, 0.574) (0.414, 0.643) (0.405, 0.605) (0.421, 0.587) (0.465, 0.667)	(0.574, 0.354) (0.643, 0.343) (0.605, 0.377) (0.587, 0.366) (0.667, 0.331)
تخمین کاملاً بیانی								
میانگین مشروط آماره تخمین MAP	0.646 0.345 0.662 0.342	0.664 0.714	0.360 0.349	0.456 0.464	- -	(0.345, 0.664) (0.347, 0.690)	(0.664, 0.360) (0.663, 0.360)	

برای دقت یا دقیق، میانگین شرطی بیز تجربی و برآورد MAP یکسان هستند.

میانگین های مشروط برای معیارهای مشترک با معیارهای اسکالار مربوطه یکسان است.

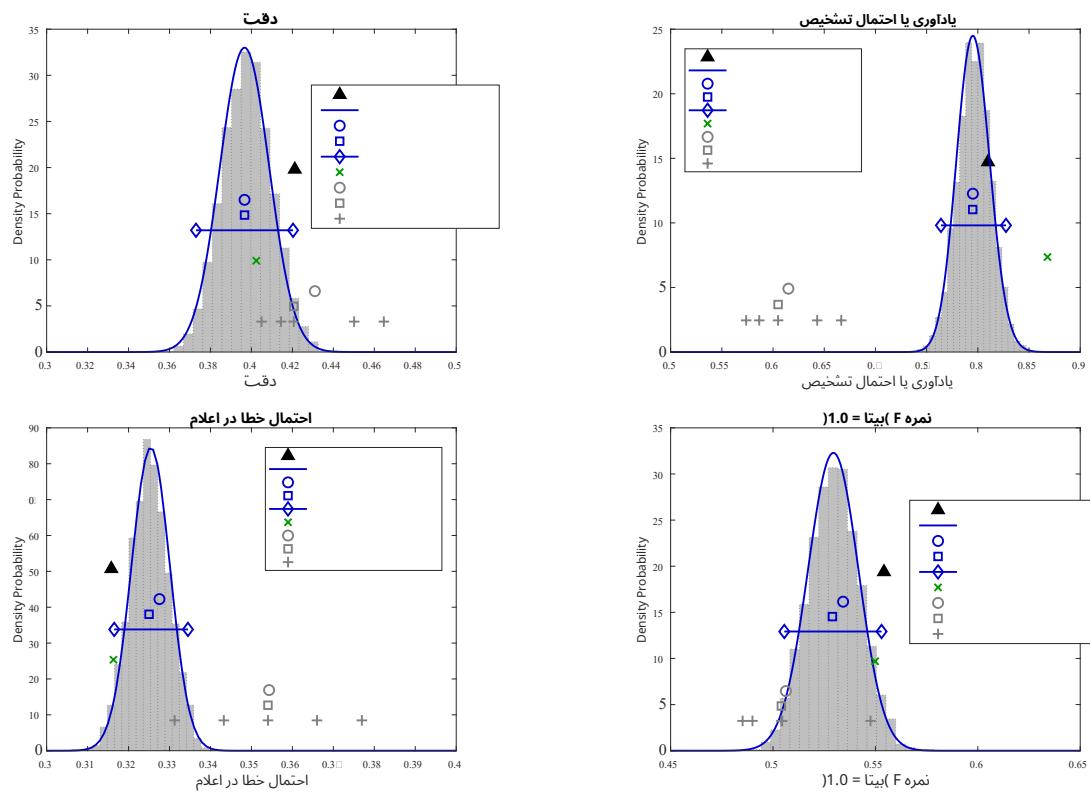
تخمین MAP یک متريک مشترک می تواند با تخمین MAP اجزای جداگانه آن متفاوت باشد.

#### جدول 10: معیارهای آزمایش برای نمونه آزمایش اصلی.

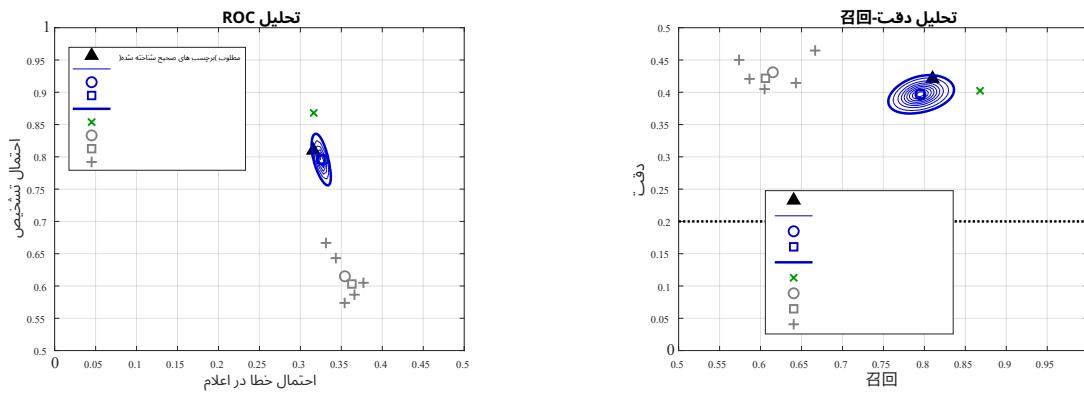
برآورد میانگین و MAP چندان دقیق نیست. این رفتار برای سایر معیارها رخ داده است، بنابراین ما روشن کاملاً بیزی را در شکل های بعدی نشان نمی دهیم.

در مرحله بعد، شکل 11 نتایج را برای سایر معیارهای اسکالار نشان می دهد. آزمایش MMSE با اس تقاده از روش تجربی بیز الگوریتم 1 تخمین هایی را در حدود 0.025 از هر اندازه سنجه تولید کرد. Fort R. ECt یا PtDt یا REct ناحیه معتبر حاوی معیار ایده آل است و برای سایر معیارها، تنها با 0.001 در خارج از منطقه معتبر قرار دارد. جدول 10 نشان می دهد که مناطق معتبر الگوریتم 2 حاوی هر معیار ایده آل هستند. روش غیررهیمه تخمین برچسب های صحیح (الگوریتم 3) برای چندین معیار کاملاً دقیق بود اما برای 0.058 کاهش داشت.

استفاده از برچسب های برچسب گذار اغلب تخمین های بسیار نادرست ایجاد می کند.



تصویر 11: مثال اصلی تست: برآورد مقیاس های متغیر عدد صحیح.



تصویر 12: نمونه اصلی تست: برآوردهای متغیرهای RV مشترک برای تحلیل ROC و P-R.

تصویر 12 نتایج تحلیل ROC و MMSE با برآورد بیایس تجربی بهترین برآوردهای مشترک را تولید کرد؛ برآوردهای نقطه‌ای آن به نزدیک ترین نقاط عملیاتی ایده‌آل است و مح دوده های ۹۵%-قابل اعتماد آن شامل نقاط عملیاتی ایده‌آل می‌شود. برآورد برچسب‌های صحیح نتایج مشترک کم دقتی را به همراه داشت و همچنان هیچ احساسی از عدم قطعیت ارائه نمی‌دهد. استفاده از برچسب‌های برچسب دهنگان نتایج بسیار بدی دارد.

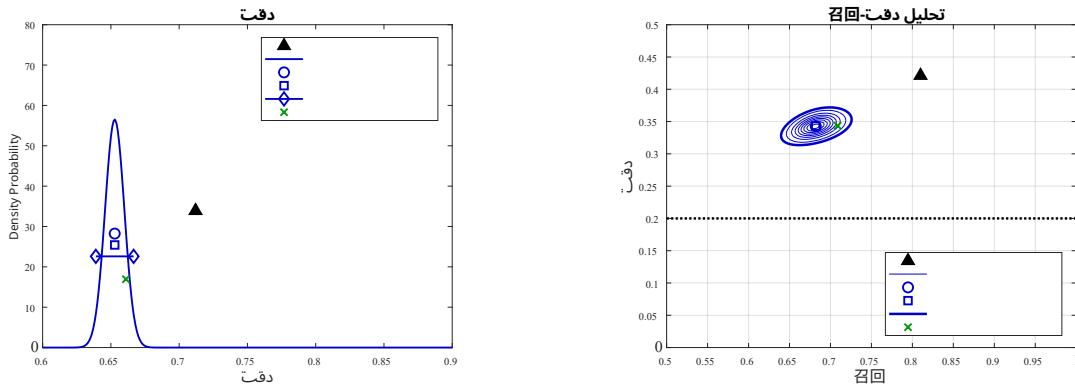
#### 5.2.2 استفاده از پرچسب های پیش بینی شده

استفاده کامل از برچسب های پیش بینی شده  $\hat{y}$  و پارامترهای  $OP(\sim Dp^{\sim} FAp)$  ویزگی مهمی از موفق ترین روش های تست است. تصویر ۳ نتایج شبیه سازی در بخش ۱.۲.۵ را نشان می دهد که در صورت استفاده کامل نشدن برچسب های پیش بینی شده، که معادل روش های تست زیر اپتیمال در بخش ۱.۶.۲ است، آنها از  $p(z_i \neq \hat{y}_i)$  استفاده می کنند به جای مدل تست  $(V(\hat{p}) - Dp^{\sim} FAp)$  با  $p(z_i \neq \hat{y}_i)$ ، که به معنای تنظیم  $z = \max$  در الگوریتم ۱، ۲ یا ۳ است. در مقایسه با تصاویر ۱ و ۲۱، احتمالات پیش زمینه و برآوردها بسیار نادرست هستند و محدوده های قابل اعتماد شامل مقیاس های ایده آل نیستند. این نتایج اهمیت افزودن برچسب های پیش بینی شده و پارامترهای  $OP$  در طول برآورد را نشان می دهد.

### آزمایش های نزدیک شدن برای الگوریتم های تکراری

بخش 2.4.4 پیش‌بینی کرده است که، هنگام تخمین ( $\hat{F}_D$ ) در تست MMSE، روش‌های بیايس تجربی (الگوریتم‌های 1 و 2) به اوح جهانی نزدیک می‌شوند، بدون توجه به پارامترهای اولیه OP. برای بررسی این امکان، ما دو آزمایش انجام دادیم که همچنین شامل روش تخمین برچسب ناکافی (الگوریتم 3) بود.

بر روی شبکه  $10 \times 10 \times 15.0.05.0 \{ .05.0 ..15.0 \} \times 95.0$  مطابق با جدول 11 نزد تایخ را خلاصه می‌کند. روش‌های بیایس تجربی هر بار به تقریباً همان پارامترهای نهایی OP مشابهی نزد یک می‌شدند، با خطاهای مطلق حداقل حدود 0.015. علاوه بر این، MAC همیشه در الگوریتم 1 رضایت داده می‌شد. روش ناکافی الگوریتم 3 خطاهای pD بزرگتری نسبت به روش‌های بیایس تجربی داشت اما خطاهای pFA کوچکتری داشت. برای این



تصویر 13: مثال هایی از برآوردهای متغیرهای مختصری اگر برچسب های پیش بینی شده  $\hat{y}$  و پارامترهای  $(\alpha_{OP}, \beta_{OP})$ ,  $(\alpha_{DFA}, \beta_{DFA})$  به طور کامل استفاده نشوند. مقایسه کنید با تصاویر 10 و 12. نمودارها نتایج برچسب گذاران را نشان نمی دهند زیرا آنها تغییر نکرده اند.

نقطه عمل یکسان	برچسب های تخمین		
	نمونه گیری	نسبتها	
مک رتبه مطلق	0.0149	0.0131	-0.0473
	$3.20 \times 10^{-4}$	$4.15 \times 10^{-4}$	$1.61 \times 10^{-2}$
	0.0153	0.0142	0.0715
$Std. dev.$	-0.0118	-0.0092	-0.0015
	$6.34 \times 10^{-5}$	$8.56 \times 10^{-5}$	$1.49 \times 10^{-3}$
	0.0119	0.0095	0.0033
میانگین معیار انحراف مکس	6.6	6.6	6.2
	1.08	1.05	2.35
	8	8	10

جدول 11: آمار خطای برآورد برای پارامترهای نهایی  $(\alpha_{OP}, \beta_{OP})$  برای روش های برآورد مکرر، برای همان نقطه ایده آل عملکرد  $(\alpha_{DFA}, \beta_{DFA})$  از 100 پارامتر اولیه نقطه ایده آل علیع ملکرد مختلف استفاده سد.

نقطه ایده آل عملکرد، همه الگوریتم ها به طور مداوم به تقریباً همان پارامترهای نهایی نقطه ایده آل عملکرد نزدیک شدند.

دوم، ما مجموعه دیگری از شیوه سازی ها را اجرا کردیم که همیشه الگوریتم های مکرر را با پارامترها  $(\alpha_{OP}, \beta_{OP})$  آغاز می کردیم، اما نقطه ایده آل اولیه پیش فرض  $(\alpha_0, \beta_0)$  را در همان شبکه  $10 \times 10$  مانند بالا تغییر می دادیم، که 100 نقطه ایده آل عملکرد مولکرد نظر  $p_{des}$  را نمایش می دهد. این شیوه سازی ها نتیجه  $\pi(0, 1) \sim \pi(0.5, 0.5)$  را دارند. نتایج در جدول 12 نشان داده می شود. روش های Bayes تجربی همیشه به نتیجه رسیدند، با خطاهای متوسط نزدیک به 0.012 و خطای مطلق حداقل کمتر از 0.045. دوباره، در الگوریتم 1 همیشه MAC رصانیت داده شد. روش ناکافی الگوریتم 3 بهتر عمل نکرد. اگرچه خطاهای متوسط آن بیشتر از 0.021 نیستند، اما انحرافات معیار آن در حدود 0.200 و خطای مطلق حداقل کثراً آن بیشتر از 0.333 است.

نقشه های ایده آل مختلف عملکرد.		نسبتها	نمونه گیری	برچسب های تخمین
		-0.0113 $1.04 \times 10^{-2}$	-0.0114 $9.56 \times 10^{-3}$	-0.0141 $1.95 \times 10^{-1}$
	مک رتبیت مطلق	0.0314	0.0310	0.3590
Error $pFA - pFA$	معیار انحراف	0.0120 $1.36 \times 10^{-2}$	0.0122 $1.14 \times 10^{-2}$	0.0210 $2.03 \times 10^{-1}$
	مک رتبیت مطلق	0.0442	0.0381	0.3369
Number of Iterations	میانگین	10.9	11.2	10.4
	Standard Deviation	3.15	3.46	3.39
	مکس	17	18	21

جدول 12: آمار خطای برآورد برای پارامترهای نهایی  $OP(\sim_{Dp}^{\sim} \mathcal{F})$  برای روس های برآورد تکراری. برای 100 نقطه عملیاتی ایده آل مختلف، از پارامترهای اولیه پیش فرض  $OP$  استفاده شده است

این نشان می دهد که اغلب در نزدیکی یک اوح محلی گیر می کند. این نتایج مردمی ای قابل توجه روش ها  
ی بیاپس تجربی را نسبت به برآورد برچسب های صحیح نشان می دهد.

#### 5.2.4 عملکرد برآورد برای نقاط عملیاتی مختلف

آزمایش‌های نزدیکی در بخش قبلی تنها خطای برآورد پارامترهای نهایی  $(OP^{\sim}, DP^{\sim}, FA^{\sim})$  را بررسی می‌کند. برای مجموعه دوم از شبیه سازی‌ها در آن بخش، ما همچنین آمارهای خطای برآورد نهایی مقیاس  $Hes_{des(p)}(95.0, 15.0, 0.05.0, 15.0 \times 10 \times 10)$  را بر روی شبکه  $\{0.0, 0.05, 0.1, 0.15, 0.2\}$  بررسی می‌کنیم.

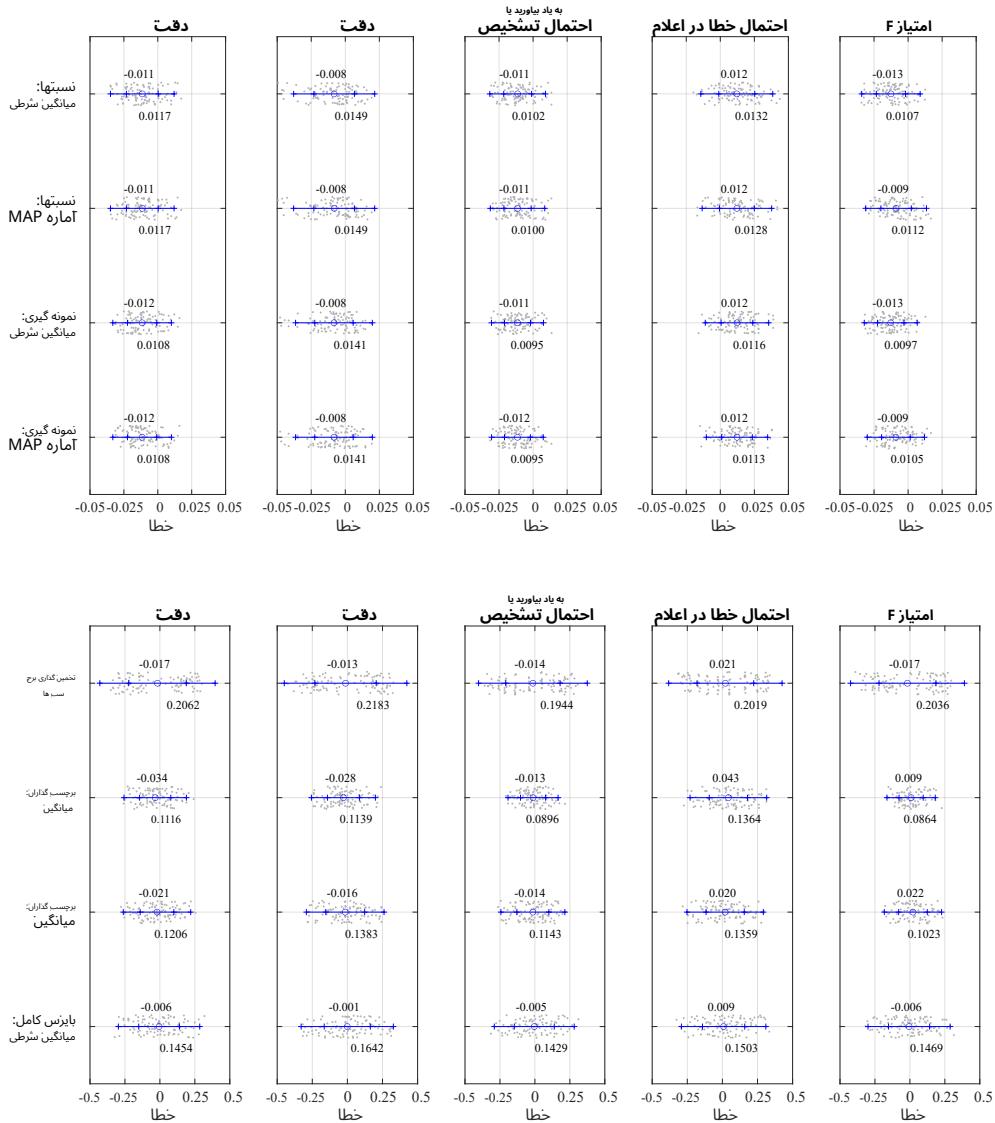
یعنی، برای 100 نقطه عملیاتی ایده‌آل مختلف.

تصویر 14 آمار خطای مقیاس های عددی برآورد شده توسط روش های مختلف آزمایش را خلاصه می کند. برای آزمایش MMSE، برآورد نقطه ای روش های بیانیس تجربی (الگوریتم های 1 و 2) خطاهای متوسطی حدود 0.012 با انحراف معیار حدود 0.011 دارند. تغییر مقیاس بین آزمایش MMSE و روش های دیگر روش های دیگر خطاهای متوسطی بین 0.001 و 0.043 دارند، اما انحراف معیارهای آنها از حدود 0.100 تا بیشتر از 0.200 متغیر است، که یک مرتبه بزرگتر از روش های بیانیس تجربی است.

تصاویر 15 و 16 آمار خطای مشترک برای تحلیل P-R و ROC را نشان می‌دهند. نتایج برای آزمایش MMSE برای برآورد میانگین شرطی و MAP مسأله بودند، با این تصور تصاویر تنها خطای برآورد میانگین شرطی الگوریتم 1 و برآورد MAP الگوریتم 2 را نشان می‌دهند. روش‌های بیانی تجربی خطاهای متوسط حدود 0.015 در هر بعد دارند. ریشه دوم اعداد ویژه ماتریس واریانس خطای بین 0.025 و 0.025 قرار دارد. تغییر مقیاس بین آزمایش MMSE و روش‌های دیگر.

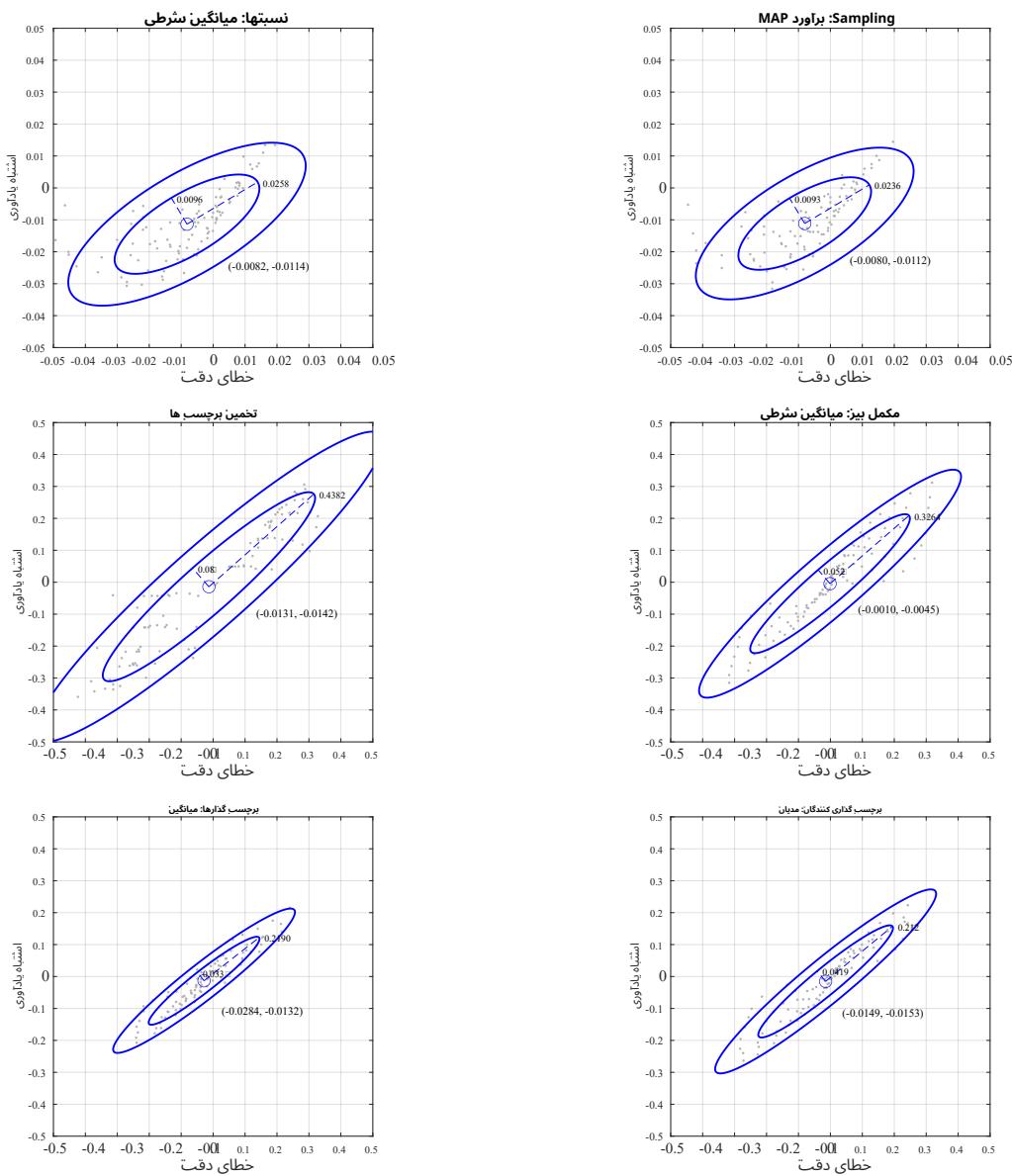
روش‌های بعدی خطاهای متوسطی بین 0.005 و 0.043 دارند، اما ریشه دوم اعداد ویره بین حدود 0.10 تا بیشتر از 0.430 متغیر است.

این نتایج عملکرد برآورد برتر روس های MMSE آزمایش را نشان می دهد، که برآوردهاییس<sup>۳</sup> معمولاً د ۰.۰۱۰ تا ۰.۰۴۰ از مقیاس های ایده آل در طیف وسیعی از نقاط عملیاتی ایده آل قرار دارند. آنها به طور قابل توجهی از روس های دیگر برتر هستند.

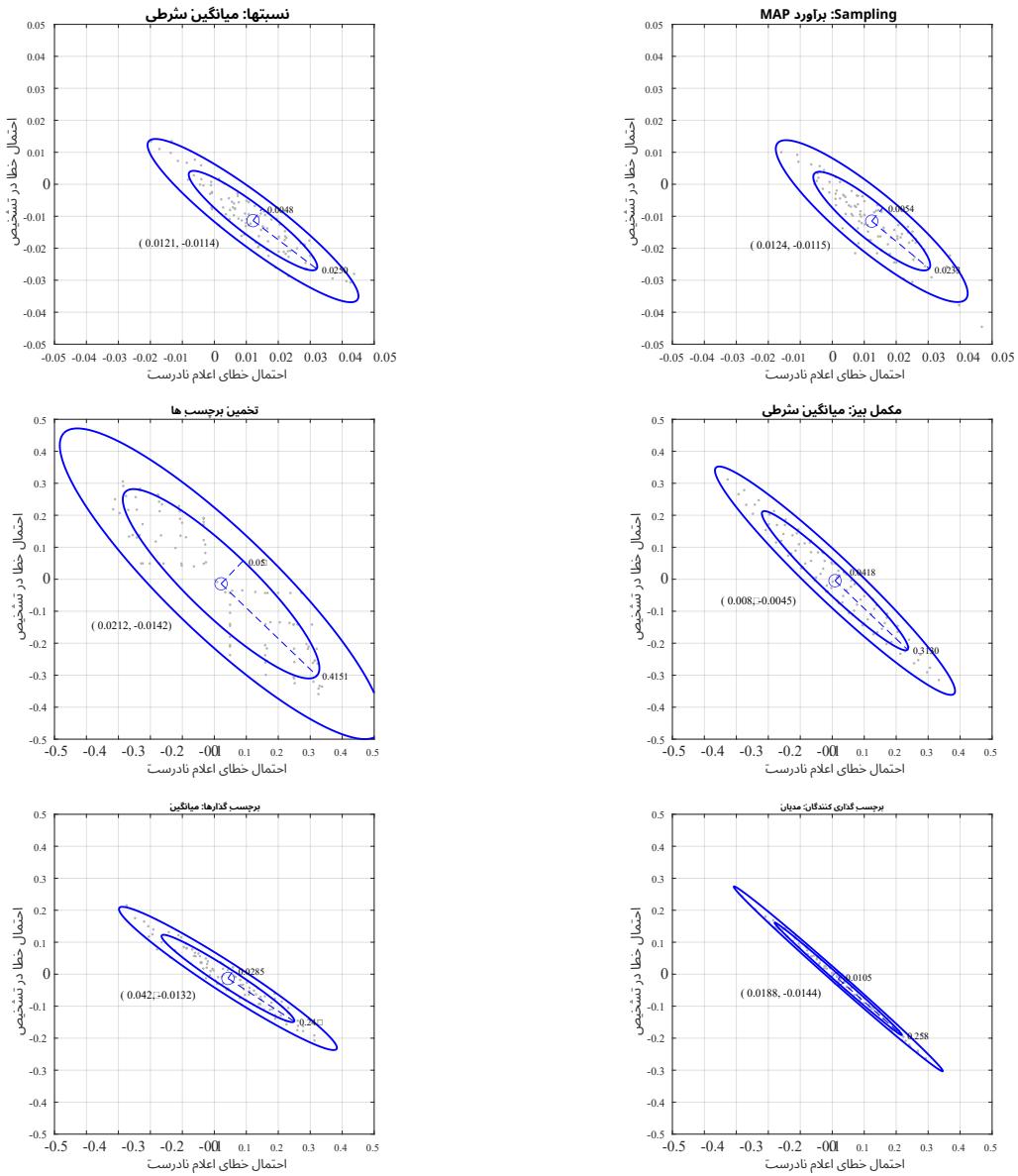


تصویر 14: خطاهای تخمین معیار اسکالر برای روش‌های مختلف تست بر روی 100 نقطه عملکرد ایده آ ل مختلف. محدوده های محور برای تست Bayes MMSE با تجربی (نقشه های بالا: منفی 0.05 تا مثبت 0.05) و روش های دیگر (نقشه های پایین: منفی 0.5 تا مثبت 0.5). نقشه های پراکنس کوچک خطاهای تخمین به عنوان نقطه های خاکس نتی طاهر می شوند. خطای متوسط با دایره و به عنوان متن بالای هر دایره مشخص شده است. تعداد چند برابر  $\pm 1$  و 2 برابر انحراف معیار خطاهای به عنوان خط های عمودی طاهر می شوند و متن زیر اولین خط عمودی انحراف معیار را نشان می دهد.

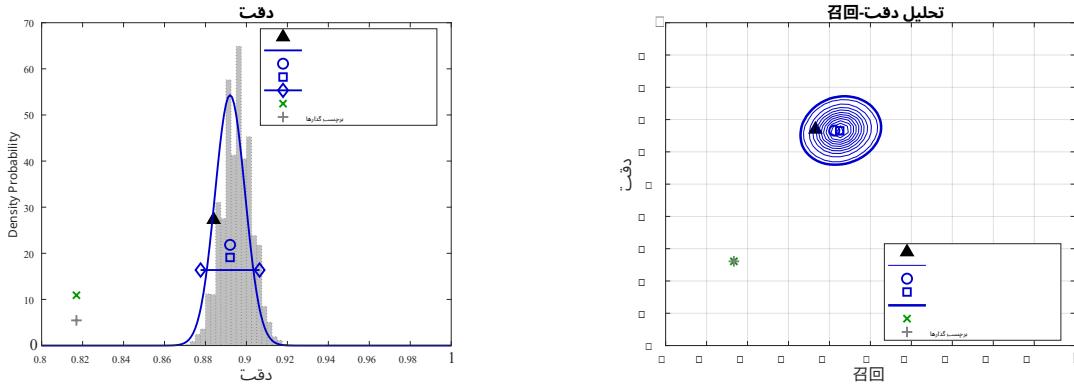
## در مسائل درستی ساری در طبقه بندی تحت نطارت



تصویر 15: خطاهای در برآورد تحلیل P-R برای روش‌های مختلف تست بر روی 100 نقطه عملکرد ایده‌آل مختلف. محدوده‌های محورها برای تست Bayes با MMSE متفاوت است (نقشه‌های بالا: منفی 0.05 تا مثبت 0.05) و روش‌های دیگر (نقشه‌های وسط و پایین: منفی 0.5 تا مثبت 0.5). نقشه‌های پراکندگی کوچک خطاهای به صورت نقطه‌های خاکستری ظاهرمی‌شوند. خطای متوسط با دایره مشخص شده و به عنوان جفت مرتب شده لیست شده است. بیضی‌ها مناطقی را نشان می‌دهند که 0.6827 و 0.9545 از توزیع دو متغیره نرمال فیت به خطاهای را پوشش می‌دهند و متنی که به محورهای نیمه بزرگ و نیمه کوچک نزدیک است، ریشه‌های مطلق ویژگی‌های ماتریس واریانس خطای نشان می‌دهد.



تصویر 16: خطاهای در برآورد تحلیل ROC برای روش‌های مختلف تست بر روی 100 نقطه عملکرد ایده‌آل مختلف. محدوده های محورها برای تست MMSE با Bayes تجربی (نقشه های بالا) منفی 0.5 تا مثبت 0 فی 0.05 تا مثبت 0.05) و روش‌های دیگر (نقشه های وسط و پایین: منفی 0.5 تا مثبت 0.5). نقشه های کوچک از خطاهای به صورت نقطه های حاکستری نمایش داده می شوند. خط ای متوسط با دایره مشخص شده و به عنوان جفت مرتب شده لیست شده است. بیضی ها مناطقی را نشان می دهند که برای 0.6827 و 0.9545 از توریع دو متغیره نرمال که به خطاهای تطبیق داده شده، حساب می شوند و متنی که در نزدیکی محورهای اصلی و کوچک قرار دارد، ریشه دوم اعداد ویژه ماتریس هم بستگی خطاهای را نشان می دهد.



تصویر 17: نمونه ای از تست برای یک برجسب گذار تنها  $T = 1$  (زمانی که احتمال خطای برجسب گذاری ثابت  $\epsilon = 1$  برای همه نمونه ها فرض می شود. میانگین و مدیان ساخته های برجسب گذارها نمایش داده نمی شود زیرا  $T = 1$ )

که اغلب خطاهای بیش از 0.100، 0.200 یا بیشتر تولید می کنند—که با توجه به اینکه ساخته ها در محدوده [0, 1] قرار دارند، غیرقابل قبول است. روش تجربی بیایس به وضوح مناسب تر از روش بیایس کامل است.

#### 5.2.5 برجسب دهنده تک و احتمال خطای برجسب دهنده ثابت

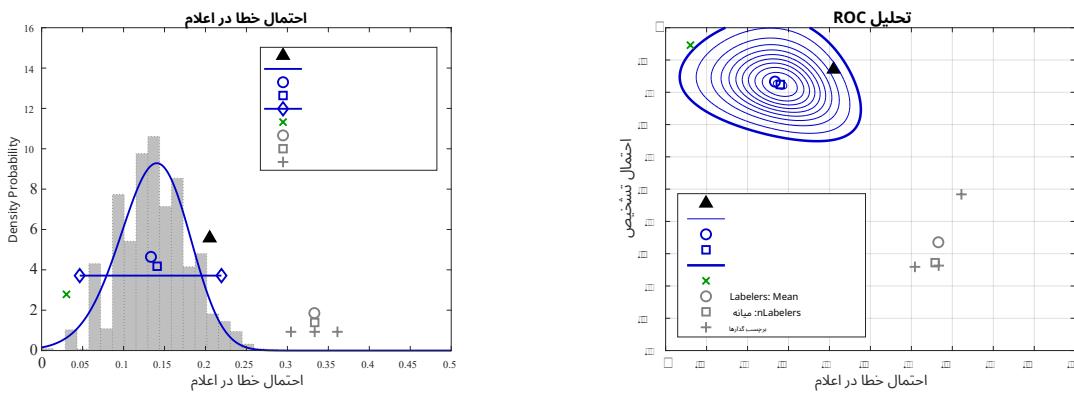
روش های تست نیز می توانند در صورت وجود یک برجسب گذار  $T = 1$  (و تنها می خواهیم بدانیم که عمل کرد تحت احتمال خطای برجسب گذاری بدترین حالت) چگونه تحت تأثیر قرار می گیرد، مفید باشند. می توان به راحتی  $\phi \equiv 0$  و  $C \equiv \phi C(0)$  را در (60) تنظیم کرد و از روش ها استفاده کرد.

تصویر 17 مثالی برای  $T = 1$  و  $\epsilon = 0.1$  نشان می دهد؛ تنظیمات دیگر شبیه ساری بودند. آزمایش MMSE<sub>ptdesDt, t ptdesFA</sub> ( ) $\pi(1) = 0.6$ ،  $\delta t_i \equiv 0.1 \cdot 0.9$ ، و  $\phi_{tt} \equiv 2.10$ . تولید تخمین های دقیق می کند و تخمین های پس زمینه و مناطق قابل اعتماد آن به یک فرد اجازه من دهد تا تعییرات ممکن ناشی از احتمال خطای برجسب گذاری فرضی را درک کند. برای نشان های مربوط به برجسب تخمین صحیح  $\hat{z}_t$  برابر با برجسب نویزده  $z_t$  است، بنابراین  $0 \leq \hat{z}_t \leq 1/2$ . این تخمین های ناکافی بسیار کم دقت تر از آن هایی هستند که از روش Bayes تجربی به دست می آیند.

#### 5.2.6 اندازه نمونه کوچک

نزدیک سازی های پشت آزمون MMSE تحت تأثیر CLT قرار دارند، بنابراین باید برای  $N$  کوچک به شرط  $N - N_1$  و  $N_1 - N$  بیشتر از یا برابر با سی و سه باشند، معتبر باشند. ما یک شبیه سازی دیگر با  $N = 3$ ،  $N_1 = 37$  و  $N - N_1 = 33$  تولید کرد؛ تنظیمات شبیه سازی دیگر شامل  $(\pi(1) = 0.5, (p_{desD}, p_{desF}) = (0.85, 0.20), \delta t_i \sim \text{Beta}(1, 2), \forall t_i, \phi_{tt} \sim U(0.2, 0.5))$  و

شکل 18 نتایج مربوط به  $PtFA(t) = PtFA(t, t)$  را نشان می دهد. قسمت های خلفی به وضوح غیر گاوی هستند و مناطق 95٪ معتبر به دلیل حجم نمونه کوچک بزرگ هستند، اما معیارهای ایده آل را در خود دارند. روش های غیربهینه دوباره تخمین های دقیق تری ارائه می کنند.



شکل 18: نمونه آزمایش برای حجم نمونه کوچک ( $tNt = 70$ ), با  $Ct = 37^\circ$ .

### 5.3 نمونه ای از آزمایش: طبقه بندی چند کلاسه

برای روش های تست چند کلاسه بخس 2.7، مثالی با  $Ct = 4$  ارائه می کنیم،

$$Nt = 2000, t \pi t = (0t.t2t.t 0t.t3t.t 0t.t1t.t 0t.t4).$$

$$\mathbf{K}^{\text{des}} = \begin{bmatrix} n & 0.75 & 0.08 & 0.10 & 0.07 \\ & 0.10 & 0.65 & 0.12 & 0.13 \\ & 0.04 & 0.06 & 0.80 & 0.10 \\ & 0.10 & 0.05 & 0.05 & 0.80 \end{bmatrix},$$

جربی بیز (الگوریتم 4) از  $Mt = 2500$ ,  $t Ct = 10t4t$  و  $t ttt Ut(0t,t 0t.t4),t ttt$  صحیح را تخمین می زند (الگوریتم 3) به طبقه بندی چند طبقه گسترش دادیم. هر دو الگوریتم پس از سه تکرار همگرا نشدند. برای تک تک برچسب ها، ماتریس های سردرگم آنها را محاسبه کردیم، ماتریس سردرگمی برچسب ها را بایت  $N/t$  (تعداد برچسب ها از برچسب گذار  $t ttt$ )، و میانگین و میانه چند بعدی ماتریس های مقیاس شده را محاسبه کردیم. به غیر از حالت ایده آل، آزمایش  $t$  شامل برچسب صحیح نمی شد.

ابتدا، جدول 13 تخمین های RV دقت را برای روش های تخمین مختلف مقایسه می کند. آزمایش MMSE دقیق ترین تخمین ها را ایجاد کرد، و مقدار دقت ایده آل دقیقاً در منطقه معنی بر 95٪ قرار دارد. تخمین برچسب های صحیح تخمین نسبتاً خوبی از دقت اما دوباره بدون احساس عدم قطعیت به دست آورد. با استفاده از برچسب های برچسب گذار جداگانه، دقت های 0.5667، 0.6586، 0.6397، 0.6698 و 0.5504 به دست می آید؛ در نظر گرفتن میانگین یا میانه این مقادیر، تخمین های بسیار نادرستی از دقت را ایجاد می کند.

دوم، جدول ۱۴ ماتریس مخلوط ایده آل و ماتریس مخلوط برآورد شده از تکنیک های مختلف را نشان می دهد. ده عنصر ماتریا از تست MMSE به عنصرهای مساوی در ماتریس مخلوط ایده آل نزدیک ترین هستند، شش عنصر از برآورد برچسب های صحیح نزدیک ترین هستند، یک عنصر از میانگین برچسب دهنگان نزدیک ترین است و یک عنصر از میانگین برچسب دهنگان نزدیک ترین است. (دو مورد همانکه وجود داشته است، بنابراین این اعداد به جای شانزده به هجده می رسند). در نهایت، جدول ۵۱ محدوده های قابل اعتماد ۵۹ درصدی عنصرهای ماتریس مخلوط از تست MMSE را نشان می دهد. هر محدوده قابل اعتماد شامل عنصر مساوی در ماتریس مخلوط ایده آل است.

در مسائل درستی ساری در طبقه بندی تحت نطارت

روش تخمین	ACC
مطلوب(برچسب های صحیح شناخته شده)	0.7350
تست MMSE: تخمین تبلور شبکه گیجی ماتریس از طریق نمونه گیری	0.7355 0.7355 (0.7282, 0.7427)
تخمین برچسب های صحیح	0.7390
ادغام معیارها از هر برچسب گذار	0.6170 0.6397

جدول 13: دقت تخمینی برای مثال طبقه بندی 4-طبقه. خط تیره نشان دهنده تخمینی است که نزدیک ترین به دقت ایده آل یا محدوده قابل اعتماد شامل دقت ایده آل است.

ایده آل (درست) برچسب ها شناخته شده)		برچسب پیش‌بینی شده			
		0	1	2	3
درست	0	290	40	40	39
برچسب	1	58	391	68	68
	2	8	10	168	27
	3	93	43	36	621

MMSE Testing	Predicted Label				مقدار تخمینی صحیح Labels	Predicted Label			
	0	1	2	3		0	1	2	3
rrect133.9Correct1	0	284.8	<b>41.3</b>	<b>36.6</b>	33.9Co	<b>288</b>	43	35	<b>37</b>
برچسب	2	63.7	<b>391.7</b>	67.0	67.6	1	61	395	70
	3	<b>7.5</b>	<b>7.3</b>	<b>169.8</b>	<b>29.0</b>	2	7	4	170
		<b>93.0</b>	<b>43.6</b>	38.7	<b>624.6</b>	3	<b>93</b>	42	<b>25</b>

میانگین Labelers	Predicted Label				میانگین Labelers	Predicted Label			
	0	1	2	3		0	1	2	3
درست1	0	238.9	60.7	65.3	54.2	0	254.1	57.4	46.8
برچسب	2	<b>60.7</b>	335.5	66.3	86.3	1	71.1	343.4	<b>67.2</b>
	3	36.7	54.2	130.1	64.8	2	26.1	35.9	145.9
		76.6	91.3	48.8	529.6	3	90.6	65.8	46.7

جدول 14: ماتریای مخلوط ایده آل و تخمینی برای مثال طبقه بندی 4-طبقه. خط تیره نشان دهنده نزدیک ترین عنصر به عنصر مثبتاً در ماتریای مخلوط مخلوط ایده آل است.

95%-مطمئن		0	1	2	3
		(277.72, 291.95)	(35.80, 46.79)	(32.27, 40.88)	(28.54, 39.22)
درست	0	(57.75, 69.56)	(384.72, 398.77)	(61.51, 72.46)	(60.74, 74.43)
	1	( 4.79, 10.23)	(3.97, 10.67)	(163.73, 175.83)	(24.03, 33.93)
برچسب	2	(86.74, 99.25)	(38.47, 48.82)	(34.34, 42.99)	(615.90, 633.21)
	3				

جدول 15: مناطق قابل اعتماد 95% برای عناصر فردی ماتریس استباه برآورد شده توسط آزمون MMSE (الگوریتم 4) برای مثال طبقه بندی 4-طبقه.

### مثال آموزش: رگرسیون لوژیستیک

این بخش از روش‌های آموزشی در بخش 3 با استفاده از رگرسیون لوژیستیک توصیح می‌دهد. ما از مجموعه داده‌های دوطبقه‌ی Ionosphere از مخزن داده‌های یادگیری ماسیون UCI استفاده می‌کنیم (الط Graff, 2017 مراجعه کنید). که شامل N = 351 نمونه است که هر کدام از 34 ویژگی عدد صحیح تشکیل شده‌اند. طبقه 0 به بارگشت رادار خوب و طبقه 1 به بارگشت رادار بد مربوط می‌شود. اینونوسفر شامل 126 بارگشت نادرست رادار است، بنابراین  $\pi = 0.359$ . ما از اعتبارسنجی نگه داری طبقه بندی شده 75%-25% استفاده می‌کنیم زیرا اعتبارسنجی چندین برابری نتایجی شلوغ تولید کرد که خواندن آن‌ها بسیار مشکل بود. در عمل، البته می‌توان از اعتبارسنجی چندین برابری استفاده کرد.

آموزش و تست برای  $T = 1, 5, 9$  و 13 انجام شده است. مجموعه داده‌ها برچسب‌های صحیح را فراهم می‌کند؛ برچسب‌های نویز شبیه بخش 5.1.2 شبیه سازی شده‌اند؛ بنابراین، RV‌های برچسب‌های نویز به صورت  $c_0$  تنظیمات به صورت زیر بودند:  $1 \equiv \delta_i, \phi_t = (\delta_i, \phi_t)$  هستند، ( $\delta_i, \phi_t$ ) مانند  $\text{conditionally independent}$  اند.  $\forall i, t \in T, \delta_i \approx \text{Beta}(1, 5), \phi_t \approx \text{U}(0, 0.5), \forall i, t \in T, \eta_i \equiv 1$  تا مجبور شود اول  $\eta_t \approx \text{U}(0.33, 1), t \in T$  باشد و  $\delta_i \approx \eta_i$ . واقعیت سخت که برای نمونه‌ها  $(..., 0.021, 0.021, 0.181, 0.181, 0.524, 0.115, 0.115, 0.367, 0.367, 0.440, 0.137, 0.137, 0.207, 0.207, 0.148, 0.148, 0.314, 0.314, 0.290, 0.290, 0.133, 0.133, 0.142, 0.142, 0.127, 0.127, 0.079, 0.079)$  بود. برای هر مقدار  $T$ ، شامل برچسب‌های نویز برای شماره‌های برچسب‌گذار از 1 تا  $T$  بود. در واقع، مثال معرفی شده در جدول 1 قطعه‌ای از  $Z$  برای  $T = 5$  است.

ما پنح کلاسیفایر در نظر می‌گیریم. کلاسیفایر ایده‌آل عملکرد ML را با ماتریس  $\{y, x\}$  و به عنوان مرجع درج شده است. طبق بخش 1a دیدگاه یکپارچه در بخش 3.2، طبق اصطلاح اولیه (32) در بخش 3.3.1، طبقه بندی کننده ML-بهینه آموزش ML را با توجه به  $\{\psi, \pi, z, x\}$  نجام می‌دهد. جزئیات در اپنديکس F.1 آورده شده است. ممکن است تابع هدف حاصل دیگر محدود نباشد، اما همچنان می‌توان از نزول جستجوی گرادیان برای یافتن اوح محلی استفاده کرد. کلاس برنده اپتیمال از آموزش MMSE بر اساس بخش 2 دیدگاه یکپارچه، شرط اصلی (40) از بخش 3.4.3 استفاده می‌کند؛ جزئیات در اپنديکس F.2 آورده شده‌اند. همچنین، کلاس برنده‌های تخمین، سارگار با زیرساخت و برآورد برچسب و رأی گیری که در بخش 3.7 توصیح داده شده اند نیز شامل می‌شود. ما تکرار نمونه را حذف کرده ایم زیرا به طور اساسی یک شکل کوانتینه شده از آموزش با (32) است. برای تمام آموزش کلاس برنده‌ها، ما از

تنظیمات تقویتی ۱۱۸۲ و روش استاندارد Broyden-Fletcher-Goldfarb-Shanno.

همه معیارهای تست با استفاده از روش Bayes تجربی الگوریتم 1 محاسبه شده‌اند، بنابراین، به استثنای طبقه بندی ایده‌آل، هیچ برچسب صحیحی در طول آموزش یا تست استفاده نشده است. با این‌که آموزش می‌تواند بهینه یا نا بهینه باشد، تست همیشه از همان تکنیک استفاده می‌کند. برای هر روش آموزشی، وزن تقویتی  $\lambda$  در محدوده  $\{0.1, 0.5, 1, 5, 10\}$  بررسی شد، که منجر به تولید یک مدل آموزش دیده شد. برای کلاس بندی ایده‌آل، ما مدل را که

بزرگترین محدوده ایده آل زیر منحنی ROC در مجموعه آزمایش نگه داری شده. برای دیگر کلاس های طبقه بندی، منحنی ROC را بر روی مجموعه آزمایش برآورد کردیم (در بخش 5.4.2 به طور مختصر توضیح داده خواهد شد) و مدل ای که بزرگترین محدوده برآورد شده زیر منحنی ROC را دارد انتخاب کردیم.

#### 5.4.1 پارامتر تصمیم ثابت

یک مدل رگرسیون ۰/۱۰۵۲ آمارتمنی آموزش دیده  $\sim \theta; y; x$  باشد. [۱] را به عنوان برآورد خود از  $y; \theta; x$  محاسبه می کند و آن را با یک پارامتر مزدی مقایسه می کند؛ برچسب پیش بینی شده  $y$  برابر با ۱ است اگر  $\sim \theta; y; x >$  باشد و در غیر این صورت ۰. این بخش نتایج مربوط به پارامتر تصمیم یکپارچه  $T = 1/2$  را ارائه می دهد.

تصویر ۱۹ نمودارهای تحلیل P-R را برای مجموعه آزمایش نگه داری شده به عنوان افزایش T نشان می دهد. برای کلاس های طبقه بندی آموزش دیده با برچسب های نویز، محدوده ها نشان دهنده احتمالات مشترک (REC PREC هستند، دایره ها نشان دهنده میانگین های شرطی هستند و خطوط مستقیم نشان دهنده محدوده های ۹۵%-قابل اعتماد هستند. عموماً برچسب های صحیح و در دسترس نمی باشند، اما چون ما از فرصت دانستن آنها برخورداریم، مثلث های معکوس هر کلاس های عملکرد واقعی خود را نشان می دهند. مثلث سیاه مس تقييم نشان دهنده نقطه عملکرد کلاس های ایده آل است که برای تمام مقادیر T یکسان است.

با  $T = 1$  اطلاعات کمی در مورد  $y$  ارائه می دهد، بنابراین احتمالات پراکنده هستند و محدوده های قابل اعتماد نسبتاً بزرگ هستند. برای هر روش آموزش، میانگین شرطی و نقطه عملکرد واقعی خیلی به هم نزدیک نیستند، اما نقطه عملکرد واقعی در محدوده قابل اعتماد قرار دارد. روش های ناکارآمدتر از کلاس های MMSE-بهینه و ML-بهینه برای این مورد عملکرد بهتری دارند، اما محدوده های قابل اعتماد به هم جنان زیاد برخورد می کنند که نمی توان این نتیجه را بدون دسترسی به برچسب های صحیح گرفت. این نمودار همچنین نشان می دهد که روش های آموزش و آزمایش ما حتی برای یک برچسب دهنده ناقص نیز قابل اعمال هستند.

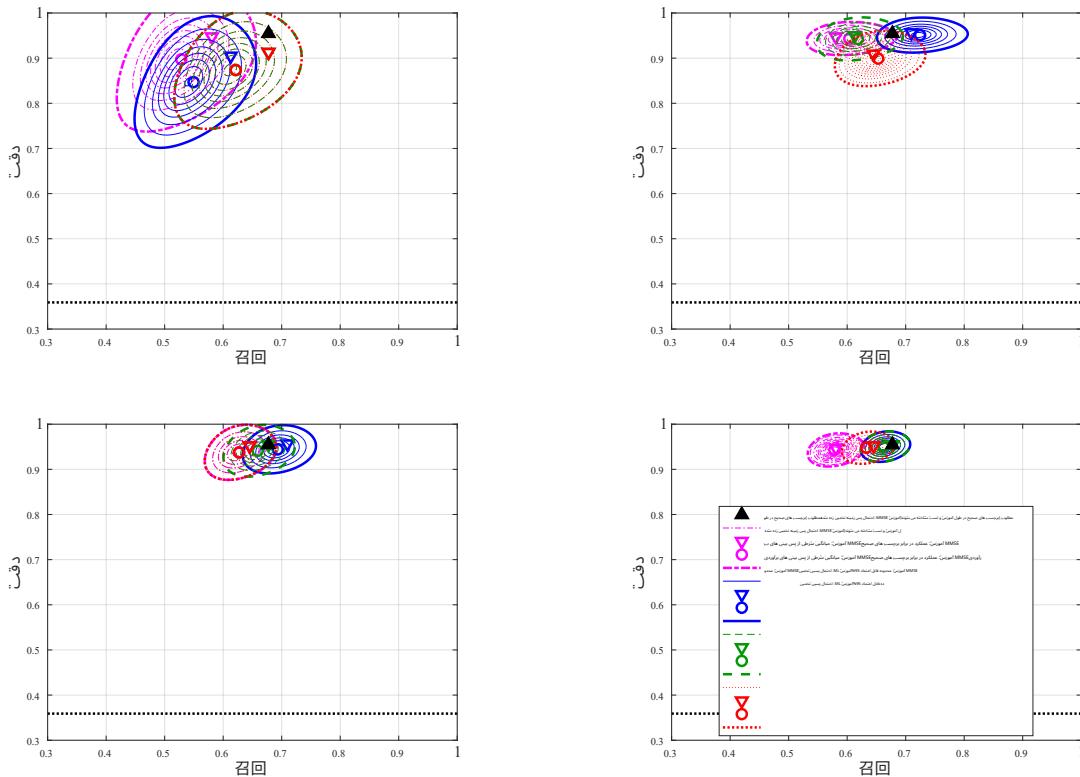
وقتی  $T = 5$  است، اطلاعات بیشتری از  $z$  در مورد  $y$  در دسترس است، بنابراین محدوده های قابل اعتماد کوچکتر می شوند و میانگین های شرطی بسیار نزدیک تر به نقطه های عملکرد واقعی برای همه کلاس ها می شوند. کلاس های ML-بهینه عملکرد بهتری از کلاس هایی که از آموزش ناکارآمد استفاده کرده اند، دارند و از احتمالات می توان انتظار داشت که اینگونه باشد. همچنین اتفاق افتاده است که عملکرد بهتری از کلاس ایده آل دارد. کلاس MMSE-بهینه عملکرد مشابهی با کلاس برآورد برچسب دارد.

برای  $T = 9$ ، روش های آموزش برآورد مشابهی از دقت ارائه می دهند، اما کلاس ML-بهینه دقت باز گشتنی بیشتری دارد. با این حال، احتمالات به طور قابل توجهی برخورد می کنند، بنابراین نمی توان این نتیجه را بدون دسترسی به برچسب های صحیح گرفت. کلاس ML-بهینه دوباره کمی بهتر از کلاس ایده آل عمل می کند. کلاس های MMSE-بهینه و آموزش دیده با روش رای گیری عملکرد یکسانی دارند.

با رسیدن  $T = 13$ ، احتمالات و محدوده های قابل اعتماد بسیار تنک تر می شوند و میانگین های شرطی برآورد دقیقی از نقطه های عملکرد واقعی ارائه می دهند. کلاس ML-بهینه و کلاس آموزش دیده با برآورد برچسب بهتر از آموزش رای گیری عملکرد دارند و نقطه های عملکرد واقعی آنها با کلاس ایده آل همخوانی دارند. برای این انتخاب پارامتر مزدی، کلاس MMSE-بهینه دقت بارگشتی کمترین دارد. در بخش های بعدی، پارامتر مزد تغییر می کند و کلاس MMSE-بهینه عملکرد رقابتی ای دارد.

#### 5.4.2 منحنی های عملکرد

همچنین، عموماً عملی معمولی است که محدوده ای ممکن  $T$  را بررسی کنیم تا منحنی های ROC یا P-R را به دست آوریم. تصویر ۲۰ منحنی های ROC برآورد شده را نمایش می دهد که شرطی هستند



شکل 19: نمونه آموزشی با تعداد متفاوت برچسب دهنگان  $T$ : نتایج تخمین تست با استفاده از الگوریتم 1 بر روی مجموعه تست نگه داری شده. خط سانس به عنوان یک خط نقطه ای سیاه نشان داده منشود.

میانگین  $FAP_{DP}$ ) از تست MMSE با استفاده از الگوریتم ۱، برای روش‌های مختلف آموزش و تعداد برجسب دهنگان. عملکرد تخمین شده به عنوان  $T$  افزایش می‌یابد، اما تعییرات زیادی  $T = 5$  وجود نداشد. وقتی  $T = 1$  است، تخمین عملکرد دشوار است که منجر به منحنی‌های ROC تخمین شده خشن می‌شود. به عنوان  $T$  افزایش می‌یابد، منحنی‌های ROC تخمین شده شروع به نمایش شکل پله ای منحنی ROC ایده‌آل من کنند.

برای هر مقدار  $T$ ، تمام روش‌های آموزش عملکرد مشابه دارند. این رفتار نشان می‌دهد که روش‌های آموزش غیر بهینه سازی شده تنظیم شده جایگزین‌های عملی برای روش‌هایی که بر اساس دیدگاهی کپارچه بهینه هستند، ارائه می‌دهند. این مشاهده با آزمایش‌های تست در بخش‌های ۵.۲.۴ و ۵.۲.۵ متفاوت است، جایی که تست MMSE (الگوریتم‌های ۱ و ۲) عملکرد بهتری از روش تخمین برچسب‌های صحیح (الگوریتم ۳) داشته است. ما فرض می‌کنیم که این تفاوت نشان دهنده اهداف بنیادی متفاوت آموزش و تست وجود ترتیم است. هدف آموزش یادگیری یک مدل پیش‌بینی کرنده است که به خوبی به داده‌های خارج از نمونه ای که فراتر از مجموعه آمورشی است، تعیین می‌یابد

$\{x, \psi, \pi, z\}$ . بنابراین، آموزش شامل تنظیم است که به روش آموزش غیر بهینه سازی شده کمک می‌کند تا از توانایی تخمین خود جبران کند. در مقابل، هدف تست دریافت معیارهای درون نمونه برای مجموعه تست  $\{y, z, \psi, \pi\}$  است. در این مورد نیازی به تنظیم نیست و تست MMSE می‌تواند عملکرد تخمین شده بهتری از روش‌های تست غیر بهینه سازی شده ارائه دهد.

شکل 21 منحنی‌های ROC واقعی را که بر اساس برچسب‌های صحیح محاسبه شده اند، نشان می‌دهد. این منحنی‌ها در صورت وجود مشکلات حقیقت یابی در عمل قابل دسترسی نخواهند بود. این منحنی‌ها نشان می‌دهند که روش‌های آموزش می‌توانند عملکرد مشابه با حالت ایده‌آل داشته باشند و مقایسه با شکل 20 نشان می‌دهد که منحنی‌های ROC تخمین شده حتی وقتی  $T = 1$  است، مناسب هستند و برای  $T = 9$  یا ۱۳ بسیار خوب هستند.

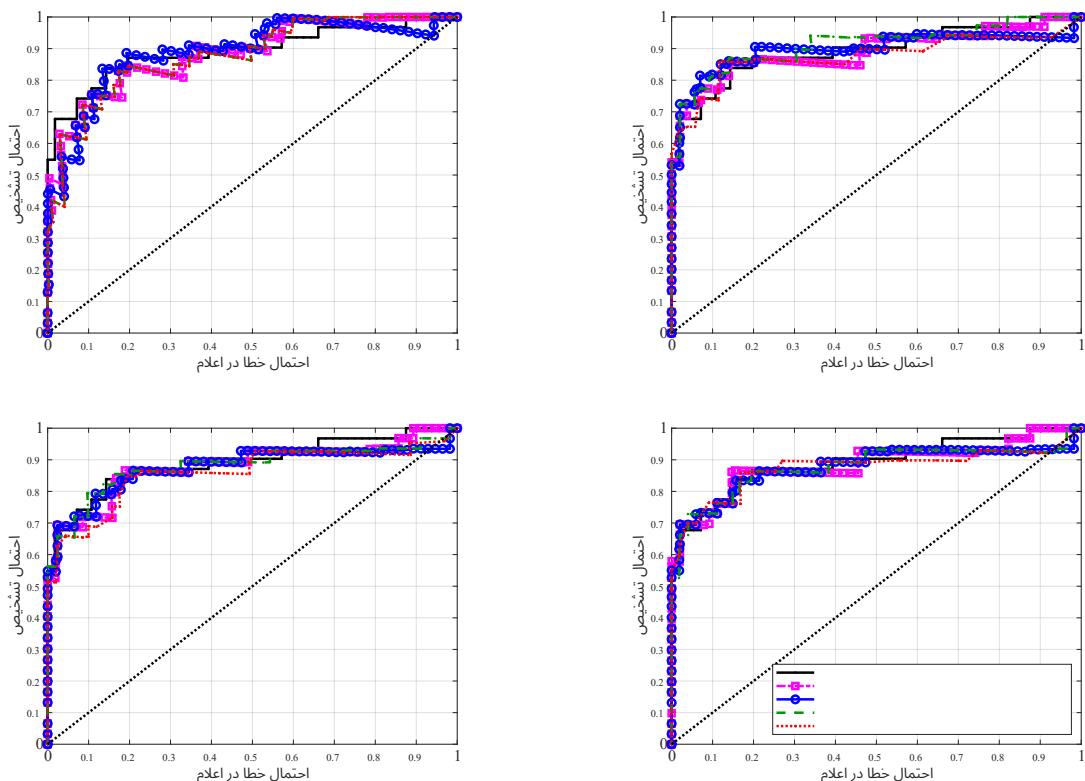
#### 5.4.3 پس زمینه منحنی عملکرد

تست MMSE به ما امکان می‌دهد تا پس زمینه مسترک  $REC\_PRE(\mathcal{C})$  یا  $FAP_{DP}$  (را تخمین بزنیم. می‌توانیم پس زمینه‌ها را بر روی همه مقادیر پارامتر میانگین به طور متوسط بگیریم تا پس زمینه منحنی ROC را به دست آوریم—این یک قابلیت مهم است وقتی مشکلات حقیقت یابی وجود دارد. شکل 22 پس زمینه منحنی‌های ROC برای کلاس بندی بهینه ML را با تغییر  $T$  با  $REC(\mathcal{C})$  می‌نماییم در محدوده  $[10^{-3}, 10^{-4}]$  با استفاده از نقشه‌های حرارتی با مقیاس رنک لوگاریتمی دهانه ۱۰ نشان داده می‌شوند.

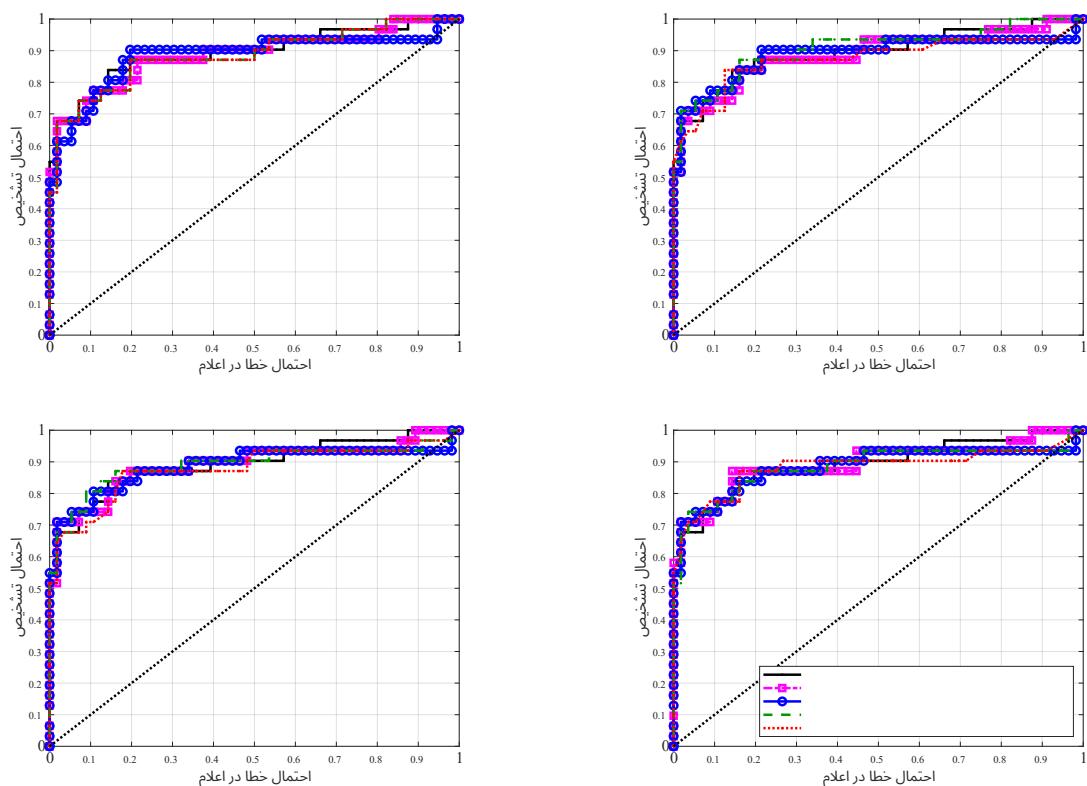
این نمودار همچنین منحنی‌های ROC تخمینی و واقعی برای این روش آموزش را نمایش می‌دهد. همچنین می‌توانیم محدوده‌های قابل اعتماد برای منحنی‌ها را محاسبه کنیم، اما آن‌ها را نمایش نمی‌دهیم تا از شلوغی تصویر جلوگیری کنیم. تصویر ۹۱ نشان دهنده خوبی محدوده‌های قابل اعتماد است.

همانطور که در شکل ۳۲ نشان داده شده است، پشتای های منحنی-R MMSE-بهینه نمایش داده شده است. وقتی  $1 < T < 2$ ، کلاس فایزر پیش‌بینی می‌کند که  $tyi = 0$ ، بنابراین یادآوری صفر است اما دقیق تعریف نشده است و هیچ یک از منحنی‌ها نقطه‌ای را نشان نمی‌دهد که  $rec = 0$  باشد.

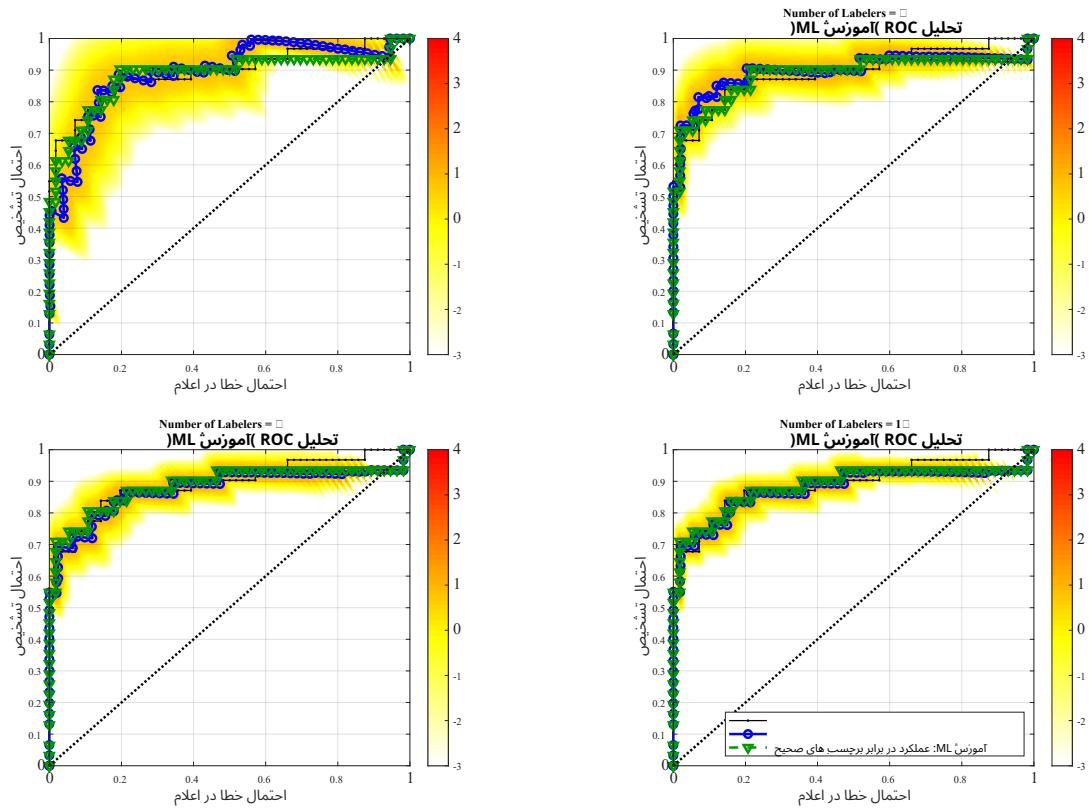
تصاویر همچنین یک مزیت دیگر از تست MMSE را آشکار می‌کنند. با اینکه عملکرد واقعی در حالت  $T = 1$  یا ۵ است، مشابه حالت ایده‌آل است، منحنی‌های برآورده شده برخی اختلافات بزرگ با منحنی‌های واقعی دارند و پشتیبانی‌های پسین گسترده‌ای دارند که نشان دهنده عدم قطعیت قابل توجهی در مورد عملکرد است. برای  $T = 9$  یا ۱۳، منحنی‌های برآورده شده به طور نزدیک با منحنی‌های واقعی همخوانی دارند و پشتیبانی‌های پسین بیشتر متمرکز شده اند که نشان دهنده عدم قطعیت کمتری در مرور عملکرد است. در حضور مشکلات حقیقت یابی، عملکرد واقعی



تصویر 20: نمونه آموزش با تعداد متفاوت برچسب گذاران  $T$ : منحنی های ROC تخمینی از تست MMSE با الگوریتم 1 بر روی مجموعه تست نگه داری شده. خط سانس به عنوان یک خط نقطه ای سه یا نماییس داده می شود.



تصویر 21: نمونه آموزشی با تعداد متفاوت برچسب گذاران  $T$ : منحنی های ROC واقعی بر روی مجموعه تست نگه داری شده. خط شناس به عنوان یک نقطه ای سیاه نمایش داده می شود.



شکل 22: نمونه آموزشی با تعداد متفاوت برچسب دهنگان برای کنندگان آموزش دیده ML:  
احتمالات منحنی ROC از الگوریتم 1 بر روی مجموعه تست نگه داری شده برآورد شده.  
احتمالات به صورت نقشه های حرارتی با مقیاس رنگی لوگاریتمی base-10 طاهر می شوند؛ خطوط مارک بارهای رنگی به عناصر اعشار مقیاس مربوط می شوند. خط سانس به عنوان یک خط نقطه دار سیاه طاهر می شود.

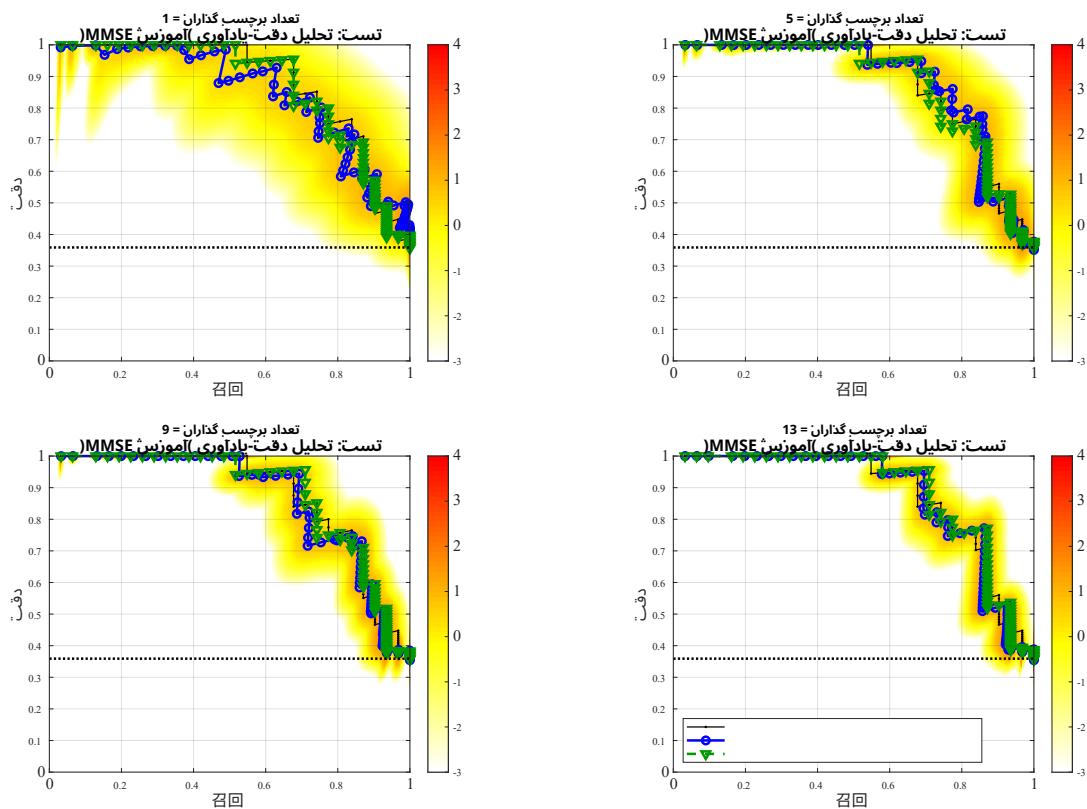
منحنی عملکرد در دسترس نخواهد بود، اما پشتیبانی منحنی احتمالات عملکرد می تواند به درک تنوع بالقوه در عملکرد که ممکن است رخداد کند.

### 5.5 مثال های اطلاعات مشترک معادل

در بخش 4، استفاده از اطلاعات مشترک به عنوان پایه ای برای مقایسه ترکیبات مختلف برچسب گذاران، نشان داد که چندین برچسب گذار متوسط می توانند به اندازه یک برچسب گذار متخصص اطلاعاتی باشند. برای بررسی این نتیجه، ما مجموعه داده های Ionosphere را بازدید کردیم و برچسب های نویزی ار با مدل (57) شبیه سازی کردیم، هم برای یک برچسب گذار خوب با  $\epsilon' = 0.05$  و هم برای نه برچسب گذار متوسط، هر کدام با  $\epsilon = 0.25$ . هر برچسب گذار برچسب نویزدار برای هر نمونه ارائه داد. سپس  $I(Z; Y) = 0.667$  و  $I(Y; T) = 0.758$  (با  $T = 9$  و  $\epsilon = 0.25$ ) بیت.

مانند بخش 5.4، ما کلاس های یهینه سازی ML و MMSE را با استفاده از رگرسیون لوگارitmیک با تنظیمات و برآورد ساخته های تست با استفاده از الگوریتم 1 برای تست MMSE آموزش دادیم.

## در مسائل درستی ساری در طبقه بندی تحت نطارت



تصویر 23: نمونه آموزش با تعداد مختلف برچسب گذاران T برای کلاس های آموزش داده شده با MMSE: برآورد احتمالات P-R از الگوریتم 1 بر روی مجموعه تست نگه داری شده. احتمالات به صورت نقشه های حرارتی با مقیاس رنگ لوگاریتمی ده گانگی نمایش داده می شوند؛ خطوط مقیاس رنگ ها به عناصر اعشار مقیاس پاسخ می دهند. خط سانس به عنوان خط سیاه و پیک سلی نمایش داده می شود.

شکل ۴۲ نمودارهای تحلیل ROC برای مجموعه آزمایش نگه داری شده را نشان می‌دهد که مشخص کننده ای که با برچسب‌های نویزدار برچسب دهنده خوب آموزش دیده، عمل می‌کند، که مفهوم را تأیید می‌کند. برچسب‌های نویزدار برچسب دهنده حرفه ای را با  $0.01 = \epsilon'$  و  $399 = I(Z; Y | T = 399, \epsilon')$  بیت و هر دو کلاس رکدام با  $0.45 = \epsilon$ ، شبیه سازی کردیم، بنابراین  $I(Z; Y | T = 399, \epsilon) = 0.863$ . احتمالات پس زمینه منحنی-R در شکل 25 نشان داده شده اند؛ وقتی که  $\tau = 0.859$  بیت، بنابراین یادآوری صفر است اما دقت تعریف نشده است و هیچ نقطه‌ای برای  $rec = 0$  رسم نشده است. منحنی‌ها نشان دهنده عملکرد قابل مقایسه هستند و دوباره مفهوم را تأیید می‌کنند.

## 6. خلاصه، نتیجه‌گیری‌ها و مسیرهای آینده

در طبقه بندی نطارت شده، ممکن است تعداد زیادی مشکلات حقیقی سازی رخ دهنده: برچسب‌های نویز؛ برچسب‌های گم شده؛ برچسب‌های چندگانه و متعارض برای همان نمونه؛ و ترکیبات مختلف بر جسب گذاران برای نمونه‌های مختلف. این وضعيت شامل سه عنصر می‌شود که هر کدام نیاز به یک مدل دارند: حقیقی سازی نیاز به مدل برچسب نویزدار، آموزش مدل پیش‌بینی کننده یاد می‌گیرد و تنسی نیاز به مدل مستقل فرمول سازی و یادگیری مدل برچسب نویز را مطالعه نکردیم که مخصوص بسیاری از کارهای مرتبط است. به جای آن، ما بر روی تنسی و آموزش تمرکز کردیم و با فرض این که یک مدل برچسب نویز خوب  $(y|t)p(\psi|y)$  موجود است، که کار ما را با کارهای مرتبط سازگار و مکمل می‌کند، شروع کردیم. روش‌های ما مدل‌هایی با برچسب گذاران وابسته را پشتیبانی می‌کنند.

### 6.1 خلاصه و نتیجه‌گیری‌ها

با اعمال اصول از نظریه برآورد بیایی، موفق به دریافت چند پاسخ امیدوارکننده و بینش بخشنده سوالاتی که در معرفی مطرح شده اند شدیم.

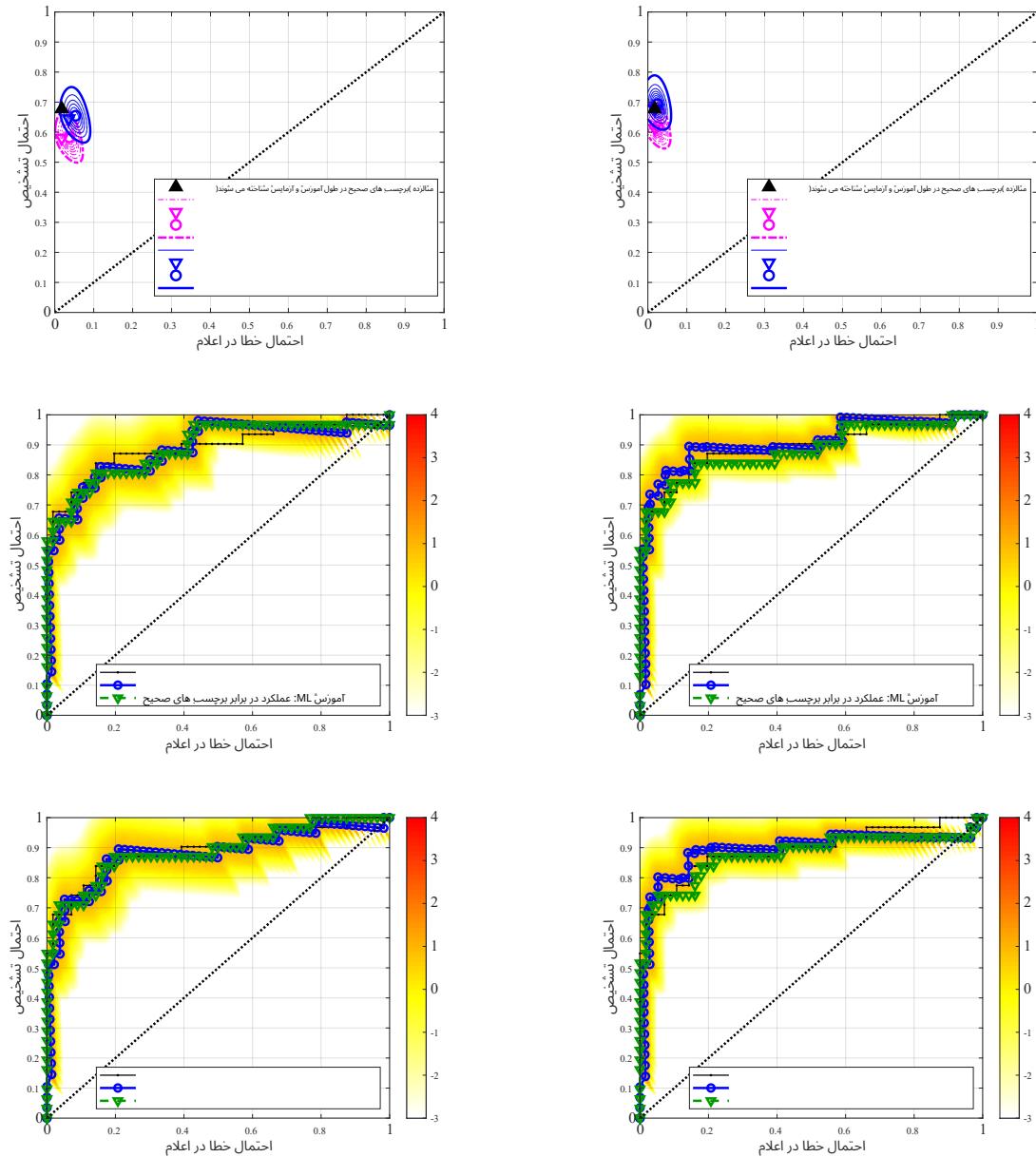
1. چگونه می‌توان یک طبقه بندی کننده را در حضور مسائل مربوط به صدقه آزمایش کرد

با توجه به برچسب‌های نویز دار  $Z$ ، برچسب‌های پیش‌بینی شده  $\hat{y}$ ، پارامترهای مدل برچسب نویز  $\psi$  و پیش‌بینی‌های کلاس  $\pi$ ، ما روش‌های تستی ایده‌آل را توسعه دادیم: آنها مقیاس‌های  $RV$  را تخمین می‌زنند به جای  $RV$  برچسب‌های صحیح، تمام اطلاعات موجود را به طور کامل بهره برداری می‌کنند و یک معیار به خوبی تعریف شده (MMSE) را بهینه سازی می‌کنند. رویکرد ما کاملاً از آموزش جدا است و در فراتر از دامنه یادگیری ماسیین قابل استفاده است. به عنوان مثال، می‌تواند برای هماهنگی تشخیص‌های پرسشکار یا دسته بندی‌هایی که توسط دانشمندان انجام می‌شود، استفاده شود.

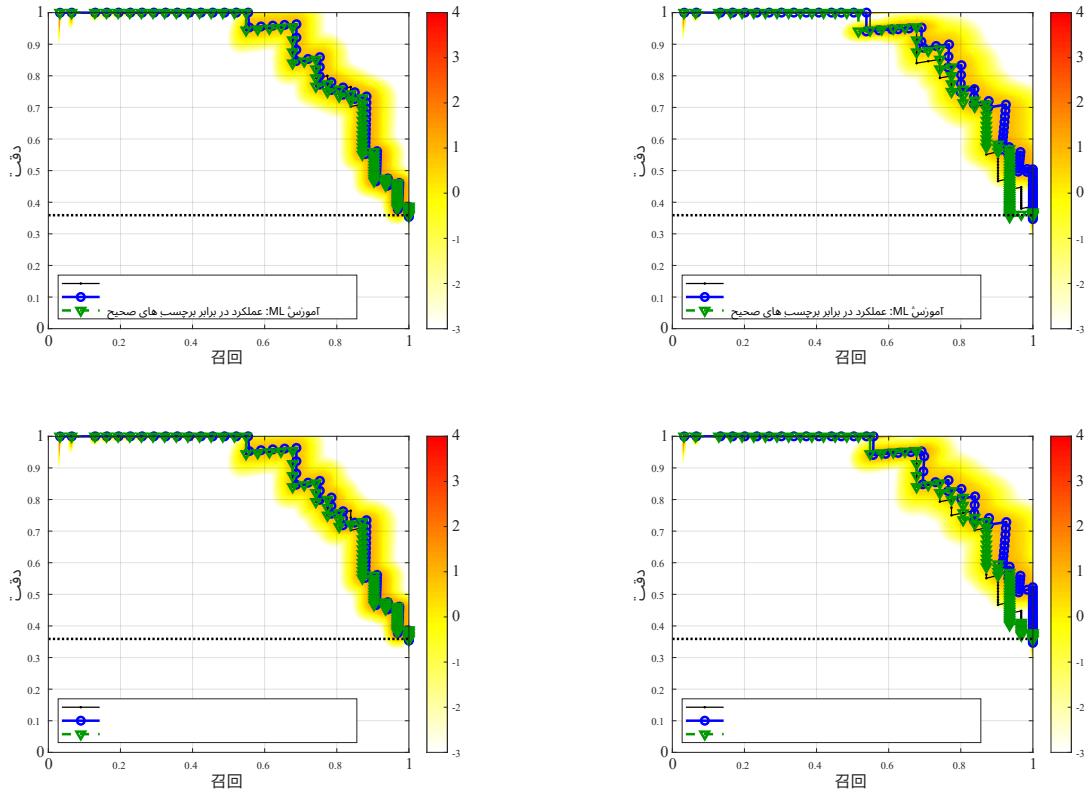
برای دستیابی به روش‌ها، ما یک مدل آزمایش نوآورانه (7) (پیشنهاد دادیم و از آن برای استنتاج احتمالات مرزی تقریبی برای چندین ساخته سکالر، به علاوه احتمالات مسترک برای تحلیل ROC و P-R استفاده کردیم. سپس ما آزمایش MMSE را معرفی کردیم و الگوریتم‌های بیاییس تجربی (ا) لگوریتم‌های 1 و 2 (را برای یافتن به طور مکرر برآورد MMSE پارامترهای مدل آزمایش از  $\{\hat{y}, Z, \pi, \psi\}$  توسعه دادیم. پس از برآورد پارامترها، ما برآوردهای بهینه بیاییس (نقاط برآورد MMSE یا MAP، یا مناطق اعتماد) برای  $RVs$  ساخته را محاسبه کردیم. در نهایت، ما این روش را به طبقه بندی چندگانه گسترش دادیم.

در آزمایش‌های ما، تست MMSE ارزیابی‌های برجسته ای از بسیاری از ساخته‌های طبقه بندی دوگانه ارائه داد. خطاهای ارزیابی آن‌ها یک مرتبه کوچکتر بودند

## در مسائل درستی ساری در طبقه بندی تحت نطارت



تصویر 24: تحلیل ROC برای یک برجسپ دهنده خوب (چپ) و چند برجسپ دهنده متوسط (راست). نمودارهای بالایی عملکرد با یک حد  $\alpha = 0.02$  را نشان می‌دهند.  
 نقشه‌های وسط، احتمالات محدوده ROC برای طبقه بندی کنندگان آموزش دیده با ML را نشان می‌دهند. نقشه‌های پایین، احتمالات محدوده ROC برای طبقه بندی کنندگان آموزش دیده با MMSE را نشان می‌دهند. احتمالات به صورت نقشه‌های حرارتی با مقیاس رنگی لوگاریتمی پایه 10 نمایش داده می‌شوند؛ خطوط مقیاس رنگ‌ها معادل اکسپوننتیون مقیاس هستند. خط سانس به صورت خط نقطه دار سیاه نمایش داده می‌شود.



تصویر 25: احتمالات محدوده P-R برای یک برجسب دهنده حرفه ای یک نفر (چپ) و بسیاری از برجسب دهنگان صنعتی (راست). نقشه های بالا نتایج طبقه بندی کنندگان آموزش دیده با ML را نشان می دهند؛ نقشه های پایین نتایج طبقه بندی کنندگان آموزش دیده با MMSE را نشان می دهند. احتمالات به صورت نقشه هایی حرارتی با مقیاس رنگی لوگاریتمی پایه 10 نمایش داده می شوند؛ خطوط مقیاس رنگ ها معادل اکس پونتیون مقیاس هستند. خط سانس به صورت خط نقطه دار سیاه نمایش داده می شود.

بهتر از روش‌های ناکارآمد مانند تخمین برچسب‌های صحیح یا میانگین‌گیری از مقیاس‌های برعسب دهنگان فردی است. برای طبقه بندی چندگانه، آزمون MMSE در تخمین دقّت و عناصر فردی ماتریس استباه بهتر از روش‌های ناکارآمد عمل کرده است.

2. چگونه می‌توان یک طبقه بندی کننده را در حضور مشکلات درستی یا بی آموزش داد؟
- با  $\pi$ ,  $\psi$ ,  $z$  و درستهای ویرگی  $x$  داده شده، یک دیدگاه یکپارچه از آموزش ارائه دادیم که زیبا و قالب فهم است. دیدگاه یکپارچه توصیح می‌دهد که چگونه می‌توان یک گستره وسیعی از طبقه بندی کننگان را آموزش داد و برخی از کارهای مرتبط را سازماندهی می‌کند. هر یک از روش‌های آموزشی ما بهینه است: از یک تابع احتمال مناسب، احتمال پس زمینه یا برآورده کننده استفاده می‌کند؛ اطلاعات موجود را به طور کامل بهره برداری می‌کند؛ و یک معیار محازات یا کاربردی به خوبی تعریف شده را بهینه می‌کند. همچو یک از روش‌ها برچسب‌های صحیح را تخمین نمی‌زند.
- برای مدل‌های احتمالاتی (یعنی مدل‌های تولیدی یا تشدیصی) (با پارامترها  $\theta$ ، نشان دادیم که اصل بهینه ساری از آموزش ایده‌آل می‌تواند حفظ شود. در آموزش ML، تابع احتمال  $p(y|tx; \theta)$  یا  $p(y, x; \theta)$  به ترتیب با  $p(z|tx; \psi, \theta)$  یا  $p(z|x; \psi, \theta)$  جایگزین می‌شود. در آموزش MAP، احتمال پس زمینه  $p(\theta|y, x)$  با  $p(\theta|z, x)$  جایگزین می‌شود. برای مدل‌های غیراحتمالاتی، ما آموزش MMSE را پیشنهاد دادیم که تابع صرر اصلی را حفظ می‌کند و حداقل برآورد  $R_{MMSE}$  از خط تجربی را بهینه می‌کند. برخی از کارهای مرتبط با همان سنتل آموزش پیشنهاد داده اند اما از نظریه ارزیابی برخی از گیزه بخشی آن استفاده نکرده اند. ما ویرگی‌های برآورده کننده MMSE را بحث کردیم و شرطی برای زمانی که برآورده کننده MMSE یک برآورده کننده مدام است ارائه دادیم.

آزمایش‌هایی که از رگرسیون لوجستیک دودویی استفاده شده است، اثربخشی روش آموزشی ما را نشان داده است، همچنین رقابتی برخی روش‌های نا بهینه مانند تخمین برچسب‌های صحیح یا رای گیری بین برچسب دهنگان که با زیرساخت‌های موجود یادگیری ماسیون سارگار هستند را نشان داده است. ما استدلال کردیم که تقویت regularization (به روش‌های نا بهینه کمک می‌کند تا کمبودهای تخمین خود را جبران کنند. آزمایش‌ها از روش‌های تست ما استفاده کرده اند، بنابراین امکان پذیری آموزش و تست با استفاده از برچسب‌های نویزدار را نشان داده اند. علاوه بر این، نشان داده اند که روش‌های تست ما می‌توانند پیش‌بینی‌های تقریبی پس زمینه و تخمین‌های بهینه از متحن ROC و متحن R می‌دانند.

3. چگونه می‌توان ترکیبات مختلف برچسب‌گذاران با توانایی‌های مختلف را مقایسه کرد؟
- روند برچسب‌گذاری با صدای بلند می‌تواند به عنوان یک کاتال پخسن‌تاویزیون در نظر گرفته شود، بنابراین اطلاعات مسترک میزان اطلاعاتی را که گروهی از برچسب‌گذاران در مورد برچسب صحیح انتقال می‌دهند، ارزیابی می‌کند و مقایسه بین ترکیب‌های مختلف برچسب‌گذاران را تسهیل می‌کند. روند برچسب‌گذاری با صدای بلند می‌تواند به عنوان یک کاتال پخسن‌تاویزیون در نظر گرفته شود، بنابراین اطلاعات مسترک میزان اطلاعاتی را که گروهی از برچسب‌گذاران در مورد برچسب صحیح انتقال می‌دهند، ارزیابی می‌کند و مقایسه بین ترکیب‌های مختلف برچسب‌گذاران را تسهیل می‌کند. به عنوان پایه ای دیگر برای مقایسه، هر ترکیب از برچسب‌گذاران می‌تواند به عنوان یک برچسب‌گذار معادل با احتمال خطای مرتبط نمایش داده شود. این مشاهده نشان می‌دهد که برچسب‌گذاران متوسط می‌توانند اطلاعاتی بزرگتر یا برابر با اطلاعاتی که از یک برچسب‌گذار کارشناس به دست می‌آید، انتقال دهند.

نکته قبلی نظری است؛ این نکته توصیح نمی‌دهد که چگونه در عمل این اطلاعات را استخراج کنیم. خوشبختانه، روش‌های آموزشی و تستی ما راهی برای انجام این کار ارائه می‌دهند. کار این مقاومت منجر به آزمایش‌های آموزشی و تستی شد که مفهوم را تأیید کردند و نشان دادند که روش‌های ما می‌توانند مزایای آن را به واقعیت تبدیل کنند.

## 6.2 روند گستردگی شده

مشکلات واقعی ساری در بسیاری از کاربردها وجود دارد. برای حل آنها، ما از یک روند کاری گستردگی حماست من کنیم که این کار و کارهای مرتبط مکمل را ترکیب من کند. آموزش باید دو مدل یاد بگیرد: یک مدل برچسب نویز و مدل پیش‌بینی مورد نظر. مدل‌ها می‌توانند به صورت جداگانه یاد گرفته شوند، به عنوان مثال از روش‌هایی که در بخش 1.2.1 برای یادگیری مدل برچسب نویز استفاده می‌شود و سپس از تکنیک‌های آموزشی در بخش 3 برای یادگیری مدل پیش‌بینی استفاده می‌شود.

به جای آن، مدل‌ها می‌توانند با استفاده از تکنیک‌هایی مانند آن‌هایی که توسط Raykar و همکاران ارائه شده‌اند، به طور مشترک یاد گرفته شوند. Khetan et al. (2018)، Tanno et al. (2019) و (2020) پس از یادگیری هر دو مدل، مدل پیش‌بینی کننده می‌تواند با استفاده از الگوریتم‌های 1، 2 یا 4 از آزمون MMSE تست شود.

## 6.3 مسیرهای آینده

با بحث در مورد حوزه‌های آینده کاربسته می‌شویم.

### 6.3.1 موضوعات مستقیم مرتبط

عدد زیادی جنبه از تست MMSE مستلزم مطالعه بیشتر هستند. اولاً، در بخش 2.2، CLT به RV‌های معمول U و V اعمال شد و دستکاری‌های بعدی باعث تولید پیشینه‌های تقریبی RV‌های مقیاس در بخش 2.3 شد؛ بررسی حالتی که یکی یا هر دو جمع بندی‌های (8) و (9) شامل تعداد کافی ای اصطلاح برای توکیه CLT نیست، مفید خواهد بود. دوم، دستیابی به بیان برای پیشینه مشترک (17) در خط شناس FAP = DP رضایت‌بخش خواهد بود. سوم، بیان‌های قالبی برای حداقل (17) و (18) تخمین MAP از نقاط عملکرد ROC و P-R را ساده‌تر می‌کند.

چهارم، بخش 2.4.4 به نزدیکی روش‌های Bayes تجربی بحث کرد، اما مطالعه دقیق تری نیاز است. در نهایت، بخش 2.7 به طبقه بندی چندگانه پرداخت و دقت و عناصر فردی ماتریس اشتباہ را در نظر گرفت. می‌توان بررسی کرد که دیگر ساختهای طبقه بندی چندگانه، توزیع مشترک ماتریس اشتباہ، تأثیر یک پیش‌بینی کلاس بسیار نابرابر و تدبیر زمانی که تعداد کلاس‌ها زیاد است، چگونه است.

در مورد آموزش مدل پیش‌بینی احتمالی، بخش 3.3 نشان داد که آموزش ML یا MAP می‌تواند برای برچسب‌های نویز تعییر کند. می‌توان فرم‌های داده شده در جدول 9 را برای تطبیق آموزش از حالت ایده‌آل به حالت مشکلات حقیقت یابی اعمال کرد.

برای آموزش MMSE مدل‌های پیش‌بینی غیراحتمالی، بخش 3.4.3 توصیح داد که نقطه‌های گرادیان آن مناسب برای تفاوت یابی خودکار است (مقایسه با (50)، بنابراین مشاهده آموزش MMSE در شبکه‌های عصبی عمیق بسیار هیجان‌انگیز خواهد بود. بخش 3.4.5 هماهنگ محاسبه گر MMSE برای RV خط‌تجربی را در نظر گرفت و ساده ترین شرایط کافی را ارائه داد. می‌توان شرایط یا محدودیت‌های دیگر بر روی خط‌تجربی را استنباط کرد. همچنین می‌توان بررسی کرد که تحت چه شرایطی آموزش MMSE هماهنگ اصل ERM را حفظ می‌کند. کارهای Khetan et al. (2012)، Cid-Sueiro et al. (2014) و (2012)، Cid-Sueiro (2012) در این زمینه می‌تواند مفید باشد.

نظریه اطلاعاتی با BSBC در بخش 4.1 نشان داده شد. می‌توان یک کانال ناظم (对称) کانالی با احتمالات خطوط اعلام نادرست متفاوت (یا کانالی با برچسب دهنگان وابسته یا کانال چندگانه) را بررسی کرد.

### 6.3.2 یادگیری و تخصیص برچسب دهنگان

این مقاله فرض کرده است که مدل برچسب های نویز شناخته شده است، اما در بسیاری از موارد باید یاد گرفته شود. این نیاز ایجاد 了一系列 مشکلات تخصیص من کند. به عنوان مثال، فرض کنید یک مجموعه کوچک  $\{x, y, z\}$  با برچسب های صحیح و نویز و یک مجموعه بزرگ  $\{x, z\}$  با تنها برچسب های نویز داریم. چگونه باید مجموعه ها را برای یادگیری مدل برچسب های نویز، آموزش مدل پیش بینی و تست مدل یادگرفته پیش بینی تقسیم کنیم؟

دو مثال از تخصیص برچسب دهنگان آورده می شود. با توجه به بودجه برچسب دهی و هزینه های برجسب دهنگان با توانایی های مختلف، روشن با بیشترین بهینه سازی هزینه برای کس برچسب ها چیست؟<sup>22</sup> چه طور مثابه، با توجه به مجموعه ای از نمونه ها با دشواری های برچسب دهی مختلف (مثلًا، تصاویر تحت شرایط مختلف نورپردازی)، چگونه باید تلاش برچسب دهی در مجموعه توزیع شود؟

### 6.3.3 گسترش به نطارت صنعتی

ایده های موجود در کارهای مرتبط و این مقاله می توانند به اشکال مختلفی از یادگیری نطارتی که شامل برچسب دهی نویز دار هستند، نیز تطبیق داده شوند. بر اساس بخش ۱.۴، این نیاز به سه مدل دارد: مدل پیش بینی معمولی، مدل برچسب دهی نویز دار برای فرآیند برچسب دهی ناقص و مدل تستی که پیش بینی ها و برچسب دهی های نویز دار را مرتبط می کند. مدل برچسب دهی نویز دار مانند مدل برجسب دهی نویز دار  $Z(\pi)$  است و احتمالات تخمین ارائه می دهد. باید به نمونه های غیر دیده سُد و خارج از مجموعه  $Z$ ، (تعمیم یابد، بنابراین با روش های یادگیری ماسیون یاد گرفته می شود. مدل تست مسابقه  $\hat{y} \neq z$ ، ( $\hat{y} \neq z$ ) در  $\hat{y}(p)$  است؛ عملکرد در مجموعه تست را توصیف می کند، بنابراین با روش های نظریه ای تخمین یاد گرفته می شود.

برای آموزش، بسیاری از جنبه های آموزش MMSE ممکن است برای نطارت صنعتی قابل اعمال باشند. بخش ۳.۴ و پیوست E.۴ به مدل عمومی و مدل برچسب دهی نویز دار اجازه می دهد. اگر مجموعه برچسب دهی  $Z$  پیوسته به جای محدود باشد، جمع آوری بر روی  $Z$  باید با تعادلات جایگزین شود (مقایسه با (37)، (40)، (41)، (43)، (45)) و نیاز به تکنیک هایی برای محاسبه یا تقریب تعادلات خواهد بود.

برای تست مدل پیش بینی، کارهای اختصاصی به معیارهای خاص نیاز خواهد بود؛ روش تست  $MSE$  می تواند استراتژی عمومی ای ارائه دهد. تست باید بر اساس اصل تخمین مع يار  $RV$  در مجموعه تست به جای برچسب های صحیح باشد. اول، باید یک مدل تست مناسب و پارامترها<sup>23</sup> مقایسه با بخش ۲.۱ (شناسایی کرد و معیارها را به عنوان  $RV$  ها بیان کرد) بخشن ۲.۲. دوم، باید احتمالات پارامترها و  $RV$  های معیار را به دست آورد (بخشن ۲.۳)؛ اگر نمونه ها مستقل باشند، CLT ممکن است مفید باشد. سوم، باید الگوریتم هایی برای تخمین پارامترها (بخشن ۲.۴) توسعه داد. در نهایت، پس از تخمین پارامترها، احتمالات  $RV$  های معیار من توانند برای یافتن تخمین های بهینه معیارها استفاده شوند (بخشن ۲.۵).

## تشکرها

.Sheng et al. (2008). به عنوان مثال، بینید

برای طبقه بندی دودویی، پارامترهای  $OP$  و  $DP$  به احتمال تشخیص و احتمال خطا در تشخیص مربوط می شوند که این ساختارها معیارهای استانداردی هستند، اما طبقه بندی چندگانه ماتریس مخلوط condition را معرفی کرد که معمولاً استفاده نمی شود.

نویسنده با احترام از Anna Yanchen و Rajmonda Caceres، John Holodnak، Jason Matterer ko برای نظرات و علاقه مندی شان به این کار تشکر می کند، همچنین از Paul Monti Jason Thornton و cciolo، Vijay Gadepally، Tim Dasey کار قدردانی می کند. نویسنده همچنین از ویراستاران عملی و ناشناس برای توجه دقیق، سوالات دقیق و پیشنهادات مفیدشان تشکر می کند.

این ماده بر اساس کار پشتیبانی شده توسط معاون وزیر دفاع برای تحقیق و مهندسی تحت قرارداد هوافضای شماره FA8702-15-D-0001 است.

### پیوست A. شاخص ها به صورت RV معمول

برای احتمال خطأ نادرست، فرم آن

$$P_{\text{FA}} = \frac{\sum_{i=1}^N \mathbb{1}(\hat{y}_i = 1 \text{ and } Y_i = 0)}{\frac{1}{N} \sum_{i=1}^N \mathbb{1}(Y_i = 0)}.$$

(نمايانگر) اين است numerator

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \mathbb{1}(\hat{y}_i = 1 \text{ and } Y_i = 0) &= \frac{1}{N} \sum_{i:\hat{y}_i=1} \mathbb{1}(Y_i = 0) \\ &= \frac{1}{N} \sum_{i:\hat{y}_i=1} (1 - \mathbb{1}(Y_i = 1)) \\ &= \frac{1}{N} \sum_{i:\hat{y}_i=1} 1 - \frac{1}{N} \sum_{i:\hat{y}_i=1} \mathbb{1}(Y_i = 1) \\ &= \frac{\hat{N}_1}{N} - U, \end{aligned}$$

(نمايانگر) اين است denominator

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \mathbb{1}(Y_i = 0) &= \frac{1}{N} \sum_{i=1}^N (1 - \mathbb{1}(Y_i = 1)) \\ &= 1 - \frac{1}{N} \sum_{i=1}^N \mathbb{1}(Y_i = 1) \\ &= 1 - (U + V). \end{aligned}$$

فرم RV دقت

$$\begin{aligned}
ACC &= \frac{1}{N} \sum_{i=1}^N \mathbb{1}(Y_i = \hat{y}_i) \\
&= \frac{1}{N} \sum_{i:\hat{y}_i=0} \mathbb{1}(Y_i = 0) + \frac{1}{N} \sum_{i:\hat{y}_i=1} \mathbb{1}(Y_i = 1) \\
&= \frac{1}{N} \sum_{i:\hat{y}_i=0} (1 - \mathbb{1}(Y_i = 1)) + U \\
&= \frac{1}{N} (N - \hat{N}_1) - V + U \\
&= U - V - \hat{N}_1/N + 1.
\end{aligned}$$

فرم RV دقت

$$\begin{aligned} PREC &= \frac{\frac{1}{N} \sum_{i=1}^N \mathbb{1}(\hat{y}_i = 1 \text{ and } Y_i = 1)}{\frac{1}{N} \sum_{i=1}^N \mathbb{1}(\hat{y}_i = 1)} \\ &= \frac{\frac{1}{N} \sum_{i:\hat{y}_i=1} \mathbb{1}(Y_i = 1)}{\frac{1}{N} \hat{N}_1} \\ &= \frac{N}{\hat{N}_1} U. \end{aligned}$$

است که با گرفتن  $F\beta$

$$F_\beta = (1 + \beta^2) \frac{PREC \cdot REC}{\beta^2 PREC + REC},$$

جایگزینی جملات مربوط به PREC و REC در جدول 8 و انجام یک مقدار حساب داری به دست می آید.

#### **بیوست B. نسبت های متغیرهای تصادفی مشترک نرمال**

این بخشن مراحل مربوط به محاسبه توزیع یا تقریب لحاظ نسبت از متغیرهای تصادفی مشترک نرم‌ال

اجازه دهید  $Z' \sim \text{Marginal}(Z|W)$  باشد. تراکم  $Z'|W$  به صورت زیر به دست می‌آید:

1.  $h = \pm \sigma Z$ ,  $r = \sigma W/h$  اجازه دهد  $\rho^2$ ; علامت  $h$  به زوایی تعیین خواهد شد. همچنین اجازه دهد  $Z$  باشد.

2. علامت h را انتخاب کنید به طوری که a و b دارای علامت مسماطه باشند. سپس متغیر تصادفی W'(Z') = 'T' - ns (من تواند به صورت  $X + a + b$ ) بیان شود، جایی که X و Y متغیرهای تصادفی مستقل(0) هستند. تراکم 'T' به صورت زیر است ،

$$p(t') = \frac{e^{-(a^2+b^2)/2}}{\pi(1+t'^2)} \left( 1 + q(t') e^{(q(t'))^2/2} \int_0^{q(t')} e^{-x^2/2} dx, \right),$$

که  $\text{erf}(y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-t^2} dt$  و این intégral می تواند با استفاده از تابع خطای  $q(t') = (b + at')/(1 + t'^2)$  محاسبه شود.

نماد	PtDt,t REC	PtFA	Ftβ
—	در	$(1 + \beta^2)U$	
tV+	به صورت	$1 t (tUt + t Vt)$	$\beta * 2 * (U + V) + ^N1 / N$
—	—	$(1 + \beta^2)U$	$(1 + \beta^2)U$
μtWt	$\mu tUt + t \mu V$	$1 t (t\mu tUt + t \mu tVt)$	$\beta t2t(t\mu tUt + t \mu tVt) + ^tNt1t/N$
σt2tWt	$\sigma t2tUt + t \sigma t2tV$	$\sigma t2tUt + t \sigma t2tV$	$\beta t4t(t\sigma t2tUt + t \sigma t2tVt)$
cov(tZtt, Wt t)	—	—	$\beta t2t(1 + t \beta t2t)t\sigma t2tU$
r	$\sqrt{\frac{U}{\sigma^2 U + \sigma^2 V}}$	$\sqrt{\frac{V}{\sigma^2 U + \sigma^2 V}}$	$\sqrt{\frac{U}{\sigma^2 U + \sigma^2 V}}$

جدول 16: پارامترها برای مقیاس های سکولار RV با پس زمینه هایی برابر با نسبت W' / Z' از RV های تقریباً نرمال مشترک Z و W.

$$3. \quad )) s \notin (r(\mathcal{D}) | \cdot \cdot \cdot ) \notin (\mathcal{D} / WZ = T / WZ \text{ در نتیجه}$$

/Cauchy (t', p شامل تراکم استاندارد 1 + t^2) بنابراین ب) تراکم (t')

ه) طور فنی، مستقل های T به ترتیب بزرگتر از صفر وجود ندارند؛ یعنی،  $\int_{-\infty}^{\infty}$  برای ا در {1, 2, ...}. با این حال، (Marsaglia 2006, §4) اشاره می کند که در عمل ممکن است بتوان فرض کرد که نزخ Y + b به صفر نزدیک شود با احتمال ناچیز، که این امکان را فراهم می کند تا لحطات (با فرض این فرض) محاسبه شود.

برای مثال، اگر b = 4 باشد، سپس Marsaglia (1955) این پژوهی را تقریباً گزارش می دهد که، با فرض 4 > b (یا -4 < Y)، میانگین و واریانس T

$\Pr(b + Y \leq 0) = \Pr(Y \leq -4) \approx 3.17 \times 10^{-5}$ . میانگین و واریانس T

$$\text{و } \mu^2 T - 3.795 = (a^2 + 1)/(b^2 + 0.108b) \text{ است.}$$

در نهایت، Marsaglia همچنین مشاهده می کند که اگر a < 2.256 (و b > a) (یا  $a < 2.256$  و b > a) می تواند به طور معقولی با توزع نرمال تقریب داده شود.

### C. بررسی برآورد MMSE

برای بخش ۱.۴.۱ f(A) = f(D(A)) باید A را متغیر RV غیر مسأله شده در نظر بگیریم. B را متغیر RV مشاهده شده، و بگذارید  $f_D(A) = A^T$  یک درستور بعدی از توابع عددی از A، با  $A^T = \int_{-\infty}^{\infty} f_E(A) dA$ ، جایی که:  $D = 1, \dots, D$  و  $B(hMSE) = hD(B)^T$  را تعریف یک تخمین دهنده برای A (A به عنوان B) برابر با  $MhMMSE$  estimator با  $A(f_j) B(j) [(E = \sum_{j=1}^D f_j) A(f_j) B(hnse)]$  تعریف کنید. هدف یافتن  $MSE = \arg \min_h$  است.

برای بخش ۲.۶.۲ f(A) = f(D(A)) کاملاً بی ایستیمتری یک RV مقیاسی مانند دقت،  $A = ACC$ ،  $B = (Y^A, Z)$ ،  $f(A) = f(A)$  است. برآورد کنندگان (h2(B) = hFA(Y^A, Z, ψ, ~p(j-1)D, ~p(j-1)F و h1(B) = hD(Y^A, Z, ψ, ~p(j-1)D, ~p(j-1)FA) هستند. در بخش ۲.۶.۲ پارامترهای غیر تصادفی هستند. در بخش ۳.۴.۱ f(A) = f(A) = J\_pri(θ; x, Y) و یک برآوردگر J(Y) = J\_pri(θ; x, Y) است. برآورد Z

E.4.  $A = Y_i, B = Z_i, f(A) = L(\tilde{g}(tx_i; \theta), Y_i)$ , تابع از دسته‌ی ضرایب درینش  $.h(Z_i) = \tilde{i}(tx_i, \theta, Z_i) = \tilde{i}(Z_i)$ ، و یک برآورده است.

### ج. 1. نتایج استاندارد

1. MMSE  $htMMSEt(tBt) = Et[tft(tAt)t|tBt]$ ,  $ort htMMSEt(tBt) = Et[tftdt(tAt)t|tBt], t dt =$  فوچ العاده اولین، دومین، سومین و غیره، تابع از دسته‌ی ضرایب درینش است.  $t|tBt$  را برآورد کنند زیرا تابع از  $Bt$  است. با توجه به  $Bt = bt$ , برآورد MMSEt مقدار غیر تصادفی  $htMMSEt(tbt) = Et[tft(tAt)t|tBt = bt]$ ,  $ort htMMSEt(tbt) = Et[tftdt(tAt)t|tBt = bt]$ . Papoulis, 1991, t7-5, t8( ), Van Trees, 1968, t2.4.1( ) رجوع کنید به  $t|tBt = bt$ ,  $t|tBt = bt$ , Kay, 1993, nt10.3, (Kamen and Su, 1999, Theorems 3.1, 3.4), (Oppenheim, 2011) t8.1, t8.2).
2. برآورده MMSE بی طرف است:  $E[tft(tAt)] = E[tft(tAt)]$  (پاپولیس، 1991، 1993)، (کامن و سو، 1999)، (t3.3).
3. خطای تخمین  $htML SEt(tBt)t ft(thAt)$  دارای ویژگی‌های زیر است.
  - (ا)  $htMMSEt(tBt)t ft(tAt)n$  میانگین صفر است:  $E[htMMSEt(tBt)t ft(tAt)n] = 0$  از نتیجه 2.
  - (ب) واریانس کل برابر با  $MSE$  است که  $htMMSEt(tBt)t ft(tAt)$  را بینند:  $tD_{jt=1}^D var(htMMSEtj(tBt)t ftjt(tAt)) = mse(htMMSEt(tBt)t ft(tAt))$  (Kay, 1993, t11.6); اصل راستگونی (خطا راستگون) به هر تابع  $w(B)$  است:  $wt1t(tBt)t twtDt(tBt) Pa; tTt: Et[(htMMSEt(tBt)t ft(tAt)t Ttwt(tBt)t)t = 0]$  (Poulis, 1991, t8-3), (Kamen and Su, 1999, Theorems 3.2, 3.3), (Oppenheim and Verghese, 2015, t8.2.1).
  - (ج) تخمین کر  $MMSE$  را می‌توان به قانون واریانس کل مرتبط کرد، که بیان می‌کند که،  $E[var(tGt|tBt)]na, var(tGt) = E[var(tGt|tBt)] + var(E[tGt|tBt])$  با واریانس محدود،  $Gt$  با  $tD$  توصیح داده شده است | اغلب  $Gt$  به ترتیب: دیده شده Blitzsstein and Hwang, 2019, t9.5).

برابر است با کل واریانس غیرقابل توصیح  $htMMSEt(tBt)t$  (oft ft (tAt):t msen) (الف) واریانس کل  $htMMSEt(tBt)t$  (Kay, 1993, t10.4)، (Van Trees, 1968, t2.4)، (رجوع کنید به tTt: Et[(htMMSEt(tBt)t ft(tAt)t Ttwt(tBt)t)t = 0])

واریانس کل برآورده  $MMSE$  برابر با کل واریانس توضیح داده شده است  $tD_{jt=1}^D E[f_j(A)|tBt] = mse(htMMSEtj(tBt)n)$  (ب) فوت (ج) جمع جبری ز برای با 1، متغیر  $f_j(A)$  است که به این  $tD_{jt=1}^D var(htMMSEtj(tBt)n)$  معنی است که برآورده  $MMSE$  تا حد امکان تغییرات  $(A)$  را با توجه به  $B$  در نظر می‌گیرد.

بنابراین، واریانس کل  $f(A)$  برابر است با مجموع  $MMSE$  (واریانس غیر توصیح دهنده) و (ج) واریانس برآورده  $MMSE$  (واریانس توصیح دهنده):

$$\sum_{j=1}^D var(f_j(A)) = \underbrace{mse(h^{MMSE}(B), f(A))}_{\sum_{j=1}^D E[var(f_j(A)|B)]} + \underbrace{\sum_{j=1}^D var(h_j^{MMSE}(B))}_{\sum_{j=1}^D var(E[f_j(A)|B])}.$$

5. نتایج 2 تا 4 به برآورده  $MMSE$  مربوط می‌شوند. برای برآورده  $MMSE$  داده شده

$V$  و Oppenheim (mse(hMMSE(B), f(A) || B = b) =  $\sum D_{j=1}^D var(f_j(A) | B = b)$ ) برابر است با مجموع واریانس‌های شرطی توابعی که در حال تخمین زدن هستند:  $B = b$ , MSE

(mse(hMMSE(b), f(A) || b) =  $\sum D_{j=1}^D var(f_j(A) | b)$ ) برابر است با مجموع واریانس‌های شرطی توابعی که در حال تخمین زدن هستند:  $b$ , MSE = .) Vergheze, 2015, § 8.1 (mse(hMMSE(b), f(A) || b) =  $\sum D_{j=1}^D var(f_j(A) | b)$ ) برابر است با مجموع واریانس‌های شرطی توابعی که در حال تخمین زدن هستند:  $b$ , MSE = .) Vergheze, 2015, § 8.1

## C.2 شواهد و استنتاج های نتایج استاندارد

استنتاج 1 فرض می کند که  $B$  پیوسته است اما به راحتی قابل تغییر است اگر  $B$  ناگهانی باشد.  
گسترش MSE به صورت زیر است

$$\begin{aligned}
mse(\mathbf{h}(\mathbf{B}), \mathbf{f}(\mathbf{A})) &= \sum_{j=1}^D \left\{ \mathbb{E}[h_j^2(\mathbf{B})] - 2\mathbb{E}[h_j(\mathbf{B})f_j(\mathbf{A})] + \mathbb{E}[f_j^2(\mathbf{A})] \right\} \\
&\stackrel{(a)}{=} \sum_{j=1}^D \left\{ \mathbb{E}[h_j^2(\mathbf{B})] - 2\mathbb{E}\left[\mathbb{E}[h_j(\mathbf{B})f_j(\mathbf{A})|\mathbf{B}]\right] + \mathbb{E}[f_j^2(\mathbf{A})] \right\} \\
&= \sum_{j=1}^D \left\{ \int p(\mathbf{b})h_j^2(\mathbf{b}) d\mathbf{b} - 2 \int p(\mathbf{b})\mathbb{E}[h_j(\mathbf{B})f_j(\mathbf{A})|\mathbf{B}=\mathbf{b}] d\mathbf{b} + \mathbb{E}[f_j^2(\mathbf{A})] \right\} \\
&\stackrel{(b)}{=} \sum_{j=1}^D \left\{ \int p(\mathbf{b})h_j^2(\mathbf{b}) d\mathbf{b} - 2 \int p(\mathbf{b})h_j(\mathbf{b})\mathbb{E}[f_j(\mathbf{A})|\mathbf{B}=\mathbf{b}] d\mathbf{b} + \mathbb{E}[f_j^2(\mathbf{A})] \right\},
\end{aligned}$$

جایی که  $a$ ) از انتظارات تکراری است و  $b$ ) به دلیل اینکه  $B$ ) غیر تصادفی است و برابر با  $jh$ ) هنگامی که تحت شرط  $b = B$  قرار من گیرد، است. برای  $d = \dots \dots D$  متشق جزئی معادله بالا را در برابر  $h$  بگیرید، آن را برابر با صفر قرار دهید و برای  $d$  حل کنید. این کار پایان می شود که

$$\begin{aligned} \frac{\partial \text{mse}}{\partial h_d} &= 2 \int p(\mathbf{b}) h_d(\mathbf{b}) d\mathbf{b} - 2 \int p(\mathbf{b}) E[f_d(\mathbf{A}) | \mathbf{B} = \mathbf{b}] d\mathbf{b}, \quad d = 1, \dots, D, \\ &= 2 \int p(\mathbf{b}) [h_d(\mathbf{b}) - E[f_d(\mathbf{A}) | \mathbf{B} = \mathbf{b}]] d\mathbf{b}, \quad d = 1, \dots, D. \end{aligned}$$

ار آنجایی که  $p_t(tbt)$  یک توزیع احتمال است، این عبارت صفر است  $E_t[tftdt(tAt)|tBt] = 0$ .  
 برای هر  $t$ ،  $validt$  برابر با  $tBt$  است.  $tBt$  تعریف می‌انگین شرطی است. همچنین  $tBt$  برای  $t$ ،  $tBt = t$  است. بنابراین راه حل حداقل منحصر به فرد است. بنابراین، تخمین گر ۲

برای نتیجه ۲، جایی که  $E[tat|tAt] = E[tat] - E[tAt]$  است. تکراری را اعمال می‌کند.

برای نتیجه چای، گوشت پخته  

$$htMMSEt(tBt), t, ft(tAt), t, (tt), t = tttdjh = 1t Et[ttt[(Tahtmustj) (tBt)ttftjt(tAt) )t2t]tBt]t =$$

$$ntD$$
  
 از تعاریف تکراری، شرط بندی، ها و بنامه های، کاربردی، استفاده می کنند. نتیجه 4b به این دلیل است

htMMSEtj، t، بنابراین واریانس آنها برابر است. اعمال قانون واریانس کل  $tD$  به هر ترم در نتیجه  $c_4$  را به دست می‌دهد.

(MSE ont Bt = t bt msen برای نتیجه 5، شرطی کردن htMMSEt(tBt), t, ft (tAt) = 4) tBt = tbt  
 $tD^{jt=1}t Et[(tfjt(tAt)t | t E[tftjt(tAt)t | tBt]t) t2ttBt = t bt]t = t tD^{jt=1}t Et[(tfjt(tAt)t | t E[tftjt(tAt)t | tBt = t bt]t) t2ttBt)$   
 $tD^{jt=1}t (at | t) t2ttBt = t bt$  استفاده می کند.

#### پیوست D. بررسی پرآورد MPE و MAP

## صمیمه E. آموزش یا مسائل مربوط به حقیقت

### E.1 مدل تمایز، پارامترهای تصادفی

کیس ایده آل:  $\text{Find}_{\theta t} \text{ برای } \text{pt}(t|\theta)$  به حداکثر رساندن  $\text{pt}(t|\theta)$ , داده شده توسط Bayes قانون استفاده می کند و  $(\theta|29)$  است.

مسائل مربوط به حقیقت:  $\text{Find}_{\theta t} \text{ برای } \text{pt}(t|\theta)$  به حداکثر رساندن  $\text{pt}(t|\theta)$  و نوشت:  $\text{pt}(t|\theta) = \text{pt}(t|\theta)$

جایی که  $t$   $\text{pt}(t|\theta)$  استفاده می کند و  $\text{pt}(t|\theta) = \text{pt}(t|\theta)$  از قانون بیز استفاده می کند و  $\text{pt}(t|\theta)$  لایه ها (31).

#### E.2 مدل مولد، پارامترهای غیر تصادفی

=  $npt(tyt, th xt; t \theta t)$  Findt  $\theta t$  برای به حداکثر رساندن تابع احتمال  $pt(tyt, t xt; t \theta t)$ ، که می‌تواند به صورت

$t \text{ pt}(txt | tyt; t \theta t) tpt(tyt) = t tNit=1 t \text{ pt}(txt | tyt; t; t \theta t)$

$$\sum_{\substack{i=1 \\ j=tN \\ j=tNt}}^{\infty} yitYt \text{pt}(tztit, ytit, t xtit; t \psi tit, t \theta t) = \sum_{i=1}^{\infty} ytitYt \text{pt}(txtit | t ztit, ytit; t \psi tit, t \theta t) \text{pt}(tztit | t ytit; t \psi tit, t \theta t) \text{pt}(tytit; t \psi tit, t \theta t)$$

(tat)=  
(tbt)  
(tbt)

مشکلات حقیقت یابی: به جای آن،  $\theta$  را بیدا کنید تا  $p(tz, x; \psi, \theta)$  را به حداقل برسانید

شان جمع

### E.3 مدل مولد، پارامترهای تصادفی

کیس ایده‌آل:  $\underline{\theta_t}$  برای به حداکثر رساندن قسمت خلفی  $t \text{pt}(t\theta_t | t\text{yt}, t\text{xt})$  (داد) سپس  $\text{Findt}(\theta_t | t\text{yt}, t\text{xt})$  استفاده می‌کند (با اینکه  $t \text{pt}(t\text{yt}, t\text{xt} | t\theta_t)$   $t \text{pt}(t\theta_t | t\text{yt}, t\text{xt})$  است). از قانون Bayes داریم  $t \text{pt}(t\theta_t | t\text{yt}, t\text{xt}) = \frac{t \text{pt}(t\text{yt}, t\text{xt} | t\theta_t) t \text{pt}(t\theta_t)}{t \text{pt}(t\text{yt}, t\text{xt})}$ . با توجه به اینکه  $t \text{pt}(t\text{yt}, t\text{xt})$  مقدار ثابتی است،  $t \text{pt}(t\theta_t | t\text{yt}, t\text{xt})$  برابر با  $t \text{pt}(t\text{yt}, t\text{xt} | t\theta_t) / t \text{pt}(t\text{yt}, t\text{xt})$  خواهد بود.

مسائل مربوط به حقیقت Findt  $t\theta$  را برای به حداکثر رساندن  $\text{pt}(t\theta t|tzt,t xt;t \psi)$  پیدا کنید. نوشتن (داد)  $t\ pt(tzt,t xt|t\theta t;t \psi)\text{ptp}(t\theta t)$

#### E.4 مدل های غیر احتمالی: برآورد RV MMSE تجربی-ریسک

به صورت اختصاری، این تابع از دست دادن  $RV$  را مشخص کنید. همچنین اجازه دهد  $tRt(tZt) = n^tRt$  باشد. سپس برآورده  $\hat{t}Rt$  تخمینگر  $tRt(tZt)$  با  $E[(\hat{t}Rt(tZt)t t Rt(tYt))]$  باشد.  $mset(\hat{t}Rt(tZt), Rt(tYt)) = E[(\hat{t}Rt(tZt)t t Rt(tYt))]$  باشد.

نتیجه استاندارد (پیوست C دریا) این است که برآورده MMSE رسانایی است که اغلب  $Rt(tYt)$  داد  
 $\text{TheS}^tRtMMSEt(tZt) = E[tRt(tYt)|tZt] = \underline{nNt tNit}$  (سفید 35).  
 $tZtit(t) = 1t E[t' the(tYtit)|tZt] = t Nit t1t tNit = 1t E[t' t(tYtit)t | tZtit] =$  (36) است.  
 آخرين عبارت به اين دليل است که، برای يک تخمين گر  $\hat{t}' tit(tZtit) = \hat{t}' tit(txtit, t\theta, tZtit)$  loss-nfunc از  $t' tit(tZtit)$  =  
 $t' tit(ttit(tZtit), ttit(tYtit)) = arg min t' tit(mset(t' tit(tZtit), ttit(tYtit)))$  است.  
 حداقل  $MMSEt(tZtit) = arg min Z^2 R_{it}^t$  با ياراين:  $Z^2 R_{it}^t = \sum_{i=1}^n Y(i, t) \hat{Y}(i, t) - \bar{Y}(t) \hat{Y}(t)$  (برابر با ميانگين برآورد  
 كنندگار  $MMSE$  از  $RV$  هاي، فرمي، تابع (ناري، است).

نموده هستند، بنابراین  $E[\var(t' \cdot t) MMSEti | tZtit] = E[\var(t' \cdot t) MMSEti | tZtit]$  است. میانگین خطاهای تخمین برآوردهای تابع صرر MMSE است. خطای تخمین هر یک از این برآوردها دارای  $var(\hat{t} \cdot t MMSEti | tZtit) = 0$  است. نموده ها مستقل هستند، بنابراین  $E[\var(t' \cdot t) MMSEti | tZtit] = E[\var(t' \cdot t) MMSEti | tZtit] + E[\var(t' \cdot t) MMSEti | tZtit]$  است. نموده های تخمین برآوردهای تابع صرر MMSE نیز مستقل هستند. پس از آن CLT، خطای تخمین  $\hat{t} \cdot t MMSEti | tZtit$  در توزیع به یک  $RV$  معمولی یا میانگین صفر و واریانس  $E[\var(t' \cdot t) MMSEti | tZtit]$  من شود.

معادلات آموزنی، رگرسیون، لگاریتمی، استانداردسازی، شده

[1] ابعاد ویژگی ها را به  $D$  تعریف می کنیم، بنابراین  $\theta$  دارای ابعاد  $D + 1$  است. تعریف  $x_i^T x_j$  به صورت زیر است: و ساختارها از صفر به جای یک شروع می کنند. رگرسیون لوزیستیک از مدل  $(\sim g(x_i; \theta))$  استفاده می کند، بنابراین مدل فرض می کند که  $x_i^T \theta = p(y_i | x_i)$ ; جایی که  $\sim g(x_i; \theta) = 1/(1 + e^{x_i^T \theta})$  است. ما از اصلاح کننده  $J = 1/2N \sum (y_i - \sim g(x_i; \theta))^2$  استفاده می کنیم؛ وزن قطبی  $\theta$  در معرض اصلاح نیست و از جمع محو می شود.

## در مسائل درستی ساری در طبقه بندی تحت نطارت

Machine Learning آموزش F.1

از (26) و (30)، آموزش ML در حالت ایده‌آل به دنبال کمینه سازی  $\theta$  است

$$J_{\text{ideal}}(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{y}) = \underbrace{-\frac{1}{N} \log p(\mathbf{y}|\underline{\mathbf{x}}; \boldsymbol{\theta})}_{J_{\text{pri}}(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{y})} + \lambda J_{\text{reg}}(\boldsymbol{\theta}) = -\frac{1}{N} \sum_{i=1}^N \log p(y_i|\mathbf{x}_i; \boldsymbol{\theta}) + \frac{\lambda}{2N} \sum_{j=1}^D \theta_j^2,$$

و برخی معادلات جبرایی به دست آید

$$J_{\text{ideal}}(\boldsymbol{\theta}; \underline{\mathbf{x}}, \mathbf{y}) = -\frac{1}{N} \sum_{i=1}^N \left( (1-y_i) \log \frac{1}{1+e^{+\tilde{\mathbf{x}}_i^T \boldsymbol{\theta}}} + y_i \log \frac{1}{1+e^{-\tilde{\mathbf{x}}_i^T \boldsymbol{\theta}}} \right) + \frac{\lambda}{2N} \sum_{j=1}^D \theta_j^2. \quad (62)$$

برای گرادیان، مشتق‌های جزئی در مورد  $\theta$  به ترتیب

$$\frac{\partial J_{\text{ideal}}}{\partial \theta_j} = \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{1+e^{-\tilde{\mathbf{x}}_i^T \boldsymbol{\theta}}} - y_i \right) \tilde{\mathbf{x}}_i(j) + \frac{\lambda}{N} \theta_j \mathbb{1}(j \neq 0), \quad j = 0, 1, \dots, D. \quad (63)$$

از (28) و (32)، آموزش ML با مشکلات واقع سازی به معنای کمینه سازی

$$\begin{aligned} J_{\text{ML}}(\boldsymbol{\theta}; \underline{\mathbf{x}}, \underline{\mathbf{z}}, \underline{\boldsymbol{\psi}}, \boldsymbol{\pi}) &= \underbrace{-\frac{1}{N} \log p(\underline{\mathbf{z}}|\underline{\mathbf{x}}; \underline{\boldsymbol{\psi}}, \boldsymbol{\theta})}_{J_{\text{pri}}(\boldsymbol{\theta}; \underline{\mathbf{x}}, \underline{\mathbf{z}}, \underline{\boldsymbol{\psi}}, \boldsymbol{\pi})} + \lambda J_{\text{reg}}(\boldsymbol{\theta}) \\ &= -\frac{1}{N} \sum_{i=1}^N \log \sum_{y_i=0}^1 p(y_i|\mathbf{x}_i; \boldsymbol{\theta}) p(z_i|y_i; \boldsymbol{\psi}_i) + \frac{\lambda}{2N} \sum_{j=1}^D \theta_j^2 \\ &\stackrel{(a)}{=} -\frac{1}{N} \sum_{i=1}^N \log \left( \sum_{y_i=0}^1 p(y_i|\mathbf{x}_i; \boldsymbol{\theta}) \prod_{t:z_{i,t} \neq \emptyset} p(z_{i,t}|y_i; \delta_i, \phi_t) \right) + \frac{\lambda}{2N} \sum_{j=1}^D \theta_j^2 \\ &\stackrel{(b)}{=} -\frac{1}{N} \sum_{i=1}^N \log \left( \frac{1}{1+e^{+\tilde{\mathbf{x}}_i^T \boldsymbol{\theta}}} \prod_{t:z_{i,t} \neq \emptyset} p_{Z_{i,t}|Y_i}(z_{i,t}|0; \delta_i, \phi_t) \right. \\ &\quad \left. + \frac{1}{1+e^{-\tilde{\mathbf{x}}_i^T \boldsymbol{\theta}}} \prod_{t:z_{i,t} \neq \emptyset} p_{Z_{i,t}|Y_i}(z_{i,t}|1; \delta_i, \phi_t) \right) + \frac{\lambda}{2N} \sum_{j=1}^D \theta_j^2, \end{aligned}$$

جایی که (a) اعمال می‌شود  $\boldsymbol{\psi} = \emptyset$  (با توجه به (61)، و (b) به دلیل  $y_i$  و  $z_{i,t}$  در  $\boldsymbol{\theta}_{i,t}$  هستند).

برای محاسبه‌ی جریان، دیفرانسیل گیری و برخی معادلات جبری منجر به

$$\begin{aligned} \frac{\partial J_{\text{ML}}}{\partial \theta_j} &= \frac{1}{N} \sum_{i=1}^N \frac{\prod_{t:z_{i,t} \neq \emptyset} p_{Z_{i,t}|Y_i}(z_{i,t}|0; \delta_i, \phi_t) - \prod_{t:z_{i,t} \neq \emptyset} p_{Z_{i,t}|Y_i}(z_{i,t}|1; \delta_i, \phi_t)}{\prod_{t:z_{i,t} \neq \emptyset} p_{Z_{i,t}|Y_i}(z_{i,t}|0; \delta_i, \phi_t) + e^{+\tilde{\mathbf{x}}_i^T \boldsymbol{\theta}} \prod_{t:z_{i,t} \neq \emptyset} p_{Z_{i,t}|Y_i}(z_{i,t}|1; \delta_i, \phi_t)} \\ &\quad \cdot \frac{1}{1+e^{-\mathbf{x}_i^T \boldsymbol{\theta}}} \tilde{\mathbf{x}}_i(j) + \frac{\lambda}{N} \theta_j \mathbb{1}(j \neq 0), \quad j = 0, 1, \dots, D. \end{aligned}$$

## آموزش آنلاین F.2 MMSE

برای آموزش آنلاین، اولین اصطلاح در (62) را به عنوان خطر تجربی (33) بازنویسی می کنیم با تنظیم  $\log y \rightarrow \tilde{y}$  و  $\log(1y - (1 -)) \rightarrow y$  (62) را تعریف می کنیم  $L(\tilde{x}^T \theta; \tilde{y}) = \frac{1}{2} (\tilde{y} - \tilde{x}^T \theta)^2$

$$\frac{\partial}{\partial \theta_j} [L(\tilde{g}(\tilde{x}; \theta), \tilde{y})] = \left( \frac{1}{1 + e^{-\tilde{x}^T \theta}} - \tilde{y} \right) \tilde{x}(j). \quad (64)$$

برای محاسبه ای جریان، جایگذاری این معادله در (48) و (49) دوباره منجر به (63) می شود.  
برای آموزش MMSE با مشکلات حقیقی، (40) را اعمال می کنیم و آن را به حداقل می رسانیم

$$\begin{aligned} J_{\text{MMSE}}(\theta; \underline{x}, \underline{z}, \underline{\psi}, \pi) = & -\frac{1}{N} \sum_{i=1}^N \left( p_{Y|Z}(0|z_i; \psi_i) \log \frac{1}{1 + e^{+\tilde{x}_i^T \theta}} \right. \\ & \left. + p_{Y|Z}(1|z_i; \psi_i) \log \frac{1}{1 + e^{-\tilde{x}_i^T \theta}} \right) + \frac{\lambda}{2N} \sum_{j=1}^D \theta_j^2. \end{aligned}$$

برای محاسبه ای جریان، جایگذاری (64) (در) (50) و ساده سازی منجر به

$$\begin{aligned} \frac{\partial J_{\text{MMSE}}}{\partial \theta_j} = & \frac{1}{N} \sum_{i=1}^N \left( p_{Y|Z}(0|z_i; \psi_i) \frac{1}{1 + e^{-\tilde{x}_i^T \theta}} - p_{Y|Z}(1|z_i; \psi_i) \frac{1}{1 + e^{+\tilde{x}_i^T \theta}} \right) \tilde{x}_i(j) \\ & + \frac{\lambda}{N} \theta_j \mathbb{1}(j \neq 0), \quad j = 0, 1, \dots, D. \end{aligned}$$

## منابع

. CRC Press, Boca Raton. مقدمه ای بر احتمالات Jessica Hwang و Joseph K. Blitzstein

استیو برانسون، گرانت ون هورن و پیترو پرونزا. جمع سپاری ناب: Com-Int IEEE Computer Vision اتصال انسان و ماشین در یک سیستم آنلاین و تشخیص الگو(t)، صفحات 61096118، doi: 10.1109/ CVPR.2017.647. URLT [https://openaccess.thecvf.com/content\\_cvpr\\_2017/html/nanson\\_lean\\_crowdsourcing\\_combining\\_cvpr\\_2017\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2017/html/nanson_lean_crowdsourcing_combining_cvpr_2017_paper.html).

لئو بریمن. مدل سازی آماری: دو فرهنگ

M. P. Burl, M. C. Burl, U. M. Fayyad, P. Perona, P. Smyth و خودکارسازی شکار آتشفت. شان ها در زهره. بینایی IEEE کامپیوتر و تشخیص الگو(t)، pagesn302309, 1994. doi: 10.1109/ CVPR.1994.323844.

مارک جی. کارلوتو. تأثیر اشتباہات در صدق زمینی بر دقت طبیه بندی. Intl. J. Remote Sensing, t, 2009, 30(18):48314849. doi: 10.1080/01431160802672864.

George Casella. روش های بیایس تجربی را نشان می دهد. Chemometrics and Intelligent Laboratory Systems, 1992, 125-107:2(16), اکتبر 1992. doi: 10.1016/0169-7439(92)80050-E.

ژسوس سید-سوئیرو. صنرهای مناسب برای یادگیری از برچسب‌های جزئی. *Int Neural Information Proce*. Curran Associates, Inc. 15651573. صفحات (NeurIPS) 1-ssing Systems. دسامبر 2012. URL <https://dl.acm.org/doi/10.5555/2999134.2999309t>.

خسوس سید-سوئیرو، داریو گارسیا-گارسیا و رائول سانتوس-رودریگز. سارگاری زیان برای یادگیری از برچسب‌های صنیعه. یادگیری ماشین و کشف دانش بین المللی در پایگاه داده، جلد 8724 از یادداشت های سخنرانی در علوم کامپیوتر، صفحات 13-10.1007/978-3-662-44848-9. Springer. 2014. doi: 197210.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, New York, NY, USA, 1st edition, 1991. doi for 2nd edition: 10.1002/047174882X.

الف. فیلیپ داوید و آلن ام. اسکن. برآورد حداقل احتمال نزخ خطای مشاهده گر با استفاده از الگوریتم Series C (AppliednStatistics) 28(1):2028, 1979. do M.t i: 10.2307/2346806.

آرتور پی. دمپستر، نان ام. لیرد و دونالد بی. روین. برآورد حداقل احتمال از داده‌های ناقص از طریق الگوریتم Series B (Methodological) 39(1):138, 1977. URL <https://www.jstor.org/stable/2984875T>.

Krishnakumar Balasubramanian, Pinar Donmez, Guy Lebanon و. یادگیری بدون نظرت بر نطارت: تخمین خطاهای طبقه بندی و پیش‌بینی بدون برچسب. مجله تحقیقات یادگیری ماشین (JMLR)، شماره 11، 1351-1323: 44. اوریل 2010. URL <https://www.jmlr.org/papers/v11/donmez10a.html>.

<https://archive.org/details/CaseyGraff> و <https://archive.DheeruDua> مخزن یادگیری ماشین Casey Graff و Dheeru Dua UCI. URL <https://archive.org/details/MachineLearningWorkshop> مخزن یادگیری ماشین Casey Graff و Dheeru Dua UCI. URL <https://archive.org/details/lecturenotesincomputerscience> چالش‌های یادگیری (MLCW) 3944 از MLCW Workshop، ۱۷۶-۱۱۷ صفحات. Springer. 2006. doi: 10.1007/117367908.

Machine و همکاران. چالش‌های تصویری اشیاء PASCAL در سال ۲۰۰۲. در Mark Everingham lecture notes in computer science چالش‌های یادگیری (MLCW) 3944 از MLCW Workshop، ۱۷۶-۱۱۷ صفحات. Springer. 2006. doi: 10.1007/117367908.

Michel Verleysen و Benoît Frénay. طبقه بندی در حضور نویز برچسب: یک مرور. IEEE Trans. Neural Networks and Learning Systems, 25(5):845-869, 2014. doi: 10.1109/TNNLS.2013.2292894.

Robert E. Schapire و Yoav Freund. یک تعمیم نظریه تصمیم‌گیری به یادگیری آن‌لاین و کاربرد آن در تقویت. Computer and System Sciences, 55(1):119-139. آگوست ۱۹۹۷. doi: 10.1006/jcss.1997.1504.

Donald B. Rubin و Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari. تحلیل داده‌های بیزیان. Chapman and Hall-CRC, Boca Raton, FL. 2013. doi: 10.1201/b16018.

جان ق. هولووناک، جیسون ق. ماتر و ویلیام دبلیو استرایلین. تخمین دقیق طبقه بندی کننده با استفاده از برچسب‌های خبره پرس و صدا. گزارش فنی 1225 آزمایشگاه MIT لینکلن، M Lexington, A، ایالات متحده، ۲۰۱۸. آنویه ۲۰۱۸.

AmazonnM. Jing Wang, Panagiotis G. Ipeirotis, Foster Provost t, nWorkshop(SIGKDD) کشف دانش و داده کاوی(echanical Turk. Int ACM Intl. Conf .doi: 10.1145/1837885. 2010, 6467, صفحات 1837906. on Human Computation (HCOMP)

ایشان جیندال، متیو نوکلین و ژووون چن. یادگیری شبکه های عمیق از برچسب های نویزین با منظم سازی حذف. بین المللی IEEE Conf. داده کاوی(Conf. ICDM). IEEE Intl. 2016, pagesn967972. doi: 10.1109/ICDM.2016.0121.

ادوارد دبلیو. کامن و جاناتان ک. سو. مقدمه ای بر برآورد بهینه. Springer-nVerlag, London, UK, 1999. doi: 10.1007/978-1-4471-0417-9.

دیوید آر. کارگر، سونک اوه و دواورات شاه. تخصیص و طایف بهینه بودجه برای سیستم های جمع سپاری قابل اعتماد. Operations Researcht, 62(1):124. doi: 10.1287/opre.2013.1235.

استیون M. Kay. مبانی پردازش سیگنال آماری: نظریه تخمین. PrenticeHall PTR رودخانه فوقانی زین، نیوجرسی، ایالات متحده آمریکا، 1993.

استیون M. Kay. مبانی پردازش سیگنال آماری: نظریه تشخیص. PrenticeHall PTR رودخانه فوقانی سادل، نیوجرسی، ایالات متحده آمریکا، 1998.

صدرا با برچسب تکی. بین المللی بین المللی بارتمایی های یادگیری(ICLR). Conf. دسامبر 2018. آدرس https://bnrfo.net/vrmd=h1sohjb0zt.arxiv.1712.04577/10.48550

دافنه کولر و نیر فریدمن. مدل های گرافیکی احتمالی: اصول و فناوری. MIT Press, کمبریج، ایالات متحده آمریکا، 2009.

چاک پی لام و دیوید جی استورک. ارزیابی طبقه بندی کننده ها با استفاده از داده های آزمایش با برچسب های نویزدار. بین المللی بین المللی کنفرانس مستشرک هوش مصنوعی(IJCAI). صفحات 513518 Kaufmann. 2003. URLt http://ijcai.org/Proceedings/03/Papers/076.pdf.

tPatern Recognitiont, 25(1):7987, nJanuary 1992. doi: 10.1016/0031-3203(92)90008-7.

جورج مارسالیا. نسبت های متغیرهای نرمال و نسبت های مجموع متغیرهای یکنواخت.J. Statistical Assoc.t, 60(309):193204, March 1965. doi: 10.2307/2283145

.t J. Statistical Software, 16(4):110, May 2006. doi: 10.18637/jss.v016.i04.

ناگاراجان ناتاتاراجان، ایندرجیت اس دیلوون، پرادیپ راویکور، و آمیogh تواری. یادگیری با برچسب های پر سر و صدا سیستم های پردازش اطلاعات عصبی داخلی (نوروپ ها) Curra, 1196n1204. t, صفحات 29999745T. URLt https://dl.Acm.org/doi/10.5555/n29999611. 2013..n Associates Inc

Ambuj Tewari و Nagarajan Natarajan, Inderjit S. Dhillon, Pradeep Ravikumar .J. Machine Learning Research (JMLR)18(155). ری حساس به هزینه با برچسب های نویز. URLt https://jmlr.org/papers/v18/15-226.htmlt. 2018, آوریل 133

اندرو Ng Y. و مایکل I. جردن. در مورد طبقه بندی کننده های افتراقی در مقابل مولد: مقایسه ای از رگرسیون لجستیک و بیز ساده لوح. Int Neural Information Processing Systemsn(NeurIPS)t, pages 841848. MIT Press, January 2001. URLt <https://dl.acm.org/doi/n10.5555/2980539.2980648t>.

کورتیس نورثکات، لو حیانک و آیزاك چوانک. یادگیری مطمئن: تخمین برچسب های مجموعه داده عدم قطعیت ل. تحقیقات هوش مصنوعی(JAIR) t, 70:13731411, 2021a, مه 2021. ISSN 1076-9757. URLt <https://doi.org/10.1613/jair.1.12125t>.

کورتیس جی نورثکات، آنیس آتالیه و جوناس مولر. خطاهای برچسب فراگیر در مجموعه های آزمایشی، معیارهای یادگیری ماشین را بثبات من کند. Int Neural Information Processing Systems n ردیابی مجموعه داده ها و معیارها. Curran Associates, Inc. 2021b دسامبر (NeurIPS) URLt <https://doi.org/10.48550/arXiv.2103.14749t>.

Alan V. Oppenheim and George C. Verghese.t Signals, Systems and Inferencet. PearsonnEducational, 2015, لندن، انگلستان.

آтанاسیوس پاپولیس. احتمال، متغیرهای تصادفی، و فرآیند تصادفی. McGraw-nHill، نیویورک، نیویورک، ایالات متحده آمریکا، چاپ سوم، 1991.

امانویل آ. پلاتانیوس، آوبناوا دوبی، و نام میچل. برآورد دقیق از داده های بدون برچسب: یک رویکرد بیزی بین المللی بین المللی Conf. یادگیری ماشین(ICML) t, pagesn14161425, ژوئن 2016. URLt <https://proceedings.mlr.press/v48/platanios16.pdf>.

الکساندر راتنر، استیون باخ، هنری ارنبرگ، جیسون فرایز، سن وو و کریستوفرن ری. Snorkel: ایجاد سریع داده های آموزشی با نطارت ضعیف. بین المللی بین المللی VLBD(t), volume 11 (3), pages 269282. VLDB Endowment, November 2017. doi: 10.14778/3157794.3157797.

الکساندر جی. راتنر، کریستوفرن ام. دی سا، سن وو، دانیل سلسام و کریستوفرن ری. برترانه نویسی داده: ایجاد مجموعه های آموزشی بزرگ، به سرعت. Int Neural Information Processing Systems t, صفحات 35673575 (NeurIPS) 2016, دسامبر Curran Associates, Inc. URLt <https://dl.acm.org/doi/10.5555/3157382.3157497t>.

Vikat C. Raykar, Linda H. Zhao, Gerod Hemoslo Vaadez, Charles Florin, Luka Boyoni, Linda Mo. Learants from chroms.TJ. URLt <https://j.csail.mit.edu>

تاكايا سایتو و مارک رسمایر هنگام ارزیابی طبقه بندی کننده های بازنی بر روی مجموعه داده های نام تعادل، نمودار فراخوانی دقیق آموزنده تر از نمودار thenROC است. tPublic Library of Science (PLOS) ONE, 10(3):121, March 2015. doi: 10.1371/journal.pone.0118432.

t Machine Learningt, 5:197227, Junen1990. doi: 10.1007/BF00116037.

ویکتور اس. شنک، فاستر پرووست، و پانگیوتیس جی. ایپریوتیس. برچسب دیگری دریافت کنید که کی فیت داده و داده کاوی را با استفاده از چندین برچسب پر سر و صدا بهبود می بخشد. Int ACM Intl SIGKDD Conf. کشف دانش و داده کاوی (SIGKDD) t, صفحات 622-614. doi: 2008. 10.1145/1401890.1401965.

پادرأئیک اسمیت، اسامه فیاد، مایکل بدل، پیترو پروننا، و پیر بالدی. استنتاج حقیقت زمینه ای از برچسب گذاری ذهنی تصاویر زهره. In Neural Information Processing Systems (NeurIPS) t, p. URLt <https://dl.acm.org/doi/10.1109/NeurIPS46550.2023.9753302>. MIT Press Arj/de/10.khkhkh/2998687.asqaa.

.Rob Fergus و Sainbayar Sukhbatar, Joan Bruna, Manohar Paluri, Lubomir Bourdev آموزش شبکه های کانولوشنال با برچسب های نویزدار. بین المللی بین المللی Conf. نمایندگی های یادگیری (CLR) t, آوریل 2015. doi: 10.48550/arXiv.1406.2080

ریوتارو تانو، اردوان سعیدی، سوامی سانکارا رایانان، دانیل سی الکساندر و ناتان سیلبرمن: یادگیری از برچسب های پرس و صدا با تخمین منظم سردرگم حاشیه نویسی. بینایی کامپیوترا و تشخیص الگو CVPR t, صفحات 11236n11245. 2019. doi: 10.1109/CVPR.2019.01150

گرانت ون هورن، استیو برانسون، اسکات لواری، سرژ بلونگی و پیترو پرنونا. جمع سپاری چند طبقه ناب. IntIEEE Computer Vision and Pattern Recognition (CVPR) t, صفحات 27142723. 2018. doi: 10.1109/CVPR.2018.00287. URLt [https://openaccess.thecvf.com/content\\_cvpr\\_2018/papers/Van\\_Horn\\_Learn\\_Multiclass\\_Crowdsourcing\\_CVPR\\_2018\\_paper.pdf](https://openaccess.thecvf.com/content_cvpr_2018/papers/Van_Horn_Learn_Multiclass_Crowdsourcing_CVPR_2018_paper.pdf).

Robert C. Williamson و Brendan van Rooyen. نظریه ای در مورد یادگیری با برچسب ها. URLt <https://jmlr.org/papers/v18/16-315.html>. J. Machine Learning Research (JMLR) 50-1(18), 228(18), 2018, ژوئیه 2018

Harry L. Van Trees. نظریه تشخیص، تخمین و مدوله سازی. بخش اول: تشخیص، تخمین و نظریه مدوله سازی خطی. جان ولی و ایلی و سانز، نیویورک، نیویورک، ایالات متحده، 1968. doi: 10.1002/0471221082

ولادیمیر واپنیک. اصول کاهش ریسک برای نظریه یادگیری. در سیستم های اطلاعات عصبی (NeurIPS) t, صفحات 831-838. 1991. موگان-کافمن، دسامبر 1991. URLt <https://dl.acm.org/doi/10.5555/2986916.2987018>.

Cun-Hui Zhang, median و Yehuda Vardi. چندمتغیر L1 و عمقداده مرتبط. Proc. Natl. Acad. Sci. USA (PNAS) 97.4.1423. doi: 10.1073/pnas.2000.1426-1423(97)497, 2000, فروردین 1426-1423

Pietro Perona, Peter Welinder و Pietro Perona. crowdsourcing و Peter Welinder آنلاین: ارزیابی annotators و دریافت برچسب های مقرن به صرفه. در (CVPR)، IEEE Computer Vision and Pattern Recognition (CVPR) t, صفحات 32-325. 2010, ژوئن 2010. doi: 10.1109/CVPRW.2010.5543189.

Pietro Perona, Peter Welinder, Steve Branson, Serge Belongie Curran, Neural Information Processing Systems (NeurIPS) t, صفحات 2432-2432. 2010. URLt <https://dl.acm.org/doi/10.5555/2997046.2997166>.

Paul L. Ruvolo و Jacob Whitehill, Ting-fan Wu, Jacob Bergsma, Javier R. Movellan. چه رای هایی باید بیشترین اهمیت را داشته باشد: یکپارچه سازی ایده آل از برچسب های برگزاران با تخصص ناشناخته. در سیستم های پردازش اطلاعات عصبی (NeurIPS) t, صفحات 2035-2043. 2023, Curran.

در مسائل درستی ساری در طبقه بندی تحت نظر از

شرکای، Inc.. دسامبر 2009 URL <https://dl.acm.org/doi/10.5555/298401>  
2984321

Nihar B. Shah. Entrop<sub>9</sub>, Dengyong Zhou, Qiang Liu, John C. Platt, Christopher Meek  
crowdsourcing. URL <https://arxiv.org/abs/1503.07240> محدود شده برای minimax ال condition  
۵۱۰۲. مارس ۰۳. ۰۷۲۴۰