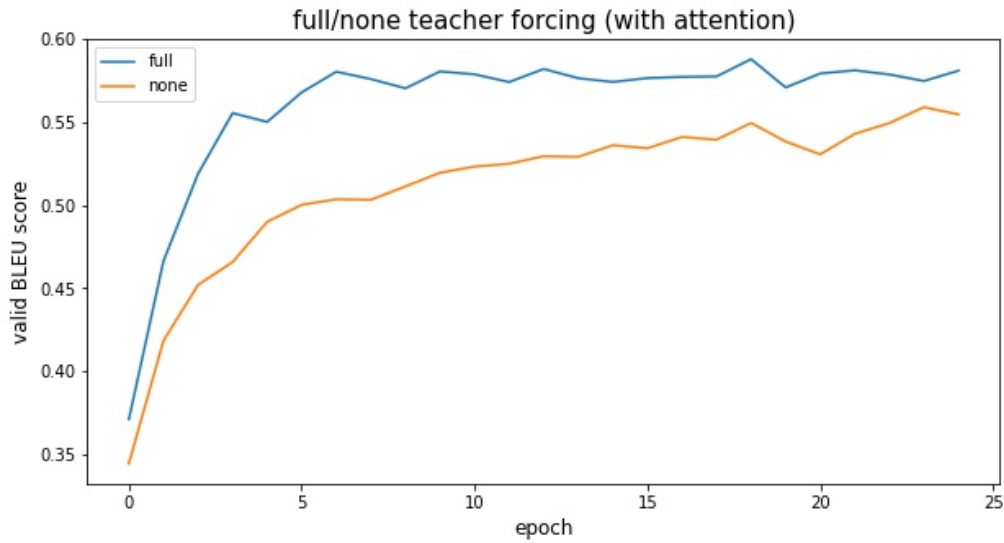


## 1. (20%) Teacher Forcing:

請嘗試移除 Teacher Forcing，並分析結果。

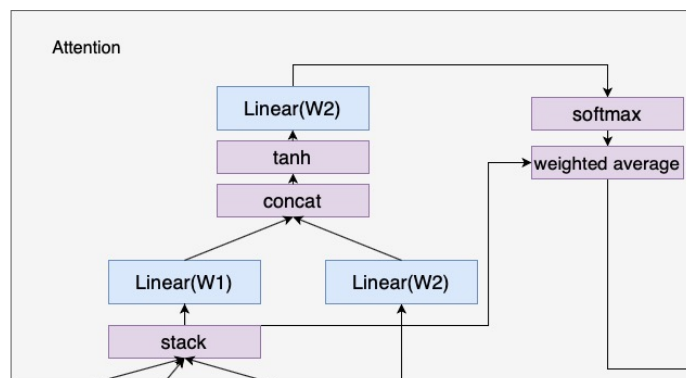


	full	none
Best valid BLEU	0.588	0.559
Test BLEU	0.564	0.533

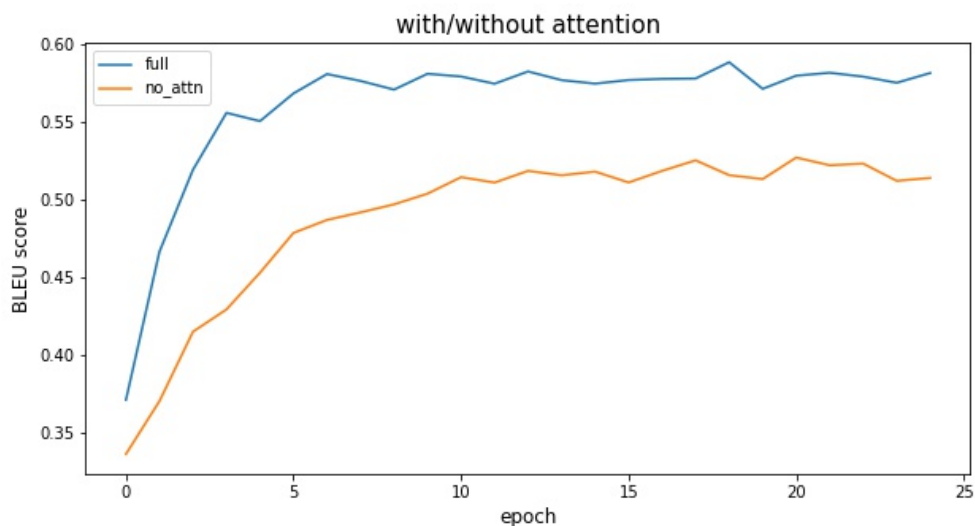
可以觀察到完全沒有 teacher forcing 的話 model 收斂的比較慢，而且最後的結果比較差。

## 2. (30%) Attention Mechanism:

請詳細說明實做 attention mechanism 的計算方式，並分析結果。



我實做的是 additive/concat attention，架構如上圖所示。訓練的結果如下：

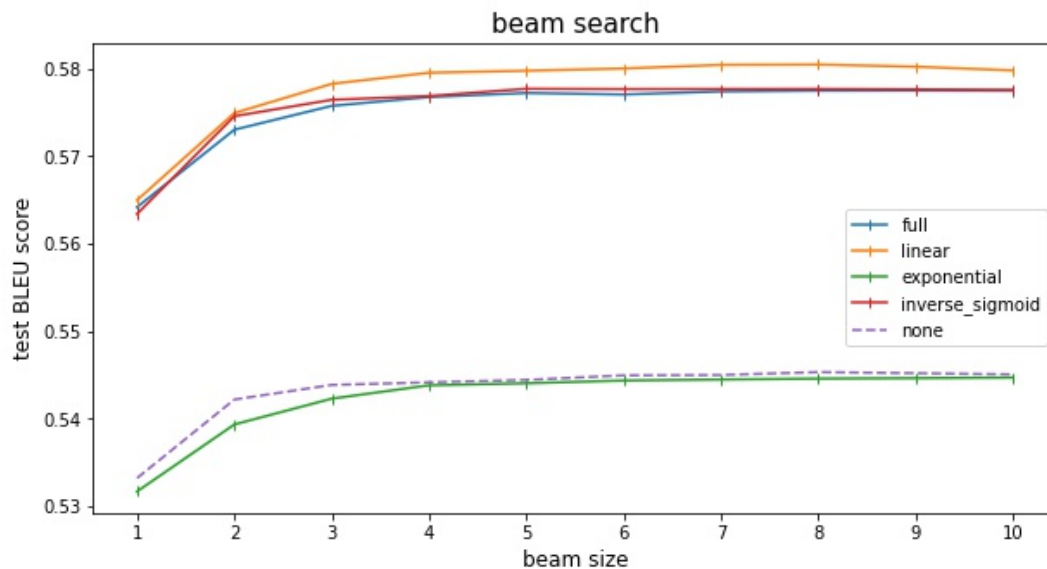


	Full(with attn)	no_attn
Best valid BLEU	0.588	0.528
Test BLEU	0.564	0.502

可以觀察到 attention mechanism 可以大大提升 model 的 BLEU score。

### 3. (30%) Beam Search:

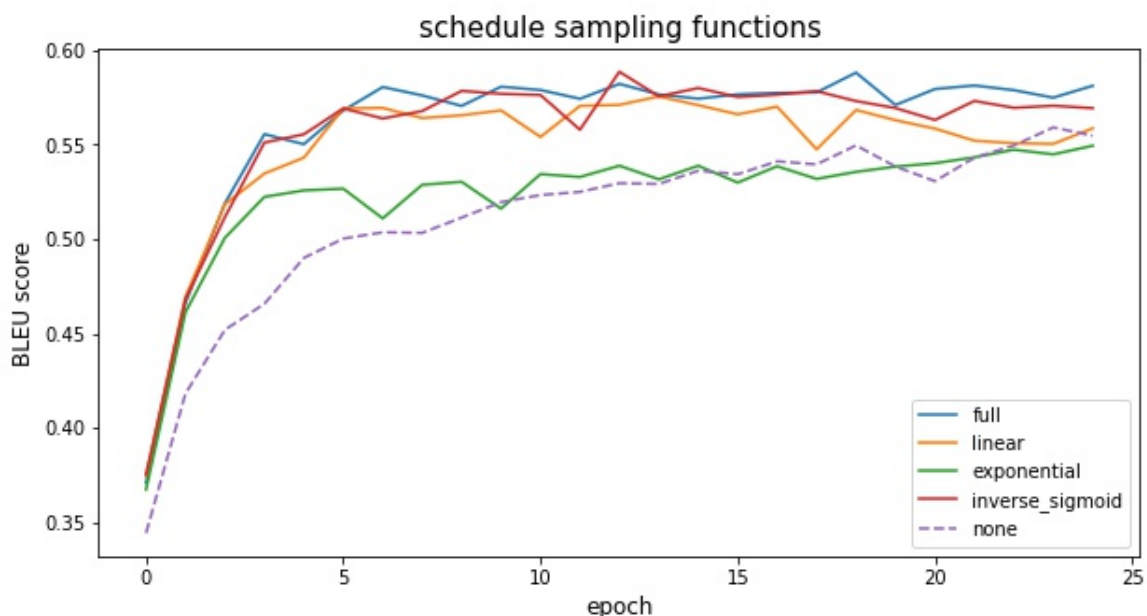
請詳細說明實做 beam search 的方法及參數設定，並分析結果。



可以觀察到對於各種不同的 model，beam search 都可以些微提升原本的 BLEU score。

### 4. (20%) Schedule Sampling:

請至少實做 3 種 schedule sampling 的函數，並分析結果。



	full	linear	exponential	inv_sigmoid	none
Best valid BLEU	0.588	0.575	0.549	0.588	0.559
Test BLEU	0.564	0.565	0.531	0.563	0.533

其中 full 為全用 teacher forcing，none 為完全沒有 teacher forcing。可以觀察到在 valid BLEU score 上 full, linear 跟 inv\_sigmoid 都表現得差不多，但在 test dataset 上 linear 做的最好（在做 beam search 後差距更大，可參考 p3 的圖表）。而 exponential 的表現就遠不如其他 schedule sampling function 的表現，甚至比完全不做 teacher forcing 來得差（雖然我覺得有一部分是這個 dataset 比較好 learn）。