

LayerNorm

2021年9月23日 星期四 上午11:58

$$\begin{bmatrix} a_{我我}V_我 + a_{我宣}V_宣 + a_{我你}V_你 + 0 \\ a_{宣我}V_我 + a_{宣宣}V_宣 + a_{宣你}V_你 + 0 \\ a_{你我}V_我 + a_{你宣}V_宣 + a_{你你}V_你 + 0 \\ a_{0我}V_我 + a_{0宣}V_宣 + a_{0你}V_你 + 0 \end{bmatrix}$$

shape: (seq_len, d_model)

线性变换、Add

$$\begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1, d_{model}} \\ z_{21} & z_{22} & \cdots & z_{2, d_{model}} \\ z_{31} & z_{32} & \cdots & z_{3, d_{model}} \\ z_{41} & z_{42} & \cdots & z_{4, d_{model}} \end{bmatrix}$$

shape: (seq_len, d_model)

BN

LN

$$\begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1, d_{model}} \\ z_{21} & z_{22} & \cdots & z_{2, d_{model}} \\ z_{31} & z_{32} & \cdots & z_{3, d_{model}} \\ z_{41} & z_{42} & \cdots & z_{4, d_{model}} \end{bmatrix}$$

$$\begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1, d_{model}} \\ z_{21} & z_{22} & \cdots & z_{2, d_{model}} \\ z_{31} & z_{32} & \cdots & z_{3, d_{model}} \\ z_{41} & z_{42} & \cdots & z_{4, d_{model}} \end{bmatrix}$$

$BN(z_{i1}) = \frac{z_{i1} - \mu_{i1}}{\sigma_{i1}}$

$$\begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1, d_{model}} \\ z_{21} & z_{22} & \cdots & z_{2, d_{model}} \\ z_{31} & z_{32} & \cdots & z_{3, d_{model}} \\ z_{41} & z_{42} & \cdots & z_{4, d_{model}} \end{bmatrix}$$

$$\begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1, d_{model}} \\ z_{21} & z_{22} & \cdots & z_{2, d_{model}} \\ z_{31} & z_{32} & \cdots & z_{3, d_{model}} \\ z_{41} & z_{42} & \cdots & z_{4, d_{model}} \end{bmatrix}$$

$LN(z_{ij}) = \frac{z_{ij} - \mu_{ij}}{\sigma_{ij}}$

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \rightarrow \begin{bmatrix} \frac{1-2}{1} & \frac{2-3}{1} \\ \frac{3-2}{1} & \frac{4-3}{1} \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}$$

 $\mu=2, \sigma^2=1$ $\mu=3, \sigma^2=1$ Batch Norm

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{matrix} \mu=1.5, \sigma^2=0.25 \\ \mu=3.5, \sigma^2=0.25 \end{matrix} \rightarrow \begin{bmatrix} \frac{1-1.5}{\sqrt{0.25}} & \frac{2-1.5}{\sqrt{0.25}} \\ \frac{3-3.5}{\sqrt{0.25}} & \frac{4-3.5}{\sqrt{0.25}} \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}$$

Layer Norm

$$\begin{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \\ \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \end{bmatrix}$$

你 好 谢 谢

$\mu=1.75, \sigma^2=0.6875, \sigma=0.8292$