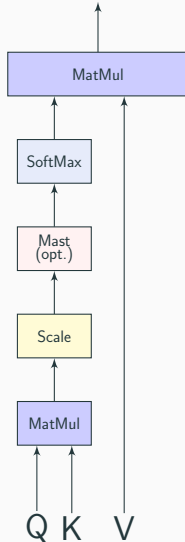
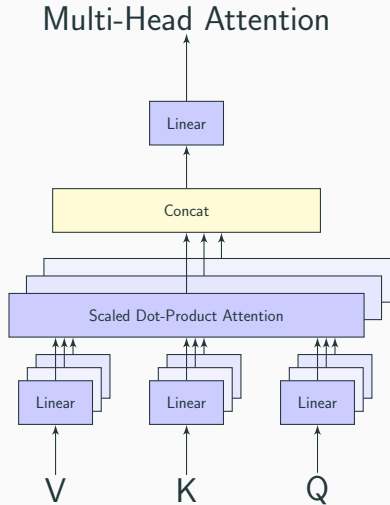
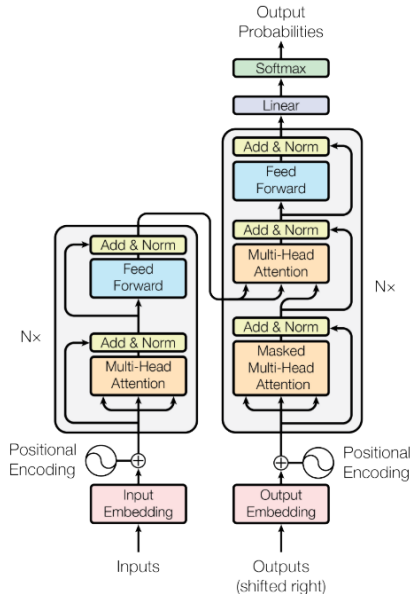


Scaled Dot Product Attention





The Transformer - model architecture





$$PE_{(pos, 2i)} = \sin(pos/1000^{2i/d_{model}})$$

$$PE_{(pos, 2i+1)} = \cos(pos/1000^{2i/d_{model}})$$