

一、Transformer 基本结构介绍

https://blog.csdn.net/longxinchen_ml/article/details/86533005

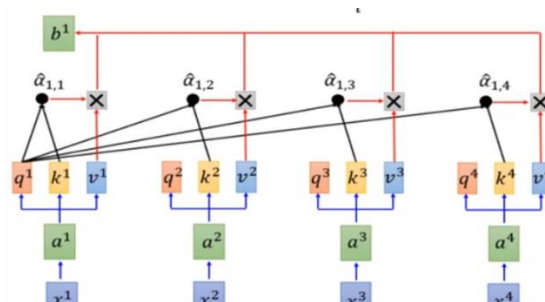
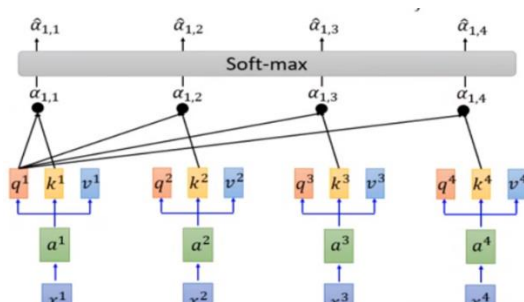
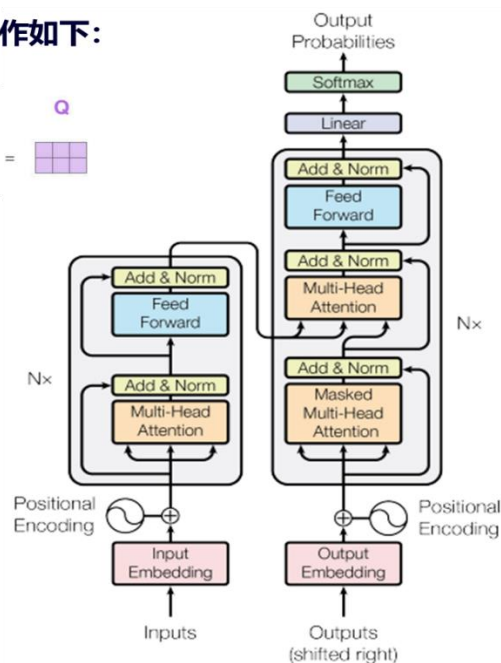
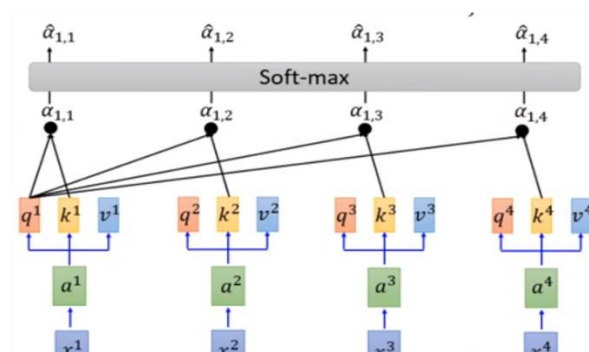
该方法最重要的部分是Self-attention，具体操作如下：

Query: to match others

Key: to be matched

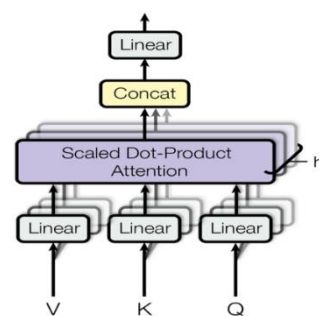
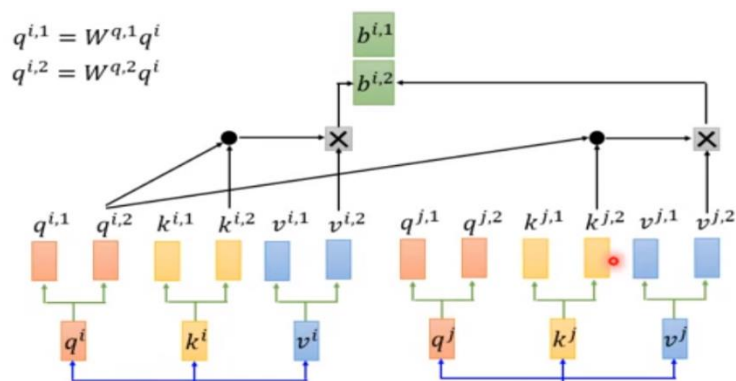
Value: to be extracted

$$X \times W^Q = Q$$



$$B = \text{softmax}(QK^T / \sqrt{d_k}) \cdot V$$

$$\text{softmax}\left(\frac{Q \times K^T}{\sqrt{d_k}}\right) \cdot V = Z$$



二、Transformer 代码详解

<http://fancyerii.github.io/2019/03/09/transformer-codes/>

三、Vision Transformer 介绍

https://blog.csdn.net/qq_37541097/article/details/118242600

