

## به نام ایزد منان



تمرین اول درس مبانی هوش محاسباتی، «شبکه‌های عصبی»

استاد درس: دکتر عبادزاده



بهار ۱۴۰۲ - دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر

نکاتی در مورد این تمرین نیاز به توجه و دقت دوستان دارد.

- ۱- هرگونه کپی کردن باعث عدم تعلق نمره به تمامی افراد مشارکت کننده در آن می‌شود.
- ۲- آخرین مهلت ارسال تمرین، ساعت ۲۳:۵۵ دقیقه روز ۱۵ فروردین ۱۴۰۲ می‌باشد. این زمان با توجه به جمع‌بندی‌های صورت گرفته، شرایط و با توجه به سایر تمرین‌ها در نظر گرفته شده است و قابل تمدید نمی‌باشد.
- ۳- دوستان فایل ارسالی خود را به صورت فشرده و به صورت «شماره دانشجویی\_HW3» مانند HW3\_97310000 نام گذاری کنید.
- ۴- در صورت هرگونه سوال یا مشکل می‌توانید با تدریس‌یاران درس از طریق آیدی‌های زیر یا ایمیل در ارتباط باشید.

@Amirparsa\_sal

@Farshid\_Nooshi

@amirrezarajabB

@anotherbrickinthewall

@Shakibaam79

aut.ci.spring2023@gmail.com

(<sup>۱</sup>) فرض کنید  $n$  متغیر boolean داریم که ورودی‌های ما هستند و می‌خواهیم حاصل XOR بین این متغیرها را حساب کنیم. همچنین فرض کنید در تمامی قسمت‌های زیر توابع فعال‌سازی همه نورون‌ها تابع پله است.

(a) آیا می‌توان تنها با یک پرسپترون این کار را انجام داد؟ توضیح دهید.

(b) با کمک گرفتن از شبکه مورد نیاز برای تابع دو متغیره XOR ساختار شبکه‌ای را ترسیم کنید که تعداد نورون‌های آن از مرتبه  $\theta(n)$  و تعداد لایه‌های پنهان آن از مرتبه  $\theta(\log n)$  باشد.

(c) با کمک گرفتن از شبکه مورد نیاز برای تابع دو متغیره XOR ساختار شبکه‌ای را ترسیم کنید که تعداد نورون‌ها و تعداد لایه‌های پنهان آن از مرتبه  $\theta(n)$  باشد.

(d) آیا می‌توان با یک شبکه عصبی MLP با یک لایه مخفی این کار را انجام داد؟ اگر بله حداقل تعداد نورون‌های لایه مخفی چقدر خواهد بود؟ (راهنمایی: از فرم  $sop^1$  تابع منطقی XOR استفاده کنید)

(<sup>۲</sup>) فرض کنید مطابق شکل زیر یک شبکه عصبی MLP با چهار ورودی داریم. لایه پنهان اول آن به نام A دارای چهار نورون، لایه پنهان دوم آن به نام D دارای سه نورون و لایه سوم پنهان آن به نام E دارای دو نورون است. همچنین در لایه خروجی یک نورون به نام Y داریم. (دقت شود که خروجی نورون A2 علاوه بر لایه D به E1 نیز متصل است) لازم به ذکر است که تابع فعال‌سازی در لایه‌های پنهان به ترتیب از نوع  $\tanh$  و  $\tanh$  و  $\swish$  با پارامتر یک و در لایه خروجی از نوع sigmoid است.

(a) با مقدار دهی اولیه بایاس‌ها به صفر و وزن‌ها به ۰.۵، با ورودی‌های  $X_0 = 1$  و  $X_1 = 1$  و  $X_2 = 1$  و  $X_3 = 1$  مقدار خروجی شبکه عصبی را حساب کنید.

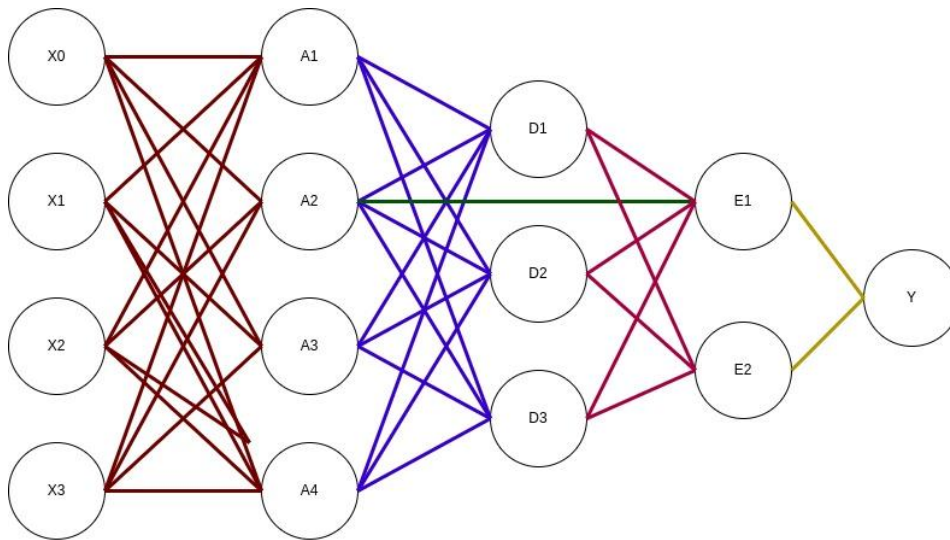
(b) اگر خروجی مورد انتظار شبکه با ورودی‌های بالا برابر یک باشد با تابع هزینه کمترین مربعات خطا<sup>۲</sup> و با نرخ یادگیری<sup>۳</sup> 0.1 وزن‌های خروجی A2 را به روز رسانی کنید.

---

<sup>1</sup> Sum Of Product

<sup>2</sup> Mean Squared Error

<sup>3</sup> Learning Rate

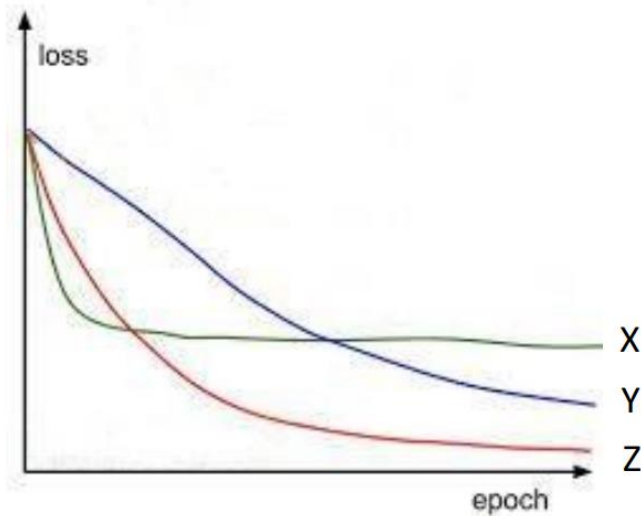


(۳) به سوالات زیر پاسخ دهید.

- (a) تفاوت الگوریتم‌های stochastic gradient descent و batch gradient descent در چیست و هر کدام باید در چه مواردی استفاده شوند؟ فرق این دو الگوریتم با الگوریتم mini-batch gradient descent چیست؟
- (b) توضیح دهید با افزایش مقدار لامبدا (پارامتر منظم‌سازی<sup>۴</sup>) در عملکرد مدل چه تغییری حاصل شده و تاثیر آن در بیش‌برازش چیست؟ در چه مواقعی از منظم‌سازی L1 و در چه مواقعی از منظم‌سازی L2 استفاده میکنیم؟
- (c) یکی از روش‌های جلوگیری از بیش‌برازش<sup>۵</sup> روش Dropout است. در مورد این روش تحقیق کنید و نحوه عملکرد آن را توضیح دهید. اگر حذف برخی از نورون‌ها در این تکنیک منجر به عملکرد بهتر مدل می‌شود، چرا از اول از یک شبکه عصبی ساده‌تر با لایه‌های کمتر و نورون‌های کمتر استفاده نمی‌کنیم؟
- (d) نمودار خطا بر حسب شماره epoch زیر از آموزش یک شبکه عصبی یکسان با نرخ یادگیری‌های متفاوت با مقادیر X, Y, Z بدست آمده است. این مقادیر را با هم مقایسه کنید (رابطه هر سه مقدار را با هم توضیح دهید).

<sup>4</sup> Regularization

<sup>5</sup> Overfitting



۴) همانطور که می‌دانید در شبکه‌های RBF وزن‌های وارد شونده به لایه میانی همان مراکز نواحی تحت پوشش شبکه هستند که فواصل ورودی‌های شبکه نسبت به آن‌ها محاسبه می‌شود و وارد تابع فعال سازی نورون‌های میانی می‌شود. نحوه محاسبه فاصله بین وزن‌ها و ورودی‌های شبکه شکل ناحیه تحت پوشش شبکه را مشخص می‌کند. یک خانواده معروف از توابع فاصله، خانواده مینکوفسکی<sup>۶</sup> است که به شکل زیر تعریف می‌شود: (X, Y دو بردار هستند)

$$d_k(x, y) = \left( \sum_{i=1}^n (x_i - y_i)^k \right)^{\frac{1}{k}}$$

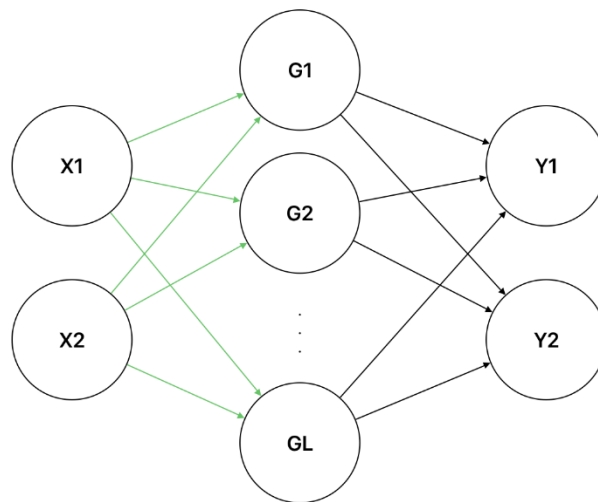
a) مشخص کنید که به ازای  $k=1$  و  $k=2$  این تابع برابر با کدام یک از توابع فاصله معروف می‌شود و شکل ناحیه تحت پوشش شبکه در این حالت چگونه خواهد بود.

b) اگر  $k$  به بی‌نهایت میل کند تابع فاصله حاصل چه چیزی خواهد بود؟ شکل ناحیه تحت پوشش شبکه را در این حالت توصیف کنید.

---

<sup>6</sup> Minkowski

۵) شکل زیر یک شبکه RBF را نشان می‌دهد که دارای  $L$  نورون در لایه میانی است. فرض کنید وزن‌های ورودی به لایه میانی که همان مراکز شبکه هستند (رنگ سبز) از قبل مشخص شده‌اند. اگر تعداد داده‌های آموزشی  $m$  باشد، ماتریس وزن‌های ورودی به لایه آخر شبکه ( $W$ ) را برحسب ماتریس مقادیر لایه میانی شبکه ( $G$ ) و ماتریس مقادیر خروجی مطلوب ( $Y$ ) محاسبه کنید. در محاسبات خود از منظم‌سازی  $L2$  استفاده کنید و ابعاد ماتریس‌ها را نیز ذکر کنید.



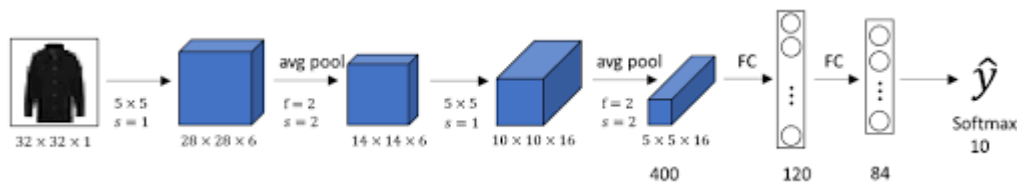
۶) به سوالات زیر پاسخ دهید.

a) خروجی لایه کانولوشنی زیر را حساب کنید.

Input					kernel		
1	2	3	3	*	1	-1	1
0	1	2	2		-1	1	-1
1	2	3	0		1	-1	1
1	2	3	0				

(b) اگر خروجی واقعی بالا ماتریس همانی باشد و بخواهیم با توجه با تابع هزینه کمترین مربعات خطا و نرخ یادگیری ۰.۱ وزن های کرنل خود رو به روز رسانی کنیم بعد از یک مرحله به چه وزن هایی می رسیم؟

(c) فرض کنید برای پیاده سازی سه لایه آخر شبکه عصبی زیر به جای استفاده از لایه های تماماً متصل<sup>۷</sup> از لایه های کانولوشنی استفاده کنیم پیشنهاد شما چیست ؟



راهنمایی: به ابعاد ورودی و خروجی توجه کنید.

(۷) مجموعه داده ای از ۵۰۰ نقطه را در یک فضای دوبعدی در نظر بگیرید، با مختصات X و Y که به طور تصادفی از توزیع یکنواخت بین ۰ و ۱ تولید می شوند. یک SOM با یک شبکه 5x5 از نورون ها برای خوشه بندی این مجموعه داده طراحی کنید. وزن نورون ها را به طور تصادفی آغاز کنید. از یک توزیع یکنواخت بین ۰ و ۱، و از تابع همسایگی گاوسی با نرخ یادگیری رو به کاهش استفاده کنید. در مورد این شبکه به سوالات زیر پاسخ دهید.

- مفهوم تابع همسایگی در SOM و چگونگی تأثیر آن بر فرآیند یادگیری را توضیح دهید.
- نقش نرخ یادگیری در SOM چیست و چگونه در طول آموزش به روز می شود؟
- در مورد مقدار دهی اولیه وزن ها در SOM و اهمیت آن در فرآیند یادگیری توضیح دهید.
- توضیح دهید که چگونه اندازه شبکه SOM بر عملکرد خوشه بندی و پیچیدگی محاسباتی تأثیر می گذارد.
- ماتریس وزن برای شبکه 3x3 SOM با ۴ ویژگی ورودی خواهد بود:

Neuron 1 = (0.1, 0.3, 0.5, 0.7)  
 Neuron 2 = (0.2, 0.4, 0.6, 0.8)  
 Neuron 3 = (0.3, 0.5, 0.7, 0.9)  
 Neuron 4 = (0.4, 0.6, 0.8, 0.1)  
 Neuron 5 = (0.5, 0.7, 0.9, 0.2)  
 Neuron 6 = (0.6, 0.8, 0.1, 0.3)  
 Neuron 7 = (0.7, 0.9, 0.2, 0.4)  
 Neuron 8 = (0.8, 0.1, 0.3, 0.5)  
 Neuron 9 = (0.9, 0.2, 0.4, 0.6)

<sup>7</sup> Fully Connected

3 نقطه داده نمونه برای تطبیق به مانند زیر هستند:

$$x1 = [0.2, 0.6, 0.1, 0.8]$$

$$x2 = [0.4, 0.2, 0.9, 0.5]$$

$$x3 = [0.7, 0.3, 0.4, 0.6]$$

برای حل این بخش باید به ترتیب زیر پیش بروید:

الف) فاصله اقلیدسی بین هر بردار ورودی و بردارهای وزن نورون ها را برای تعیین بهترین واحد تطبیق (BMU) برای هر ورودی محاسبه کنید BMU. ها را برای هر ورودی یادداشت کنید.

ب) با استفاده از نرخ یادگیری  $\alpha=0.1$  و تابع همسایگی با شعاع  $r=1$ ، وزن نورون ها را با استفاده از معادله زیر به روز کنید:

$$w(i,j) = w(i,j) + \alpha * h(i,j,t) * (x - w(i,j))$$

که در آن  $w(i,j)$  بردار وزن نورون در موقعیت  $(i,j)$  است،  $\alpha$  نرخ یادگیری است،  $h(i,j,t)$  تابع همسایگی است که در مرکز BMU در زمان  $t$ ،  $x$  است. بردار ورودی، و  $t$  تکرار جاری است.

ج) مرحله الف و ب را برای مجموع 3 تکرار تکرار کنید.

\*\*تابع همسایگی را تابع گاوسی در نظر بگیرید که فرمول آن در زیر آمده است:

$$h(i,j,t) = \exp(-d(i,j)^2 / (2 * \sigma(t)^2))$$

سیگما در این فرمول برابر عبارت زیر است:

$$\sigma(t) = \sigma_0 * \exp(-t/\tau)$$

که در آن  $\sigma_0$  مقدار اولیه سیگما است و  $\tau$  یک ثابت زمانی است که سرعت کاهش سیگما را کنترل می کند. در این سوال فرض کنید که  $\sigma_0 = 1$  و  $\tau = 3$  هستند.

با این دو مقدار  $\sigma(1) = 0.606$  خواهد بود.

## سوالات امتیازی:

برای سوالات ۸ و ۹ فقط کافی است یک سوال را انتخاب کرده و جواب دهید. اگر هر دو را جواب دهید همچنان نمره یک سوال را میگیرید.

۸) شبکه‌های عصبی عملکرد قابل توجهی در کاربردهای مختلف نشان داده‌اند، اما اغلب به دلیل «Black Box» بودن که تفسیر آنها دشوار است مورد انتقاد قرار می‌گیرند. برخی از روش‌هایی را که می‌توان برای به دست آوردن بینش در مورد عملکرد درونی یک شبکه عصبی مورد استفاده قرار داد، مانند تکنیک‌های تجسم، نقشه‌های برجسته<sup>۸</sup> و انتشار ارتباط لایه‌ای<sup>۹</sup> توضیح دهید. در مورد محدودیت‌ها و چالش‌های هر رویکرد بحث کنید و مثال‌هایی از نحوه استفاده از آنها در برنامه‌های کاربردی دنیای واقعی ارائه دهید.

۹) شبکه‌های عصبی مستعد حملات خصمانه هستند، جایی که اغتشاشات کوچک در ورودی می‌تواند باعث پیش‌بینی نادرست شبکه شود. برخی از تکنیک‌هایی را که می‌توان برای دفاع در برابر حملات خصمانه استفاده کرد، مانند آموزش دشمن<sup>۱۰</sup>، پیش‌پردازش ورودی و تنظیم مدل<sup>۱۱</sup> توضیح دهید. درباره اثربخشی و محدودیت‌های هر تکنیک بحث کنید و مثال‌هایی از نحوه استفاده از آنها در سناریوهای دنیای واقعی، مانند رانندگی مستقل و تشخیص چهره ارائه دهید.

۱۰) کیفیت خوشه‌بندی را برای SOM آموزش‌دیده ارزیابی کنید و در مورد محدودیت‌ها یا مسیرهای آینده برای بهبود دقت خوشه‌بندی بحث کنید.

موفق باشید

تیم تدریسیاری

---

<sup>8</sup> Saliency Map

<sup>9</sup> Layer-wise relevance propagation

<sup>10</sup> Adversarial Training

<sup>11</sup> Input preprocessing and model regularisation