



به نام خدا
درس یادگیری عمیق
تمرین سری اول
استاد درس : دکتر محمدرضا محمدی
دستیاران : نفیسه احمدی، علی سبحانی
دانشگاه علم و صنعت ایران، دانشکده مهندسی کامپیوتر
نیمسال دوم تحصیلی ۱۴۰۳ - ۱۴۰۴

مهلت تحویل : ۱۴۰۳/۱۲/۲۲

لطفا به نکات موجود در سند قوانین انجام و تحویل تمرین ها دقت فرمایید.

سوالات تئوری



۱. فرض کنید $f: \mathbb{R}^n \rightarrow \mathbb{R}$ یک تابع مشتق پذیر پیوسته به فرم زیر است:

$$f(x) = \exp(-\|Ax - b\|^2) + \sin\left(\sum_{i=1}^n c_i x_i^2\right)$$

که:

• $A \in \mathbb{R}^{n \times n}$ یک ماتریس با مرتبه کامل است.

• $b \in \mathbb{R}^n$ یک بردار بایاس است.

• $c_i \in \mathbb{R}$ ضرایب ثابت نامنفی یا صفر هستند.

ثابت کنید که برای هر مقدار $\epsilon > 0$ حداقل یک شبکه پرسپترون چند لایه $F(x)$ با تعداد محدودی نورون وجود دارد که بتواند تابع $f(x)$ را به طور دلخواه در یک دامنه فشرده $D \subset \mathbb{R}^n$ تقریب بزند به طوری که: (۱۵ نمره)

$$\|f(x) - F(x)\|_{L_2} < \epsilon$$

راهنما: برای حل این مسئله ابتدا تابع $f(x)$ را به دو بخش نمایی و سینوسی تقسیم کنید. سپس

با استفاده از مبانی تئوری و شبکه‌های عصبی چندلایه، هر بخش را به‌طور جداگانه تقریب بزنید. می‌توانید از منابع زیر برای اثبات خود کمک بگیرید:

- **Multilayer Feedforward Networks are Universal Approximators**
- **Approximation by superpositions of a sigmoidal function**

۲. اثبات کنید که افزودن یک ترم L_2 به تابع هزینه:

$$R_{\lambda}(F) = R(F) + \lambda \|W\|_2^2$$

واریانس مدل را کاهش داده و به بهبود تعمیم‌دهی کمک می‌کند. تأثیر پارامتر λ بر کران تعمیم (generalization bound) را استخراج کنید (۱۰ نمره).

۳. تصور کنید چند سال از فارغ‌التحصیلی شما گذشته و حالا در یک شرکت مشغول به کار هستید. به این نتیجه رسیده‌اید که به‌جای صعود در مسیر شغلی دیگران، کسب‌وکار شخصی خود را راه‌اندازی کنید. شما یک وب‌سایت آموزشی مفید و الهام‌بخش ساخته‌اید که حالا بازدیدکنندگان زیادی دارد و می‌خواهید از طریق تبلیغات آنلاین درآمد کسب کنید (۱۵ نمره).

برای کسب حداکثر درآمد از تبلیغات، به‌جای نمایش تصادفی تبلیغات، از یک سیستم حراجی استفاده می‌کنید که بهترین تبلیغات را برای هر موقعیت انتخاب می‌کند. اطلاعات تبلیغات در جدولی ثبت می‌شود که شامل موارد زیر است:

- **adv_id**: شناسه تبلیغ‌دهنده
- **cam_id**: شناسه کمپین تبلیغاتی
- **bid**: مبلغ پیشنهادی برای هر کلیک یا اقدام
- **type**: نوع درآمد (کلیک یا اقدام خاص)
- **pos_id**: موقعیت تبلیغ در سایت
- **ad_id**: شناسه تبلیغ
- **views**: تعداد نمایش تبلیغ

• **clicks**: تعداد کلیک‌ها روی تبلیغ

• **actions**: تعداد اقدامات انجام‌شده پس از کلیک

• **week_id**: زمان جمع‌آوری داده‌ها (بر اساس هفته)

week id	actions	clicks	views	ad id	pos id	type	bid	cam id	adv id
۱۷۳۵۱۱۵۵۶۳	۰	۱۵	۱۵۴۳۶	۸۹۷۵	۱۰	Click	۱۰۰۰	۶۵۷۵	۱۲۳۴
۱۷۳۵۱۱۵۵۶۳	۰	۱۰	۱۸۴۶۶	۶۷۳۵	۱۳	Click	۱۰۰۰	۶۵۷۵	۱۲۳۴
۱۷۳۵۱۲۴۵۶۹	۲	۲۰	۱۰۳۲۱	۷۱۸۵	۷۸	Action	۹۰۰۰	۹۸۷۶	۴۳۲۱
۱۷۳۵۱۱۴۵۶۳	۰	۲۵	۲۱۰۰۰	۱۰۲۴	۵	Click	۵۰۰	۴۵۳۲	۵۶۷۸
۱۷۳۵۱۱۸۵۶۳	۰	۳۰	۱۸۰۰۰	۲۳۴۱	۲۰	Click	۲۰۰۰	۳۴۵۶	۲۳۴۵
۱۷۳۵۱۳۴۵۶۳	۵	۱۸	۱۲۰۰۰	۶۵۲۳	۲۵	Action	۸۰۰۰	۷۶۵۴	۷۸۹۰

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. به این روابط دقت کنید و سعی کنید درک کنید که چرا این فرمول‌ها می‌توانند به انتخاب تبلیغات با بیشترین درآمد مورد انتظار کمک کنند.

فرمول بهینه‌سازی تبلیغات کلیکی

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

فرمول بهینه‌سازی تبلیغات اکشنی

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

- $A(p)$: 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 بر روی تبلیغ ad در جایگاه p (نرخ کلیک).
- cvr_{ad} : احتمال انجام اکشن توسط کاربر پس از کلیک روی تبلیغ.

فرض کنید مدل‌های آموزش داده شده 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.

الف) یکی از مهم‌ترین مراحل در فرآیند آموزش هر مدل یادگیری ماشین، تقسیم‌بندی داده‌ها به مجموعه‌های Train، Dev، Train-Dev و Test است. این کار به ما کمک می‌کند تا عملکرد مدل را بر روی توزیع داده‌های مختلف بررسی کنیم. در این مسئله خاص چطور این کار را باید انجام داد و در انتخاب این مجموعه‌ها به چه نکاتی باید توجه داشت؟

ب-۱) در هر یک از سناریوهای زیر، به طور مختصر مشکل را معرفی کرده و راه‌حلی برای آن ارائه دهید:

(آ) از قطعیت داده‌های ورودی اطمینان داریم ولی خطای آموزش مدل (Training Error) بالا است.

(ب) خطای آموزش مدل پایین است ولی خطای آن روی مجموعه (Train-Dev) همچنان بالا است.

(ج) خطای مدل در مجموعه‌های Train و Train-Dev پایین است ولی روی مجموعه Dev خطا زیاد است.

(د) خطای Dev پایین است ولی روی مجموعه Test خطا همچنان زیاد است.

ب-۲) آیا در اولین سناریوی مطرح‌شده در قسمت قبل، افزایش سایز داده‌های آموزش راه‌حل خوبی خواهد بود؟

ب-۳) فرض کنید مدلهایی که برای پیش‌بینی نرخ کلیک (CTR) و نرخ تبدیل (CVR) تبلیغات آموزش داده‌اید، در داده‌های آموزش 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 قابل‌توجهی بین پیش‌بینی‌های مدل و درآمد واقعی مشاهده می‌کنید. علت این خطا را شناسایی کرده و تحلیل کنید که چه عواملی ممکن است باعث این اختلاف شوند.


توجه: جدول ارائه شده صرفاً برای آشنایی با ساختار داده‌ها و فضای مسئله است و مقادیر آن فاقد اهمیت هستند.


ج) در یک سیستم بهینه‌سازی تبلیغات مبتنی بر داده‌های واقعی، یکی از چالش‌های اساسی، مواجهه با تغییرات ناگهانی در رفتار کاربران (Concept Drift) و نامتوازن بودن داده‌ها است. فرض کنید در دوره‌های زمانی خاصی (مانند مناسبت‌های خاص یا تغییر الگوریتم جستجو در موتورهای جستجو)، نرخ کلیک (CTR) و نرخ تبدیل (CVR) به‌طور ناگهانی دچار تغییرات چشمگیر می‌شوند.

- چگونه می‌توان پایداری مدل را در مواجهه با Concept Drift تضمین کرد؟
- چه روش‌هایی برای مدیریت داده‌های نامتوازن در این مسئله مناسب هستند؟

- تحقیق کنید که چگونه می‌توان با استفاده از الگوریتم‌های آنلاین یادگیری (Online Learning)، عملکرد سیستم را بهبود بخشید و به تغییرات سریع بازار واکنش نشان داد.

سوالات عملی

۴.  هدف این تمرین، پیش پردازش داده و پیاده سازی دستی ماژول های شبکه عصبی است. شما باید بخش‌های *TODO* را تکمیل کنید. لذا از هرگونه تغییر یا دستکاری ساختار اصلی کد اجتناب فرمایید. کلیه کدها باید از قبل اجرا شده باشند؛ در غیر این صورت نمره مربوطه تعلق نخواهد گرفت (۲۵ نمره).

۵.  هدف این تمرین، ایجاد ماژول های شبکه عصبی و ترکیب آنها جهت ساخت یک شبکه کامل است. شما باید بخش‌های *TODO* و سوالاتی که در نوت‌بوک بیان شده را تکمیل کنید. لذا از هرگونه تغییر یا دستکاری ساختار اصلی کد اجتناب فرمایید. کلیه کدها باید از قبل اجرا شده باشند؛ در غیر این صورت نمره مربوطه تعلق نخواهد گرفت (۳۵ نمره).