



امنیت و حریم خصوصی در یادگیری ماشین (۴۰۸۱۶)
نیم سال اول سال تحصیلی ۱۴۰۴-۱۴۰۳
استاد درس: دکتر امیرمهدی صادقزاده

طراحان: علی جاوری، محمدرضا میرباقری، رئوف زارع

تمرین اول

مهلت تحویل: ساعت ۲۳:۵۹ یکشنبه ۲۹ مهر ۱۴۰۳

نکات و قواعد

۱. سوالات خود را زیر پیام مربوطه در Quera مطرح نمایید.
۲. لطفا مطابق تاکید پیشین، حتما آداب نامه‌ی انجام تمرین‌های درسی را رعایت نمایید. در صورت تخطی از آیین نامه، در بهترین حالت مجبور به حذف درس خواهید شد.
۳. در صورتی که پاسخ‌های سوالات نظری را به صورت دست‌نویس آماده کرده‌اید، لطفا تصاویر واضحی از پاسخ‌های خود ارسال کنید. در صورت ناخوانا بودن پاسخ ارسالی، نمره‌ای به پاسخ ارسال شده تعلق نمی‌گیرد.
۴. همه‌ی فایل‌های مربوط به پاسخ خود را در یک فایل فشرده و با نام SPML_HW\StdNum_FirstName_LastName ذخیره کرده و ارسال نمائید.

سوال ۱ مروری بر پیشنیازها (۵۰ نمره)

نرم اسپکترا یا نرم ۲-الحاقی ماتریس A را به صورت $\|A\|_2$ نشان می‌دهیم و به شکل زیر تعریف می‌کنیم:

$$\|A\|_2 = \max_{\mathbf{x} \neq 0} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$$

نرم فروبنیوس ماتریس $A \in \mathbb{R}^{m \times n}$ عبارت است از $\|A\|_F = \sqrt{\text{Tr}\{A^T A\}}$ (اثر یا Tr ماتریس، جمع درایه‌های رو قطر اصلی آن ماتریس می‌باشد).
به یک تابع حقیقی $f: \mathbb{R} \leftarrow \mathbb{R}$ پیوسته لیپشیتز می‌گوییم در صورتی که عدد حقیقی و مثبت k وجود داشته باشد به طوری که برای هر x_1, x_2 داشته باشیم:

$$|f(x_1) - f(x_2)| \leq K|x_1 - x_2|$$

و به این k ثابت لیپشیتز می‌گوییم.

۱. عبارت‌های زیر را ثابت کنید. (در اینجا منظور از $\sigma_i(A)$ مقادیر تکین، $\sigma_{\max}(A)$ بزرگترین مقدار تکین و λ_i ها همان مقادیر ویژه ماتریس A هستند.)

(الف)

$$\text{null}(A) = \text{null}(A^T A)$$

(ب)

$$\|A\|_2 = \sup_{\|U\|=\|V\|=1} U^T A V, \quad A \in \mathbb{R}^{m \times n}$$

(ج)

$$\|A\|_2 = \sigma_{\max}(A)$$

(د)

$$\|A^T A\|_2 = \|AA^T\|_2 = \|A\|_2^2$$

(ه)

$$Tr(A^T B) = \sum_{i=1}^m \sum_{j=1}^n A_{ij} B_{ij}, A, B \in \mathbb{R}^{m \times n}$$

(و)

$$\|A\|_F = \sqrt{\sum \sigma_i^2(A)}$$

(ز)

$$Tr(A) = \sum \lambda_i$$

(ح)

$$\|A + X\|_F^2 = \|A\|_F^2 + \|X\|_F^2 + 2Tr(A^T X)$$

(ط)

$$\det(A) = \prod_{i=1}^n \lambda_i$$

(ی)

$$\|A\|_F \geq \|A\|_2 \geq \frac{1}{\sqrt{n}} \|A\|_F \geq \frac{Tr(A)}{n} \geq \sqrt[n]{\det(A)}$$

۲. ثابت کنید تابع $f(x) = \log(1 + \exp x)$ یک تابع $lipschitz$ $1 -$ می باشد. (راهنمایی: ابتدا به وسیله قضیه مقدار میانگین ثابت کنید اگر مشتق تابع محدود باشد تابع $lipschitz$ خواهد بود و سپس حکم را اثبات کنید.)

۳. توابع $\{f_i\}_i^n$ ، $p_i - lipschitz$ هستند. ثابت کنید $f_1 \circ f_{n-1} \circ \dots \circ f_n$ نیز $lipschitz$ است. ضریب آن را بیابید.

۴. دو تابع f و g ، $lipschitz$ هستند. ثابت کنید $f \times g$ و $f + g$ نیز $lipschitz$ می باشند. ضرایب آن ها را نیز با فرض اینکه ضرایب لیپشیتز f و g به ترتیب L_f و L_g می باشد، بیابید (با فرض کراندار بودن توابع در حالت ضرب).

سوال ۲ توابع فعالساز (۲۵ نمره)

تابع یک شبکه دو لایه از فرم زیر را در نظر بگیرید

$$y_k(\mathbf{x}, \mathbf{w}) = \sigma \left(\sum_{j=1}^M w_{kj}^{(2)} h \left(\sum_{i=1}^D w_{ji}^{(1)} x_i + w_{j0}^{(1)} \right) + w_{k0}^{(2)} \right).$$

که در آن توابع فعال سازی غیرخطی واحد پنهان^۱ $h(\cdot)$ با استفاده از توابع سیگموئید لوجستیک به شکل زیر تعریف شده اند:

$$\sigma(a) = \{1 + \exp(-a)\}^{-1}.$$

نشان دهید که یک شبکه معادل وجود دارد که دقیقاً همان تابع را محاسبه می کند، اما توابع فعال سازی واحد پنهان آن $\tanh(a)$ می باشد که به شکل زیر تعریف شده است:

$$\tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$$

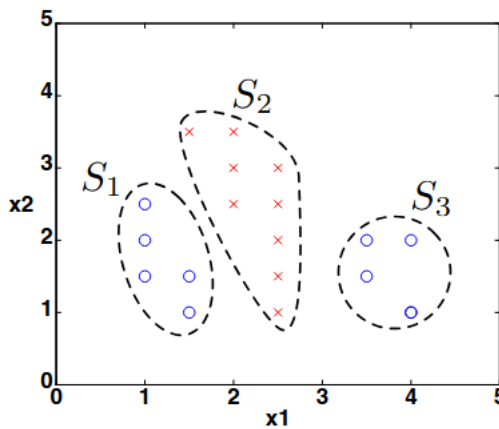
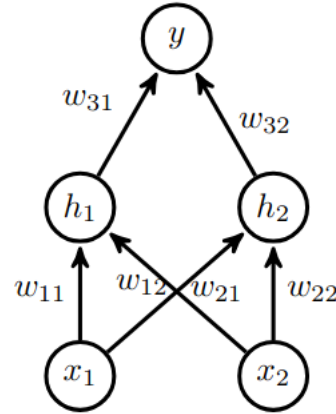
راهنمایی: ابتدا رابطه بین $\sigma(a)$ و $\tanh(a)$ را پیدا کنید و سپس نشان دهید پارامترهای شبکه ها به وسیله تبدیلات خطی باهم تفاوت دارند.

^۱Hidden-unit nonlinear activation functions

سوال ۳ شبکه‌های عصبی (۲۵ نمره)

در این سوال ما قصد داریم از یک شبکه عصبی برای دسته بندی علامت های ضرب (×) و علامت های دایره ای (○) نشان داده شده در شکل ۱ استفاده کنیم. اگرچه علامت‌های ضرب و دایره خطی تفکیک پذیر نیستند، می‌توانیم آن‌ها را به سه دسته S_1 ، S_2 و S_3 (همانطور که در شکل ۱ نشان داده شده) تقسیم بندی کنیم، به گونه‌ای که S_1 خطی تفکیک پذیر از S_2 و S_2 خطی تفکیک پذیر از S_3 باشد. سپس از این روش برای طراحی وزن‌های شبکه عصبی نشان داده شده در شکل ۱، در جهت کلاس بندی مجموعه آموزشی استفاده خواهیم کرد. برای تمامی گره‌ها از تابع فعال سازی $threshold$ که به صورت زیر تعریف شده، استفاده می‌کنیم.

$$\phi(z) = \begin{cases} 1 & z > 0 \\ 0 & z \leq 0. \end{cases}$$

(a) The dataset with groups S_1 , S_2 , and S_3 .

(b) The neural network architecture

شکل ۱

۱. در قسمت اول، پارامترهای w_{11} ، w_{12} و b_1 نورون با برجسب h_1 را طوری تنظیم می‌کنیم که خروجی آن $h_1(x) = \phi(w_{11}x_1 + w_{12}x_2 + b_1)$ یک جداکننده خطی بین مجموعه‌های S_2 و S_3 تشکیل دهد.

وزن‌های مربوط به w_{11} ، w_{12} و b_1 را به نحوی مشخص کنید که $h_1(x) = 0$ برای همه نقاط در S_3 و $h_1(x) = 1$ برای همه نقاط در S_2 باشد.

۲. در این مرحله، پارامترهای w_{21} ، w_{22} و b_2 نورون با برجسب h_2 را به گونه‌ای تنظیم می‌کنیم که $h_2(x) = \phi(w_{21}x_1 + w_{22}x_2 + b_2)$ یک جداکننده خطی بین مجموعه‌های S_1 و S_2 تشکیل دهد.

وزن‌های مربوط به w_{21} ، w_{22} و b_2 را به نحوی مشخص کنید که $h_2(x) = 0$ برای همه نقاط در S_1 و $h_2(x) = 1$ برای همه نقاط در S_2 باشد.

۳. اکنون دو طبقه بندی کننده h_1 (برای طبقه بندی S_2 از S_3) و h_2 (برای طبقه بندی S_1 از S_2) داریم. وزن‌های نورون نهایی شبکه عصبی را بر اساس نتایج h_1 و h_2 تنظیم می‌کنیم تا تلاقی‌های دایره‌ها را طبقه بندی کنیم. فرض کنید $h_3(x) = \phi(w_{31}h_1(x) + w_{32}h_2(x) + b_3)$ باشد.

(الف) عبارت های w_{31} ، w_{32} و b_3 را به گونه‌ای محاسبه کنید که $h_3(x)$ کل مجموعه داده را به درستی طبقه بندی کند.

(ب) رمز تصمیم مورد نظر خود را رسم کنید.

۴. در مثال بالا، ما باید با توجه به داده‌ها وزن‌ها را یاد بگیریم. در اولین مرحله، باید گرادیان‌های پارامترهای شبکه عصبی را بدست آوریم.

فرض کنید ما داده‌های m نمونه x_i با برجسب y_i داریم، که $i \in [1, m]$. x_i یک بردار $d \times 1$ است و $y_i \in \{0, 1\}$. ما از داده‌ها برای آموزش یک شبکه عصبی با یک لایه مخفی استفاده می‌کنیم:

$$h(x) = \sigma(W_1x + b_1)$$

$$p(x) = \sigma(W_2h(x) + b_2)$$

که در آن $\sigma(x) = \frac{1}{1+\exp(-x)}$ تابع سیگموئید است، W_1 یک ماتریس $n \times d$ و b_1 یک بردار $n \times 1$ است. W_2 یک ماتریس $1 \times n$ و b_2 یک بردار 1×1 است.

ما از تابع خطای آنتروپی متقابل استفاده کرده و منفی لگاریتم درست‌نمایی را برای آموزش شبکه عصبی به حداقل می‌رسانیم:

$$l = \frac{1}{m} \sum_i l_i = \frac{1}{m} \sum_i -(y_i \log p_i + (1 - y_i) \log(1 - p_i)),$$

که در آن $h_i = h(x_i)$ ، $p_i = p(x_i)$ هستند.

(الف) وقتی m بزرگ است، معمولاً از یک نمونه کوچک از کل داده‌ها برای محاسبه گرادیان استفاده می‌کنیم. این روش گرادیان کاهشی تصادفی (SGD) نامیده می‌شود. توضیح دهید که چرا از SGD به جای گرادیان کاهشی استفاده می‌کنیم.

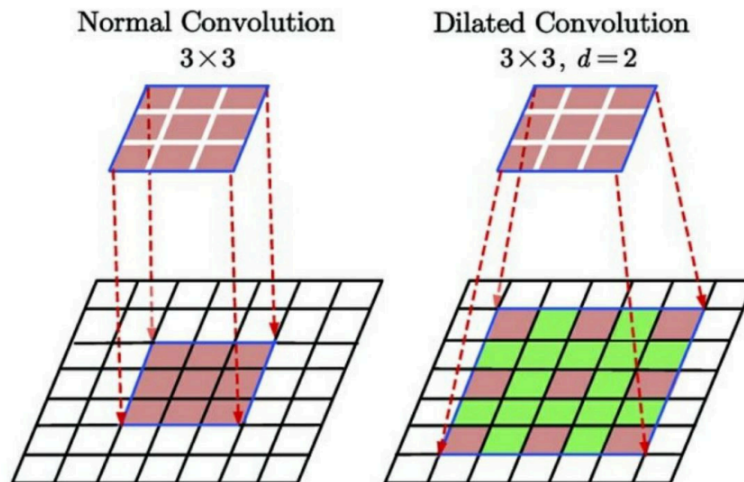
(ب) گرادیان‌های زیر را محاسبه کنید.

$$\frac{\partial l}{\partial p_i}, \quad \frac{\partial l}{\partial W_2}, \quad \frac{\partial l}{\partial b_2}, \quad \frac{\partial l}{\partial h_i}, \quad \frac{\partial l}{\partial W_1}, \quad \frac{\partial l}{\partial b_1}.$$

در هنگام محاسبه گرادیان مربوط به پارامترها در لایه‌های پایین‌تر، می‌توانید فرض کنید گرادیان در لایه‌های بالاتر در دسترس است (یعنی می‌توانید از آن‌ها در معادله استفاده کنید). برای مثال، هنگام محاسبه $\frac{\partial l}{\partial W_1}$ ، می‌توانید فرض کنید که $\frac{\partial l}{\partial p_i}$ ، $\frac{\partial l}{\partial W_2}$ ، $\frac{\partial l}{\partial b_2}$ ، $\frac{\partial l}{\partial h_i}$ را می‌دانید.

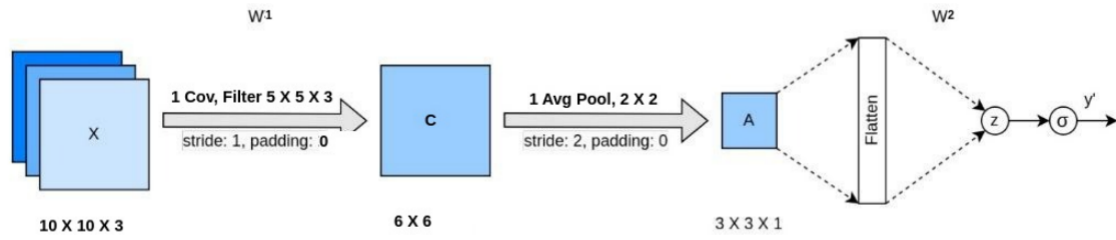
سوال ۴ شبکه‌های پیچشی (۲۵ نمره)

یکی از فیلترهای استفاده‌شده در شبکه‌های عصبی، فیلترهای dilated می‌باشد. تفاوت این فیلترها با فیلترهای معمولی این است که پیکسل‌ها را با فاصله ورودی می‌گیرند، نحوه عملکرد یکی از این فیلترها در شکل پایین دیده می‌شود. حال به سوالات زیر پاسخ دهید.



شکل ۲

- تصویر با ابعاد $i \times i \times 3$ داریم که ابتدا یک لایه کانولوشن معمولی با فیلتر $d_1 \times k_1 \times k_1 \times 3$ بر روی آن اعمال می‌شود و در لایه بعد d_2 لایه معمولی $k_2 \times k_2 \times d_1$ بر روی خروجی لایه قبل اعمال می‌شود و سپس فیلتر $d_3 \times k_3 \times k_3 \times d_2$ dilated با پارامتر گسترش n بر روی خروجی لایه دوم اعمال می‌شود. برحسب پارامترهای داده شده در صورت سوال ابعاد خروجی لایه آخر را حساب نمایید.
- وقتی از کانولوشن dilated استفاده می‌شود، مبحث محدودی دید پررنگ‌تر مطرح می‌شود. اگر در لایه اول یک فیلتر $k \times k$ بر روی ورودی اعمال شود، که پارامتر گسترش آن d_1 است و در لایه دوم یک فیلتر $k \times k$ با پارامتر گسترش d_2 اعمال شود در لایه سوم یک فیلتر $k \times k$ با پارامتر گسترش d_3 اعمال شود، برای عنصرهای i و j ام خروجی محدودی دید را بر حسب پارامترهای داده‌شده در سوال محاسبه نمایید.



شکل ۳

سوال ۵ تمرین عملی (۲۵ نمره)
نوت‌بوک `nn_numpy.ipynb` را کامل کنید.

موفق باشید.