



## یادگیری عمیق

پاییز ۱۴۰۲  
استاد: دکتر فاطمی زاده

گردآورندگان: -

تمرین دوم شبکه های پرسپترون، رگولاسیون، بهینه سازها مهلت ارسال: دوشنبه ۲۲ آبان

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- در طول ترم امکان ارسال با تاخیر پاسخ همه ی تمرین تا سقف ۵ روز و در مجموع ۲۰ روز، وجود دارد. پس از گذشت این مدت، پاسخ های 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 پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد. (دقت کنید در صورت تشخیص مشابهت غیرعادی برخورد جدی صورت خواهد گرفت.)
- 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.
- نتایج و پاسخ های خود را در یک فایل با فرمت zip به نام HW۲-Name-StudentNumber در سایت [Quera](#) قرار دهید. برای بخش عملی تمرین نیز لینک گیت هاب که تمرین و نتایج را در آن آپلود کرده اید قرار بدهید. دقت کنید هر سه فایل نوتبوک تکمیل شده بخش عملی را در گیت هاب قرار دهید.
- لطفا تمامی سوالات خود را از طریق کوثرای درس مطرح بکنید (برای اینکه تمامی دانشجویان به پاسخ های مطرح شده به سوالات دسترسی داشته باشند و جلوی سوالات تکراری گرفته شود، به سوالات در بسترهای دیگر پاسخ داده نخواهد شد).
- دقت کنید کدهای شما باید قابلیت اجرای دوباره داشته باشند، در صورت دادن خطا هنگام اجرای کدتان، حتی اگر خطا بدلیل اشتباه تایپی باشد، نمره صفر به آن بخش تعلق خواهد گرفت.

### سوالات نظری (۲۰۰ نمره)

۱. (۴۰ نمره)

- (آ) نموداری رسم کنید که محور افقی آن تعداد داده در mini-batch و محور عمودی آن تعداد گام لازم برای همگرایی توسط الگوریتم SGD جهت رسیدن به مقدار از پیش تعیین شده ای از خطای آموزش باشد. بخش آغازین نمودار (بسته های کوچک داده) و بخش پایانی نمودار (بسته های بزرگ داده) را با ذکر دلیل توجیه نمایید.
- (ب) آیا می توان گفت که در لایه Batch Normalization و در هنگام آموزش، مقداری نویز به توابع فعالیت لایه های مخفی تزریق می شود؟ چرا؟
- (ج) شبکه ی عصبی Fully Connected ای را در نظر بگیرید که تمام توابع فعالیت آن سیگموئید باشد و وزن های اولیه آن مقادیر مثبت بزرگ باشد. آیا این شبکه مناسبی برای طبقه بندی می باشد؟ چرا؟
- (د) شبکه ی عصبی Fully Connected ای با ۵ لایه مخفی، که در هر کدام از لایه ها، ۱۰ نورون وجود دارد را در نظر بگیرید. ورودی این شبکه ۲۰ بعدی و خروجی آن اسکالر می باشد. تعداد کل پارامتر های قابل آموزش را در این شبکه محاسبه کنید.

(ه) یک مسئله طبقه بندی باینری می تواند با دو روش زیر حل شود:  
روش اول: Logistic Regression ساده (یک نورون)

$$\hat{y} = \sigma(W_l x + b_l) \text{ خروجی:}$$

اگر  $\hat{y} \leq 0.5$  کلاس صفر در غیر این صورت کلاس یک طبقه بندی می شود.

روش دوم: Softmax Regression ساده (دو نورون)

$$\hat{y} = \text{softmax}(W_s x + b_s) = [\hat{y}_1, \hat{y}_2]^T \text{ خروجی:}$$

اگر  $\hat{y}_1 \geq \hat{y}_2$  کلاس صفر در غیر این صورت کلاس یک طبقه بندی می شود.

روش دوم دو برابر روش اول پارامتر دارد. آیا می توان گفت که روش دوم مدل های پیچیده تری نسبت به روش اول یاد می گیرد؟

اگر بله، پارامترهای  $(W_s, b_s)$  تابعی که روش دوم می تواند آن را مدل کند مثال بنزید در غیر این صورت نشان دهید که  $(W_s, b_s)$  همیشه می تواند برحسب  $(W_l, b_l)$  نوشته شود.

۲. (۳۰ نمره) می دانیم حتی یک شبکه عصبی یک لایه نیز می تواند طبقه بندی ارقام را با دقت خوبی انجام دهد. راه های متعددی برای ارتقای دقت مدل وجود دارد. این مقاله روشی ساده برای ارتقای عملکرد مدل، بدون تغییر در ساختار آن پیشنهاد می دهد. این پیشنهاد آموزش چند مدل مشابه است. مقاله را بخوانید و به سوالات زیر پاسخ دهید:

(آ) کمیته<sup>۱</sup> چیست و چطور به بهبود عملکرد مدل کمک می کند؟

(ب) پیش پردازش انجام شده در مقاله را شرح دهید. چگونه این پیش پردازش از وابستگی زیاد خطای مدل ها جلوگیری می کند؟

۳. (۳۰ نمره)

(آ) یک شبکه ی عصبی با ورودی  $x$  را در نظر بگیرید. برای بدست آوردن خروجی محاسبات زیر بر روی  $x$  انجام می شود.

$$z = wx + b$$

$$y = \sigma(z)$$

$$L = \frac{1}{2} (y - t)^2$$

$$R = \frac{1}{2} w^2$$

$$L_{reg} = L + \lambda R$$

گراف محاسباتی این مسئله را رسم کنید و مشتقات  $L_{reg}$  را نسبت به همه متغیرها بدست آورید.

(ب) پارامترهای یک شبکه عصبی در ابتدا به صورت تصادفی و با مقادیر کوچک مقداردهی می شوند. توضیح دهید در صورت عدم رعایت این دو ویژگی در مقداردهی چه مشکلاتی بروز پیدا می کند.

(ج) وزن های شبکه ی عصبی بدست آمده در قسمت اول را با مقادیر تصادفی دلخواه مقداردهی کنید و برای یک ورودی دلخواه، با توجه به مشتقاتی که در قسمت اول بدست آوردید، با اعمال بهینه سازی گرادیان کاهشی برای یک ایپاک با نرخ یادگیری  $0.1$ ، وزن های شبکه را آپدیت کنید.

۴. (۲۵ نمره)

الگوریتم آدام<sup>۲</sup> برای آموزش وزن های یک شبکه عصبی به صورت تکراری گام های زیر را اجرا می کند:

$$\begin{aligned} g_t &\leftarrow \nabla_{\theta} f_t(\theta_{t-1}) \\ m_t &\leftarrow \beta_1 m_{t-1} + (1 - \beta_1) g_t \\ v_t &\leftarrow \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \\ \hat{m}_t &\leftarrow \frac{m_t}{1 - \beta_1^t} \\ \hat{v}_t &\leftarrow \frac{v_t}{1 - \beta_2^t} \\ \theta_t &\leftarrow \theta_{t-1} - \frac{\alpha \hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \end{aligned}$$

(آ) الگوریتم بالا را خط به خط توضیح دهید.

(ب) نشان دهید چرا مقادیر  $m_t$  به سمت صفر بایاس دارند؟ چرا مقدار  $\hat{m}_t$  که به شکل  $\frac{m_t}{1 - \beta_1^t}$  محاسبه می شود (در  $t$  های با مقادیر کمتر)، با این مشکل روبرو نمی شود؟ (توجه کنید که مقدار اولیه  $m_0 = 0$  است)

۵. (۵۰ نمره) تصور کنید که تابع هدف یک مدل یادگیری ماشین به صورت  $w^T H w$  باشد که اگر از تجزیه مقادیر ویژه استفاده کنیم خواهیم داشت:

$$H = Q \lambda Q^T$$

(آ) اگر از روش گرادیان کاهشی با طول گام  $\epsilon$  استفاده کنیم، فرمول یادگیری ضرایب به چه صورت است؟

(ب) با شروع از حالت اولیه  $w$  ضرایب در گام  $t$  به چه صورت خواهد بود؟

(ج) تحت چه شرایطی این الگوریتم همگرا می شود؟

(د) حال بررسی کنید اگر از روش نیوتن استفاده کنیم، یادگیری به چه صورت خواهد بود؟ چند گام طول می کشد تا همگرا شویم؟

(ه) چرا با وجود اینکه روش مرتبه ۲ نیوتن از روش مرتبه ۱ گرادیان کاهشی بسیار سریع تر همگرا می شود، در آموزش شبکه های عمیق از آن استفاده نمی شود؟

۶. (۲۵ نمره) تابع خطا در یک شبکه با اعمال Dropout گوسی-جمعی به شکل زیر است:

$$J_1 = \frac{1}{2} \left( y_d - \sum_{k=1}^n (w_k + \delta_k) x_k \right)^2$$

که در آن  $\delta_k \sim N(0, \alpha w_k^2)$  می باشد.

(آ) مقدار امید ریاضی گرادیان تابع هدف نسبت به متغیر  $w_k$  را محاسبه و تا حد امکان ساده کنید.

$$E\left[\frac{\partial J_1}{\partial w_i}\right]$$

(ب) آیا می توانید تعبیری از رگولاسیون با استفاده از این نوع Dropout ارائه دهید؟

Adam<sup>۲</sup>

۱. (۱۵۰ نمره) در این تمرین شبکه عصبی Fully-Connected را پیاده سازی خواهید کرد. در این پیاده سازی با طراحی شبکه، انتشار رو به عقب (backpropagation)، انواع بهینه سازها، Dropout برای رگولاسیون شبکه آشنا خواهید شد. فایل zip. داده شده شامل یک نوتبوک و یک فایل fully-connected-networks.py می باشد. برای هر قسمت شما باید فایل پایتون py. را کامل کنید و با تست پیاده سازی خود در فایل نوتبوک از صحت پیاده سازی درست خود اطمینان حاصل کنید.

**توجه!**

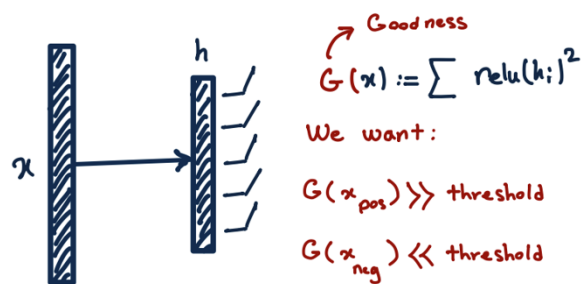
(آ) هیچ کدی را خارج از بلوک های مشخص شده در نوتبوک و فایل پایتون ننویسید یا تغییر ندهید.

(ب) سلول های نوتبوک را اضافه یا کم نکنید.

(ج) در نهایت هر دو فایل py. و نوتبوک را به عنوان پاسخ نهایی ارسال نمایید.

۲. (۱۵۰ نمره) هدف از این مسئله پیاده سازی الگوریتم  $forward - forward$  و آموزش یک شبکه عصبی به وسیله آن بر روی دیتاست MNIST است. این الگوریتم در مقاله ای از ارائه جفری هینتون در NeurIPS ۲۰۲۲ شرح داده شده است که شما از طریق این **لینک** می توانید آن را مطالعه کنید.

به طور خلاصه، الگوریتم  $forward - forward$  جایگزینی برای EBP است که از دو مسیر  $forward$  استفاده میکند. ایراد EBP عدم تطابق آن با مدل مغز، عدم پیاده سازی بهینه آن برای کاربرد های  $real time$  و نیاز آن به وجود یک مدل کامل است. از این جهات الگوریتم  $Forward - Forward$  می تواند جایگزین EBP باشد. در این الگوریتم دیتا به همراه لیبیل درست تبدیل به یک وکتور و همچنین دیتا به همراه لیبیل غلط به یک وکتور دیگر تبدیل می شوند سپس به کمک یک تابع هزینه محلی وزن ها تعیین می شوند. این تابع هزینه محلی برای هر لایه تعریف می شود و در نتیجه دیگر نیازی به  $back propagation$  نمی باشد. هم چنین مفهومی به نام  $goodness$  تعریف می شود که لازم است مقدار آن برای دیتاهای صحیح بیشینه و برای دیتا های غلط کمینه باشد. دقت کنید خروجی هر لایه نیاز به نورمالایز شدن دارد تا از تحریک لایه های بعد جلوگیری شود.



در ابتدا می خواهیم این الگوریتم را به روش Supervised پیاده کنیم. برای پیاده سازی این الگوریتم به ترتیب گام های زیر را انجام دهید:

(آ) **لود کردن دیتاست:** دیتاست مورد استفاده در این بخش دیتاست MNIST است. با تعریف یک dataloader مناسب داده ها را به شیوه مناسب بخوانید.

(ب) **تولید داده:** برای این کار ابتدا لیبیل ها را به یک بردار one-hot تبدیل کرده و آن را در تصویر هر دیتا جایگذاری نمایید. از لیبیل های اشتباه برای تولید دیتای منفی استفاده کنید.

(ج) **پیاده سازی شبکه:** می دانیم هر لایه به صورت محلی محاسبات مربوط به آپدیت وزن های و خطا را انجام می دهد. با تعریف مناسب تابع خطا به نحوی که شروط لازم  $goodness$  برای داده های مثبت و منفی برقرار باشد، هر لایه را به صورت مناسب تعریف کنید.  $loss$  برابر است با:

$$loss = mean(\log(1 + e^{[(threshold - positedata), (negativedata - threshold)]}))$$

توضیح دهید چرا تعریف loss به شکل بالا میتواند شروط لازم goodness را برآورده کند در نهایت با استفاده از لایه های پیاده سازی شده، شبکه یادگیری را تعریف کنید. توجه داشته باشید که شبکه باید قابلیت آموزش و پیش بینی داشته باشد. دقت کنید در هر اپیاک باید از batch دیتای درست، دیتای غلط را تولید و به همراه دیتای درست forward کنید.

(د) **گزارش نتایج:** در انتها مقدار خطا و دقت مدل را برای داده های آموزش و تست بدست آورده و گزارش دهید. هم چنین توضیحات لازم راجع به الگوریتم و نحوه پیاده سازی آن را در گزارش خود ذکر کنید.

حال این الگوریتم را به روش Unsupervised پیاده سازی می کنیم. دقت کنید در این روش مجاز به استفاده از لیبیل ها در فرآیند آموزش شبکه FF نمی باشید:

(آ) **تعریف mask مناسب:** با استفاده از روش توضیح داده شده در مقاله mask مناسب را برای تولید داده های منفی تعریف کنید.

(ب) **تولید داده:** به کمک mask ای که در قسمت قبل تعریف کردید، تابعی برای تولید داده های هیبرید که به عنوان داده های منفی استفاده می شوند بنویسید.

(ج) **پیاده سازی شبکه:** مجدداً با تعریف مناسب تابع loss به نحوی که شروط لازم goodness برای داده های مثبت و منفی برقرار باشد، لایه را تعریف کنید loss برابر است با:

$$loss = mean(\log(1 + e^{[(threshold - positedata), (negativedata - threshold)]}))$$

در نهایت شبکه یادگیری را به نحوه مناسب پیاده سازی کنید. بدیهتاً این شبکه باید قابلیت آموزش و پیش بینی را داشته باشد. سپس با استفاده از تابعی که برای داده های منفی تعریف کرده بودید، شبکه FF را آموزش دهید. دقت کنید در هر اپیاک باید از batch دیتای درست، دیتای غلط را تولید و به همراه دیتای درست forward کنید. توضیح دهید خروجی شبکه شما چیست و چه ارتباطی با لیبیل ها دارد؟

(د) **پیاده سازی طبقه بند خطی:** یک linear classifier تعریف کنید. دقت کنید فقط مجاز به استفاده از شبکه FF در حالت inference هستید. چرا باید طبقه بند خطی تعریف و آموزش داده شود؟

(ه) **گزارش نتایج:** در انتها مقدار خطا و دقت مدل را برای داده های آموزش و تست بدست آورده و گزارش دهید. هم چنین توضیحات لازم راجع به الگوریتم و نحوه پیاده سازی آن را در گزارش خود ذکر کنید.