

基于ARIMA的时序信号分析

应用—金融市场分析

Speaker: 杨阔

Email: yangkuohao@shu.edu.cn

ARIMA

差分自回归移动平均模型



平稳性

平稳性：



平稳性就是要求经由样本时间序列所得到的拟合曲线在未来的一段时间内仍能顺着现有的形态“惯性”地延续下去



平稳性要求序列的均值和方差不发生明显变化

严平稳与弱平稳：



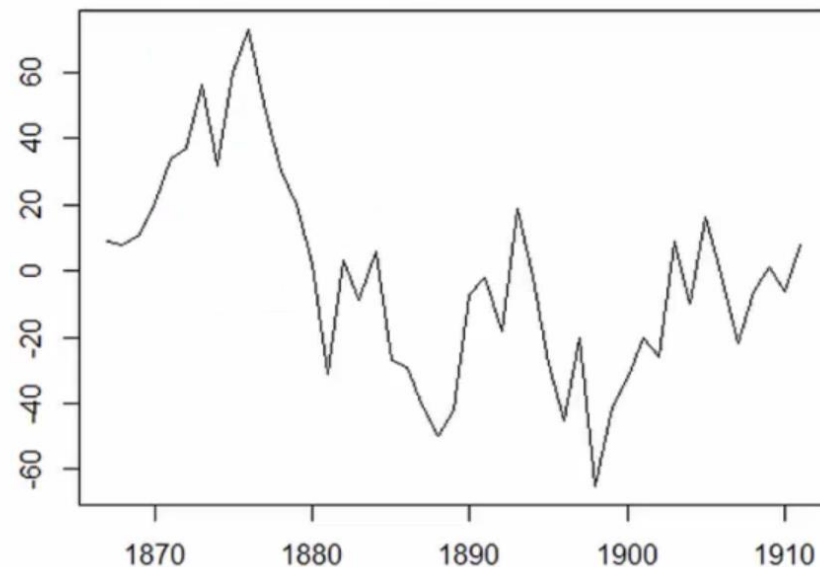
严平稳：严平稳表示的分布不随时间的改变而改变。

如：白噪声（正态），无论怎么取，都是期望为0，方差为1



弱平稳：期望与相关系数（依赖性）不变如

未来某时刻的 t 的值 X_t 就要依赖于它的过去信息，所以需要依赖性

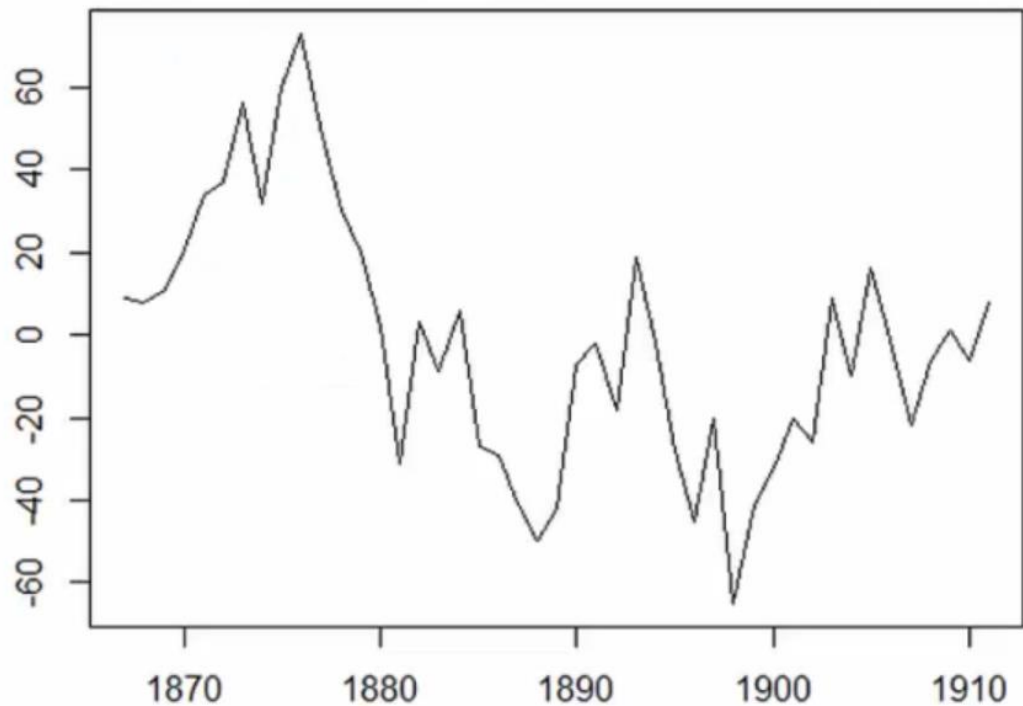




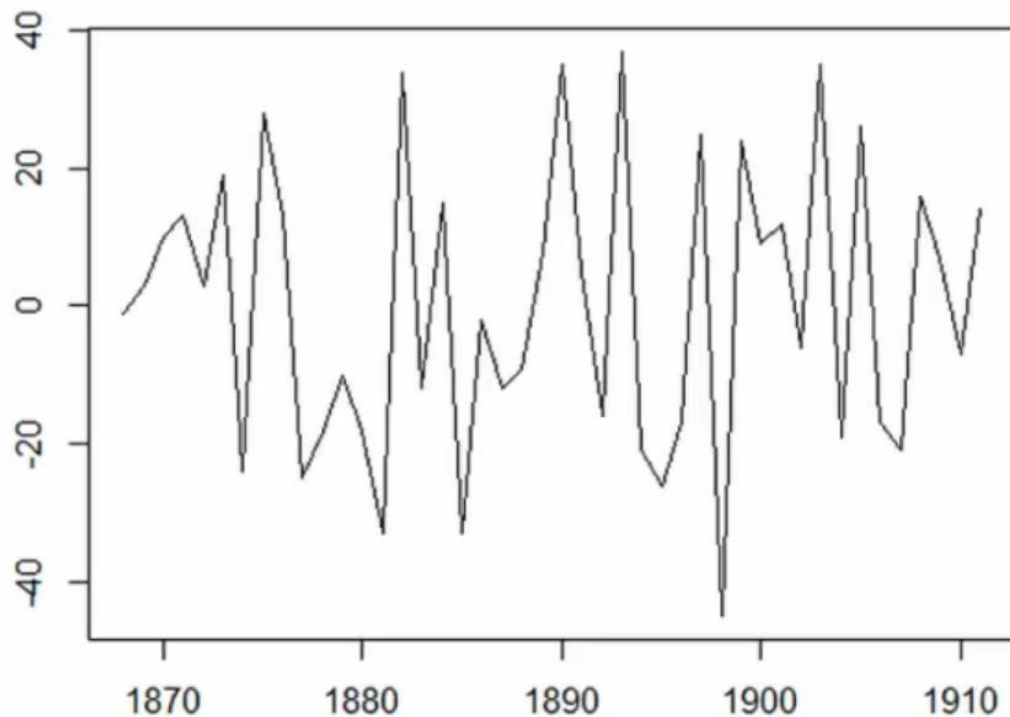
差分法



差分法: 时间序列在 t 与 $t-1$ 时刻的差值



原数据



一阶差分



AR



自回归模型(AR)

- ✎ 描述当前值与历史值之间的关系，用变量自身的历史时间数据对自身进行预测
- ✎ 自回归模型必须满足平稳性的要求
- ✎ p阶自回归过程的公式定义： $y_t = \mu + \sum_{i=1}^p \gamma_i y_{t-i} + \epsilon_t$
- ✎ y_t 是当前值 μ 是常数项 P 是阶数 γ_i 是自相关系数 ϵ_t 是误差



✓ 自回归模型的限制

✎ 自回归模型是用自身的数据来进行预测

✎ 必须具有平稳性

✎ 必须具有自相关性，如果自相关系数(φ_i)小于0.5，则不宜采用

✎ 自回归只适用于预测与自身前期相关的现象



✓ 移动平均模型 (MA)

✎ 移动平均模型关注的是自回归模型中的误差项的累加

✎ q阶自回归过程的公式定义： $y_t = \mu + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i}$

✎ 移动平均法能有效地消除预测中的随机波动

✓ 自回归移动平均模型 (ARMA)

✎ 自回归与移动平均的结合

✎ 公式定义： $y_t = \mu + \sum_{i=1}^p \gamma_i y_{t-i} + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i}$



ARIMA

- ✓ ARIMA(p, d, q)模型全称为差分自回归移动平均模型 (Autoregressive Integrated Moving Average Model, 简记ARIMA)
- ✎ AR是自回归, p 为自回归项; MA为移动平均
 q 为移动平均项数, d 为时间序列成为平稳时所做的差分次数
- ✎ 原理: 将非平稳时间序列转化为平稳时间序列然后将因变量仅对它的滞后值以及随机误差项的现值和滞后值进行回归所建立的模型



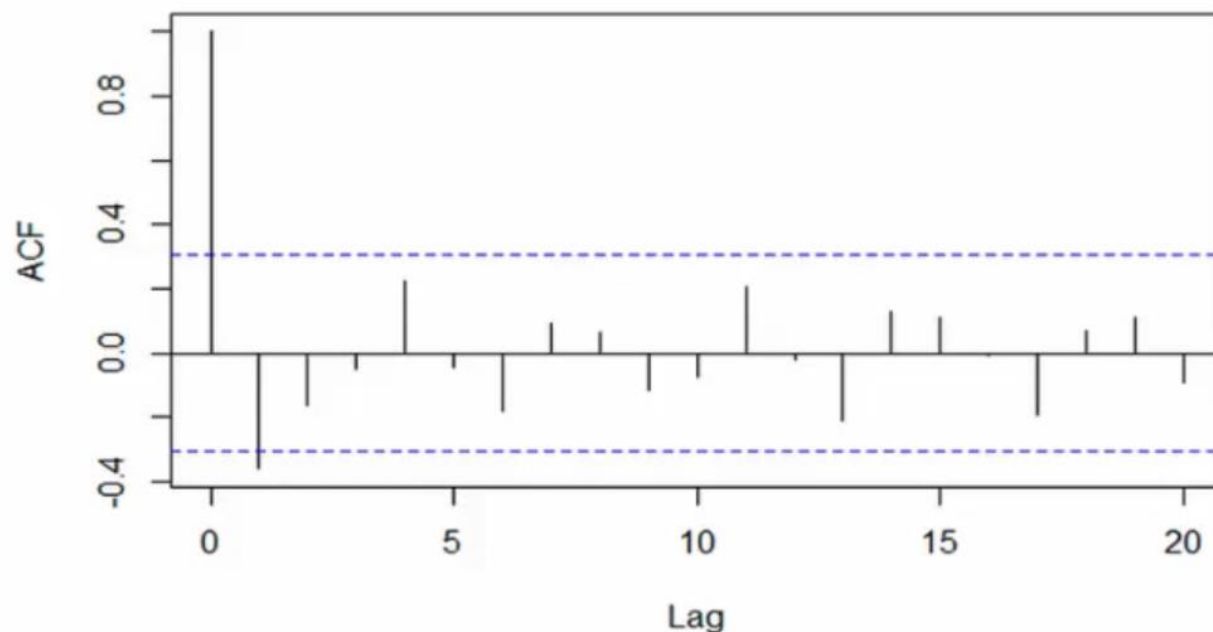
ARIMA

✓ 自相关函数ACF(autocorrelation function)

✎ 有序的随机变量序列与其自身相比较
自相关函数反映了同一序列在不同时序的取值之间的相关性

✎ 公式： $ACF(k) = \rho_k = \frac{\text{Cov}(y_t, y_{t-k})}{\text{Var}(y_t)}$

✎ ρ_k 的取值范围为 $[-1, 1]$





✓ 偏自相关函数(PACF)(partial autocorrelation function)

✎ 对于一个平稳AR(p)模型，求出滞后k自相关系数 $p(k)$ 时实际上得到并不是 $x(t)$ 与 $x(t-k)$ 之间单纯的相关关系

✎ $x(t)$ 同时还会受到中间 $k-1$ 个随机变量 $x(t-1)$ 、 $x(t-2)$ 、.....、 $x(t-k+1)$ 的影响而这 $k-1$ 个随机变量又都和 $x(t-k)$ 具有相关关系所以自相关系数 $p(k)$ 里实际掺杂了其他变量对 $x(t)$ 与 $x(t-k)$ 的影响

✎ 剔除了中间 $k-1$ 个随机变量 $x(t-1)$ 、 $x(t-2)$ 、.....、 $x(t-k+1)$ 的干扰之后 $x(t-k)$ 对 $x(t)$ 影响的相关程度。

✎ ACF还包含了其他变量的影响而偏自相关系数PACF是严格这两个变量之间的相关性



模型概述

1 ARIMA模型

ARIMA(p,d,q)模型是差分运算和ARMA模型的组合。ARMA模型是自回归移动平均模型，可以细分为AR模型、MA模型、ARMA模型。

2 d阶差分运算

相距一期的两个序列值之间的减法运算称为一阶差分运算，所得结果再次差分运算，如此d次，则是d阶差分运算。

3 差分平稳序列

差分运算具有强大的确定性信息提取能力，许多非平稳序列差分后会显示出平稳序列的性质，这时称这个非平稳序列为差分平稳序列。

4 AR模型

随机变量 X_t 的取值 x_t 是前p期的多元线性回归，认为 X_t 主要受过去p期的序列值的影响。误差项是当期的随机干扰 et ，为零均值白噪声序列。

5 MA模型

随机变量 X_t 的取值 x_t 是前q期随机扰动的多元线性回归，认为 X_t 主要受过去q期的误差项的影响。误差项是当期的随机干扰 et ，为零均值白噪声序列， μ 是序列 $\{X_t\}$ 的均值。



建模流程

统计量	平稳AR(p)模型 的性质	平稳MA(q)模型 的性质	平稳ARMA(p,q)模型 的性质
均值	常数	常数	常数
方差	常数	常数	常数
自相关系数ACF	拖尾	'q阶截尾	拖尾
偏自相关系数PACF	'p阶截尾	拖尾	拖尾

◆拖尾

相关系数呈指数的速度衰减，始终有非零取值，几何型或震荡型。

◆截尾

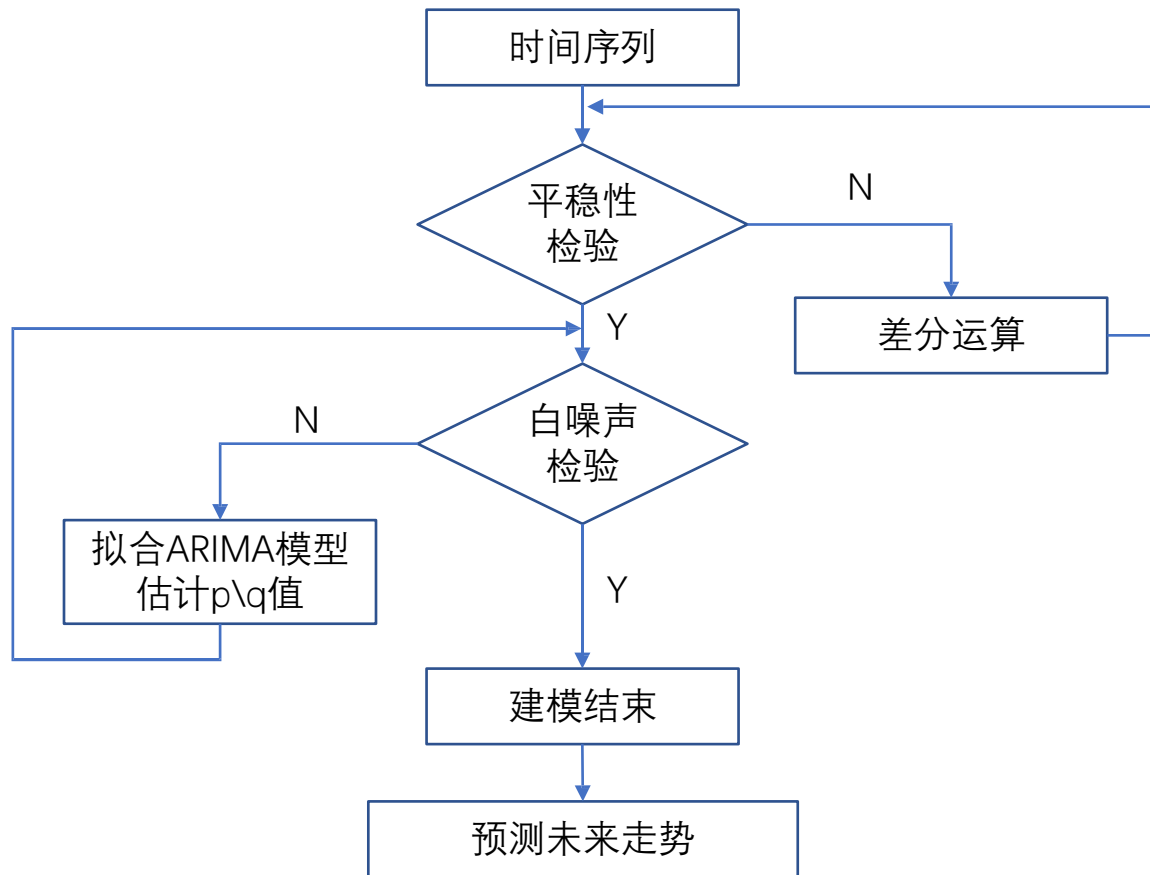
落在置信区间内（95%的点都符合该规则）

◆ACF与PACF的差异

ACF衡量了 X_t 与 X_{t-k} 之间的关系，但相关系数掺杂了其他变量对 X_t 与 X_{t-k} 的影响。为单纯测量 X_t 与 X_{t-k} 的关系，引进了PACF的概念。



建模流程



差分平稳时间序列ARIMA模型建模步骤



模型评估

✓ 模型选择AIC与BIC: 选择更简单的模型

✎ AIC:赤池信息准则 (Akaike Information Criterion , AIC)

$$AIC = 2k - 2\ln(L)$$

✎ BIC:贝叶斯信息准则 (Bayesian Information Criterion , BIC)

$$BIC = k\ln(n) - 2\ln(L)$$

✎ k为模型参数个数，n为样本数量，L为似然函数



知识点回顾



- 建模流程
- 模型简介
- 模型识别
- 检验

Python实战



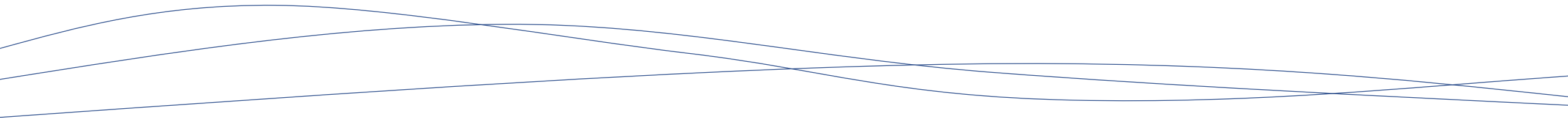
- 导入工具及数据
- 原始序列的检验
- 一阶差分序列的检验
- 定阶（参数调优）
- 建模与预测

参考材料《Python数据分析与挖掘实战》

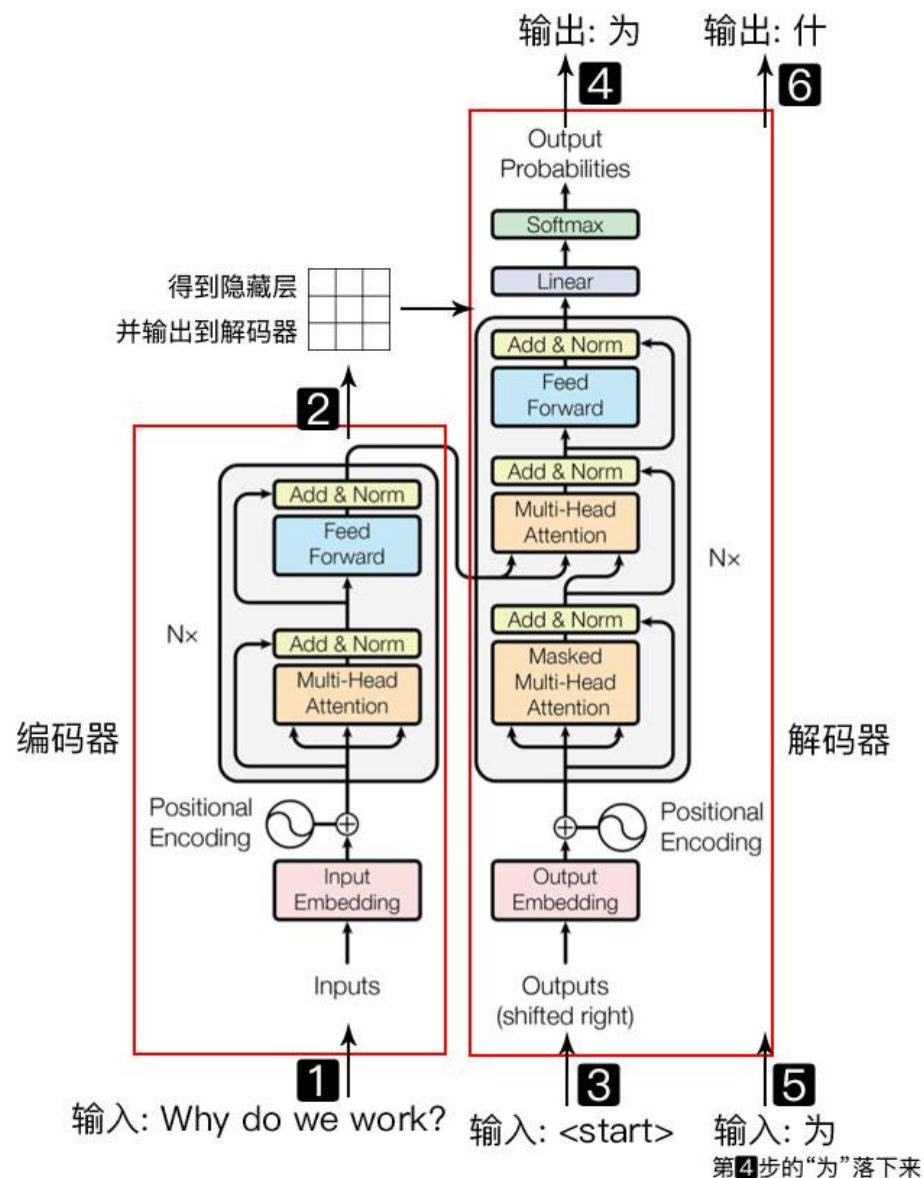
基于注意力机制的时序信号分析深度 学习模型

Transformer

Attention is all you need



Transformer



01

输入部分

- 源文本嵌入层及其位置编码器
- 目标文本嵌入层及其位置编码器

02

输出部分

- 线性层
- softmax层

03

编码器部分

- 由 N 个编码器层堆叠,每个编码器层由两个子层连接
- 一个多头自注意力子层和规范化层以及一个残差连接
- 一个前馈全连接子层和规范化层以及一个残差连接

04

解码器部分

- 一, 一个多头自注意力子层和规范化层以及一个残差连接
- 二, 一个多头注意力子层和规范化层以及一个残差连接
- 三, 个子层连接结构包括一个前馈全连接子层和规范化层以及一个残差连接

研究方法

YOUR ENGLISH TITLE

01

1. Positional Encoding

$$PE(pos, 2i) = \sin(pos/10000^{2i/d_{\text{model}}})$$
$$PE(pos, 2i + 1) = \cos(pos/10000^{2i/d_{\text{model}}})$$

02

2. Self Attention Mechanism

它需要三个指定的输入Q(query), K(key), V(value), 然后通过公式得到注意力的计算结果

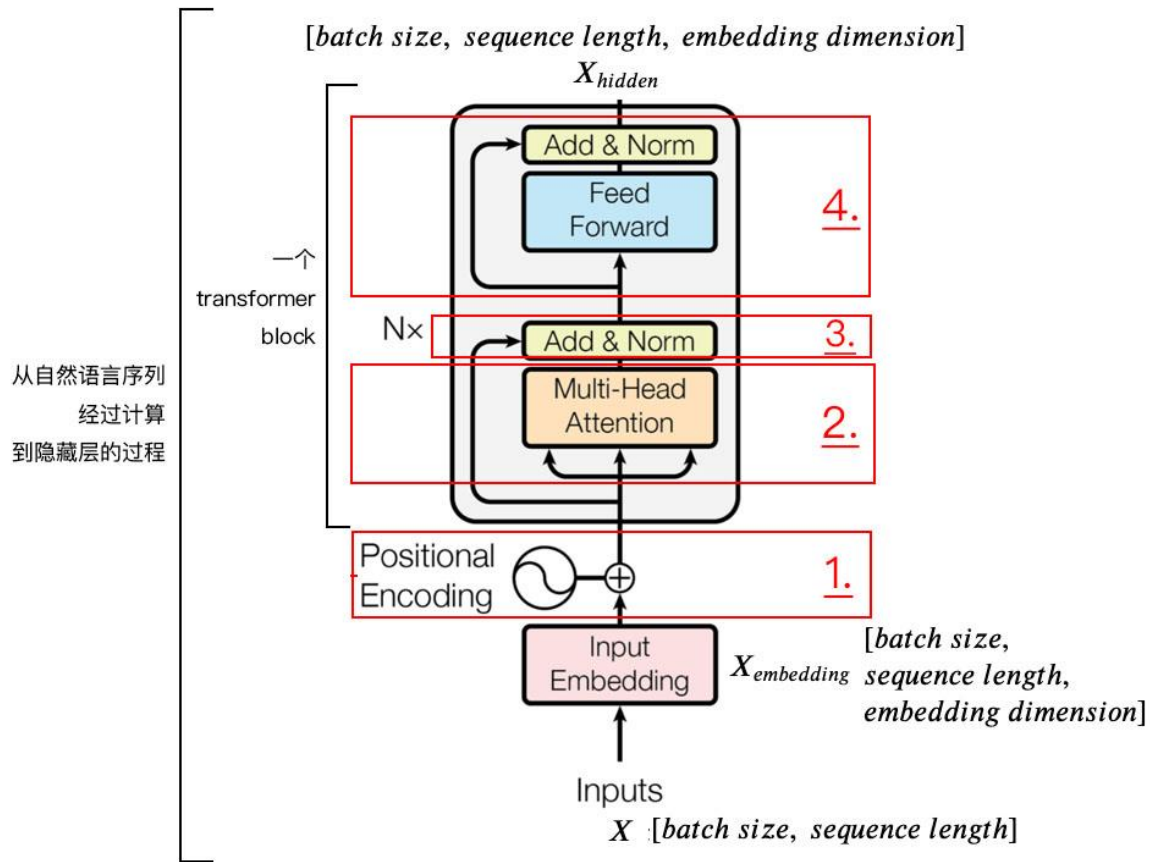
03

残差连接和 Layer Normalization

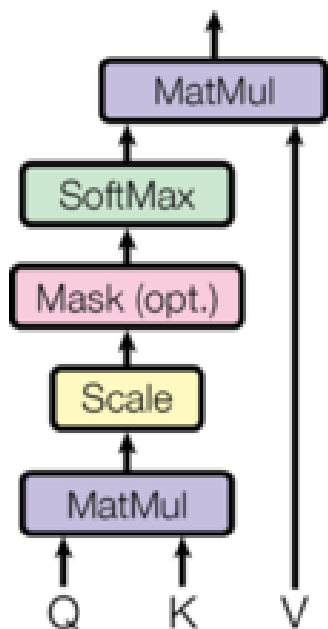
我们在上一步得到了经过 self-attention 加权之后输出, 然后把他们加起来做残差连接, Layer Normalization 的作用是把神经网络中隐藏层归一为标准正态分布

04

前馈全连接层



Attention



$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

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

A green 3x4 matrix X is multiplied by a purple 4x3 matrix W^Q to produce a purple 3x3 matrix Q.

$$X \times W^K = K$$

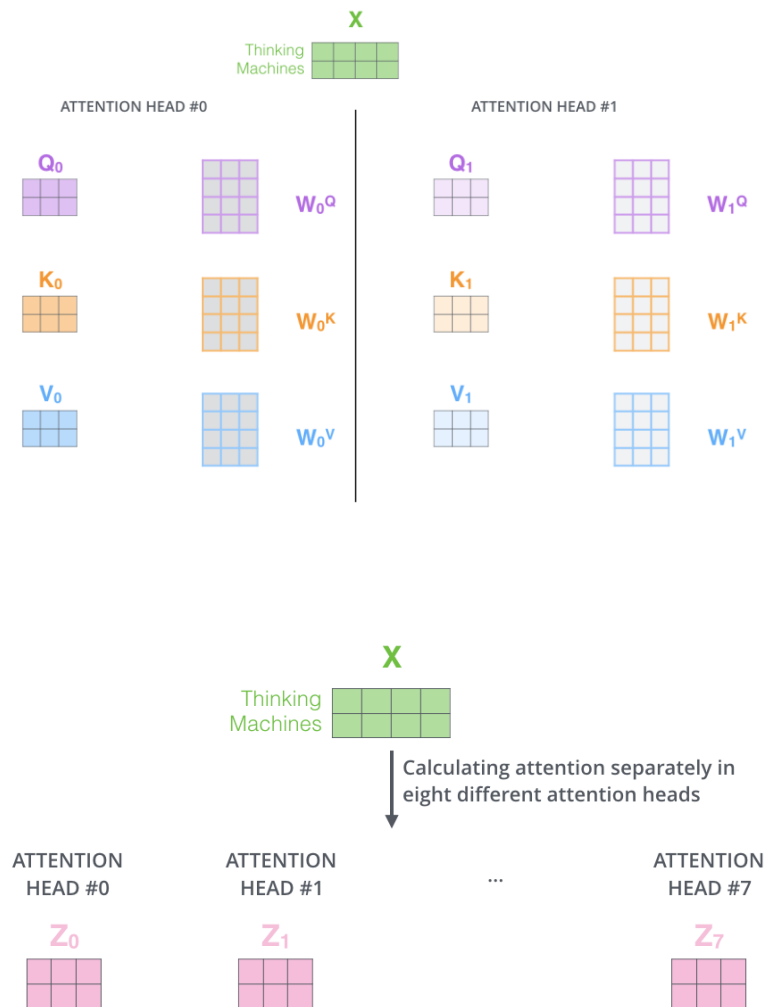
A green 3x4 matrix X is multiplied by an orange 4x3 matrix W^K to produce an orange 3x3 matrix K.

$$X \times W^V = V$$

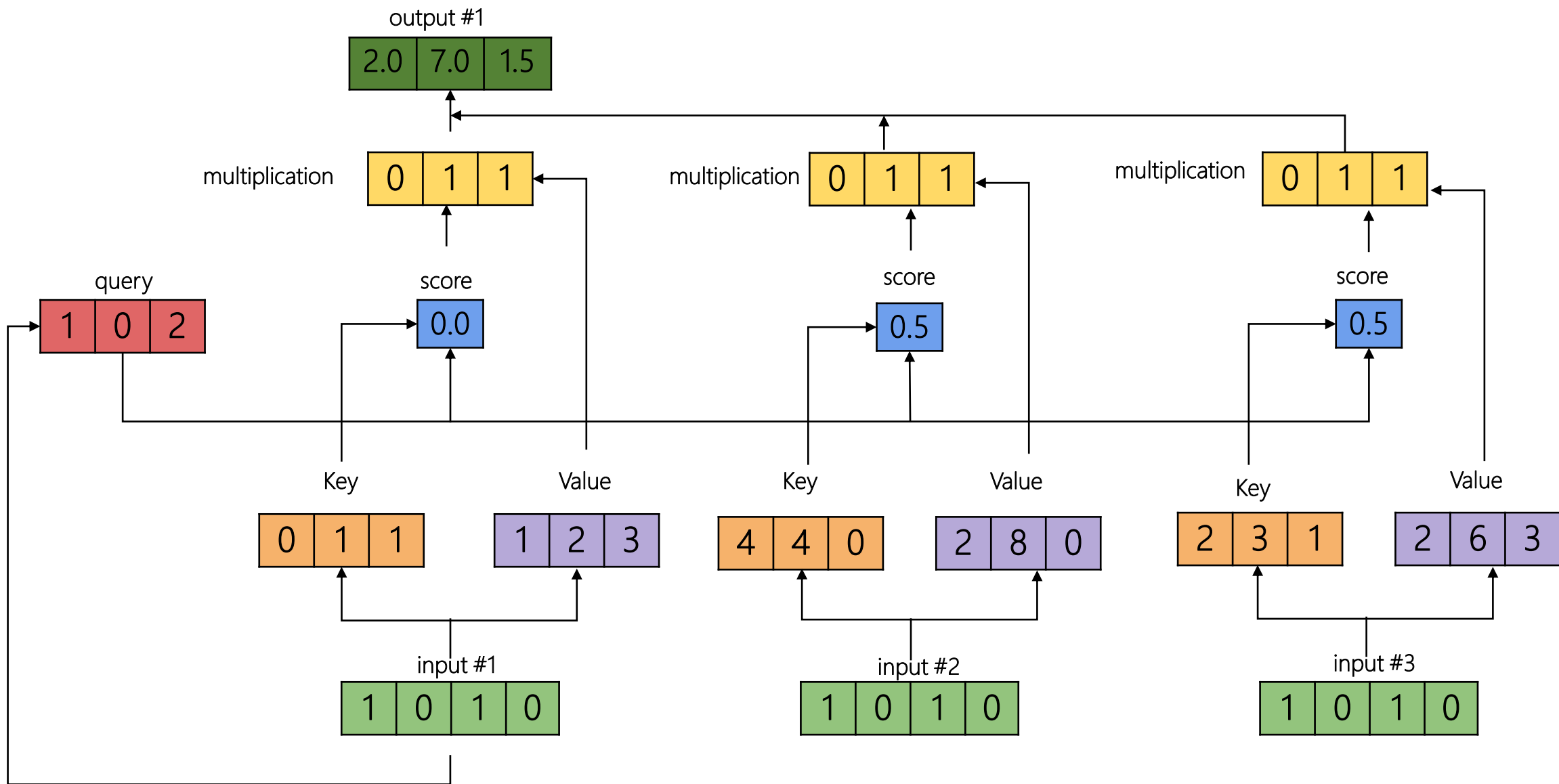
A green 3x4 matrix X is multiplied by a blue 4x3 matrix W^V to produce a blue 3x3 matrix V.

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

The final step shows the softmax of the scaled dot product of Q and K^T multiplied by V, resulting in a pink 3x3 matrix Z.

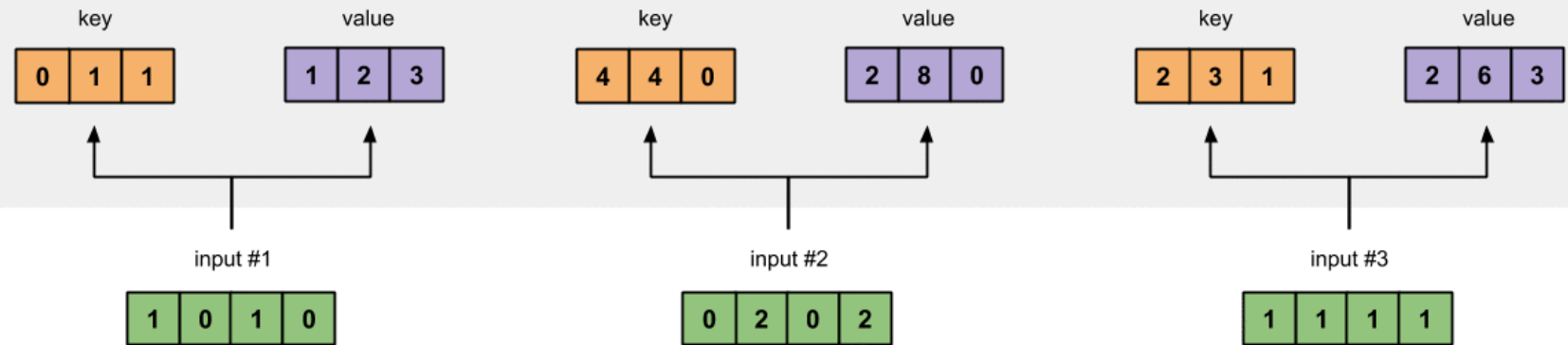


self-attention



Attention

Self-attention



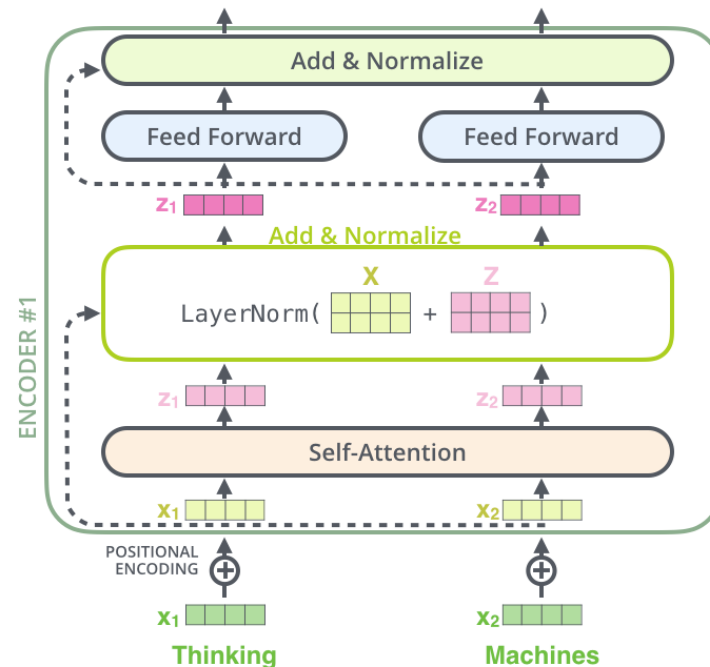
残差连接和 Layer Normalization

残差连接

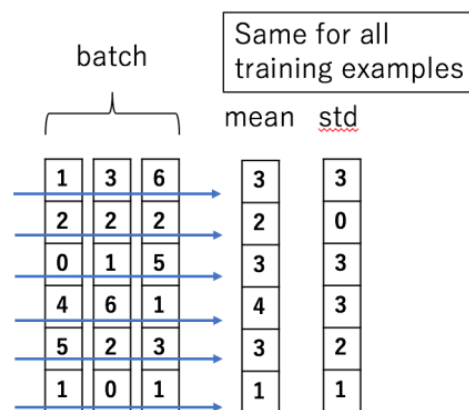
我们在上一步得到了经过 self-attention 加权之后输出，也就是 $\text{Attention}(Q, K, V)$,

Layer Normalization

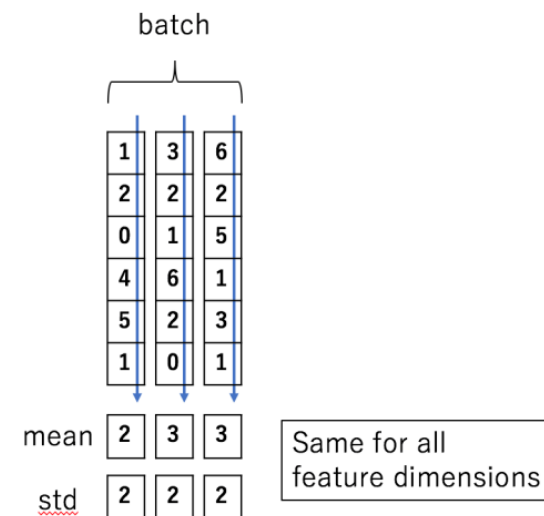
Layer Normalization 的作用是把神经网络中隐藏层归一为标准正态分布，也就是独立同分布，以起到加快训练速度，加速收敛的作用。用每一列的每一个元素减去这列的均值，再除以这列的标准差，从而得到归一化后的数值，加 是为了防止分母为 0



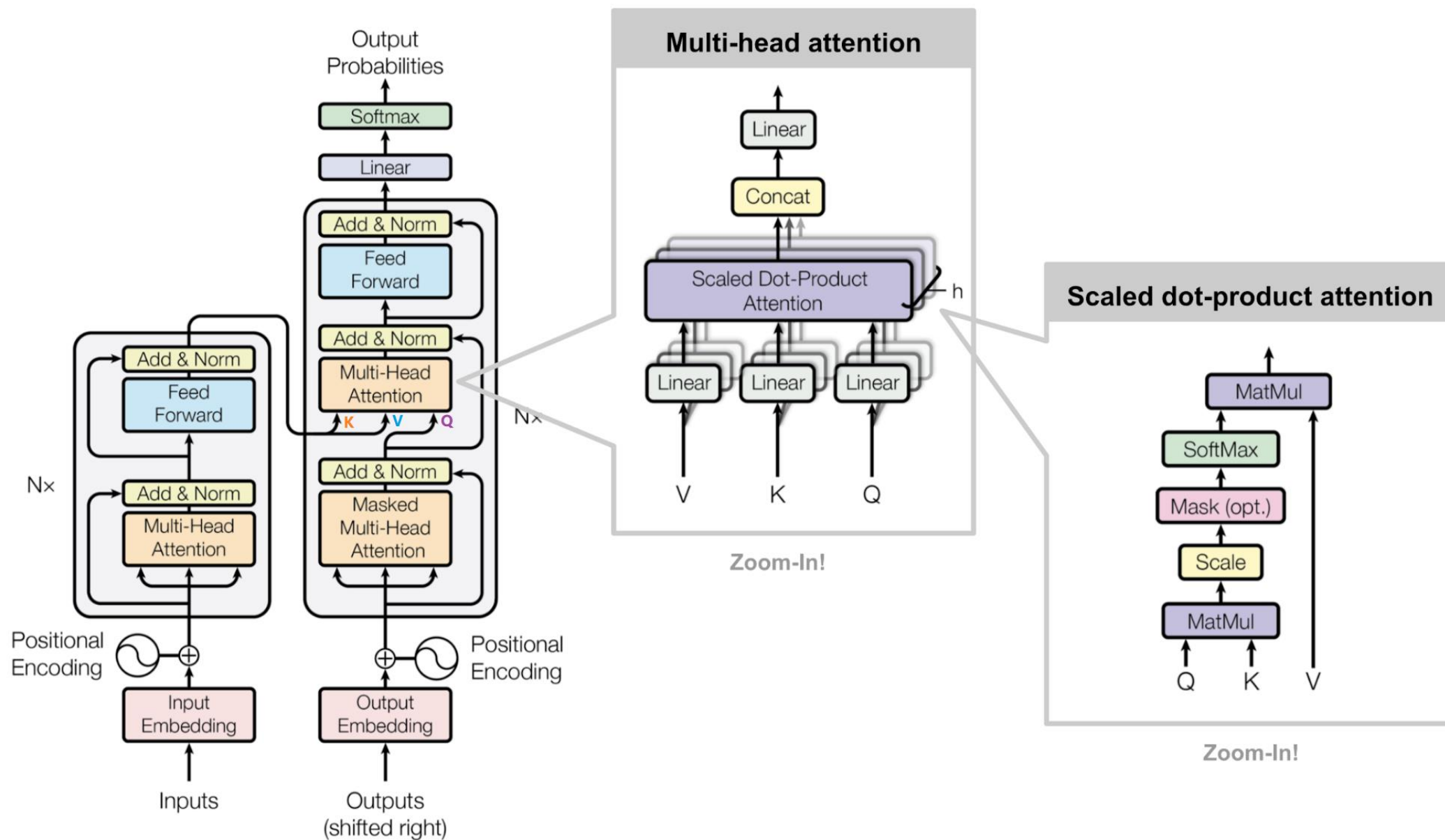
Batch Normalization



Layer Normalization



Transformer结构



THANKS