تکلیف نهم درس شناسایی الگو

وحید ملکی
شماره دانشجویی: ۴۰۳۱۳۰۰۴

۲۹ آذر ۱۴۰۴

سوال ۱۲

در این سوال یک معماری شبکه عصبی کانولوشنی (CNN) ساده ارائه شده است. روابط حاکم بر شبکه به شرح زیر است:

- **لایه اول (کانولوشن):** شامل یک فیلتر با سایز ۳ است. وزن‌های فیلتر با v_1, v_2, v_3 نمایش داده می‌شوند.

$$h_i = s(u_i) = s\left(\sum_{j=1}^{3} v_j x_{i+j-1}\right)$$

که در آن $s(x)$ تابع فعالیت سیگموئید است و u_i ورودی خالص به نورون h_i می‌باشد.

- **لایه دوم (تمام متصل):** خروجی z ترکیب خطی نورون‌های مخفی است.

$$z = \sum_{i=1}^{4} w_i h_i$$

- **تابع هزینه:** مربع خطا تعریف شده است.

$$R = (y - z)^2$$

الف) تعداد کل پارامترهای شبکه

در شبکه‌های کانولوشنی، اصل «اشتراک وزن» (Weight Sharing) برقرار است. این یعنی وزن‌های استفاده شده برای محاسبه هر h_i یکسان هستند.

- **لایه اول (کانولوشن):** اگرچه خطوط زیادی در شکل رسم شده است، اما تنها یک فیلتر با سایز ۳ وجود دارد. بنابراین تنها ۳ پارامتر v_1, v_2, v_3 در این لایه وجود دارد.
- **لایه دوم (تمام متصل):** هر نورون مخفی (h_1 تا h_4) با یک وزن منحصر به فرد به z متصل است. پس ۴ پارامتر w_1, w_2, w_3, w_4 در اینجا داریم.
- **بایاس:** طبق صورت سوال، ترم‌های بایاس وجود ندارند.

تعداد کل پارامترها:

$$3\ (\text{برای } v) + 4\ (\text{برای } w) = 7\ \text{پارامتر}$$

ب) محاسبه مشتق تابع هزینه نسبت به w_i

برای محاسبه $\frac{\partial R}{\partial w_i}$ از قاعده زنجیره‌ای استفاده می‌کنیم:

$$\frac{\partial R}{\partial w_i} = \frac{\partial R}{\partial z} \cdot \frac{\partial z}{\partial w_i}$$

۱. مشتق هزینه نسبت به خروجی شبکه (z):

$$\frac{\partial R}{\partial z} = \frac{\partial}{\partial z}(y-z)^2 = 2(y-z)(-1) = -2(y-z)$$

۲. مشتق خروجی نسبت به وزن w_i: با توجه به رابطه $z = \sum_{k=1}^{4} w_k h_k$:

$$\frac{\partial z}{\partial w_i} = h_i$$

۳. ترکیب روابط:

$$\frac{\partial R}{\partial w_i} = -2(y-z)h_i$$

ج) فرم برداری رابطه قبل

می‌خواهیم گرادیان $\frac{\partial R}{\partial \mathbf{w}}$ را محاسبه کنیم که در آن $\mathbf{w} = [w_1, w_2, w_3, w_4]^T$ است.

اگر بردار خروجی‌های لایه مخفی را $\mathbf{h} = [h_1, h_2, h_3, h_4]^T$ در نظر بگیریم، با توجه به اینکه عبارت $-2(y-z)$ یک اسکالر است که در تمام مولفه‌ها ضرب می‌شود:

$$\frac{\partial R}{\partial \mathbf{w}} = -2(y-z)\mathbf{h}$$

د) محاسبه مشتق تابع هزینه نسبت به v_j

این قسمت نیازمند دقت بیشتری است زیرا پارامتر v_j در محاسبه تمام h_iها مشترک است (Weight Sharing). بنابراین تغییر در v_j روی تمامی h_1 تا h_4 و در نهایت روی z اثر می‌گذارد. طبق قاعده زنجیره‌ای چند متغیره باید روی تمام مسیرهای ممکن (تمام iها) جمع ببندیم:

$$\frac{\partial R}{\partial v_j} = \frac{\partial R}{\partial z} \sum_{i=1}^{4} \left(\frac{\partial z}{\partial h_i} \cdot \frac{\partial h_i}{\partial v_j} \right)$$

اجزای رابطه را محاسبه می‌کنیم:

۱. $\frac{\partial R}{\partial z} = -2(y-z)$ (مشابه قسمت ب)

۲. $\frac{\partial z}{\partial h_i} = w_i$

۳. محاسبه $\frac{\partial h_i}{\partial v_j}$: می‌دانیم $h_i = s(u_i)$ که در آن $u_i = \sum_{k=1}^{3} v_k x_{i+k-1}$.

$$\frac{\partial h_i}{\partial v_j} = \frac{\partial s(u_i)}{\partial u_i} \cdot \frac{\partial u_i}{\partial v_j}$$

$$\frac{\partial h_i}{\partial v_j} = s'(u_i) \cdot x_{i+j-1}$$

(توضیح: ضریب v_j در عبارت u_i برابر است با x_{i+j-1}).

جایگذاری در رابطه کلی:

$$\frac{\partial R}{\partial v_j} = -2(y-z)\sum_{i=1}^{4}\left(w_i \cdot s'(u_i) \cdot x_{i+j-1}\right)$$

از آنجا که مشتق تابع سیگموئید $s'(x) = s(x)(1-s(x))$ است و $h_i = s(u_i)$، می‌توانیم بنویسیم $s'(u_i) = h_i(1-h_i)$. فرم نهایی به صورت زیر خواهد بود:

$$\frac{\partial R}{\partial v_j} = -2(y-z)\sum_{i=1}^{4} w_i h_i (1-h_i) x_{i+j-1}$$