



هوش مصنوعی

پاییز ۱۴۰۰

استاد: محمدحسین رهبان

گردآورندگان: کیمیا نوربخش، امیررضا سلیمان بیگی، آرمان زارعی

مهلت ارسال: ۳۱ خرداد

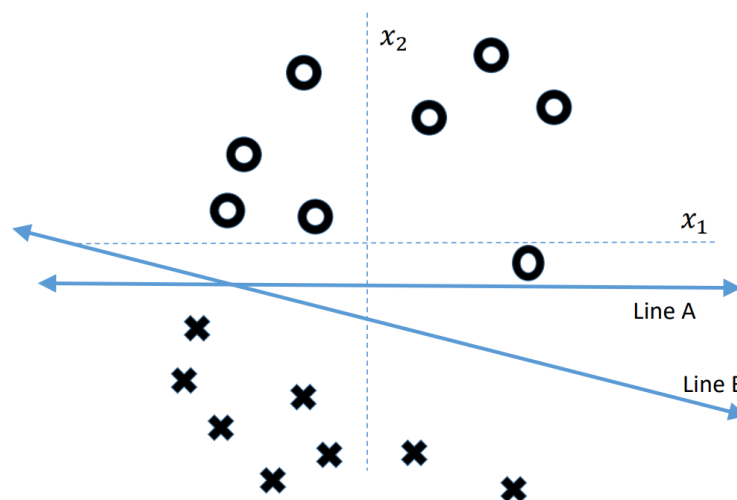
یادگیری ماشین و یادگیری عمیق

تمرین ششم

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- در طول ترم امکان ارسال با تاخیر پاسخ همه‌ی تمرین تا سقف ۷ روز و در مجموع ۲۰ روز، وجود دارد. پس از گذشت این مدت، پاسخ‌های ارسال شده پذیرفته نخواهند بود. همچنین، به ازای هر روز تأخیر غیر مجاز ۱۰ درصد از نمره تمرین به صورت ساعتی کسر خواهد شد.
- همکاری و هم‌فکری شما در انجام تمرین مانعی ندارد اما پاسخ‌های ارسالی هر کس حتماً 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.
- لطفاً تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.

سوالات نظری (۱۰۰ نمره)

۱. (۲۰ نمره) داده آموزش تصویر زیر را در نظر بگیرید:



هر داده ما دو ویژگی (feature) دارد که یکی با محور افقی و دیگری با محور عمودی نشان داده شده است. هر داده به یکی از دو کلاس تعلق دارد، کلاس ۱ همان "X" است و کلاس -۱ همان "O" است. مدل logistic regression ما به صورت زیر است که:

$$P(y = 1|x, w) = \frac{1}{1 + \exp(-w_0 - w_1x[1] - w_2x[2])}$$

الف) دو خط A, B که در تصویر نشان داده شده است، ممکن است نتیجه آموزش ما روی این داده باشند. یک مقدار ممکن برای $w = [w_0, w_1, w_2]$ که خط‌های A, B را تولید کنند را بیابید (برای هر خط یک مقدار

پیشنهاد کنید). خط A یک خط افقی است که محور عمودی را در ۱- قطع می‌کند و خط B با شیب $\frac{-1}{3}$ محور عمودی را در ۲- قطع می‌کند.

ب) نشان دهید برای مقادیر w ای که در بخش الف یافتید، مقدار $2w$ نیز همچنان همان دو خط جدا کننده را نشان می‌دهند (همان A, B).

ج) در درس آموختیم که یک loss function مناسب برای این مسئله

$$J(w) = \sum_{i=1}^n \log(1 + \exp(-y_i x_i^T w))$$

است. اگر سعی کنیم این loss را کمینه کنیم، w ای که در بخش الف یافتیم، برابر با جواب

$$\hat{w} = \operatorname{argmin} \sum_{i=1}^n \log(1 + \exp(-y_i x_i^T w))$$

نمی‌شود. چرا؟ (با توجه به بخش ب جواب دهید). چرا این موضوع یک مشکل ایجاد می‌کند؟

د) حال فرض کنید که ما به تابع loss، بخش regularizer برابر با $\lambda(|w_1| + |w_2|)$ را اضافه می‌کنیم. برای مقادیر بزرگ λ کدام یک از دو خط A, B انتظار می‌رود loss کمتری داشته باشند؟ (یعنی مقدار J در بالا بعلاوه قسمت regularizer کمتری داشته باشد؟) توجه داشته باشید که پاسخ این بخش صرفاً مفهومی باشد و محاسبه‌ای برای آن لازم نیست.

۲. (۱۵ نمره)

الف) نشان دهید تابع tanh و تابع sigmoid با رابطه زیر به هم مرتبط هستند:

$$\tanh(a) = 2\sigma(2a) - 1$$

ب) حال نشان دهید که یک ترکیب خطی کلی از توابع sigmoid به فرم

$$y(x, w) = w_0 + \sum_{j=1}^M w_j \sigma\left(\frac{x - \mu_j}{s}\right)$$

با یک ترکیب خطی از توابع tanh به فرم

$$y(x, u) = u_0 + \sum_{j=1}^M u_j \tanh\left(\frac{x - \mu_j}{s}\right)$$

برابر است. روابطی که میان پارامترهای $\{u_1, \dots, u_M\}$ با پارامترهای $\{w_1, \dots, w_M\}$ به وجود می‌آید را بنویسید.

۳. (۱۵ نمره) می‌دانیم که 2^{2^n} تابع boolean متمایز بر روی n ورودی وجود دارند. در نتیجه برای ۲ ورودی، ۱۶ تابع boolean متمایز وجود دارد. چند تا از این ۱۶ تابع را می‌توان توسط یک perceptron نشان داد؟ (چرا؟)

۴. (۱۰ نمره) در دسته بندی دودویی (binary classification) توسط شبکه عصبی، آموختیم که اگر a_n مربوط به activation داده n ام باشد، برای تابع Error داریم:

$$E(w) = - \sum_{n=1}^N \{t_n \ln y_n + (1 - t_n) \ln(1 - y_n)\}$$

نشان دهید

$$\frac{\partial E}{\partial a_k} = y_k - t_k$$

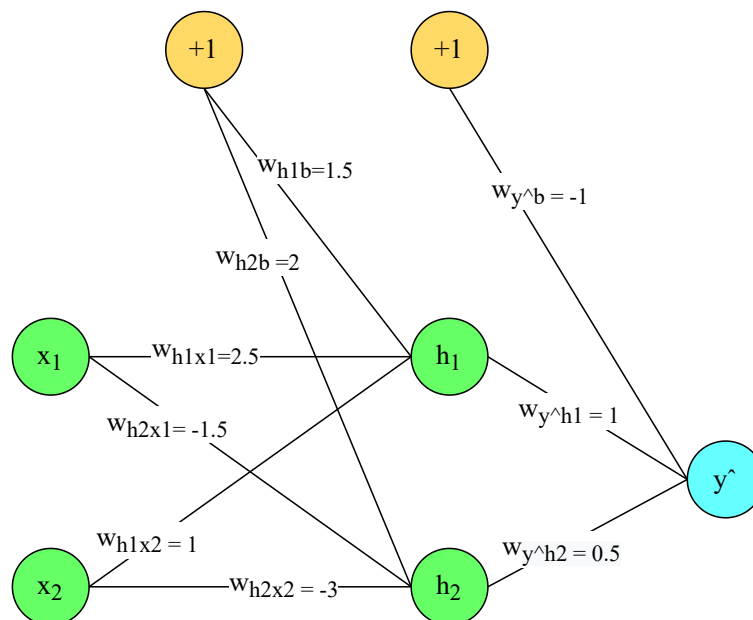
(یادآوری: y_n, t_n به ترتیب کلاس اصلی و خروجی شبکه هستند.)

۵. (۱۵ نمره) یک شبکه عصبی با تنها یک لایه پنهان (Hidden Layer) طراحی کنید که تابع

$$(A \vee \bar{B}) \oplus (\bar{C} \vee \bar{D})$$

را پیاده‌سازی کند. شبکه را رسم کنید و تمامی وزن‌ها را در آن مشخص کنید.

۶. (۲۵ نمره) تصویر زیر یک شبکه عصبی دو لایه را با دو node در لایه پنهان و یک node خروجی نشان می‌دهد. x_1, x_2 دو ورودی شبکه هستند. برای پاسخ به این سوال فرض کنید که مقدار Learning Rate برابر $\alpha = 0.1$ است. هر node یک مقدار bias ورودی به اندازه ۱ دارد. فرض کنید که activation function در لایه پنهان و در node خروجی تابع sigmoid است.



الف) مقادیر خروجی در node های h_1, h_2, \hat{y} این شبکه را برای ورودی $\{x_1 = 0, x_2 = 1\}$ محاسبه کنید.
 ب) یک مرحله از الگوریتم backpropagation را برای مثال قسمت الف و خروجی $y = 1$ اجرا کنید. وزن‌های آپدیت شده برای لایه پنهان و لایه خروجی را (در مجموع ۹ وزن آپدیت شده) با اجرای یک مرحله از gradient descent به دست آورید. تمامی مراحل محاسبات خود را در پاسخ بنویسید. (راهنمایی: واضح است که در این مثال خاص تابع error برابر با $E = \frac{1}{2}(y - O)^2$ است که O همان خروجی شبکه است.)

سوالات عملی (۱۵۰ نمره)

۱. (۵۰ نمره) برای حل این سوال به پوشه Neural Networks رجوع کرده و فایل Neural Networks.ipynb را حل کنید.

۲. (۱۰۰ نمره) این سوال دارای ددلاین جداگانه و برابر با ددلاین آخرین تمرین درس است. برای حل این سوال به فایل CIFAR10_Classification_And_Colorization.ipynb رجوع کنید. دقت کنید نمره هر کدام از قسمت‌های Classification و Colorization برابر با ۵۰ نمره است.