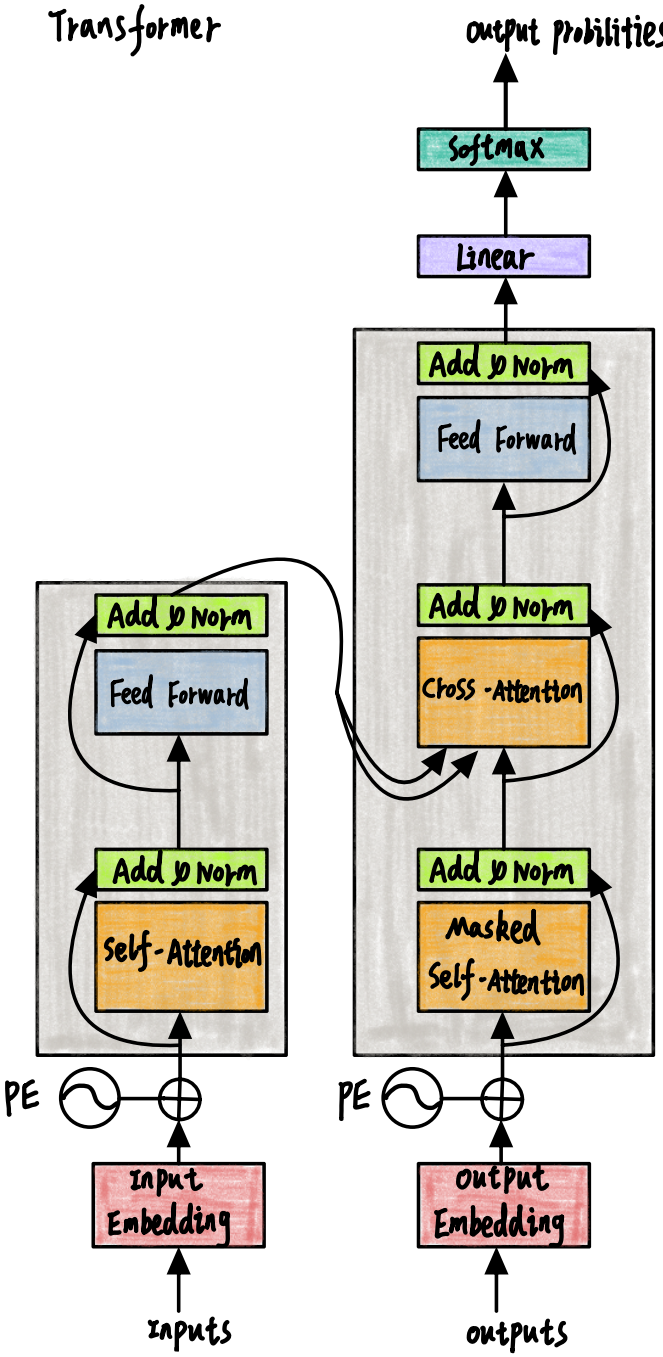
Transformer



$$(\tilde{x}^1, \tilde{x}^2 \dots \tilde{x}^t \dots \tilde{x}^n) = E_1(o^1\, o^2 \dots o^t \dots o^n)$$

$$\underset{(m,n)}{\tilde{X}} = \underset{(m,T)}{E_1}\ \underset{(T,n)}{O}$$

$$(x^1, x^2 \dots x^t \dots x^n) = (\tilde{x}^1, \tilde{x}^2 \dots \tilde{x}^t \dots \tilde{x}^n) + (p^1, p^2 \dots p^t \dots p^n)$$

$$\underset{(m,n)}{X} = \underset{(m,n)}{\tilde{X}} + \underset{(m,n)}{P}$$

$$\underset{(m,n)}{Q_1} = \underset{(m,m)}{W_{q1}}\ \underset{(m,n)}{X}$$

$$\underset{(m,n)}{K_1} = \underset{(m,m)}{W_{k1}}\ \underset{(m,n)}{X}$$

$$\underset{(m,n)}{V_1} = \underset{(m,m)}{W_{v1}}\ \underset{(m,n)}{X}$$

$$\underset{(n,n)}{A_1} = \underset{(n,m)}{K_1^T} \cdot \underset{(m,n)}{Q_1}$$

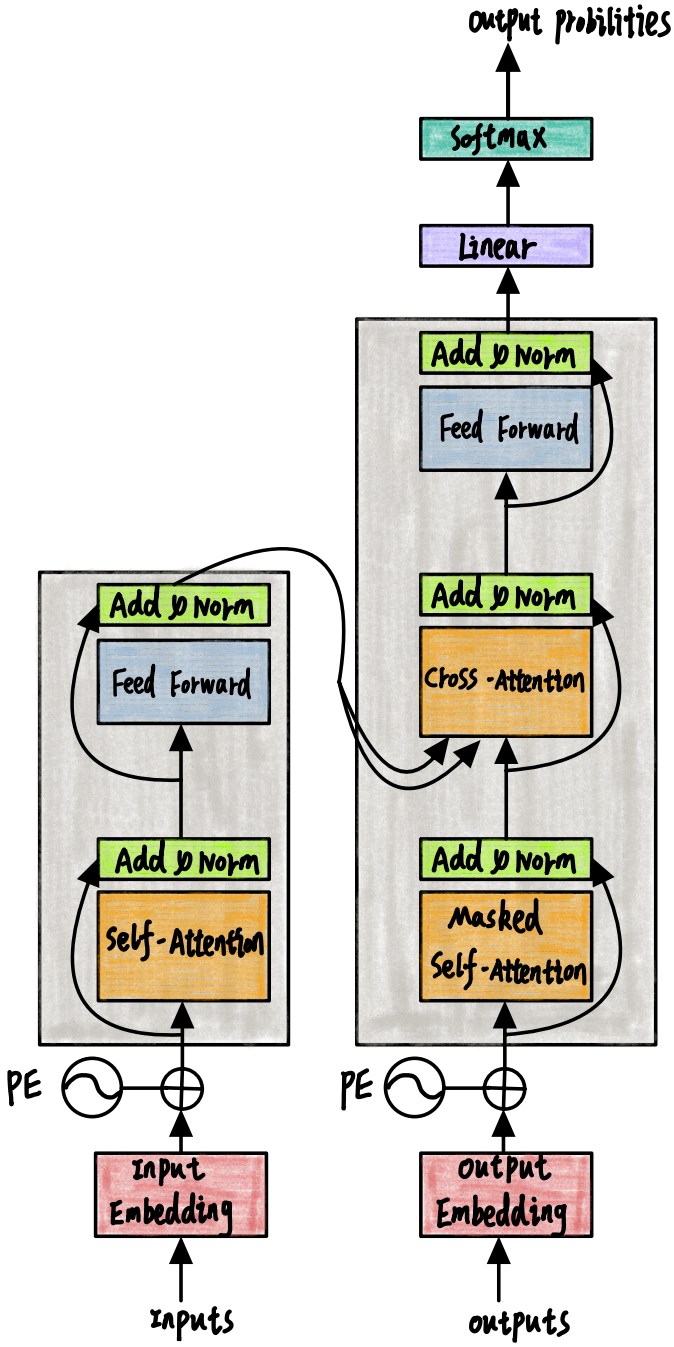
$$\underset{(n,n)}{A_1'} = \text{Softmax}\underset{(n,n)}{(A_1)}$$

$$\underset{(m,n)}{Y_1} = \underset{(m,n)}{V_1}\ \underset{(n,n)}{A_1'}$$

$$\underset{(m,n)}{Y_1'} = \text{Norm}\,(\underset{(m,n)}{Y_1} + \underset{(m,n)}{X})$$

$$\underset{(m,n)}{Y_1''} = \underset{(m,m)}{W_E}\ \underset{(m,n)}{Y_1'}$$

$$\underset{(m,n)}{S} = \text{Norm}\,(\underset{(m,n)}{Y_1''} + \underset{(m,n)}{Y_1'})$$



$$(\tilde{z}^1, \tilde{z}^2 \cdots \tilde{z}^t \cdots \tilde{z}^n) = E_2(u^1, u^2 \cdots u^t \cdots u^n)$$

$$\underset{(m,n)}{\tilde{Z}} = \underset{(m,T)}{E_2}\,\underset{(T,n)}{U}$$

$$(z^1, z^2 \cdots z^t \cdots z^n) = (\tilde{z}^1, \tilde{z}^2 \cdots \tilde{z}^t \cdots \tilde{z}^n) + (p^1, p^2 \cdots p^t \cdots p^n)$$

$$\underset{(m,n)}{Z} = \underset{(m,n)}{\tilde{Z}} + \underset{(m,n)}{P}$$

$$\underset{(m,n)}{Q_2} = \underset{(m,m)}{W_{q2}}\,\underset{(m,n)}{Z}$$

$$\underset{(m,n)}{K_2} = \underset{(m,m)}{W_{k2}}\,\underset{(m,n)}{Z}$$

$$\underset{(m,n)}{V_2} = \underset{(m,m)}{W_{v2}}\,\underset{(m,n)}{Z}$$

$$\underset{(n,n)}{A_2} = \underset{(n,m)}{K_2^T} \cdot \underset{(m,n)}{Q_2}$$

A_2 某些 score 置为 $-\infty$

$$\underset{(n,n)}{A_2'} = \mathrm{Softmax}(\underset{(n,n)}{A_2})$$

$$\underset{(m,n)}{Y_2} = \underset{(m,n)}{V_2}\,\underset{(n,n)}{A_2'}$$ (某些置为 0)

$$\underset{(m,n)}{R} = \mathrm{Norm}(\underset{(m,n)}{Y_2} + \underset{(m,n)}{Z})$$

$$\underset{(m,n)}{Q_3} = \underset{(m,m)}{W_{q3}}\,\underset{(m,n)}{R}$$

$$\underset{(m,n)}{K_3} = \underset{(m,m)}{W_{k3}}\,\underset{(m,n)}{S}$$

$$\underset{(m,n)}{V_3} = \underset{(m,m)}{W_{v3}}\,\underset{(m,n)}{S}$$

$$\underset{(n,n)}{A_3} = \underset{(n,m)}{K_3^T} \cdot \underset{(m,n)}{Q_3}$$

$$\underset{(n,n)}{A_3'} = \mathrm{Softmax}(\underset{(n,n)}{A_3})$$

$$\underset{(m,n)}{Y_3} = \underset{(m,n)}{V_3}\,\underset{(n,n)}{A_3'}$$

$$Y_3' = \underset{(m,n)}{\text{Norm}}(Y_3 + R)$$
(m,n) (m,n) (m,n)

$$Y_3'' = W_D Y_3'$$
(m,n) (m,m) (m,n)

$$Y_3''' = \text{Norm}(Y_3'' + Y_3')$$
(m,n) (m,n) (m,n)

$$L = W_L Y_3''' + b$$
(T,n) (T,m) (m,n) $(T,1)$

$$U' = \text{Softmax}(L)$$
(T,n) (T,n)

怎么自回归？

解码器的输入和输出维度

掩码

什么是 multi-head？