**Bert笔记**

BERT本质上是由多个Transformer的 Encoder堆叠而成，再加以MLM（Masked Language Model）和NSP（Next Sentence Prediction）进行预训练，后续还可以接各种fine-tune任务（如文本分类、NER）等

**Encoder – Decoder**

**Attention机制**

**Transformer**