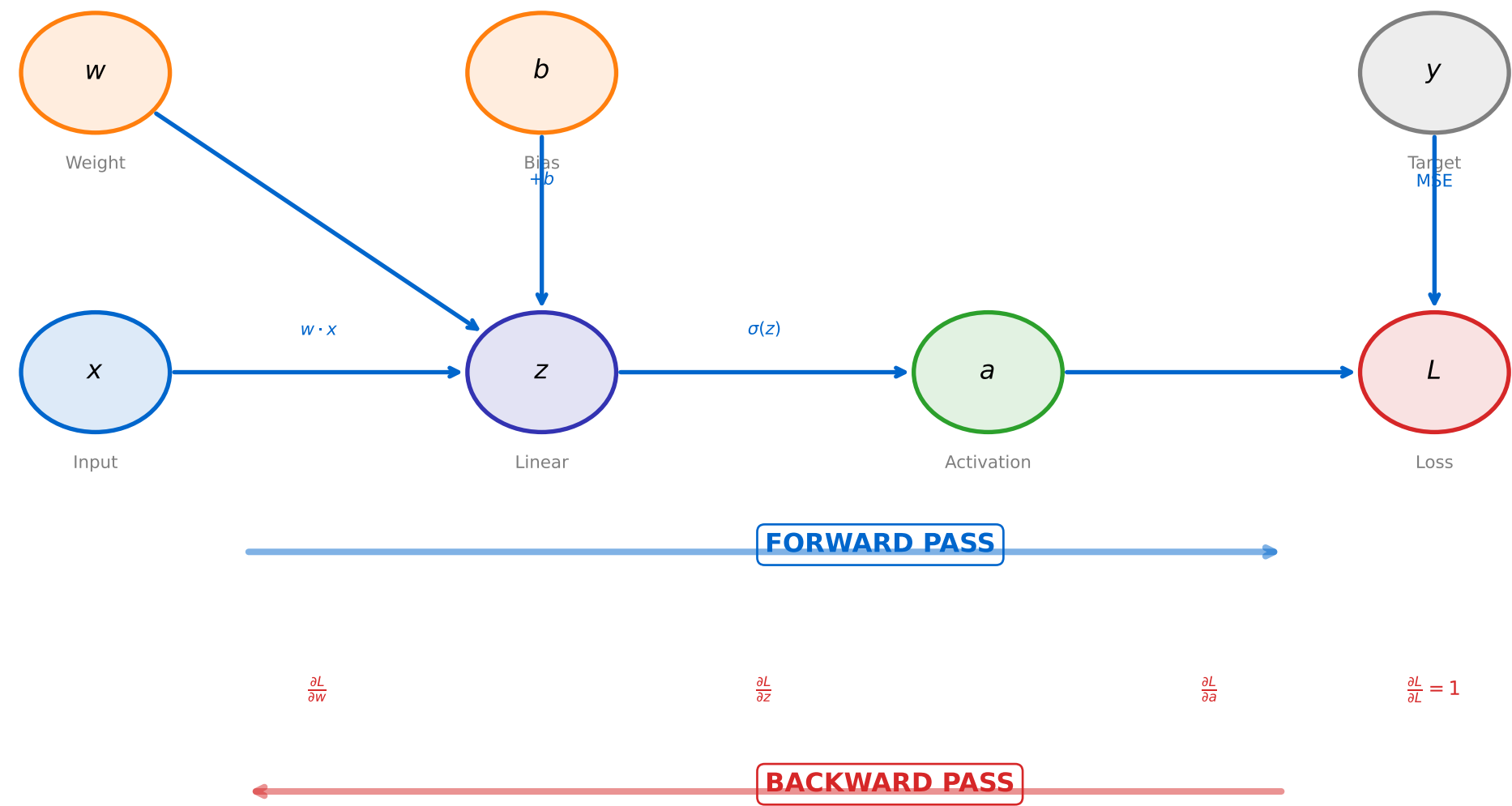


Computational Graph: Forward and Backward Pass



Chain Rule:
$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} \cdot \frac{\partial z}{\partial w}$$

Gradient computation:

- Start from loss: $\frac{\partial L}{\partial L} = 1$
- Propagate backwards through each operation
- Accumulate gradients using chain rule