NSP Mask LM Mask LM

classification head decoder

Encoder layer – L

Encoder layer – 2

Encoder layer – 1

Embedding

Tokenizer

Masked sentence A Masked sentence B

Pre-training