CS 224n: Assignment #4 😈

1.

(j)

dot product attention $e_{t,i} = s_t^T h_i$ compared to **multiplicative attention** $e_{t,i} = s_t^T W h_i$:

Advantage : easy to compute

Disadvantage: less expressional ability, and s_t, h_i must be have the same size

multiplicative attention $e_{t,i} = s_t^T W h_i$ compared to **additive attention** $v^T tanh(W_i h_i + W_2 s_t)$

Advantage : easy to compute

Disadvantage: less expressional ability