

رسالة محمد

مبانی یادگیری عمیق

مدرس: محمدرضا محمدی

۱۴۰۰

Adam

```
first_moment = 0
second_moment = 0
for t in range(1, num_iterations):
    dx = compute_gradient(x)
```

```
first_moment = beta1 * first_moment + (1 - beta1) * dx
```

Momentum

```
second_moment = beta2 * second_moment + (1 - beta2) * dx * dx
```

```
first_unbias = first_moment / (1 - beta1 ** t)
```

Bias correction

```
second_unbias = second_moment / (1 - beta2 ** t)
```

```
x -= learning_rate * first_unbias / (np.sqrt(second_unbias) + 1e-7))
```

AdaGrad / RMSProp

- بهینه‌ساز Adam با پارامترهای $\beta_1 = 0.9$ و $\beta_2 = 0.999$ و نرخ آموزش برابر با $1e-3$ یا $5e-4$ یک نقطه شروع خوب برای بسیاری از مدل‌ها است

نرخ آموزش

- تمام بهینه‌سازهای SGD، SGD+Momentum، Adagrad، RMSProp و Adam دارای ابرپارامتر نرخ آموزش هستند

- می‌توانیم نرخ آموزش را در طول زمان کاهش دهیم

- step decay

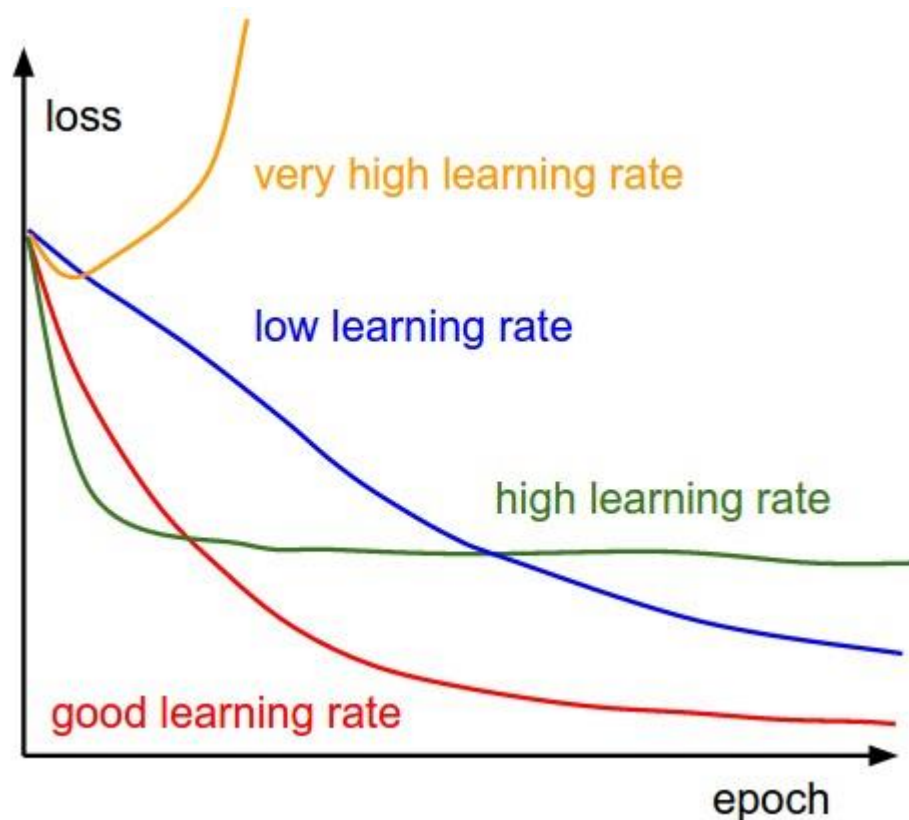
- نرخ آموزش بعد از چند epoch در ضریبی مانند ۰.۵ ضرب می‌شود

- exponential decay

$$\alpha = \alpha_0 e^{-kt}$$

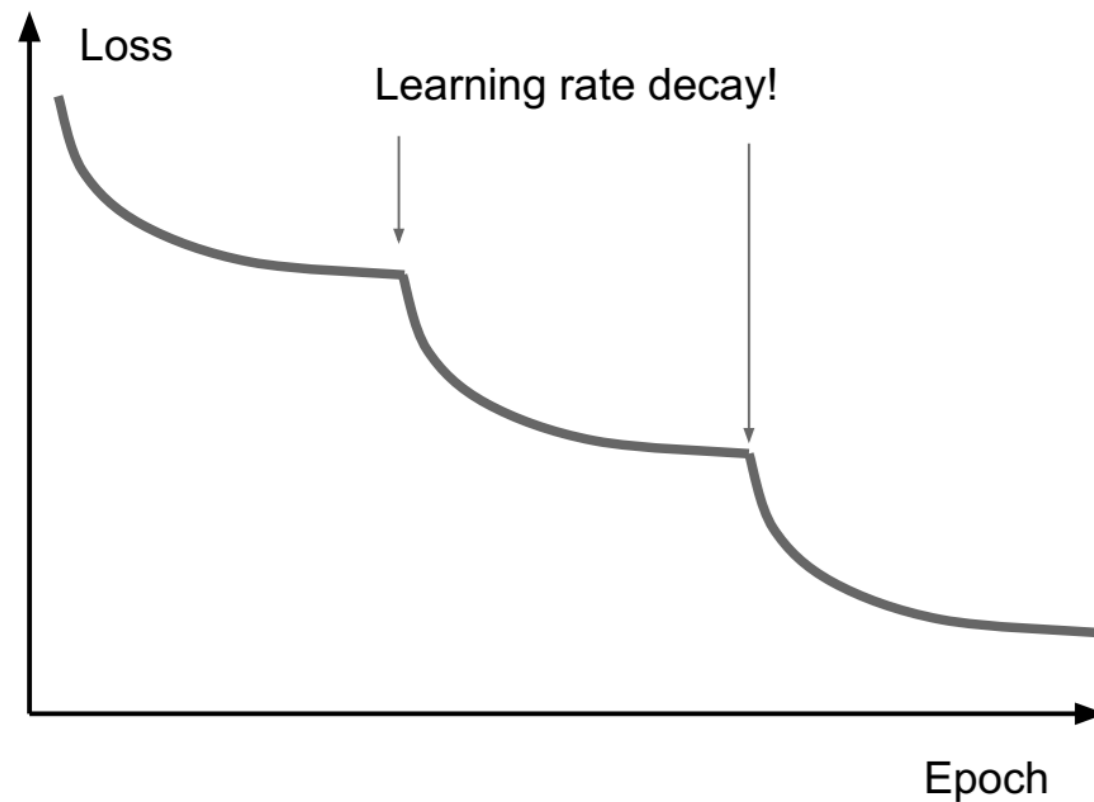
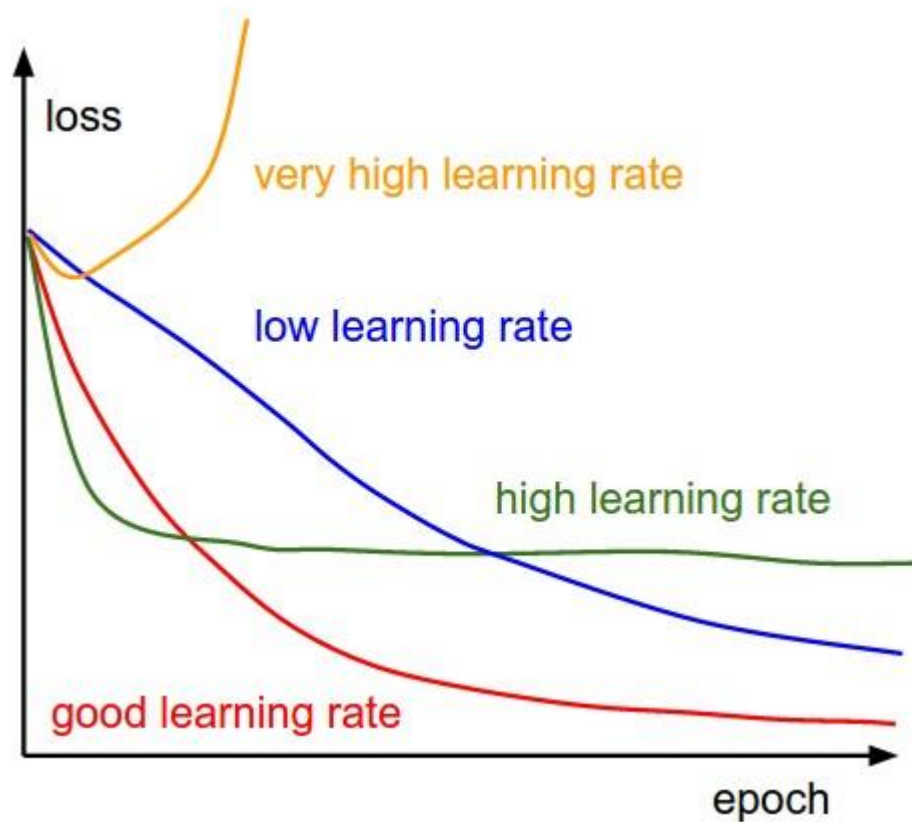
- 1/t decay

$$\alpha = \alpha_0 / (1 + kt)$$



کاهش نرخ آموزش

- برای SGD+Momentum حیاتی تر است، برای Adam کمتر رایج است



Keras Callbacks

Using custom callbacks

Creating new callbacks is a simple and powerful way to customize a training loop. Learn more about creating new callbacks in the guide [Writing your own Callbacks](#), and refer to the documentation for the `base callback` class.

Available callbacks

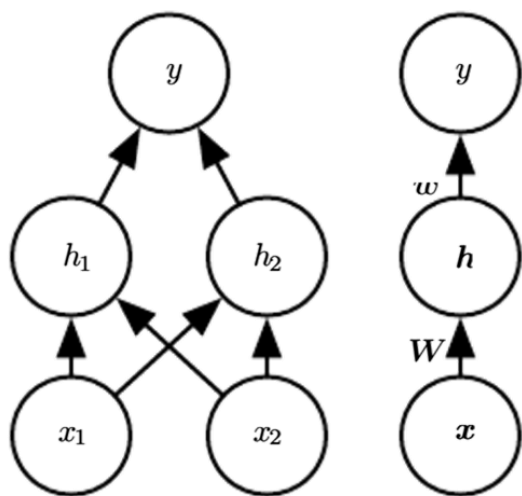
- `Base Callback` class
- `ModelCheckpoint`
- `TensorBoard`
- `EarlyStopping`
- `LearningRateScheduler`
- `ReduceLROnPlateau`
- `RemoteMonitor`
- `LambdaCallback`
- `TerminateOnNaN`
- `CSVLogger`
- `ProgbarLogger`

محاسبه گرادیان

$$h_i = g(\mathbf{x}^T \mathbf{W}_{:,i} + c_i) = f^{(1)}(\mathbf{x})$$

$$y = f^{(2)}(f^{(1)}(\mathbf{x})) = f(\mathbf{x}; \mathbf{W}, \mathbf{c}, \mathbf{w}, b) = \mathbf{w}^T \max\{0, \mathbf{W}^T \mathbf{x} + \mathbf{c}\} + b$$

$$J(\boldsymbol{\theta}) = \frac{1}{N} \sum (f^*(\mathbf{x}) - f(\mathbf{x}; \mathbf{W}, \mathbf{c}, \mathbf{w}, b))^2$$



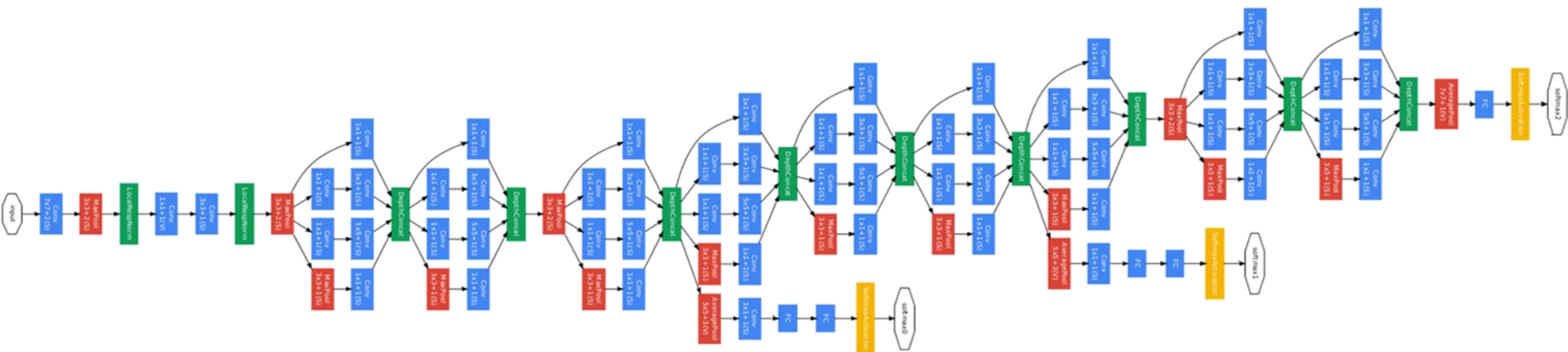
- اگر بتوانیم $\frac{\partial J}{\partial \mathbf{W}}$ ، $\frac{\partial J}{\partial \mathbf{c}}$ ، $\frac{\partial J}{\partial \mathbf{w}}$ و $\frac{\partial J}{\partial b}$ را محاسبه کنیم، می‌توانیم پارامترهای مدل را آموزش بدهیم

- اگر بخواهیم تابع ضرر را تغییر بدهیم

- نیاز است تا معادلات دوباره از ابتدا استخراج شوند

محاسبه گرادیان

- بسیار خسته کننده است و نیاز به محاسبات ماتریسی فوق العاده زیادی دارد



پس انتشار (Backpropagation)

$$f(x, y, z) = (x + y)z$$

e.g., $x = -2$, $y = 5$, $z = -4$

$$q = x + y \Rightarrow f = qz$$

we want $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ and $\frac{\partial f}{\partial z}$

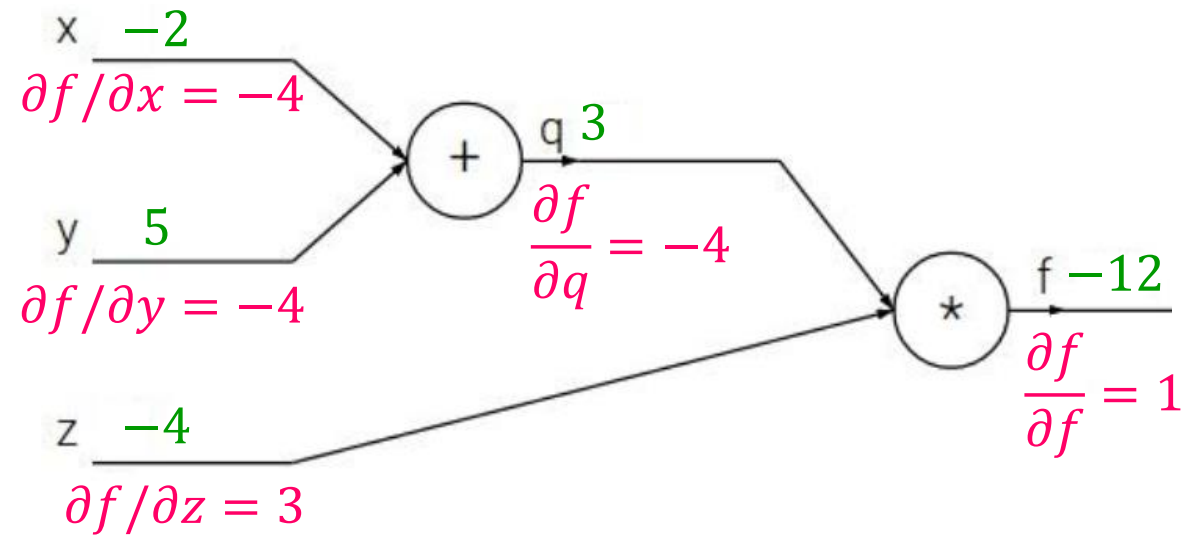
$$\frac{\partial f}{\partial q} = z \quad \frac{\partial f}{\partial z} = q$$

Chain rule

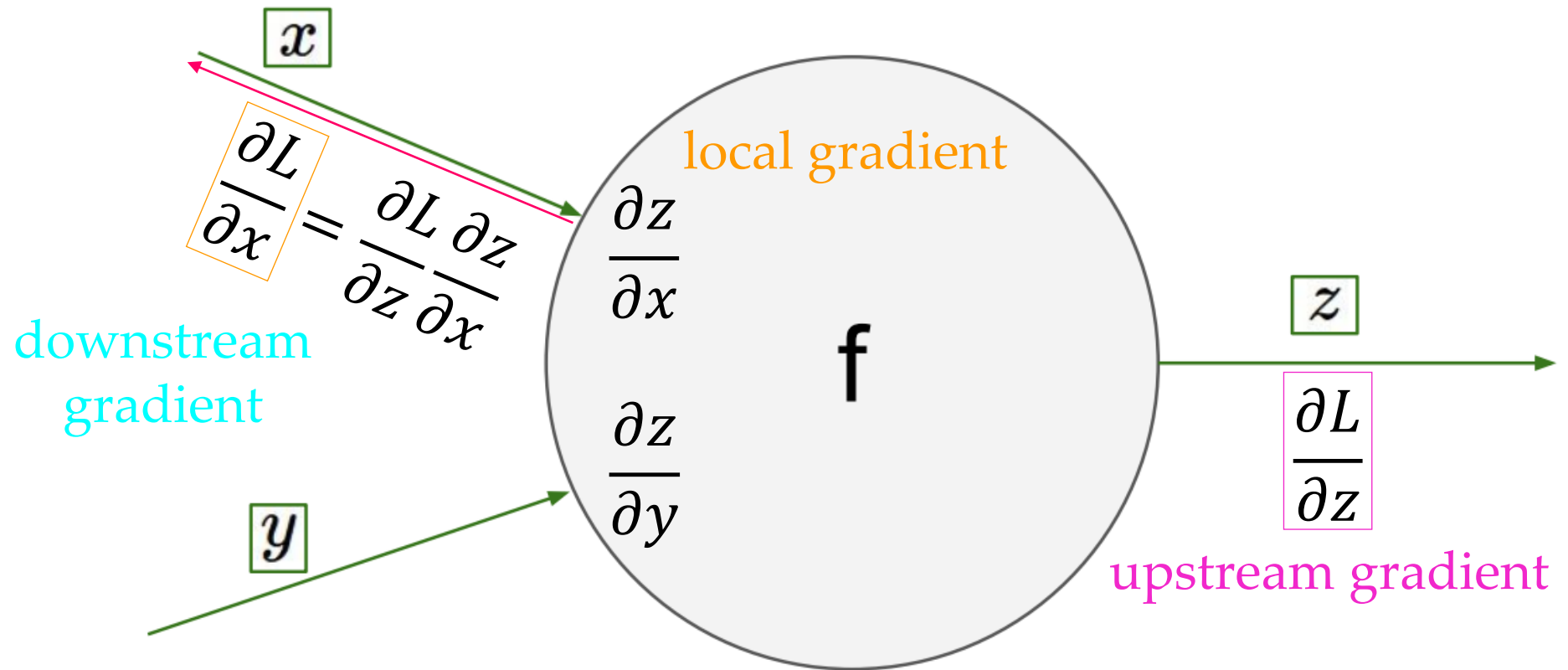
$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} \frac{\partial q}{\partial x}$$

$$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial q} \frac{\partial q}{\partial y}$$

$$\frac{\partial q}{\partial x} = 1 \quad \frac{\partial q}{\partial y} = 1$$

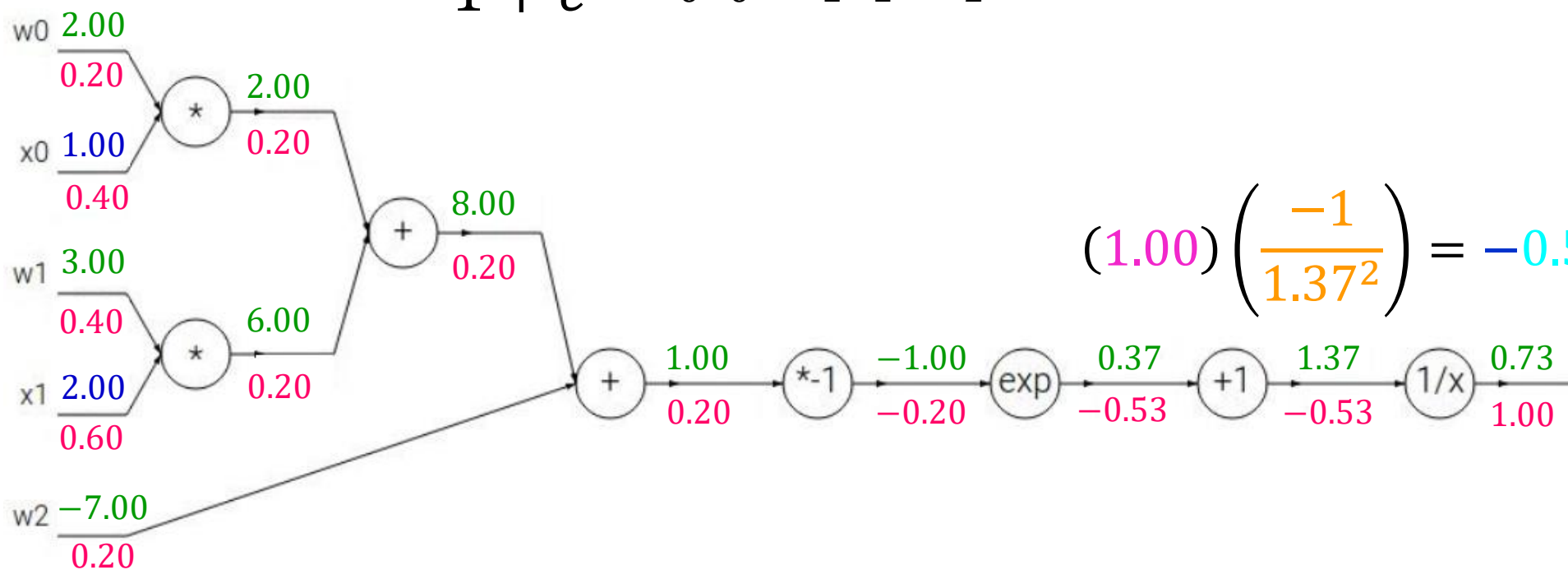


پس انتشار (Backpropagation)



مثال: Linear + Sigmoid

$$f(\mathbf{w}, \mathbf{x}) = \frac{1}{1 + e^{-(w_0x_0 + w_1x_1 + w_2)}}$$



$$\frac{\partial(ax)}{\partial x} = a$$

$$\frac{\partial(c + x)}{\partial x} = 1$$

$$\frac{\partial(e^x)}{\partial x} = e^x$$

$$\frac{\partial(1/x)}{\partial x} = \frac{-1}{x^2}$$