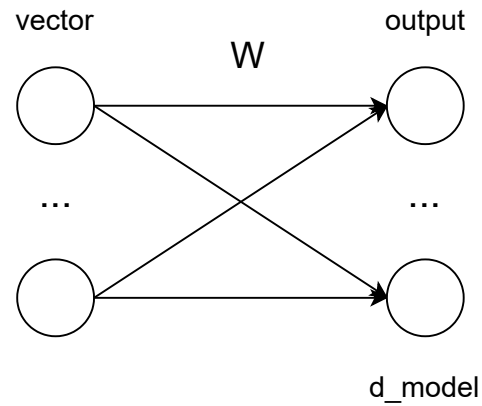
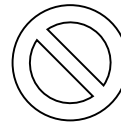


# Encoder Layer

## Multihead Attention Layer



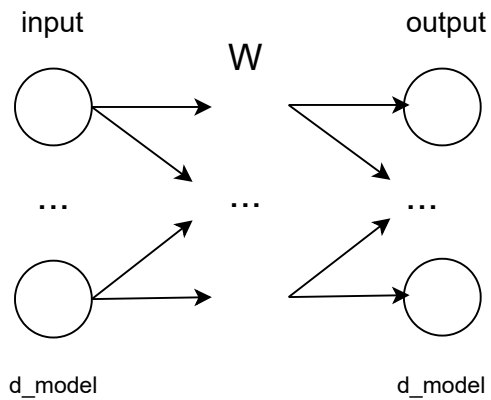
dropout  
0.1



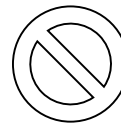
Norm  
Layer

$$x'_i = \frac{x_i - \bar{X}}{S}$$

## Feed Forward Network



dropout  
0.1



Norm  
Layer

$$x'_i = \frac{x_i - \bar{X}}{S}$$

output