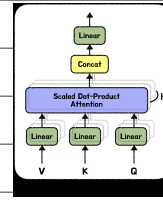
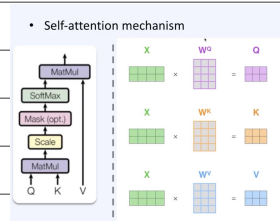


# Transformer

복습

- self-attention

- Multi-Head-Attention



## Transformer

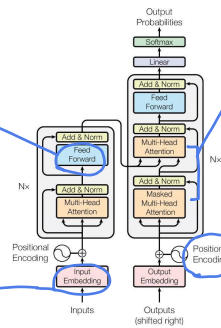
- 아래의 과정을 여러번 진행하여 값을 가져온다.
- 이렇게 사용된 Transformer은 Vision Transformer로 바뀌어서 사용한다.

### Feed Forward

- attention 블록 이후에 완전 연결 계층(Dense Layer)구성됨
- 좀더 복잡하게 만들기 위해

### Input Embedding

- 단어나 토큰을 고정된크기의 벡터로 표현한다.
- 보통 사전 훈련된 (word2vec, glove)를 사용한다



### Masked

- Decoder 부분이라서 깊게 안했다.
- 일반적인 Multi 부활에 추가적으로 Mask를 사용하여 미처된 정보에 접근금지시킬려고

### Positional Encoding

- 문장을 보면 순서가있다. 단어벡터 앞에 위치 마크를 추가해준다.
- 순서대로 1,2,3 해도 되지만 값이 커져 버리는문제 때문에 sin과 cos의 주기를 늘려서 사용한다.