



هوش مصنوعی

بهار ۱۴۰۲

استاد: محمدحسین رهبان

دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر

گردآوردندگان: پارسا حسینی، یاسمن زلفی، امیرحسین محمد رضایی، علی نظری

مهلت ارسال: -

شبکه‌های عصبی

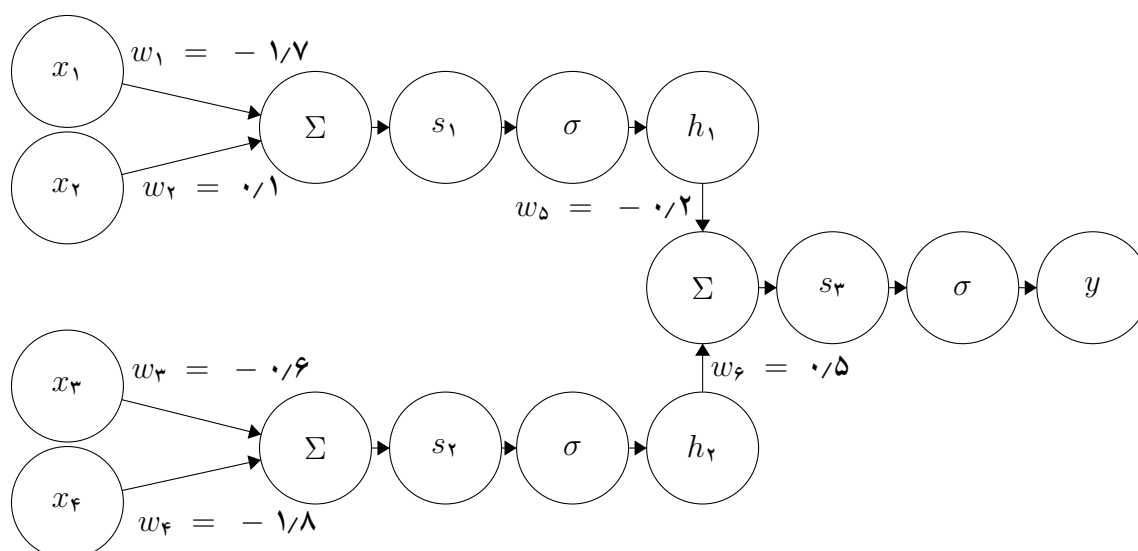
تمرین ششم

سوالات نظری (۱۴۰ نمره)

۱. (۲۰ نمره) شبکه‌ی زیر را در نظر بگیرید که متشکل از برخی متغیرها و برخی تابع‌ها است. منظور از σ همان logistic function است که با کمک تابع زیر محاسبه می‌شود:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

به عنوان نمونه، مقدار میانی h_1 در شبکه زیر برابر با $h_1 = \frac{1}{1 + e^{-x_1 w_1 - x_2 w_2}}$ است.



برای loss function هم L2 loss را در نظر بگیرید که به شکل $L(y, \hat{y}) = ||y - \hat{y}||^2$ محاسبه می‌شود. فرض کنیم ورودی $(x_1, x_2, x_3, x_4) = (-0/7, 1/2, 1/1, -2)$ است و مقدار واقعی خروجی هم 0.5 باید باشد. با استفاده از backpropagation مقدار $\frac{\delta L}{w_1}$ را محاسبه کنید.

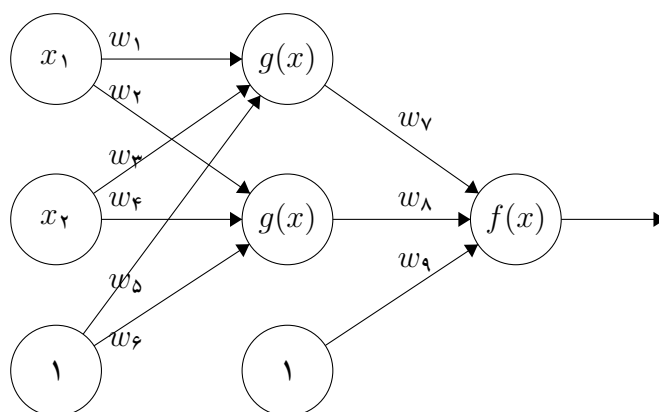
۲. (۲۰ نمره) فرض کنید که یک تابع convex مانند f روی بازه بسته $[-b, b]$ داریم. f نشان دهنده مشتق تابع f است. نرخ یادگیری را هم با α نشان می‌دهیم. برای رساندن تابع به کمترین مقدار خود، یکی از روش‌ها استفاده از کاهش گرادیان است. به عنوان مثال از $x_0 = 0$ شروع می‌کنیم و از روش کاهش گرادیان استفاده می‌کنیم. اگر مقدار پس از به روز رسانی مقدار زیر $-b$ بود، آن را برابر با $-b$ قرار می‌دهیم و اگر مقدار بعد از به روز رسانی بالای b بود هم مقدار را برابر با b می‌گذاریم. حال به یک الگوریتم optimization مانند کاهش گرادیان، $\epsilon - converges$ گفته می‌شود که اگر در نقطه‌ای مقدار x به فاصله کمتر از ϵ به مقدار واقعی برسد.

(آ) برای $\alpha = 0/1$ و $b = 1$ و $\epsilon = 0/001$ یک تابع convex پیدا کنید که با اجرای کاهش گرادیان روی آن، $\epsilon - converges$ رخ ندهد. به طور خاص جوری جلو بروید که $x_0 = 0$ و $x_1 = b$ و $x_2 = -b$ و $x_3 = b$ و $x_4 = -b$ و ...

(ب) برای $\alpha = 0.1$ و $b = 1$ و $\epsilon = 0.001$ یک تابع convex پیدا کنید که $\epsilon - \text{converges}$ داشته باشد ولی بعد از حداقل 10000 گام اجرای کاهش گرادیان.

(ج) یک الگوریتم optimization دیگر طراحی کنید که این ویژگی را داشته باشد که همیشه $\epsilon - \text{converges}$ کند برای هر تابع convex و در $\log_2(2b/\epsilon)$ قدم هم این اتفاق بیفتد.

۳. (۲۰ نمره) شبکه عصبی زیر برای دسته‌بندی binary در نظر گرفته شده است که یک لایه نهان دارد. در واحدهای نهان از تابع فعال‌سازی خطی $g(x) = cx$ و در واحد خروجی از تابع فعال‌سازی sigmoid که به صورت $f(x) = \frac{1}{1+e^{-x}}$ است، استفاده شده است تا تابع $P(y = 1|x, w)$ یاد گرفته شود. $x = (x_1, x_2)$ ورودی و $w = (w_1, w_2, w_3, \dots, w_9)$ وزن‌های شبکه عصبی هستند.

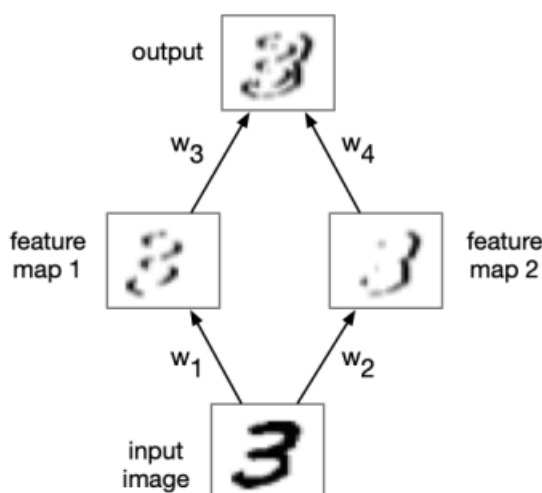


(آ) خروجی شبکه بالا را بر حسب c ، ورودی‌های x_i و وزن‌های w_i بیان کنید. مرز (boundary) مربوط به classifier نهایی چیست؟

(ب) شبکه عصبی معادل شبکه بالا که فاقد لایه نهان است را طراحی کنید و وزن‌های شبکه جدید را بر حسب c و w_i مشخص کنید.

(ج) آیا هر multi layered neural net با توابع فعال‌سازی خطی را می‌توان به صورت یک neural net بدون لایه نهان نشان داد؟ پاسخ خود را توضیح دهید.

۴. (۲۰ نمره) در این سوال به دنبال طراحی شبکه convolutional هستیم که مرزهای عمودی تصویر را تشخیص دهد. معماری شبکه در زیر نشان داده شده است.



تابع فعال‌سازی اعمال‌شده به لایه convolution اول ReLU و تابع فعال‌سازی اعمال‌شده به لایه خروجی همانی است.

$$ReLU(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases}$$

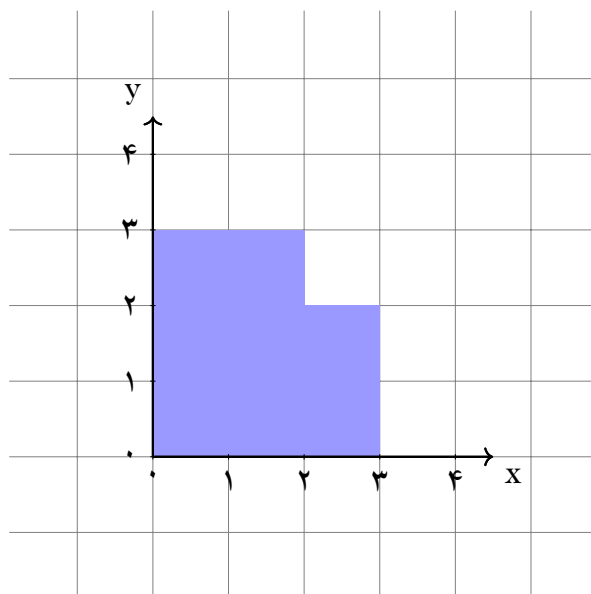
در این سوال از رنگ سفید برای نشان دادن صفر و از رنگ‌های تیره‌تر برای نشان دادن مقادیر بزرگ‌تر استفاده شده است.

(فرض کنید که w_1 و w_2 دو کرنل $3 \times 3 \times 1$ هستند که پس از اعمال آن‌ها، دو کانال به شکل نشان داده شده ایجاد می‌شود. سپس با اعمال یک کرنل $3 \times 3 \times 2$ به نام w که از الحاق w_3 و w_4 با ابعاد $3 \times 3 \times 1$ تشکیل شده است، خروجی ایجاد می‌شود.)

(آ) دو کرنل w_1 و w_2 با اندازه $3 \times 3 \times 1$ را طراحی کنید. یکی از آن‌ها مرز سیاه/سفید و دیگری مرز سفید/سیاه را تشخیص می‌دهد.

(ب) حال w_3 و w_4 با اندازه $3 \times 3 \times 1$ را به گونه‌ای معرفی کنید که خروجی مورد نظر را تولید کند.

۵. (۲۰ نمره) شبکه عصبی طراحی کنید که با ورودی گرفتن دو عدد حقیقی x و y مشخص کند که آیا نقطه با مختصات داده شده در ناحیه رنگی قرار دارد یا ندارد. وزن لایه‌های شبکه و توابع فعال‌سازی را که در شبکه عصبی استفاده کرده‌اید را مشخص کنید.



۶. (۲۰ نمره) به سوالات زیر پاسخ دهید.

(آ) مشکل vanishing gradient و exploding gradient در شبکه‌های عصبی را شرح دهید و برای مقابله با هر کدام راهکاری پیشنهاد دهید.

(ب) مزایا و معایب اضافه کردن لایه‌های بیشتر به یک شبکه عصبی چه می‌تواند باشد؟

(ج) کدام یک از توابع زیر را می‌توان به عنوان تابع فعال‌ساز در یک شبکه عصبی استفاده کرد؟

$$f(x) = -\min(2, x)$$

$$f(x) = \begin{cases} \max(x, 0/x), & x \geq 0 \\ \min(x, 0/x), & x < 0 \end{cases}$$

$$f(x) = \begin{cases} \min(x, \cdot/\cdot x), & x \geq \cdot \\ \min(x, \cdot/\cdot x), & x < \cdot \end{cases}$$

۷. (۲۰ نمره) به سوالات زیر پاسخ دهید.

- (آ) یک شبکه عصبی دولایه برای رگرسیون با p متغیر ورودی x_1, \dots, x_p ، یک لایه مخفی با اندازه U ، تابع فعال‌سازی $h: \mathbb{R} \rightarrow \mathbb{R}$ و خروجی $\hat{y} \in \mathbb{R}$ رسم کنید. تعداد پارامترهای مدل چندتا است؟
(ب) مدل قسمت قبل را می‌توان به صورت زیر نوشت:

$$q_i = h\left(b_i^{(1)} + \sum_{j=1}^p W_{ij}^{(1)} x_j\right), \quad i = 1, \dots, U \quad (1)$$

$$\hat{y} = b^{(2)} + \sum_{l=1}^U W_l^{(2)} q_l \quad (2)$$

همانطور که می‌دانید، برای پیاده‌سازی این مدل باید معادلات را به فرمت برداری/ماتریسی داشته باشیم. معادلات بالا را به کمک متغیرهای زیر برداری کنید و در نهایت به فرمت $\hat{y} = f(\mathbf{x})$ بنویسید. دقت کنید که جواب نهایی نباید هیچ سیگما یا حلقه‌ای داشته باشد.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}, \mathbf{q} = \begin{bmatrix} q_1 \\ \vdots \\ q_U \end{bmatrix}, \mathbf{b}^{(1)} = \begin{bmatrix} b_1^{(1)} \\ \vdots \\ b_U^{(1)} \end{bmatrix}, \mathbf{W}^{(1)} = \begin{bmatrix} W_{11}^{(1)} & \dots & W_{1p}^{(1)} \\ \vdots & & \vdots \\ W_{U1}^{(1)} & \dots & W_{Up}^{(1)} \end{bmatrix}, \quad (3)$$

$$\mathbf{b}^{(2)} = [b^{(2)}], \mathbf{W}^{(2)} = [W_1^{(2)}, \dots, W_U^{(2)}]$$

- (ج) حالا فرض کنید $\{\mathbf{x}_i\}_{i=1}^n$ داده‌های ما باشند. برای هر کدام از این داده‌ها رابطه قسمت قبل برقرار است، یعنی:

$$\hat{y}_i = f(\mathbf{x}_i), \quad i = 1, \dots, n \quad (4)$$

این n معادله را هم به فرمت ماتریسی بنویسید، یعنی به فرمت $\hat{\mathbf{Y}} = f(\mathbf{X})$. از متغیرهای زیر کمک بگیرید.

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix}, \hat{\mathbf{Y}} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{bmatrix}, \mathbf{Q} = \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} \quad (5)$$

- (د) فرض کنید تابع فعال‌سازی همانی باشد یعنی $h(x) = x$. در این صورت نشان دهید مدل بالا معادل با یک مدل رگرسیون خطی می‌شود. به طور خاص، باید نشان دهید تمام پارامترهای شبکه عصبی به فرمت یک مدل رگرسیون خطی می‌شود:

$$\hat{y} = \theta_0 + \sum_{j=1}^p \theta_j x_j \quad (6)$$