

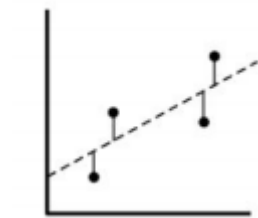
بسمه تعالی

تکلیف اول درس هوش محاسباتی

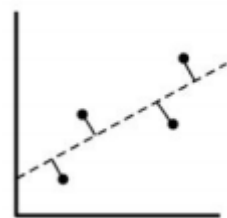
موعد تحویل: ۱۴ فروردین ۱۴۰۰

رگرسیون خطی:

- ۱- مساله رگرسیون خطی را در نظر بگیرید، کدامیک از اشکال زیر به درستی میتواند، محاسبه فاصله مجموعه نقاط از خط تخمین زده شده را نشان دهد؟ دلیل انتخاب خود را توضیح دهید.



فاصله عمودی



فاصله قائم

- ۲- فرض کنید میخواهیم مدلی برای تخمین قیمت املاک با استفاده از ویژگی های مختلف از جمله تعداد اتاق نشیمن، ارزش محله طراحی کنیم. اگر قسمت تخمینی خانه را زیاده از حد معمول اعلام کنیم ممکن است خانه به فروش نرسد ولی ضرر مالی کمی میبینیم چون میتوان در آینده قیمت آن را بالا برد. در صورتی که اگر قیمت خانه خیلی پایین تر از قیمت واقعی آن اعلام شود، خانه بلافاصله به فروش میرسد و ضرر مالی جبران ناپذیر خواهد بود. بنابراین برای اینکار نیازمند مدلی هستیم که در بیشتر مواقع مقدار پیشبینی شده بزرگتر از مقدار واقعی باشد یا اگر کمتر است فاصله خیلی کمی وجود داشته باشد. به عبارتی دیگر نیازمند مدلی هستیم که مقادیر underestimate را بیشتر از مقادیر overestimate جرمه کند. سعی کنید یک تابع هزینه asymmetric طراحی کنید که این کار را انجام دهد، سپس نمودار این تابع را نسبت به تابع هزینه معمول در روش رگرسیون خطی رسم کنید.

- ۳- تابع زیر را با دو متغیر در نظر بگیرید،

$$f(x, y) = \left(\frac{3}{4}x - \frac{3}{2}\right)^2 + (y - 2)^2 + \frac{1}{4}xy$$

بردار گرادیان این تابع را نسبت به پارامتر های آن محاسبه کنید و با استفاده از روش gradient descent مقدار مینیمم آن را با شروع از نقطه (۴، ۵) پیدا کنید. برای اینکار نرخ یادگیری را در سه مرحله از مجموعه مقادیر [۰/۷۵، ۰/۵، ۰/۰۱] انتخاب کنید و اگر به مقدار بهینه نزدیک شدید تعداد مراحل طی شده برای رسیدن به مقدار بهینه را گزارش کنید.

- ۴- الگوریتم gradient descent با یک نرخ یادگیری ثابت را نظر بگیرید. برای ۴ تابع مختلف، مقادیر تابع در هر مرحله گزارش شده است. در هر یک از حالات زیر بگویید مقدار نرخ یادگیری خیلی بزرگ، خیلی کوچک یا مناسب است.

الف) $f_1 = 100, 99, 98, 97, 96, 95, 94, 93, 92, 91$

ب) $f_2 = 100, 50, 75, 60, 65, 45, 75, 110, 90, 85$

ج) $f_3 = 100, 80, 65, 50, 40, 35, 31, 29, 28, 27.5, 27.3$

د) $f_4 = 100, 80, 60, 40, 20, 0, -20, -40, -60, -80, -100$

رگرسیون لاجیستیک:

۵- با رسم شکل، مساله ای را مثال بزنید که توزیع نمونه های موجود در دیتاست به گونه ای است که قادر به استفاده از رگرسیون خطی برای دسته بندی آنها نیستیم. چند دلیل برای حمایت از ادعای خود ذکر کنید، سپس تفاوت مهم **logistic regression** نسبت به **linear regression** که باعث میشود این روش سازگاری بیشتری با مسائل شبیه به مثال شما داشته باشد را شرح دهید.

۶- یک روش برای حل مسائل دسته بندی استفاده از رگرسیون خطی به همان شکلی که برای خروجی های حقیقی^۱ استفاده میکردیم است. استفاده از این روش باعث کاهش دقت تصمیمگیری میشود و اساسا داشتن خروجی هایی کمتر از ۰ و بیشتر از ۱ با فرض مساله یعنی $y \in \{0,1\}$ تناقض دارد. برای حل این مشکل از یک تابع دیگر برای نگاشت خروجی روش قبل به بازه ۰ تا ۱ استفاده میکنیم. یکی از توابعی که میتوان برای اینکار استفاده کرد تابع سیگموئید $g(z) = \sigma(\theta^T x)$ است که در کلاس درس معرفی شد. الف) مشتق تابع سیگموئیدی $g(z)$ نسبت به z را بر حسب خود تابع به دست بیاورید. این رابطه یکی از ویژگی های مهم تابع سیگموئید است که در تحلیل های قسمت بعدی سوال به شما کمک میکند.

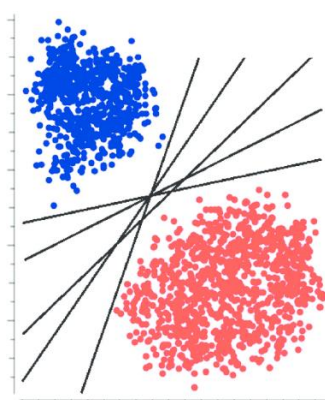
ب) در مورد دیگر توابع که برای اینکار استفاده میشوند تحقیق کنید و با تحلیل ویژگی های هر کدام، بهترین تابع را برای رگرسیون لاجیستیک انتخاب کنید. سعی کنید در تحلیل های خود از رسم نمودار این توابع کمک بگیرید.

(راهنمایی: به این توابع در شبکه های عصبی، توابع فعالسازی گفته میشود. برای مثال دو ویژگی از ویژگی هایی که میتواند باعث نتایج بهتر شود **zero-centered** بودن تابع فعالسازی و از بین نرفتن گرادیان در طی مراحل **gradient descent** هنگام استفاده از آن تابع است.)

۷- **منظم سازی**^۲: با اضافه کردن عبارت منظم سازی در رگرسیون لاجیستیک، تابع هدف میتواند به شکل رابطه زیر باشد:

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m \left[-y^i (\log(h_{\theta}(x^i))) - (1 - y^i) (\log(1 - h_{\theta}(x^i))) \right] + \lambda \theta_j^2$$

در این مساله خاص، $j \in \{0,1,2\}$ مجموعه مقادیر ممکن برای پارامتر j است. مشتق تابع هزینه را نسبت به هر یک از پارامتر های وزن θ_j ، و به دست آورید، سپس با توجه به داده های آموزشی نمودار زیر و مرزهای تصمیم گیری احتمالی، توضیح دهید بعد از منظم سازی خطای آموزش برای مقادیر بزرگ λ ، به ازای هر یک پارامتر های وزن بیشتر میشود یا تغییر نمیکند؟ دلایل خود را به صورت مختصر و برای هر وزن به صورت جداگانه توضیح دهید.



¹ Real values

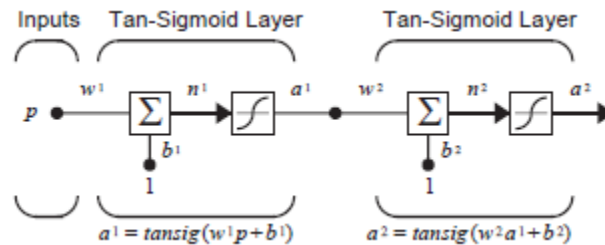
² Regularization

۸- **الگوریتم عقبگرد**^۳: فرض کنید در شبکه دولایه ای که در شکل زیر میبینید، در مرحله ای از یادگیری، وزن ها و بایاس ها به صورت زیر انتخاب شده اند و نرخ یادگیری ۱ است.

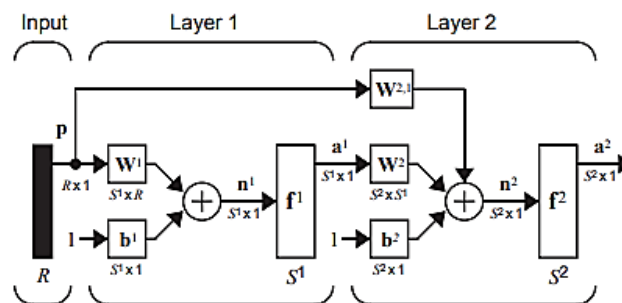
$$w^1 = -1, b^1 = 1$$

$$w^2 = -2, b^2 = 1$$

با انتشار ورودی (p=-1, t=1) در شبکه و تابع هزینه cross entropy، خروجی هر لایه را به دست آورده و با به دست آوردن مشتقات، مقادیر وزن ها و بایاس ها را به روزرسانی کنید. (این کار را تنها برای یک مرحله انجام دهید).



۹- شبکه عصبی زیر، یک شبکه feed-forward دولایه معمولی است، با این تفاوت که یک ارتباط از ورودی به لایه دوم اضافه شده است.



الف) هدف از قسمت اول این تمرین به دست آوردن تمامی آپدیت های الگوریتم عقبگرد برای تمامی پارامترهای شبکه است. با فرض اندازه داده های آموزشی m و تابع هزینه cross entropy، سعی کنید روابط به روزرسانی پارامترها را به صورت ماتریسی ارائه دهید. مسیر پیشرو در شبکه به صورت زیر است:

$$n^1 = W^1 p + b^1$$

$$a^1 = f^1(n^1) = f^1(W^1 p + b^1)$$

$$n^2 = W^2 a^1 + W^{2,1} p + b^2$$

$$a^2 = f^2(n^2) = f^2(W^2 a^1 + W^{2,1} p + b^2)$$

ب) در این قسمت با اضافه کردن ترم ضریب منتظم سازی به تابع هزینه نشان دهید معادلات آپدیت چه تغییری خواهند کرد.

معادلات آپدیت ماتریسی را بازنویسی کنید.

موفق باشید

نسرتین صالحی

³ Backpropagation