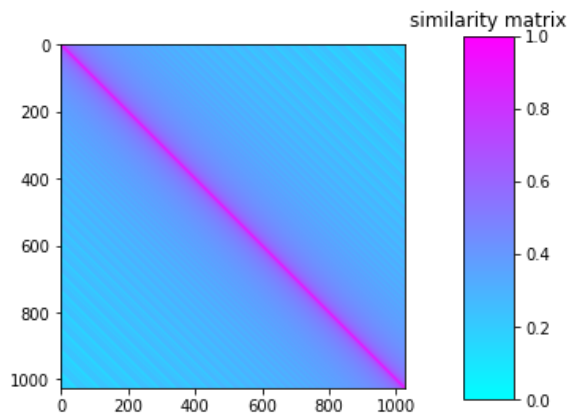


ML HW5 Report

B09901066 謝承修

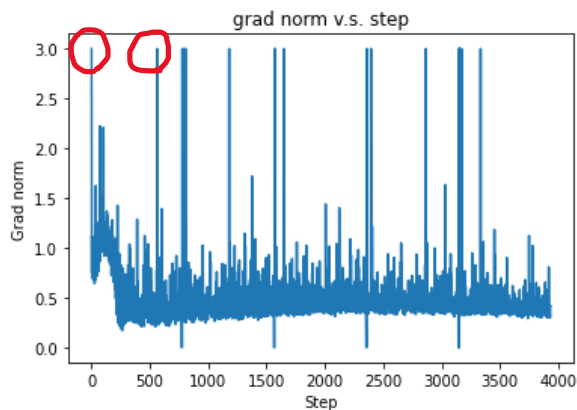
1. Visualize the similarity between different pairs of positional embedding and briefly explain the result.

As seen in the figure, the diagonal elements have the largest similarity as we expected, since the similarity between the embedding of close positions is much stronger.



2. Clip gradient norm and visualize the changes of gradient norm in different steps. Circle two places with gradient explosion.

As seen in the following figure, there exists several gradient explosion like the marked ones.



p.s. 我畫出的版本不是我交出去的版本，因為我忘記存 grad norm 了哭啊。