

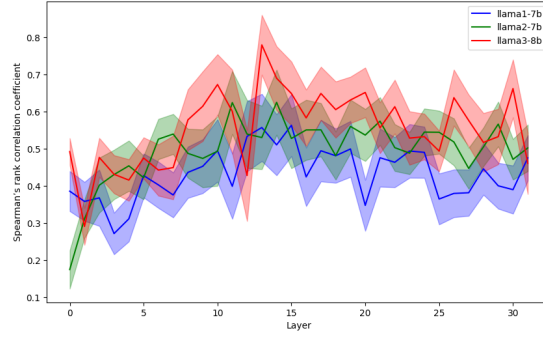


Figure 1: Spearman's rank correlation coefficient between the similarity and the attention score of two tokens. The thick line represents the mean μ calculated from different samples. The upper and lower boundaries of the line are $\mu + \sigma$ and $\mu - \sigma$, respectively, where σ is the standard deviation of different samples.

```

1  """the python code for calculate the
2  low bound base for a context length of 32k"""
3  import torch
4  import numpy as np
5  def get_BMtheta_expectation(base,context_size=2**15,dim=128):
6      realdim = dim / 2
7      d= torch.arange(0, realdim, 1)
8      theta = base ** (-2*d/dim)
9      dist= torch.outer(torch.arange(0,context_size),theta).cos()
10     return dist.sum(dim=1) / realdim
11  search_base = []
12  for x in range(3,10):
13     for i in range(1,10):
14         for j in range(10):
15             search_base.append((i+j/10)* (10**x))
16  for base in search_base:
17     ans = get_BMtheta_expectation(base)
18     if True not in (ans<0):
19         print("Find!Base=", base)
20         break
21     idx = np.argmax(ans < 0)
22     print('base', base, 'first zero position', idx)

```