



تحلیل هوشمند تصاویر زیست پزشکی

نیم سال اول ۰۴-۰۳

مدرس: محمدحسین رهبان

سوالات

تمرین پنجم

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- در طول ترم امکان ارسال با تاخیر تمرین ها بدون کسر نمره تا سقف ۱۵ روز وجود دارد. محل بارگزاری جواب تمرین ها بعد از ۴ روز بسته خواهد شد و پس از گذشت این مدت، پاسخ های ارسال شده پذیرفته نخواهند شد.
- هم فکری در انجام تمرین مانعی ندارد، فقط توجه داشته باشید که پاسخ تمرین حتما باید توسط خود شخص نوشته شده باشد. همچنین در صورت هم فکری در هر تمرین، در ابتدای جواب تمرین نام افرادی که با آن ها هم فکری کرده اید را حتما ذکر کنید.
- برای پاسخ به سوالات نظری در صورتی که از برگه خود عکس تهیه می کنید، حتما توجه داشته باشید که تصویر کاملا واضح و خوانا باشد. در صورتی که خوانایی کافی را نداشته باشد، تصحیح نخواهد شد.
- در صورت وجود هرگونه ابهام یا مشکل، در کوئرا یا گروه تلگرام درس آن مشکل را بیان کنید و از پیغام دادن مستقیم به دستیاران آموزشی خودداری کنید.

بخش نظری (۶۰ نمره)

سوال اول - یادگیری فدرال (۱۰ نمره)

مقاله [Federated Learning on Non-iid Data via Local and Global Distillation](#) را مطالعه کنید و به سوالات زیر پاسخ دهید.

(آ) بیشتر الگوریتم های یادگیری فدرال بر اساس روش FedAvg طراحی شده اند، که در آن کلاینت ها مدل های local خود را آموزش می دهند، آن ها را به سرور ارسال می کنند، و سرور با میانگین گیری پارامترهای این مدل ها، یک مدل global ایجاد می کند. این روش در شرایطی که داده ها به طور یکسان توزیع شده باشند (iid) به خوبی عمل می کند، اما عملکرد آن در شرایط داده های ناهمگون (Non-IID) به شدت کاهش می یابد. چرا الگوریتم FedAvg در مواجهه با داده های Non-IID دچار چالش می شود؟

(ب) برای مقابله با چالش های داده های Non-IID در یادگیری فدرال و غلبه بر محدودیت های روش های پیشین، این مقاله الگوریتم جدیدی به نام FedSND را پیشنهاد می کند. ایده اصلی این الگوریتم استفاده از Knowledge Distillation برای بهینه سازی فرآیند آموزش از طریق Self-Distillation در سمت کلاینت و Noise-Distillation در سمت سرور است. این تکنیک ها به صورت هماهنگ کار می کنند تا مدل های local را تصحیح کرده و آن ها را به صورت global هم تراز کنند. بر این اساس، به سوالات زیر پاسخ دهید:

- در سمت کلاینت ماژول self-distillation چگونه آموزش local را تقویت می کند و مزایای آن برای یادگیری فدرال چیست؟
- نمونه های noise چگونه تولید می شوند و چرا به جای داده های واقعی از noise استفاده می شود؟ این نمونه ها چه مزایایی برای یادگیری فدرال فراهم می کنند؟
- ماژول noise-distillation چگونه مدل های clients را هم تراز می کند و چرا این روش در کاهش مشکل weight-shift مؤثر است؟

✓ سوال دوم - تفسیرپذیری (۱۰ نمره)

۱. روش Example-based یکی از روش های تفسیرپذیری می باشد. در این روش تاثیر داده های آموزش روی تصمیم گیری مدل هنگام دریافت یک نمونه داده، در زمان تست بررسی می شود. برای اینکار یک Score بین تمام داده های آموزش و نمونه داده تست محاسبه می شود.

بر اساس مقاله [Understanding Black-box Predictions via Influence Functions](#) به سوالات زیر پاسخ دهید.

(آ) جزئیات محاسباتی رویکرد مطرح شده در بخش دوم مقاله را توضیح کنید.

(ب) چالش های محاسبه کارآمد influence بیان کرده و راهکارهای موجود برای حل آن را توضیح دهید. (برای این منظور می توانید به بخش سوم مقاله مراجعه کنید)

(ج) در تعریف مسئله پایه فرض شده که تابع هزینه دوبار مشتق پذیر و محدب می باشد درحالیکه در شبکه های عصبی چنین فرضیاتی در عمل ممکن نیست. آیا محاسبه influence functions در این حالت همچنان می تواند معنی دار باشد؟ بر اساس بخش ۴ مقاله توضیح دهید.

۲. یکی از روش های تفسیرپذیری Attributions Methods است. پیرامون این روش به سوالات زیر پاسخ دهید.

(آ) میزان اثرگذاری هر مولفه ورودی با چه معیاری اندازه گیری می شود و دلیل منطقی بودن این معیار را بیان کنید. در شبکه های عصبی با چه روشی معیار مذکور محاسبه می شود. با ذکر یک مثال عددی ساده محاسبه مذکور را نشان دهید.

(ب) سه چالش معیار فوق را بیان کرده و برای هر چالش، راهکاری را جهت رفع آن ذکر کنید.

سوال سوم - بخش بندی و تشخیص (۱۵ نمره)

۱. مقاله [Attention U-Net: Learning Where to Look for the Pancreas](#) را مطالعه کرده و به سوالات زیر پاسخ دهید.

(آ) چگونه بکارگیری Attention Gate ها در معماری U-Net توانایی مدل را در مدیریت تنوع شکل و اندازه ی اندامها در تسک بخش بندی (Segmentation) بهبود می بخشد؟

(ب) این مقاله به توجه افزایشی (additive attention) به عنوان یک مکانیزم برای محاسبه ضرایب گیت (gating coefficients) می پردازد. فرمول بندی ریاضی این مکانیزم را توصیف کرده و توضیح دهید که چرا توجه افزایشی در این زمینه نسبت به توجه ضربی (multiplicative attention) ترجیح داده شده است.

(پ) براساس نتایج ارائه شده در مقاله، چه نتیجه گیری هایی می توان در مورد مزایا و محدودیت های استفاده از Attention U-Net برای بخش بندی تصاویر پزشکی نسبت به رویکردهای multi-modal انجام داد؟

۲. به سوالات زیر در رابطه با U-Net پاسخ دهید.

(آ) طراحی معماری مدل U-Net را توضیح دهید. به طور خاص، نقش encocer-decoder و skip connections را شرح داده و بگویید این عناصر چگونه به اثربخشی U-Net در وظیفه بخش بندی تصاویر (Image Segmentation) کمک می کنند.

(ب) چه چیزی معماری U-Net را از سایر سیستم‌های بخش‌بندی تصویر مانند شبکه‌های کاملاً پیچشی (Fully Con- volutional Networks - FCN) متمایز می‌کند؟ سپس چند مورد از مزایا و معایب U-Net در مقایسه با FCN را نام ببرید.

(پ) نقش تابع هزینه را در فرآیند آموزش U-Net تحلیل کنید. تابع هزینه Cross-Entropy و تابع هزینه Dice را برای وظایف بخش‌بندی مقایسه و تفاوت‌های آن‌ها را توضیح دهید. در چه شرایطی یکی از این دو تابع را به دیگری ترجیح می‌دهید؟ یک تابع هزینه ترکیبی پیشنهاد دهید و مزایای بالقوه آن را توضیح دهید.

(ت) U-Net چگونه وظایف بخش‌بندی تصویر چندکلاسی (multi-class picture segmentation) را مدیریت می‌کند؟ دو مورد از چالش‌های مرتبط با این تسک را نام برده و راه حل پیشنهادی خود را ارائه دهید.

(ث) معماری U-Net را در نظر بگیرید:

- تصور کنید که ابعاد تصویر ورودی یک شبکه U-Net، 256×256 می‌باشد. حال فرض می‌شود که در این معماری هر لایه در انکدر ابعاد را به نصف کاهش می‌دهد و در دیکدر دو برابر می‌کند. در پایین‌ترین لایه (عمیق‌ترین لایه) این معماری، فضای ویژگی ما چند پیکسل خواهد داشت؟
- فرض کنید انکدر دارای لایه‌هایی با $256, 128, 64$ و 512 فیلتر است. اگر هر لایه کانولوشن از کرنل‌های 3×3 استفاده کند، تعداد پارامترهای لایه کانولوشن دوم انکدر را محاسبه کنید.

سوال چهارم - تحلیل خطا (۱۵ نمره) Claude, DeepSeek

امروز شما را به آینده می‌بریم؛ آینده‌ای که دیگر خبری از استرس امتحان‌های میان‌ترم و جزوه‌نویسی دقیقه نودی نیست. شما از دانشگاه فارغ‌التحصیل شده‌اید و چند سالی است که در شرکت‌های مختلف مشغول به کار هستید. پس از مدتی به این نتیجه رسیده‌اید که بهتر است به‌جای صعود از نردبان ترقی دیگران، نردبان خودتان را بسازید. به همین دلیل کسب‌وکار خودتان را راه‌اندازی کرده‌اید: یک وب‌سایت پرمحتوا و آموزنده که الهام‌بخش هزاران کاربر است.

اما حالا که بازدیدکننده‌ها به سایت شما علاقه‌مند شده‌اند، به فکر درآمدزایی از آن افتاده‌اید. راه‌حل؟ تبلیغات آنلاین! شما می‌خواهید تبلیغات را به بهینه‌ترین شکل ممکن نمایش دهید و از علم دوران دانشگاه استفاده کنید. در هر جایگاه سایت، به‌جای نمایش تصادفی تبلیغات، یک حراجی (Auction) برگزار می‌کنید تا تبلیغی که بیشترین درآمد را برای شما دارد، نمایش داده شود.

برای این منظور، ابتدا داده‌هایی از تبلیغات به‌صورت تصادفی نمایش داده شده در سایت جمع‌آوری کرده‌اید. ساختار داده‌ها به‌صورت زیر است:

adv_id	cam_id	bid	type	pos_id	ad_id	views	clicks	actions	week_id
1234	6575	1000	Click	10	8975	15436	15	0	1735115563
1234	6575	1000	Click	13	6735	18466	10	0	1735115563
4321	9876	9000	Action	78	7185	10321	20	2	1735124569
5678	4532	500	Click	5	1024	21000	25	0	1735114563
2345	3456	2000	Click	20	2341	18000	30	0	1735118563
7890	7654	8000	Action	25	6523	12000	18	5	1735134563

شرح جدول:

- **adv_id**: شناسه تبلیغ‌دهنده‌ای که کمپین‌های خود را در سایت شما اجرا می‌کند.
- **cam_id**: شناسه کمپین تبلیغاتی که شامل مجموعه‌ای از تبلیغات است.
- **bid**: مبلغی که تبلیغ‌دهنده برای هر کلیک یا اکشن خاص پرداخت می‌کند (به تومان).

• **type**: نوع کسب درآمد شما از کمپین:

– **Click**: مبلغ برای هر کلیک پرداخت می‌شود.

– **Action**: مبلغ برای انجام یک اکشن خاص پرداخت می‌شود.

• **pos_id**: شناسه جایگاه تبلیغ در سایت.

• **ad_id**: شناسه تبلیغی که نمایش داده شده است.

• **views**: تعداد دفعات نمایش تبلیغ.

• **clicks**: تعداد کلیک‌های انجام شده روی تبلیغ.

• **actions**: تعداد اکشن‌های خاصی که پس از کلیک انجام شده است.

• **week_id**: زمان شروع هفته‌ای که داده‌ها در آن جمع‌آوری شده‌اند.

با استفاده از این داده‌ها و فرمول‌های بهینه‌سازی که در ادامه ارائه شده است، می‌توانید انتخاب تبلیغات را در جایگاه‌های مختلف بهینه کنید تا درآمد شما حداکثر شود (البته در واقعیت این فرآیند پیچیده‌تر است، اما به هر حال!). به این روابط دقت کنید و سعی کنید درک کنید چرا این فرمول‌ها می‌توانند به انتخاب تبلیغاتی با بیشترین درآمد مورد انتظار کمک کنند.

فرمول بهینه‌سازی تبلیغات کلیکی

For each position p select: $\arg \max_{ad \in \mathcal{A}(p)} (bid_{ad} \times ctr_{(ad,p)})$

فرمول بهینه‌سازی تبلیغات اکشنی

For each position p select: $\arg \max_{ad \in \mathcal{A}(p)} (bid_{ad} \times ctr_{(ad,p)} \times cvr_{ad})$

• $\mathcal{A}(p)$: مجموعه تبلیغاتی که امکان نمایش در جایگاه p دارند.

• bid_{ad} : مبلغی که تبلیغ‌دهنده برای هر کلیک پرداخت می‌کند.

• $ctr_{(ad,p)}$: احتمال کلیک کاربر روی تبلیغ ad در جایگاه p (نرخ کلیک).

• cvr_{ad} : احتمال انجام اکشن توسط کاربر پس از کلیک روی تبلیغ.

فرض کنید مدل‌های آموزش داده‌شده در مواجهه با ورودی‌های ناشناخته (مانند تبلیغات، کمپین‌ها یا تبلیغ‌دهندگان جدید)، به صورت میانگین‌گیری سیستماتیک عمل می‌کنند. به طور مشخص، اگر یک تبلیغ جدید باشد اما کمپین مرتبط با آن قبلاً در سیستم دیده شده باشد، CTR و CVR آن تبلیغ به صورت میانگین وزنی از مقادیر مربوط به تبلیغات قبلی آن کمپین محاسبه می‌شود.

۱. چگونگی تقسیم داده‌ها

یکی از مهم‌ترین مراحل برای آموزش هرمدلی، تقسیم داده‌ها به مجموعه‌های $Train$ ، Dev ، $Test$ و $Train-Dev$ است. این کار به ما کمک می‌کند تا عملکرد مدل را بر روی توزیع داده‌های مختلف بررسی کنیم. بطور اینکار را باید انجام داد و در این انتخاب این مجموعه‌ها به چه نکاتی باید توجه داشت؟

۲ تحلیل خطا

۱. در هر یک از سناریوهای زیر، به طور مختصر مشکل را معرفی کنید و راهکاری برای آن ارائه دهید؟
- از قطعیت داده‌های ورودی اطمینان داریم ولی خطای آموزش مدل (Training Error) بالاست.
 - خطای آموزش مدل پایین است ولی خطای آن روی مجموعه (Train-Dev) همچنان بالاست.
 - خطای مدل در مجموعه‌های Train و Train-Dev پایین است ولی روی مجموعه Dev خطا زیاد است.
 - خطا Dev پایین است ولی روی مجموعه Test خطا همچنان زیاد است.
۲. آیا در اولین سناریو مطرح شده در قسمت قبل افزایش سائز داده‌های آموزش راه حل خوبی خواهد بود؟

۳. فرض کنید مدلهایی که برای پیش‌بینی نرخ کلیک (CTR) و نرخ تبدیل (CVR) تبلیغات آموزش داده‌اید، در داده‌های آموزشی به دقت بالایی دست یافته‌اند. با این حال، زمانی که این مدل‌ها بر روی داده‌های Dev اعمال می‌شوند و شما قصد دارید درآمد را بهینه کنید، اختلاف قابل توجهی بین پیش‌بینی‌های مدل و درآمد واقعی مشاهده می‌کنید. علت این خطا را شناسایی کنید و تحلیل کنید که چه عواملی ممکن است باعث این اختلاف شوند. توجه: جدول ارائه شده در صورت سؤال صرفاً برای آشنایی با ساختار داده‌ها و فضای مسئله است و مقادیر آن فاقد اهمیت هستند.

۳ استفاده از انتقال یادگیری

با اضافه کردن تبلیغات اکشنی در سایت، با مشکلی به نام کمبود داده‌های مثبت (Positive Data) در پیش‌بینی CVR مواجه می‌شوید. اجازه دهید این مسئله را با یک مثال شهودی بررسی کنیم.

فرض کنید میانگین نرخ کلیک (CTR) در سایت شما حدود 0.001 (یک هزارم) و میانگین نرخ تبدیل (CVR) حدود 0.01 (یک صدم) باشد. این بدین معناست که به طور متوسط از هر ۱۰ هزار نمایش یک تبلیغ در یک جایگاه، تنها یک اکشن ثبت خواهد شد. اگر مدل پیش‌بینی نرخ تبدیل شما برای هر جفت تبلیغ و جایگاه (ورودی مدل) به حداقل ۳۰ اکشن نیاز داشته باشد تا داده‌های آن قابل اعتماد و معنادار باشد، آنگاه تنها نمونه‌هایی معتبر خواهند بود که حداقل ۳۰۰ هزار نمایش داشته باشند. این محدودیت به شدت حجم داده‌های آموزشی برای مدل پیش‌بینی نرخ تبدیل را کاهش می‌دهد.

با این شرایط، برای جمع‌آوری یک مجموعه داده معتبر با چند ده هزار نمونه مناسب برای آموزش، به چند ده میلیارد نمایش تبلیغ نیاز خواهید داشت که در مقیاس فعلی سایت شما ممکن است غیرعملی باشد. با در نظر گرفتن این مشکل، دو انتخاب پیش روی شماست:

- صرف نظر کردن از ساخت چنین مدلی و صبر کردن تا زمانی که حجم داده‌های شما به مقیاس پلتفرم‌های بزرگی مثل یوتیوب برسد.
- استفاده از روش‌های جایگزین، مانند انتقال یادگیری (Transfer Learning) و (Multi-Task Learning).

حال به سوالات زیر در رابطه با Transfer Learning و یادگیری چندوظیفه‌ای (Multi-Task Learning) پاسخ دهید:

تحت چه شرایطی می‌توان از Transfer Learning استفاده کرد؟ توضیح دهید که این روش چه زمانی مناسب است و چگونه می‌تواند در این مسئله کاربرد داشته باشد.

اگر Transfer Learning در این مسئله مناسب باشد، چه تسک جایگزینی را برای حل مشکل پیشنهاد می‌کنید؟ دلایل خود را برای انتخاب تسک جایگزین ارائه دهید.

آیا استفاده از Multi-Task Learning در این مسئله مفید است؟ بررسی کنید که چگونه یادگیری چندوظیفه‌ای می‌تواند بهبود عملکرد را ممکن سازد. اگر استفاده از این روش را مفید نمی‌دانید، دلایل محکمی برای رد آن ارائه دهید. در صورتی که این روش را مفید می‌دانید، تسک‌های مناسب برای پیاده‌سازی آن را پیشنهاد کنید.

	very similar dataset	very different dataset
very little data	Use Linear Classifier on top layer	You're in trouble... Try linear classifier from different stages
quite a lot of data	Feature a few layers	Feature a larger number of layers or start from scratch

ML_CV_3

سوال پنجم - دیفیوژن (۱۰ نمره)

در مدل های Diffusion-based دو مسیر Forward و Backward وجود دارد. در مسیر Forward به داده ها نویز گوسی اضافه می شود تا در T مرحله نهایتاً به نویز کاملاً گوسی تبدیل شوند. در مسیر Backward سعی می شود تا با تخمین مقدار نویز افزوده شده در هر گام، این نویز حذف شود تا تصویر اولیه حاصل شود.

در مسیر Forward در یک زنجیره مارکوف طی T گام با طول $\{\beta_t \in (0, 1)\}_{t=1}^T$ ، به تصویر نویز اضافه می شود و نمونه های x_1, x_2, \dots, x_T را می سازد:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t, \sqrt{1 - \beta_t}x_{t-1}, \beta_t I)$$
$$q(x_{1:T}|x_0)$$

(آ) می دانیم با توجه به خاصیت نویز گوسی نیازی به اضافه کردن نویز به صورت تکرار شونده نیست. در واقع می توان با یک مرحله به هر کدام از بازنمایی ها رسید. با استفاده از خواص نویز گوسی این فرایند را اثبات کنید. (راهنمایی: x_1 و x_2 را بر حسب β_1 و β_2 و سعی کنید رابطه بدست آمده را تعمیم دهید)

(ب) در روش score matching سعی می شود به کمک score function و یافتن مسیر مناسب برای حرکت در فضای داده برای بدست آوردن توزیع داده، تابع هدف بهینه شود. اگر $score\ function = \nabla \log(p(x))$ و مدل شبکه عصبی ما برای تخمین تابع امتیاز باشد، باید مقدار Divergence Fisher میان این دو مقدار کمینه شود:

$$\mathbb{E}_{p(x)}[||\nabla \log(p(x)) - s_\theta||_2^2]$$

اما در عمل مشاهده می شود که با کمینه کردن این مقدار تصاویر تولید شده همچنان نویزی هستند. دلیل این مشکل را بیان کنید و توضیح دهید راه حل ارائه شده چگونه این مشکل را برطرف می کند.

بخش عملی (۴۰ نمره)

سوال اول - سگمنتیشن (۲۰ نمره)

این نوتبوک تمرینی برای پیاده سازی مدل U-Net در بخش بندی تصاویر قفسه سینه طراحی شده است. در این نوتبوک، مراحل مختلفی از جمله آماده سازی داده ها، تعریف معماری مدل، آموزش آن و ارزیابی عملکرد به صورت گام به گام ارائه شده است. وظیفه شما این است که با تکمیل بخش های مشخص شده در کد، تمرین را کامل کنید. توضیحات لازم در هر بخش قرار داده شده اند تا بتوانید فرآیندهای مختلف را به درستی درک و پیاده سازی کنید. با پیش بردن این تمرین، مهارت هایی مانند پیش پردازش داده های تصویری، تنظیم پارامترهای مدل و تحلیل نتایج را کسب خواهید کرد.

سوال دوم - ViT و تفسیرپذیری (۲۰ نمره)

در این نوتبوک، ما قابلیت های ترنسفورمرهای بصری (ViT) را در طبقه بندی تصویر (Clas-sification) روی مجموعه داده تومور مغزی بررسی می کنیم. مجموعه داده تومور مغزی مورد استفاده در این نوتبوک شامل مجموعه ای از اسکن های MRI مغز است که هر اسکن به ۴ کلاس دسته بندی شده است. هدف ما آموزش یک مدل ViT برای طبقه بندی دقیق این اسکن های MRI بر اساس محتوای آن ها است. این مدل شامل یک لایه patch initial، embedding، به دنبال آن چندین لایه encoder Transformer می باشد. مکانیزم self-attention ترنسفورمر به مدل اجازه می دهد تا روابط فضایی پیچیده بین مناطق مختلف تصاویر مغزی را یاد بگیرد، که به دقت طبقه بندی آن کمک می کند.