



هوش مصنوعی

پاییز ۱۴۰۲

اساتید: محمدحسین رهبان، مهدیه سلیمانی باغشاه
گردآورندگان: محمدرضا دویران، بهار دیبائی‌نیا، علی شریفی،
پیام تائبی، بردیا محمدی و علی مهربانی

مهلت ارسال: ۱۹ دی

یادگیری ماشین و شبکه‌های عصبی

تمرین پنجم

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- در طول ترم امکان ارسال با تاخیر پاسخ همه‌ی تمرین تا سقف ۳ روز و در مجموع ۵ روز، وجود دارد. پس از گذشت این مدت، پاسخ‌های ارسال شده پذیرفته نخواهند بود. همچنین، به ازای هر روز تأخیر غیر مجاز ۲۴ درصد از نمره تمرین به صورت ساعتی کسر خواهد شد. جزئیات نحوه اعمال تاخیرها را می‌توانید در سایت درس مشاهده کنید.
- همکاری و هم‌فکری شما در انجام تمرین مانعی ندارد اما پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد.
- در صورت هم‌فکری و یا استفاده از هر منابع خارج درسی، نام هم‌فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- لطفا تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.

سوالات نظری (۹۰ نمره)

۱. (۲۰ نمره) تابع باینری زیر را در نظر بگیرید. ابتدا این تابع را با درخت تصمیم تصویر کنید. سپس مطرح کنید که آیا می‌توان با یک نورون (شبکه یک لایه) این تابع را ساخت و در مرحله آخر سعی کنید با شبکه‌ای از نورون‌ها (شبکه چند لایه) این تابع را پیاده‌سازی کنید و مقادیر بایاس شبکه را نیز تنظیم کنید (سعی کنید کمترین تعداد نورون و لایه را در شبکه به کار ببرید).

$$f(A, B, C, D) \equiv (\neg A \rightarrow B) \rightarrow ((B \wedge D) \rightarrow \neg C)$$

۲. (۲۰ نمره) در یک شبکه عصبی با L لایه، a^l نشان دهنده بردار خروجی لایه l است (l عددی از ۱ تا L است) و به‌منظور z^l نشان دهنده ورودی‌های آن لایه است؛ طبیعتاً خواهیم داشت $a^l = f^l(z^l)$ که f^l تابع اکتیویشن لایه l است از طرفی اگر وزن یال‌های بین ۲ لایه را با ماتریس W^l نشان دهیم که مربوط به یال‌های بین لایه a^l و a^{l-1} است خواهیم داشت $z^l = W^l a^{l-1} + b^l$ ؛ طبیعتاً تابع ضرر، تابعی از a^L و مقادیر واقعی خروجی است و اگر طبق روش‌های مبتنی بر گرادیان کار کنیم باید گرادیان تابع ضرر را بر حسب W و b و a به دست آوریم، پس اگر تابع ضرر را به شکل $J = g(y, a^l)$ داشته باشیم، گرادیان J را بر حسب b^i و W^i به دست آورید

راهنمایی: سعی کنید طبق تعریف رابطه‌ی $\delta^l = \frac{\partial J}{\partial z^l}$ ، آن را بازگشتی بنویسید تا δ^{l-1} بر حسب δ^l به دست آید و مسئله را حل کنید.

۳. (۲۰ نمره) به سوالات زیر پاسخ کوتاه دهید.

- (آ) مزایا و معایب اضافه کردن لایه‌های بیشتر به یک شبکه عصبی عمیق چیست؟
- (ب) تابع فعال‌سازی sigmoid را در نظر بگیرید. مقدار گرادیان این تابع برای یک ورودی بسیار بزرگ چقدر خواهد بود؟ این چه مشکلی برای آموزش شبکه‌های عصبی ایجاد می‌کند؟ چگونه می‌توان این مورد را حل کرد؟

(ج) می‌خواهیم سن افراد را از طریق regression و اسکن مغزی آن‌ها حدس بزنیم. تعداد اسکن‌های مغز موجود از افراد برابر ۱۰ می‌باشد و بردار ویژگی به دست آمده برای هر فرد شامل ۱۵ هزار ویژگی مختلف است. با توجه به مفاهیم فراگرفته شده کدام یک از روش‌های L1 Regularization و L2 Regularization در این حالت موثرتر است؟ چرا؟

(د) آیا gradient descent می‌تواند در یک مینیمم محلی گیر کند در زمانی که یک مدل logistic regression را آموزش می‌دهیم؟

۴. (۳۰ نمره) یک مدل *logistic regression* را در نظر بگیرید. می‌دانیم $y = g(w^T x)$ و تابع هزینه آن *binary cross entropy* است و

$$g(z) = \frac{1}{1 + e^{-z}}$$

یک روش تغییر داده شده از *logistic regression* ارائه می‌کنیم که در آن $g(z) = \frac{e^{-z}}{1 + e^{-z}}$ و تابع هزینه همچنان همان *binary cross entropy* است. پارامترها و پیش‌بینی‌های مدل یادگرفته‌شده جدید نسبت به مدل اولیه چه تفاوتی دارد؟ به صورت ریاضی توضیح دهید.
راهنمایی: تابع هزینه به صورت زیر است: $-(y \log(g(z)) + (1 - y) \log(1 - g(z)))$

سوالات عملی (۷۰ + ۵۰ نمره)

۱. (۷۰ نمره) برای پاسخ به این سوال به پوشه سوالات عملی بخش Q1 مراجعه کنید.
۲. (۵۰ نمره امتیازی) برای پاسخ به این سوال به پوشه سوالات عملی بخش Q2-Extra مراجعه کنید.