بسم الله الرحمن الرحیم

پرسش ۱. Matrix Differentiation (۲۰ نمره)

برای یک تابع $f : \mathbb{R}^n \to \mathbb{R}^m$ می‌توان مشتق آن را به ازای یک ورودی نظیر $x \in \mathbb{R}^n$ به صورت زیر تعریف کرد:

$$J_{i,j} = \frac{\partial f_i}{\partial x_j}, \quad J = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \cdots\cdots\cdots & \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

اولین نکته قابل توجه این است که این ماتریس $J \in \mathbb{R}^{m\times n}$ بوده و سطرهای آن ترانهاده گرادیان هر یک از ابعاد خروجی نسبت به ورودی می‌باشند. برخی منابع ترانهاده این ماتریس را به عنوان مشتق در نظر می‌گیرند و شما باید همواره به این نکته دقت داشته باشید. در صورتی که قرار باشد از یک تابع مانند $f(X) \in \mathbb{R}^{n\times m}$ بر حسب یک ماتریس همچون $X \in \mathbb{R}^{k\times p}$ مشتق بگیریم، حاصل این کار یک تنسور[1] $\frac{\partial f(X)}{\partial X} \in \mathbb{R}^{n\times m\times k\times p}$ از مرتبه چهار خواهد شد.

[1] Tensor

به صورت مشابه می‌توان برای ابعاد بالاتر نیز مشتق‌گیری را انجام داد. به یاد داشته باشید که برای مشتق‌گیری از هر تابعی با هر ابعادی نسبت به هر ورودی با هر ابعادی، کافیست نسبت به المان‌های آن‌ها، نظیر به نظیر مشتق جزئی را محاسبه نماییم و مقادیر بدست آمده را کنار هم قرار دهیم. یک روش ساده برای جلوگیری از مواجهه با تنسورها این است که در صورت نیاز، ماتریسی همچون $A \in \mathbb{R}^{m\times n}$ را به شکل یک بردار تخت نظیر $a \in \mathbb{R}^{mn}$ درآورد و نسبت به آن مشتق بگیریم. به زبان ریاضی خواهیم داشت:

$$A \in \mathbb{R}^{n\times m}, a \in \mathbb{R}^{nm} \rightarrow A_{ij} = a_{(i-1)n+j}$$

با استفاده از این تعریف تلاش کنید تا به پرسش‌های زیر پاسخ دهید و هرجا که نیاز شد، بجای ایجاد تنسور از تخت‌سازی ماتریس‌ها استفاده کنید. (ذکر دقیق مراحل برای کسب نمره ضروری است)

۱. اگر $a, x \in \mathbb{R}^n$ باشند، نشان دهید که:

$$\frac{\partial(a^\top x)}{\partial x} = \frac{\partial(x^\top a)}{\partial x} = a^\top.$$

۲. برای $x \in \mathbb{R}^n$ و $A \in \mathbb{R}^{m\times n}$، مقدار

$$\frac{\partial(Ax)}{\partial x}$$

را بیابید.

۳. برای $A \in \mathbb{R}^{n\times n}$ و $x \in \mathbb{R}^n$، عبارت

$$\frac{\partial(x^\top Ax)}{\partial x}$$

را محاسبه کنید. همچنین مشتق نسبت به A به صورت

$$\frac{\partial(x^\top Ax)}{\partial A}$$

را نیز تعیین کنید.

۴. برای $A,\, X \in \mathbb{R}^{n\times n}$، مقدار

$$\frac{\partial\, \mathrm{tr}(X^\top AX)}{\partial X}$$

را محاسبه کنید.

پرسش ۲. Backpropagation (۲۵ نمره)

در این سوال، با یک مسئله دسته بندی سه کلاسه روبه‌رو هستیم. معماری شبکه‌ی عصبی مورد استفاده به صورت زیر است:

$$l^{(1)} = \mathrm{ReLU}(W^{(1)}x), \quad l^{(21)} = \mathrm{ReLU}(W^{(21)}l^{(1)}), \quad l^{(22)} = \sigma(W^{(22)}l^{(1)}), \quad z = \max(l^{(21)}, l^{(22)})$$

علاوه بر این، از لایه‌ی Softmax برای خروجی شبکه استفاده شده است:

$$\hat{y} = \mathrm{softmax}(z)$$

ابعاد متغیرها به‌صورت زیر داده شده‌اند:

$$x \in \mathbb{R}^4, \quad W^{(1)} \in \mathbb{R}^{2\times 4}, \quad W^{(21)}, W^{(22)} \in \mathbb{R}^{3\times 2}$$

۱. گراف محاسباتی این شبکه را رسم کنید. در این گراف، هر گره نشان‌دهنده‌ی یک عملیات (نظیر ReLU، sigmoid، ضرب ماتریسی، و انتخاب ماکسیمم) است و یال‌ها وابستگی بین این مقادیر را نشان می‌دهند.

۲. در مرحله‌ی Backward Pass، گرادیان تابع هزینه $\mathcal{L}$ نسبت به وزن‌های شبکه را محاسبه کنید. توجه کنید که
در
در Forward Pass برخی مقادیر محاسبه شده و ذخیره می‌شوند و نیازی به محاسبه‌ی مجدد آن‌ها نیست.

پاسخ، از گراف محاسباتی استفاده کنید و روابط زیر را به‌دست آورید:

$$\frac{\partial \mathcal{L}}{\partial W^{(21)}}, \quad \frac{\partial \mathcal{L}}{\partial l^{(22)}}, \quad \frac{\partial \mathcal{L}}{\partial l^{(21)}}, \quad , \quad \frac{\partial \mathcal{L}}{\partial W^{(1)}} \quad \frac{\partial \mathcal{L}}{\partial l^{(1)}}, \quad \frac{\partial \mathcal{L}}{\partial W^{(22)}},$$

در این بخش، مشتق‌ها را گام‌به‌گام بر اساس زنجیره‌ی محاسباتی به‌دست آورید.

۳. مقدار خروجی و گرادیان‌ها را بر اساس مقداردهی اولیه‌ی زیر محاسبه کنید. فرض کنید که تابع هزینه‌ی مورد استفاده Cross Entropy است و داده‌ی ورودی به کلاس دوم تعلق دارد.

$$x = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \end{bmatrix}, \quad W^{(1)} = \begin{bmatrix} -1 & 0 & 1 & -1 \\ 0 & -1 & 1 & 1 \end{bmatrix}, \quad W^{(21)} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad W^{(22)} = \begin{bmatrix} 2 & -1 \\ 4 & -2 \\ -2 & 1 \end{bmatrix}$$

مراحل مورد نیاز برای محاسبه‌ی Forward Pass را به طور کامل نمایش دهید. در پایان، مقدار $\hat{y}$ را محاسبه کرده و لاجیت‌ها را تا دو رقم اعشار گرد کنید.

سپس، با استفاده از قواعد به‌دست‌آمده در قسمت قبل، Backward Pass را اجرا کنید و گرادیان‌های وزن‌ها را تعیین کنید.