

به نام خدا

هوش محاسباتی: تمرین سری دوم

تاریخ تحویل: جمعه ۱۹ آبان

۱- یک وظیفه آموزشی ثابت با  $L$  الگوی آموزشی در اختیار داریم. می‌خواهیم یک شبکه MLP آموزش دهیم که از روی بردار ورودی‌ها، خروجی متناظر آنها را تولید کند. شخصی پیشنهاد می‌دهد به جای استفاده از تعریف خطای مرسوم، از تعریف زیر به عنوان خطای یک الگوی آموزشی استفاده کنیم:

$$e^{(\ell)} = \sum_{v \in U_{out}} e_v^{(\ell)}$$
$$e_v^{(\ell)} = \begin{cases} \frac{1}{2} - \frac{1}{2} \cos\left(\frac{(o_v^{(\ell)} - out_v^{(\ell)})^2 \pi}{2\sigma}\right) & (o_v^{(\ell)} - out_v^{(\ell)})^2 \leq 2\sigma \\ 1 & otherwise \end{cases}$$

که در آن  $o$  و  $out$  به ترتیب بیانگر خروجی مطلوب و خروجی واقعی هستند و  $\ell$  شماره الگوی آموزشی بوده و  $\sigma$  یک پارامتر قابل تنظیم است.

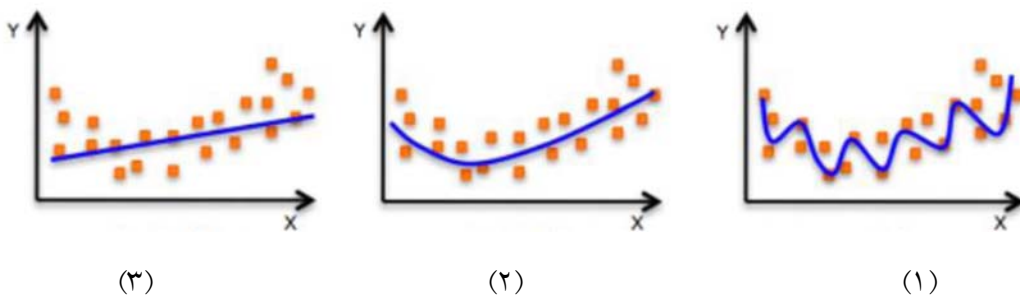
الف) این تعریف خطا، در تعریف خطای مرسوم چه تغییری ایجاد کرده است؟ این تعریف خطا چه مزیتی نسبت به تعریف خطای مرسوم دارد؟ بزرگ یا کوچک کردن پارامتر  $\sigma$  چه تاثیری دارد؟

ب) اگر از الگوریتم Gradient Descent برای آموزش شبکه استفاده کنیم، رابطه تغییر وزن یک نورون خروجی را در یک الگوریتم Batch به دست آورید.

روابط تغییر وزن مربوط به تعریف خطای مرسوم به صورت زیر است:

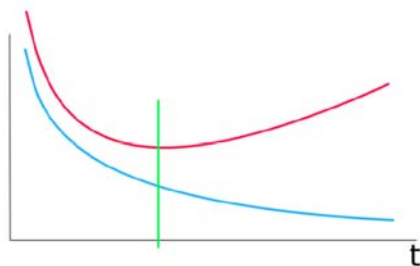
$$\nabla_{w_u} e^{(\ell)} = \frac{\partial e^{(\ell)}}{\partial w_u} = \frac{\partial e^{(\ell)}}{\partial net_u^{(\ell)}} \frac{\partial net_u^{(\ell)}}{\partial w_u}$$
$$\frac{\partial net_u^{(\ell)}}{\partial w_u} = in_u^{(\ell)}$$
$$\frac{\partial e^{(\ell)}}{\partial net_u^{(\ell)}} = \frac{\partial \sum_{v \in U_{out}} (o_v^{(\ell)} - out_v^{(\ell)})^2}{\partial net_u^{(\ell)}} = \sum_{v \in U_{out}} \frac{\partial (o_v^{(\ell)} - out_v^{(\ell)})^2}{\partial net_u^{(\ell)}}$$
$$\frac{\partial e^{(\ell)}}{\partial net_u^{(\ell)}} = -2 \sum_{v \in U_{out}} (o_v^{(\ell)} - out_v^{(\ell)}) \frac{\partial out_v^{(\ell)}}{\partial net_u^{(\ell)}} = -2\delta_u^{(\ell)}$$
$$\forall u \in U_{out}: \quad \delta_u^{(\ell)} = (o_u^{(\ell)} - out_u^{(\ell)}) \frac{\partial out_u^{(\ell)}}{\partial net_u^{(\ell)}}$$
$$\forall u \in U_{out}: \quad \nabla_{w_u} e_u^{(\ell)} = \frac{\partial e_u^{(\ell)}}{\partial w_u} = -2(o_u^{(\ell)} - out_u^{(\ell)}) \frac{\partial out_u^{(\ell)}}{\partial net_u^{(\ell)}} in_u^{(\ell)}$$
$$\forall u \in U_{out}: \quad \Delta w_u^{(\ell)} = -\frac{\eta}{2} \nabla_{w_u} e_u^{(\ell)} = \eta(o_u^{(\ell)} - out_u^{(\ell)}) \frac{\partial out_u^{(\ell)}}{\partial net_u^{(\ell)}} in_u^{(\ell)}$$

۲- الف) اصطلاحات Overfitting و Underfitting را توضیح دهید و مشخص کنید در کدام یک از شکل‌ها یکی از این دو مورد مشاهده می‌شود.



ب) دلایل ایجاد بیش‌برازش در شبکه‌های عصبی پرسپترون را توضیح دهید. در بیش‌برازش خطای آموزش و خطای ارزیابی به چه صورت است؟

ج) شکل زیر نمودار خطا را برای داده‌های آموزشی و داده‌های ارزیابی نشان می‌دهد. اولاً مشخص کنید، هر نمودار چه خطایی را نشان می‌دهد. ثانیاً مشخص کنید در چه بخشی از نمودار بیش‌برازش رخ داده است؟



د) روش‌های جلوگیری از ایجاد Underfitting را در شبکه‌های عصبی پرسپترون توضیح دهید.

ه) در مورد روش k-fold cross-validation تحقیق کنید و توضیح دهید که این روش برای جلوگیری از underfitting مناسب است یا overfitting؟ چرا؟

۳- فرض کنید که تستی برای تشخیص یک بیماری وجود دارد. در صورتی که فرد بیمار، سالم تشخیص داده شود، درمان وی شروع نمی‌شود و در پی این موضوع بیماری وی پیشرفت خواهد کرد. اما اگر فرد سالم به اشتباه بیمار شناخته شود، هزینه زیادی در بر نخواهد داشت. با توجه به توضیحات گفته شده، مینیموم کردن کدام یک از مقادیر false positive و false negative برای ما اولویت دارد؟ با توجه به این موضوع کدام یک از دو معیار specificity و sensitivity مهم‌تر است؟

۴- یک پرسپترون دو لایه با  $n$  نورون ورودی و  $m$  نورون خروجی و  $n \times m$  اتصال بین نورون‌ها را با مشخصات زیر در نظر بگیرید:

تابع ورودی: مجموع وزن‌دار ورودی‌ها

تابع فعال‌سازی: تابع softmax که به صورت زیر تعریف می‌شود:

$$f_{act,i} = \frac{e^{net_i}}{\sum_{j=1}^m e^{net_j}}$$

این تابع در واقع مقدار نرمالیزه شده یک نورون خروجی را نسبت به نورون‌های لایه خروجی محاسبه می‌کند.

تابع خروجی: همانی

تابع خطا: تابع Cross Entropy که به صورت زیر تعریف می‌شود:

$$e = - \sum_{i=1}^m o_i \log(out_i)$$

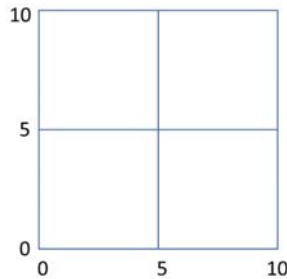
که  $o_i \in \{0, 1\}$  برچسب هدف برای نورون خروجی  $i$  و  $out_i$  خروجی نورون  $i$  است.

قانون تغییر وزن‌های اتصالات،  $w_{ab}$ ‌ها، را بدست آورید ( $a = \{1, \dots, n\}, b = \{1, \dots, m\}$ ).

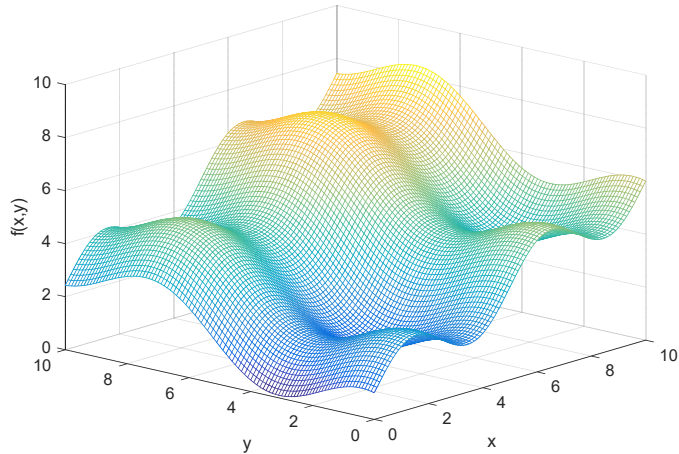
۵- در یک صفحه شطرنجی  $N \times N$  می‌خواهیم تعداد  $N$  وزیر بگنجانیم. یک شبکه عصبی MLP طراحی کنید به گونه ای که اگر وزیر‌ها همدیگر را تهدید کردند خروجی ۰ و اگر هیچ دو وزیری هم را تهدید نکردند (یعنی هیچ دو وزیر در یک سطر ستون یا قطر قرار نداشته باشند) خروجی ۱ بازگرداند.

۶- یک رویه فرضی  $f(x, y)$  مانند شکل (الف) را در نظر بگیرید (محدوده  $x$  و  $y$  بین ۰ تا ۱۰).

فرض کنید ۵۰۰ نمونه از نقاط این رویه به عنوان داده آموزش به ما داده شده است و قرار است یک شبکه عصبی MLP طراحی کنیم که به بهترین نحو این رویه را تولید کند.



(ب)



(الف)

(الف) در ساده‌ترین حالت فرض کنید کل محدوده رویه را به چهار بخش مربعی مشابه شکل (ب) تقسیم می‌کنیم و مقدار  $f$  را در هر بخش ثابت در نظر می‌گیریم. شبکه ساده‌ای طراحی و رسم کنید که رویه را با در نظر گرفتن این چهار ناحیه تخمین بزند. نوع توابع فعالسازی و مقادیر وزن‌ها و آستانه‌ها را به طور دقیق مشخص کنید یا نحوه محاسبه آنها را توضیح دهید.

(ب) فرض کنید می‌خواهیم یک شبکه MLP طراحی کنیم که خطای داده‌های آموزشی از آستانه داده‌شده  $E_m$  کمتر شود. از شبکه طراحی‌شده در بخش (ج) شروع کرده و در یک الگوریتم تکرارشونده، در هر گام یک یا چند نورون به شبکه اضافه می‌کنیم، به گونه‌ای که در هر مرحله خطای داده‌های آموزشی کاهش یابد. الگوریتم پیشنهادی را به صورت یک شبه‌کد بنویسید. نحوه تعیین مقادیر وزن‌ها و آستانه‌ها را در هر مرحله توضیح دهید. توجه کنید نورون اضافه‌شده در هر مرحله ممکن است بنا بر ضرورت الگوریتم در راستای  $x$  یا  $y$  نواحی را تغییر دهد و الزامی بر متقارن بودن نواحی در دو راستا نیست.

۷- در یک مسئله فروشنده دوره‌گرد، قرار است یک فروشنده سفر خود را از یک شهر شروع کرده، به شهرهای دیگر سفر کند و در انتهای سفر به شهر مبدأ بازگردد. سفری معتبر است که فروشنده به هر شهر دقیقاً یک بار سفر کند. در اینجا برای سادگی فرض کنید که هفت شهر داریم. یک الگوریتم خودکار برای حل مسئله ارائه داده‌ایم که یک رشته ۷ تایی از شماره شهرها را در خروجی ایجاد می‌کند که به ترتیب شماره شهرهایی که فروشنده به آنها سر زده است را مشخص می‌کند. به طور مثال ممکن است خروجی الگوریتم به صورت زیر (3571352) باشد، یعنی سفر از شهر ۳ شروع شده، بعد به ۵ رفته و... و در آخر از شهر ۲ به مبدأ باز می‌گردد.

الف) یک شبکه MLP ارائه دهید که رشته خروجی الگوریتم خودکار را به عنوان ورودی دریافت کرده و در خروجی خود مشخص کند که رشته ارائه شده پاسخی معتبر برای مسئله است یا خیر. در شبکه MLP، ۷ نورون ورودی قرار دهید که ورودی خارجی نورون  $k$ -ام شماره شهری است که در گام  $k$ -ام به آن سفر شده است. در شبکه MLP طراحی شده، نوع توابع ورودی، فعالسازی و خروجی را مشخص کنید. وزن ها و آستانه ها را نیز به طور کامل مشخص کنید. نیازی به رسم همه جزییات شبکه نیست، بخش هایی از آن را رسم کنید و در مورد بقیه قسمت ها توضیح دهید.

ب) شبکه بخش (الف) را به گونه ای تغییر دهید که در خروجی مشخص کند در رشته ورودی، به چند شهر سفر نشده است.

ج) فرض کنید خطوط ارتباطی بین شهرها فقط به صورت افقی و عمودی باشد و در نتیجه فاصله بین شهر  $i$ -ام و  $j$ -ام به صورت زیر محاسبه شود:

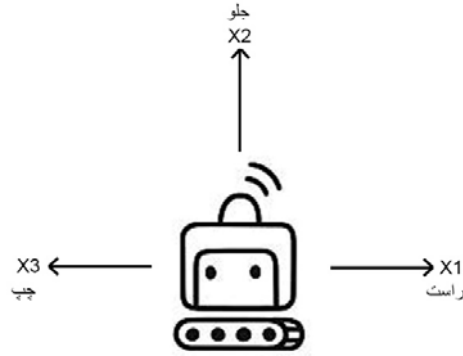
$$d(i, j) = |x_i - x_j| + |y_i - y_j|$$

که  $x_i$  و  $y_i$  مختصات شهر  $i$ -ام هستند. یک شبکه MLP طراحی کنید که مختصات شهر اول تا هفتمی را که به آنها سفر شده در ورودی دریافت کند و طول مسیر را در خروجی مشخص کند. فرض کنید بیشترین فاصله افقی یا عمودی بین دو شهر  $L_m$  باشد. در شبکه MLP طراحی شده، نوع توابع ورودی، فعالسازی و خروجی را مشخص کنید. وزن ها و آستانه ها را نیز به طور کامل مشخص کنید. نیازی به رسم همه جزییات شبکه نیست، بخش هایی از آن را رسم کنید و در مورد بقیه قسمت ها توضیح دهید.

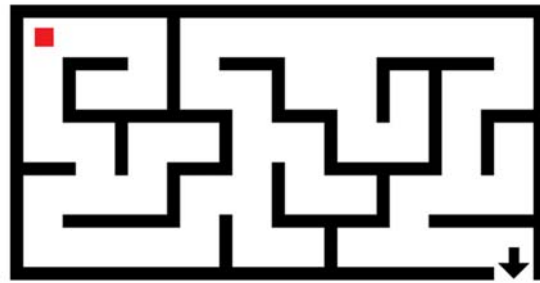
۸- تصور کنید که شما یک ربات امدادرسان حاوی کمک های اولیه برای حادثه دیده های زلزله به یک ساختمان فروپاشیده اعزام کرده اید اما به دلیل مشکلات فنی، دوربین ربات از کار افتاده است و نمی توانید آن را از دور کنترل کنید. تنها سیگنال هایی که از ربات به شما می رسند، سیگنال سنسورهای مربوط به وجود مانع در هر لحظه اطراف ربات است (مطابق شکل). با طراحی یک شبکه MLP، کد ۲ بیتی را خروجی دهید که فرمان حرکت به ۴ جهت (راست، جلو، چپ، عقب) را با توجه به موانع اطراف صادر کند. در پاسخ خود اتصالات، وزن ها، آستانه ها، و توابع را مشخص کنید.

ورودی شبکه:

$$x_i = \begin{cases} 1, & \text{if obstacle} \\ 0, & \text{o.w} \end{cases}$$



توجه کنید که شبکه شما قرار نیست مسیر بهینه را پیاده‌سازی کند، بلکه باید از برخورد ربات با موانع خطرناک جلوگیری کند و بالاخره از ماز خارج شود. دقت کنید که شبکه پیشنهادی نباید باعث شود تا ربات تا ابد در ماز گیر کند.



موانع و محل ربات (مربع قرمز)

۹- با در نظر گرفتن تابع هزینه زیر به سوالات پاسخ دهید:

$$f(x) = \frac{5}{6}x^4 - 7x^3 + \frac{115}{6}x^2 - 18x + 6$$

الف) تابع فوق در یک بازه مناسب رسم کنید.

همانطور که مشاهده می‌شود تابع دارای دو کمینه می‌باشد. می‌خواهیم تاثیر نرخ یادگیری را در الگوریتم Gradient descent بررسی کنیم.

ب) مقدار اولیه  $x$  را یک متغیر تصادفی با توزیع نرمال در نظر بگیرید و میانگین آن را طوری انتخاب کنید که فاصله آن از هر دو کمینه به صورت تقریبی برابر باشد. با در نظر گرفتن نرخ یادگیری برابر با  $\eta = 0.02$  محاسبه کنید که با چند قدم می‌توان به یکی از کمینه‌ها رسید. حد همگرایی را برابر  $f'(x) < 0.01$  قرار دهید.

قسمت (ب) را برای هر یک از مقادیر  $\eta = [0.01, 0.02, 0.04, 0.08, 0.1]$  چندین بار تکرار کنید.

ج) برای هر یک از مقادیر بالا، احتمال رسیدن به کمینه مطلق را بدست آورید و رسم کنید.

د) برای هر یک از مقادیر میانگین تعداد گام‌ها برای رسیدن به کمینه مطلق و کمینه محلی را بدست آورید و

رسم کنید.

ه) برای هر یک از مقادیر بالا نرخ واگرایی را بدست آورید و رسم کنید.