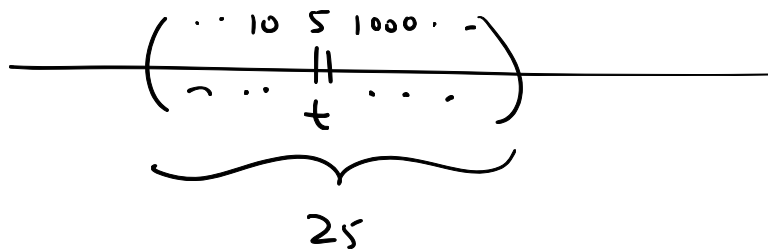


Dataset:

for all sequence
for all offsets (0, 1, 2)

S = seq of tokenized (int) kmers

[5, 15, 0, 1000, ...]

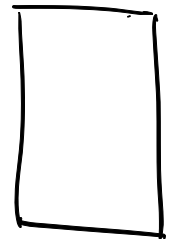


Dataloader

→ num workers = 4

→ batch = 1

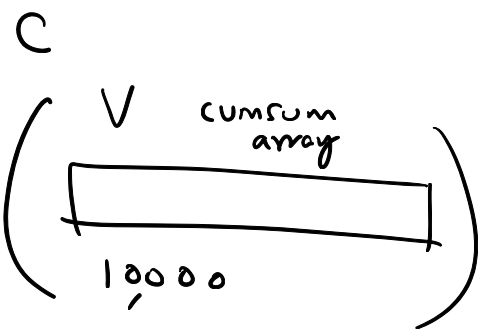
block



512x1024 pos



6x512x1024



np. searchsorted (C, $\frac{5 \times 512 \times 1024}{}$)

worker-id = 0, 1, 2, 3

num workers = 4

array of Neg sampler [0]

