

# تکلیف نهم درس شناسایی الگو

وحید ملکی  
شماره دانشجویی: ۴۰۳۱۳۰۰۴

۲۹ آذر ۱۴۰۴

## سؤال ۹

### الف- تعداد پارامترهای شبکه

در این شبکه عصبی، پارامترها شامل وزن‌ها و بایاس‌ها هستند. طبق صورت سوال، بایاس‌ها نیز به عنوان پارامتر در نظر گرفته می‌شوند (که در ابعاد بردارهای ورودی گسترش یافته لحاظ شده‌اند).

• لایه اول (ورودی به پنهان):

- تعداد ورودی‌ها (تصویر مسطح شده):  $64 \times 64 = 4096$ .
- از آنجا که بردار ورودی  $x$  گسترش یافته است (شامل بایاس)، طول آن برابر است با  $4096 + 1 = 4097$ .
- تعداد نرون‌های لایه پنهان: 1024.
- ابعاد ماتریس وزن  $V$  برابر است با  $1024 \times 4097$ .
- تعداد پارامترهای این لایه:  
 $1024 \times 4097 = 4,195,328$

• لایه دوم (پنهان به خروجی):

- خروجی لایه پنهان  $h$  نیز گسترش یافته است. تعداد نرون‌های پنهان 1024 است، بنابراین با احتساب یک بایاس اضافی، طول  $h$  برابر 1025 می‌شود.
- تعداد نرون‌های خروجی: 2 (زاویه فرمان و سرعت).
- ابعاد ماتریس وزن  $W$  برابر است با  $2 \times 1025$ .
- تعداد پارامترهای این لایه:  
 $2 \times 1025 = 2,050$

• مجموع کل پارامترها:

$$4,195,328 + 2,050 = 4,197,378$$

بنابراین، این شبکه در مجموع دارای 4,197,378 پارامتر قابل آموزش است.

## ب- ابعاد ماتریس‌ها و بردارها

با توجه به توضیحات سوال و بخش قبل، ابعاد به شرح زیر است:

•  $x$ : بردار ورودی گسترش‌یافته (شامل 4096 پیکسل + 1 بایاس).

of Dimensions  $x : 4097 \times 1$

•  $V$ : ماتریس وزن‌های لایه اول (اتصال ورودی به پنهان). تعداد سطرها برابر تعداد نرون‌های پنهان و ستون‌ها برابر ورودی گسترش‌یافته.

of Dimensions  $V : 1024 \times 4097$

•  $g$ : ورودی خالص لایه پنهان ( $g = Vx$ ).

of Dimensions  $g : 1024 \times 1$

•  $h$ : خروجی لایه پنهان گسترش‌یافته (شامل خروجی تابع فعالیت + 1 بایاس).

of Dimensions  $h : 1025 \times 1$

•  $W$ : ماتریس وزن‌های لایه دوم (اتصال پنهان به خروجی). تعداد سطرها برابر تعداد خروجی و ستون‌ها برابر خروجی پنهان گسترش‌یافته.

of Dimensions  $W : 2 \times 1025$

•  $\hat{y}$  و  $y$ : بردار خروجی شبکه و خروجی مطلوب.

of Dimensions  $\hat{y}, y : 2 \times 1$

## ج- مشتق تابع هزینه نسبت به $W_{ij}$

تابع هزینه عبارت است از:

$$J = \frac{1}{2} \|y - \hat{y}\|^2 = \frac{1}{2} \sum_{k=1}^2 (y_k - \hat{y}_k)^2$$

می‌دانیم که  $\hat{y}_i = \sum_m W_{im} h_m$  (سطر  $i$ -ام ماتریس  $W$  ضرب در بردار  $h$ ).  
برای محاسبه  $\frac{\partial J}{\partial W_{ij}}$  از قاعده زنجیره‌ای استفاده می‌کنیم:

$$\frac{\partial J}{\partial W_{ij}} = \frac{\partial J}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial W_{ij}}$$

مرحله اول (مشتق هزینه نسبت به خروجی):

$$\frac{\partial J}{\partial \hat{y}_i} = \frac{\partial}{\partial \hat{y}_i} \left( \frac{1}{2} (y_i - \hat{y}_i)^2 \right) = -(y_i - \hat{y}_i)$$

مرحله دوم (مشتق خروجی نسبت به وزن):

$$\frac{\partial \hat{y}_i}{\partial W_{ij}} = \frac{\partial}{\partial W_{ij}} \left( \sum_m W_{im} h_m \right) = h_j$$

بنابراین:

$$\frac{\partial J}{\partial W_{ij}} = -(y_i - \hat{y}_i) h_j$$

#### د- مشتق تابع هزینه نسبت به ماتریس $W$

با توجه به رابطه به دست آمده در بخش (ج)، عنصر  $(i, j)$  ماتریس گرادیان برابر است با ضرب اسکالر خطای خروجی  $i$  در ورودی  $j$  (از لایه قبل). به فرم ماتریسی، این عبارت برابر است با حاصل ضرب بردار ستونی خطا در ترانزاده بردار ورودی لایه (برداری  $h$ ). اگر بردار خطا را  $e = y - \hat{y}$  در نظر بگیریم:

$$\frac{\partial J}{\partial W} = -(y - \hat{y})h^T$$

بررسی ابعاد:

$$(2 \times 1) \times (1 \times 1025) = 2 \times 1025$$

که دقیقاً برابر ابعاد ماتریس  $W$  است.

#### ه- مشتق تابع هزینه نسبت به $V_{ij}$

وزن  $V_{ij}$  ورودی  $x_j$  را به نرون پنهان  $i$  متصل می‌کند. مسیر تاثیرگذاری به صورت زیر است:

$$V_{ij} \rightarrow g_i \rightarrow h_i \rightarrow \hat{y} \rightarrow J$$

از قاعده زنجیره‌ای استفاده می‌کنیم. از آنجا که تغییر در  $h_i$  روی تمامی خروجی‌های  $\hat{y}_k$  تاثیر می‌گذارد، باید روی  $k$  جمع ببندیم:

$$\frac{\partial J}{\partial V_{ij}} = \sum_{k=1}^2 \left( \frac{\partial J}{\partial \hat{y}_k} \cdot \frac{\partial \hat{y}_k}{\partial h_i} \right) \cdot \frac{\partial h_i}{\partial g_i} \cdot \frac{\partial g_i}{\partial V_{ij}}$$

اجزای رابطه:

$$\frac{\partial J}{\partial \hat{y}_k} = -(y_k - \hat{y}_k) \quad ۰.۱$$

$$\frac{\partial \hat{y}_k}{\partial h_i} = W_{ki} \quad ۰.۲$$

$$\frac{\partial h_i}{\partial g_i} = f'(g_i) \quad ۰.۳ \quad (\text{مشتق تابع فعالیت ReLU})$$

$$\frac{\partial g_i}{\partial V_{ij}} = x_j \quad ۰.۴$$

جایگذاری در رابطه اصلی:

$$\frac{\partial J}{\partial V_{ij}} = \left( \sum_{k=1}^2 -(y_k - \hat{y}_k) W_{ki} \right) f'(g_i) x_j$$

یا به صورت ساده‌تر با فاکتورگیری از منفی:

$$\frac{\partial J}{\partial V_{ij}} = - \left( \sum_{k=1}^2 (y_k - \hat{y}_k) W_{ki} \right) f'(g_i) x_j$$

## و- روابط بازگشتی به روزرسانی وزن‌ها

با فرض نرخ آموزش  $\eta$ ، قانون به روزرسانی گرادیان نزولی به صورت  $\theta_{new} = \theta_{old} - \eta \frac{\partial J}{\partial \theta}$  است. برای وزن‌های  $W_{ij}$ :

$$W_{ij}^{(new)} = W_{ij}^{(old)} - \eta (-(y_i - \hat{y}_i)h_j)$$

$$W_{ij} \leftarrow W_{ij} + \eta (y_i - \hat{y}_i)h_j$$

برای وزن‌های  $V_{ij}$ : با استفاده از گرادیان محاسبه شده در بخش (ه):

$$V_{ij}^{(new)} = V_{ij}^{(old)} - \eta \left( - \left[ \sum_{k=1}^2 (y_k - \hat{y}_k) W_{ki} \right] f'(g_i) x_j \right)$$

$$V_{ij} \leftarrow V_{ij} + \eta \left( \sum_{k=1}^2 (y_k - \hat{y}_k) W_{ki} \right) f'(g_i) x_j$$