

HW5

True or False

P1 True; in the noise adding step, we could replace the word with [MASK] until a all-mask sentence. If we take a relatively large dimension of word embedding, For the position embedding part, we remain them unchanged(such as cosine embedding). Thus, the final embedding sequence still contain position information. $H_t = [h_t, e_t]$, where $h_t = \sqrt{\alpha_t}h_0 + \sqrt{1 - \alpha_t}\epsilon_t$. And the denoising step is to sample $P(H_{t-1}|H_t) = \text{Transformer}(H_t)$, and finally, after sample H_0 , use nearest neighbor search to find the most similar word in the dictionary.

Q&A

P2