Beam Search

Beam search只在test的时候需要，因为训练的时候知道正确答案，并不需要在进行这个搜索。

Beam search中只有一个参数B，叫做beam width(集束宽)，用来表示在每一次挑选top B的结果。

test的时候，假设词表大小为3，内容为a , b , c。beam size为2.

Decoder解码的时候：

* 1：生成第1个词的时候，选取概率最大的2个词，假设为a,c，那么当前的两个序列就是a和c。
* 2：生成第二个词的时候，我们将当前序列a和c，分别与词表当中的所有词进行组合，得到新的6个序列aa ab ac ca cb cc，然后从中选择两个得分最高的，作为当前序列，假设为aa cb。
* 3：后面会不断重复这个过程，直到遇到结束符或者达到最大长度为止。最终输出2个得分最高的序列。