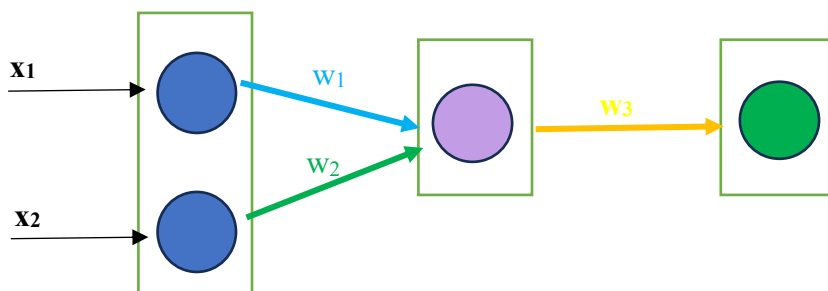


تحقیق ۱:

روش به روزرسانی هر یک از وزن‌ها در روش **backpropagation** طی هر سیکل (epoch):



Input	Hidden	Output
Hidden layer:	$hidden = w_1x_1 + w_2x_2 + b_h$	
Output layer:	$\hat{y} = output = w_3 \times hidden + b_0$	
Error:	$error = (y - \hat{y})^2 = (y - output)^2$	

فرض میکنیم، $x_1=0.6$ ، $x_2=0.9$ و $y=1$

و برای وزن‌ها $w_1=0.5$ ، $w_2=0.4$ و $w_3=0.9$ و برای نرخ یادگیری $\eta=0.1$ آنگاه داریم:

$$hidden = 0.6 \times 0.5 + 0.9 \times 0.4 = 0.66$$

$$\hat{y} = 0.9 \times 0.66 = 0.594$$

$$error = (1 - 0.594)^2 \approx 0.165$$

$$\begin{aligned} \frac{\partial error}{\partial w_3} &= \frac{\partial error}{\partial output} \times \frac{\partial output}{\partial w_3} = -2(y - output) \times hidden \\ &= -2(1 - 0.594) \times 0.66 \approx \\ &-0.536 \end{aligned}$$

$$w_3 = w_3 - \eta \times \frac{\partial error}{\partial w_3}$$

$$\rightarrow w_3 = 0.9 - 0.1 \times -0.536 \approx 0.954$$

$$\begin{aligned} \frac{\partial error}{\partial w_1} &= \frac{\partial error}{\partial output} \times \frac{\partial output}{\partial hidden} \times \frac{\partial hidden}{\partial w_1} \\ &= -2(y - output) \times w_3 \times x_1 \\ &= -2(1 - 0.594) \times 0.954 \times 0.6 \end{aligned}$$

$$\approx -0.438$$

$$w_1 = w_1 - \eta \times \frac{\partial error}{\partial w_1}$$

$$\rightarrow w_1 = 0.5 - 0.1 \times -0.438 \approx 0.544$$

$$\frac{\partial error}{\partial w_2} = \frac{\partial error}{\partial output} \times \frac{\partial output}{\partial hidden} \times \frac{\partial hidden}{\partial w_2}$$

$$= -2(y - output) \times w_3 \times x_2$$

$$= -2(1 - 0.594) \times 0.954 \times 0.9$$

$$\approx -0.697$$

$$w_2 = w_2 - \eta \times \frac{\partial error}{\partial w_2}$$

$$\rightarrow w_2 = 0.4 - 0.1 \times -0.697 \approx 0.47$$

با توجه به وزن‌های به روز شده $w_1=0.544$ ، $w_2=0.47$ و $w_3=0.954$ برای هر یک از لایه‌ها

در epoch اول داریم:

$$hidden = 0.6 \times 0.544 + 0.9 \times 0.47 = 0.749$$

$$\hat{y} = 0.954 \times 0.749 = 0.714$$

$$error = (1 - 0.714)^2 \approx 0.082$$

تحقیق ۲:

روش بهینه‌سازی NADAM و تفاوت آن با ADAM:

نادام یک الگوریتم بهینه‌سازی است که نقاط قوت گرادیان شتاب‌یافته آدام و نستروف (NAG^1) را با هم ترکیب می‌کند. این الگوریتم، مشابه آدام، نرخ یادگیری را برای هر پارامتر به صورت جداگانه تطبیق می‌دهد، اما مومنتوم نستروف را نیز در خود جای می‌دهد که به پیش‌بینی تغییر گرادیان کمک می‌کند و منجر به همگرایی سریع‌تر می‌شود. اساساً، نادام را می‌توان به عنوان آدام با مومنتوم نستروف در نظر گرفت.

:ADAM

آدام (تخمین مومنتوم تطبیقی) یک الگوریتم بهینه‌سازی است که از میانگین‌های نمایی نزولی گرادیان‌های گذشته و گرادیان‌های مربع گذشته برای تطبیق نرخ یادگیری برای هر پارامتر استفاده می‌کند.

:NAG

¹ Nesterov Accelerated Gradient

NAG یک تکنیک بهینه‌سازی مبتنی بر مومنتوم است که با "نگاه به آینده" برای پیش‌بینی موقعیت آینده پارامتر، مومنتوم استاندارد را بهبود می‌بخشد.

NADAM:

نادام قابلیت نگاه به آینده NAG را در قانون به‌روزرسانی آدام ادغام می‌کند. به طور خاص، نحوه استفاده از تخمین مومنتوم اول (یک جزء کلیدی در آدام) را تغییر می‌دهد و تخمین مومنتوم فعلی را هنگام محاسبه به‌روزرسانی در نظر می‌گیرد.

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{\hat{v}_t} + \varepsilon} m'_t$$

در این رابطه \hat{v}_t اصلاح شده دومین ممان نستروف است.

ε ثابت کوچکی است که برای پایداری عددی اضافه شده است، معمولاً 10^{-8} است.

در این رابطه m'_t تخمین اولین ممان نستروف با بایاس اصلاح شده است.

لازم به ذکر است رابطه کلی به روز رسانی وزنها در ADAM مشابه NADAM است با این تفاوت که m'_t و \hat{v}_t به ترتیب اولین و دومین ممنتوم می‌باشند و به صورت ساده‌تر محاسبه میشوند و شامل محاسبه شتاب‌یافته نستروف نمی‌شوند.

مزایا:

این ترکیب اغلب منجر به همگرایی سریع‌تر و بهبود عملکرد در مقایسه با استفاده از Adam یا مومنتوم معمولی به تنهایی می‌شود.

کاربرد:

در حالی که Adam بسیار محبوب است، Nadam می‌تواند جایگزین خوبی باشد، زمانی که Adam در حال نوسان مشاهده می‌شود یا زمانی که همگرایی سریع‌تر مورد نظر است.