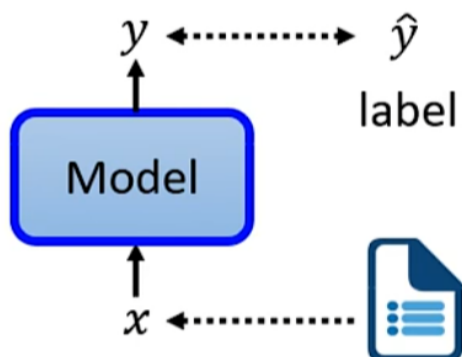
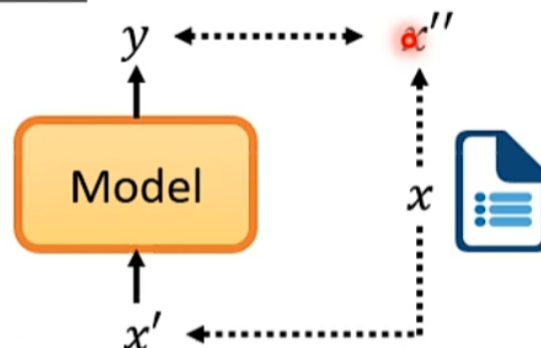


自监督式学习

Supervised



Self-supervised



BERT

BERT是Transformer的编码器。

- 输入时**随机**盖住一些token（用一个特殊字符遮盖或者随机用其他字符替代），Bert的目标是要成功预测出遮盖的内容。

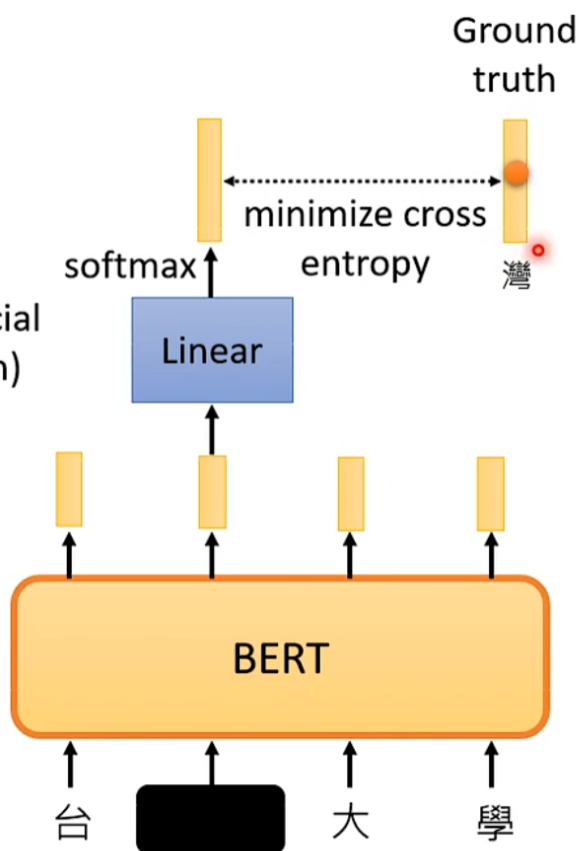
Masking Input

<https://arxiv.org/abs/1810.04805>

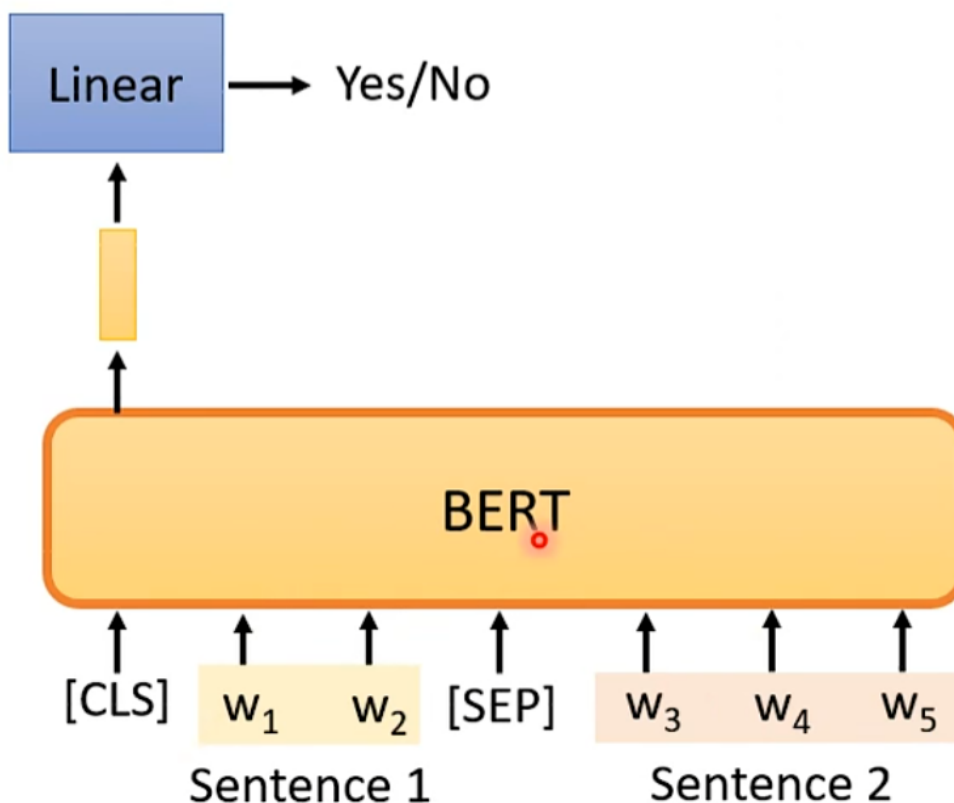


Transformer Encoder

Randomly masking some tokens



- Next Sentence Prediction: 判断CLS之后的两个句子是否相接（二元分类问题）

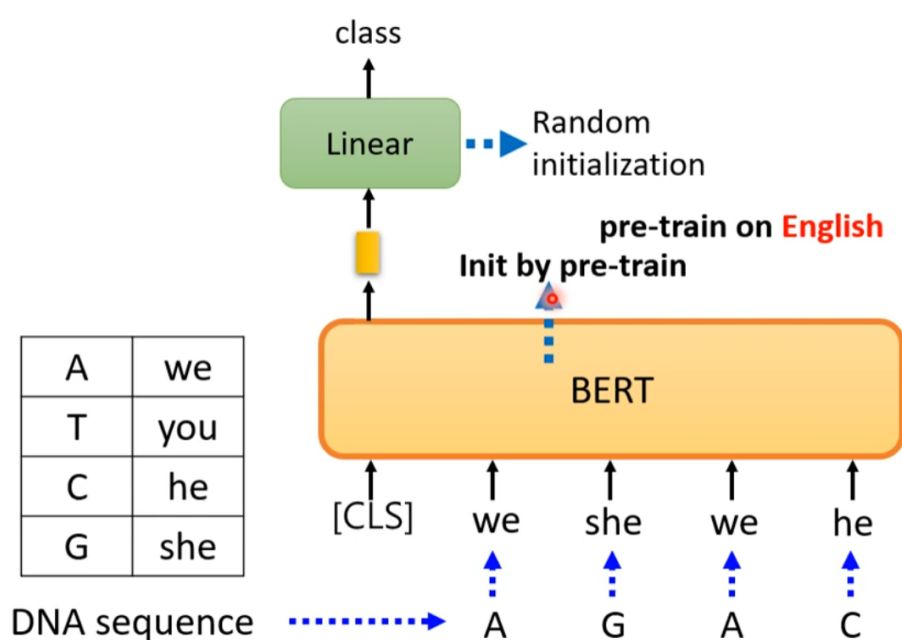


- 运用到下游任务 (Bert通过预训练之后微调)

为什么BERT有用

说法一：在预训练中可以充分了解token上下文间的关系

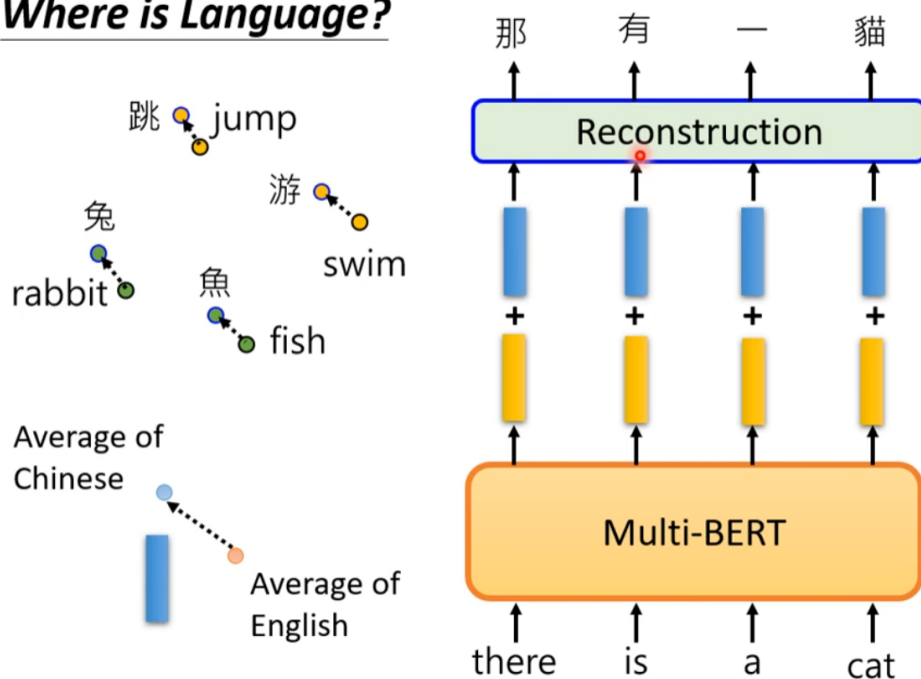
说法二：在说法一的基础上，但在基因任务中利用在英语数据集训练的BERT也能表现的很好。没有统一的说法解释。



Multi-BERT

在English的预训练参数在中文的训练上也能表现很好（前提是数据量要大）

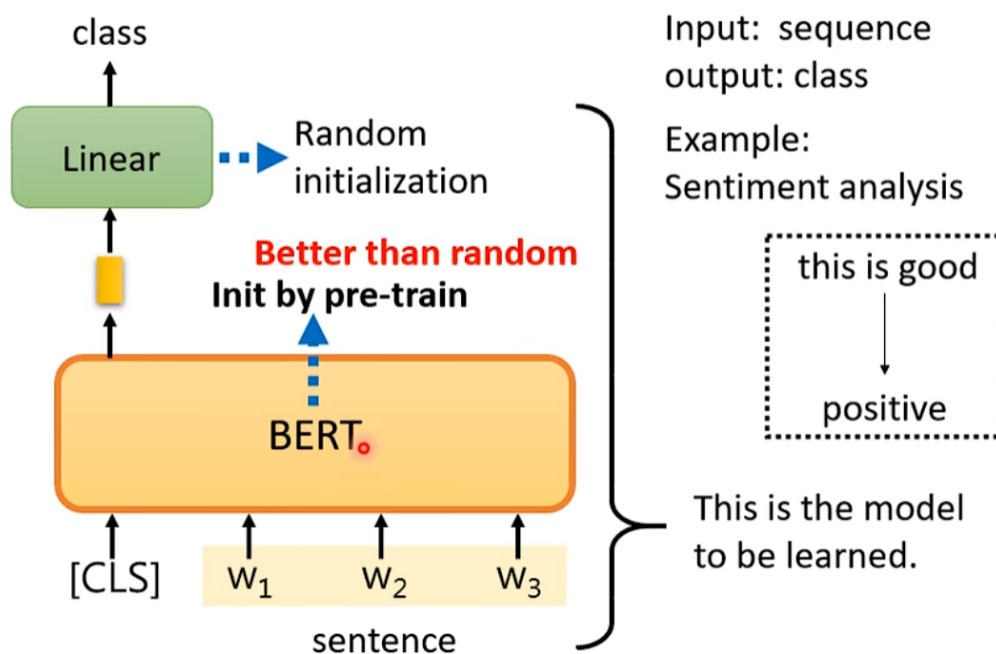
Where is Language?



但新学习的内容不是替换关系，而是藏在参数里面。当输入英语加上蓝色的语言差异，就可以直接翻译。

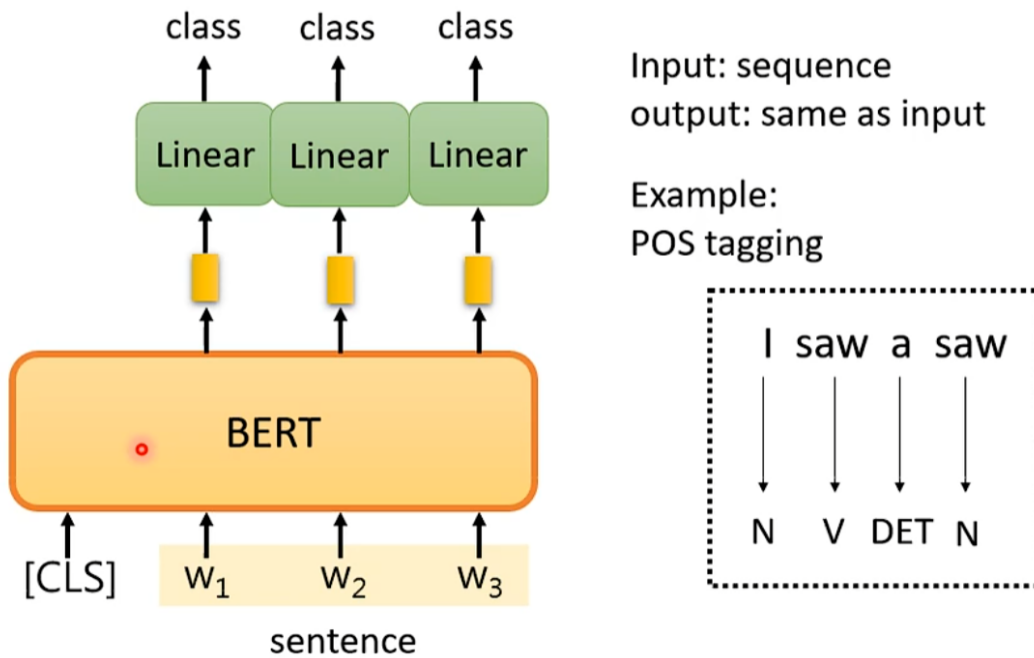
如何使用BERT

- （1）情感分析，只随机初始化CLS，其他用预训练参数（loss下降更快）



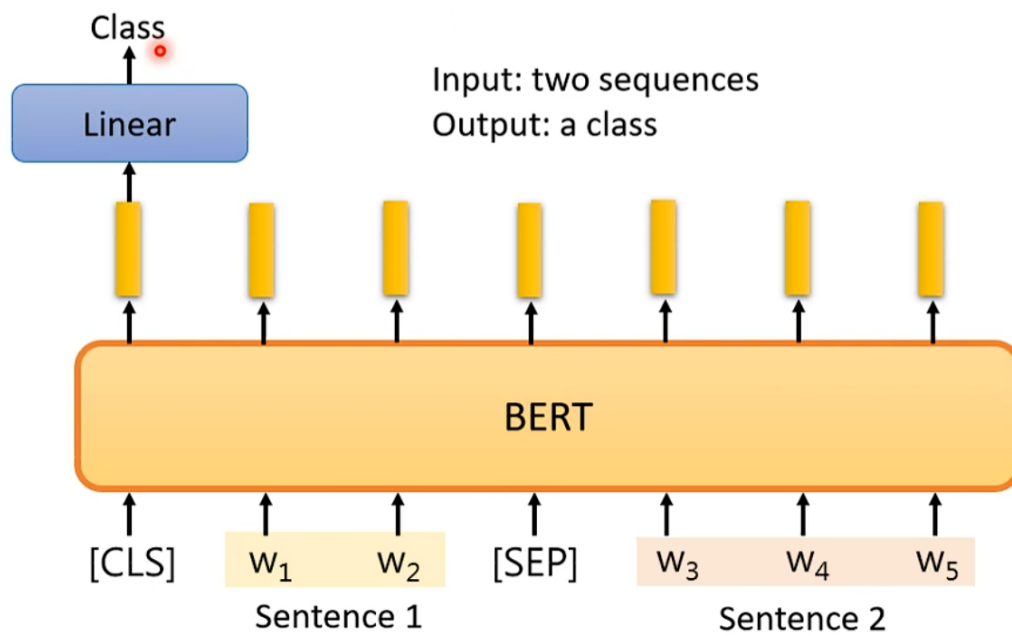
只随机初始化CLS的Linear。

- (2) 输入输出长度一样 (如: POS tagging)



只随机初始化Linear。

- (3) 输入两个句子，输出一个类别 (如: NLI)



只随机初始化CLS的Linear。

- (4) 答案一定出现在文案里的问答系统

- Extraction-based Question Answering (QA)

Document: $D = \{d_1, d_2, \dots, d_N\}$

Query: $Q = \{q_1, q_2, \dots, q_M\}$



output: two integers (s, e)

Answer: $A = \{d_s, \dots, d_e\}$

In meteorology, precipitation is any product of the condensation of 17 spheric water vapor that falls under gravity. The main forms of precipitation include drizzle, rain, sleet, snow, graupel and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals within a cloud. Short, intense periods of rain 77 at 79 cations are called "showers".

What causes precipitation to fall?

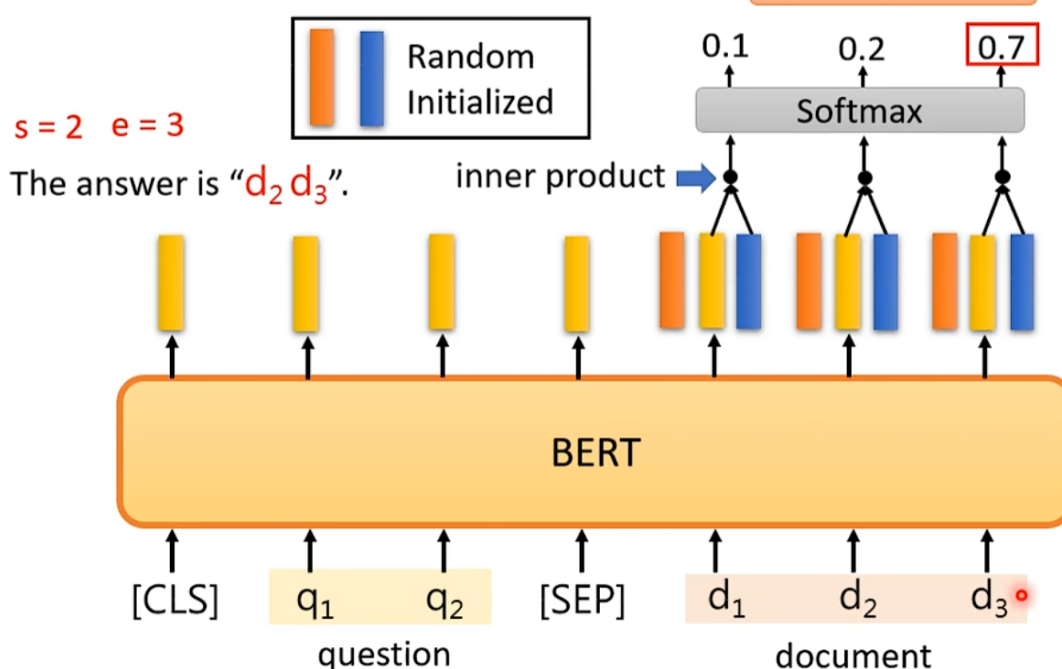
gravity $s = 17, e = 17$

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?

graupel

Where do water droplets collide with ice crystals to form precipitation?

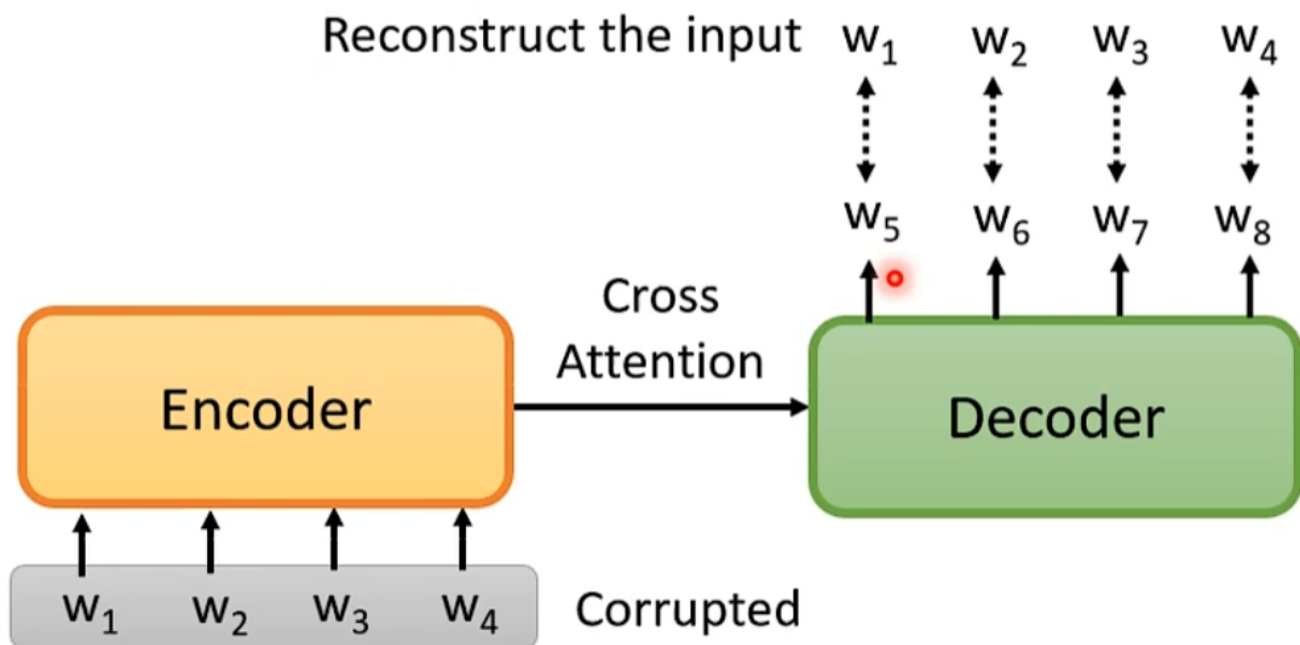
within a cloud $s = 77, e = 79$



用一个开始和终结的token随机初始化训练答案的起始和终结

预训练Transformer

故意“弄坏”输入token，输出还原token（MASS、删除、换位等）



GPT

训练方法：预测下一个token

