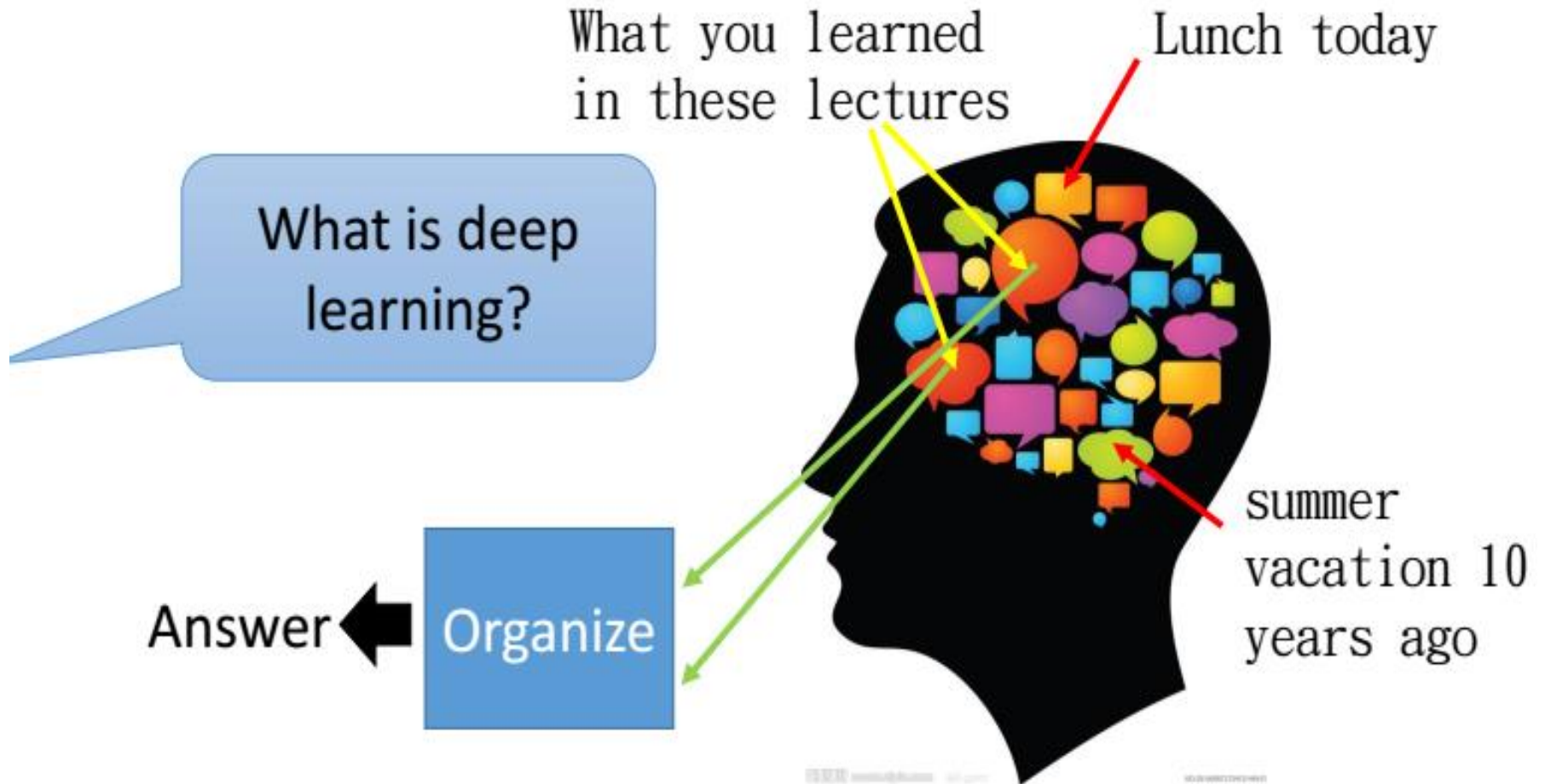


Attention Model of Memory

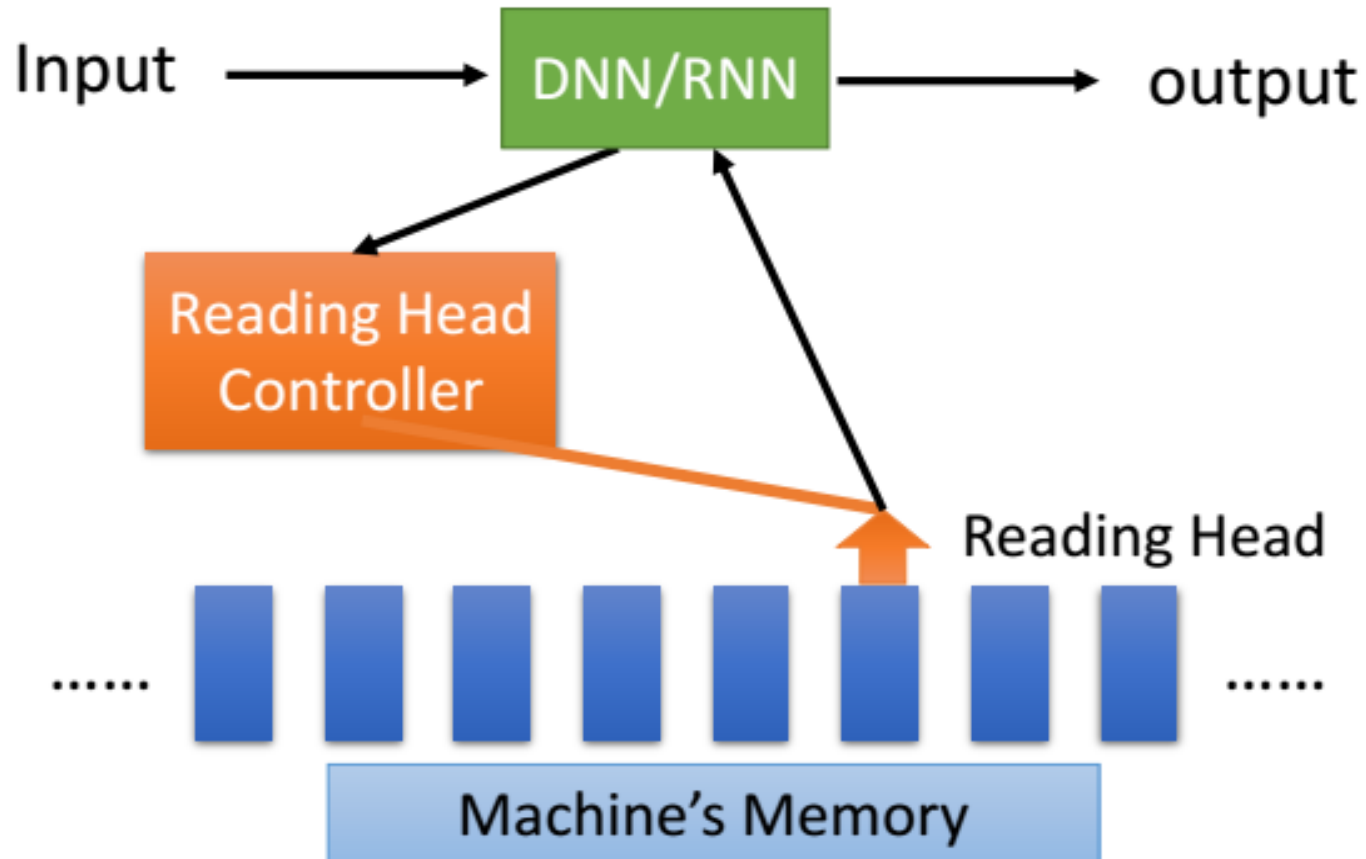
Memory

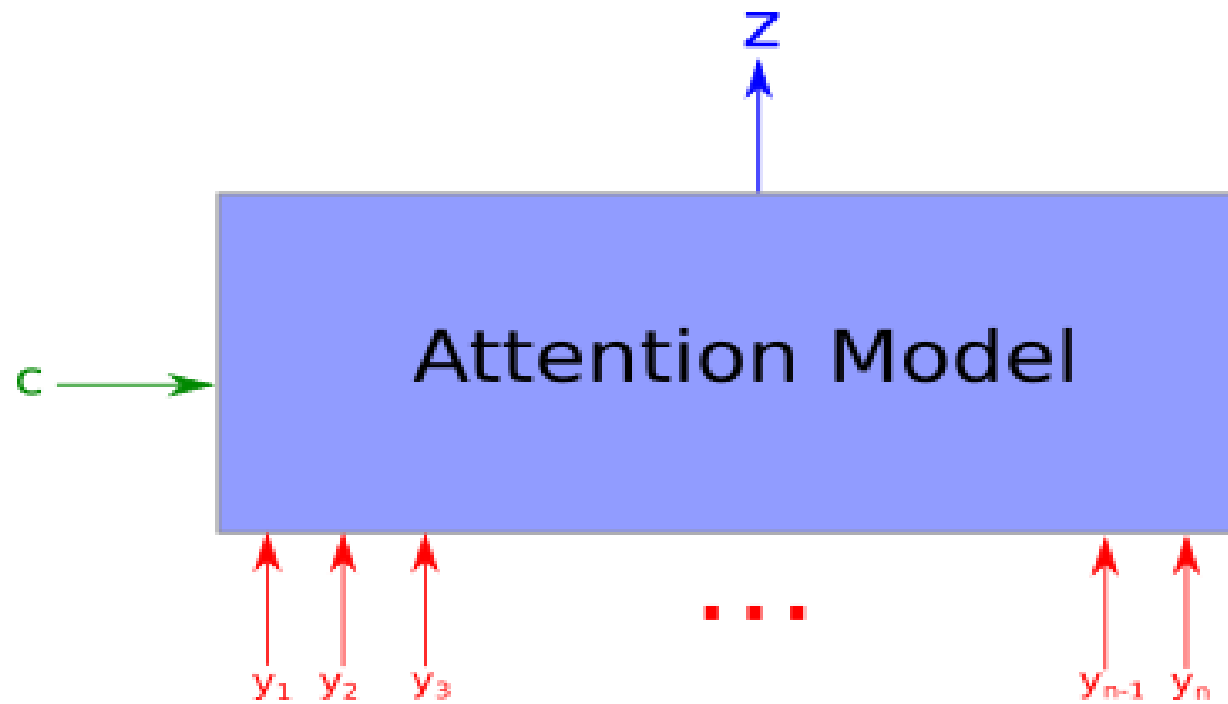


Attention Model for Memory

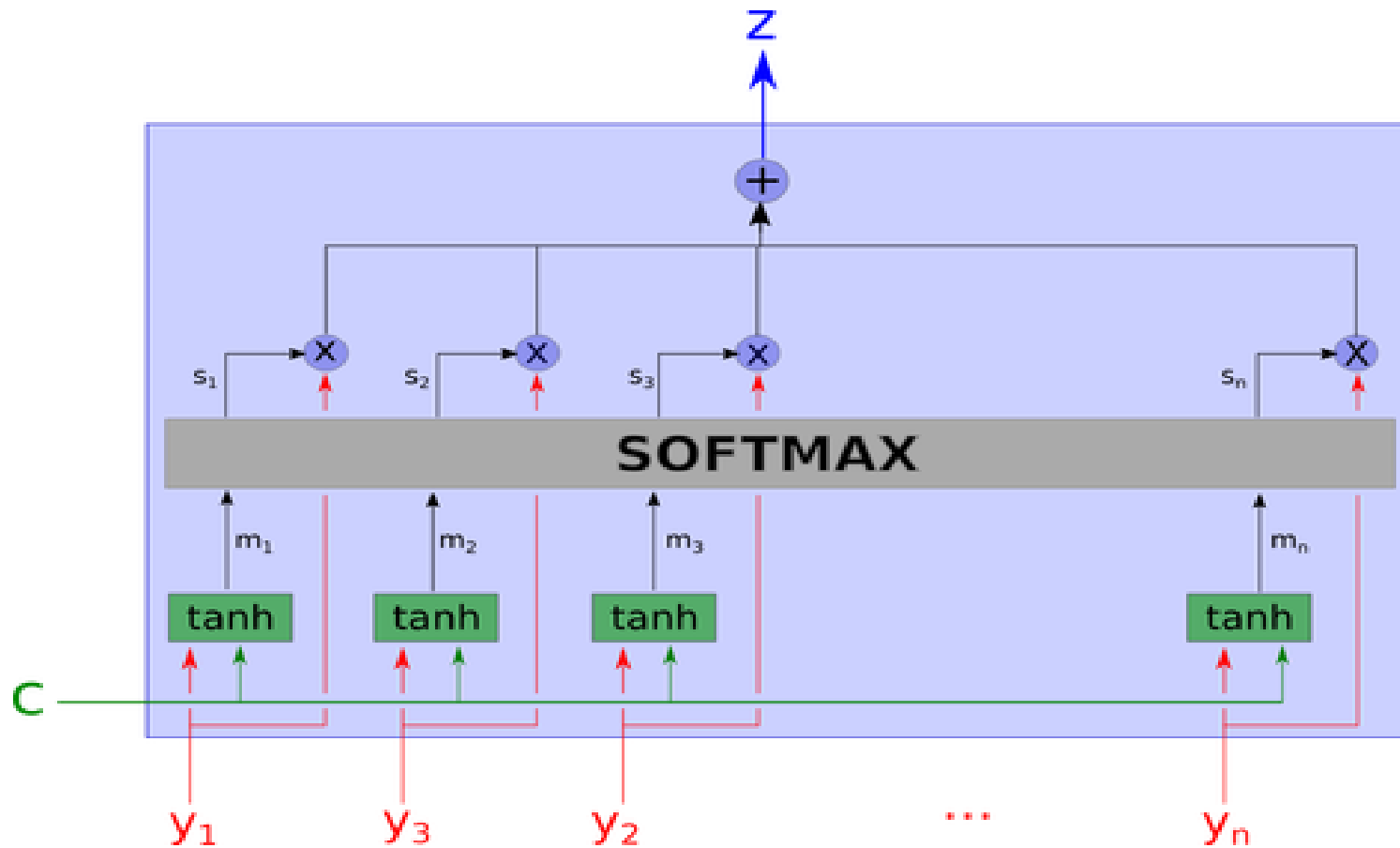
- How does the learning system know which parts of the memory are relevant for each step of target task is different? Attention solves this problem.
- The attention model « softly-choosing » the parts of memory which are most correlated with the target context.

Attention model1 in computing





Attention model1: tanh relevance



$$m_i = \tanh(W_{cm}c + W_{ym}y_i)$$

$$\sum_i s_i = 1$$

$$z = \sum_i s_i y_i$$

- It computes the relevance among y_i and context c using tanh layer and name them as m_1, m_2, \dots, m_n . An important remark here is that each m_i is computed without looking at the other y_j for $j \neq i$ (i.e., they are computed independently).
- Compute the normalized relevances of m_i s using softmax layer, call them as s_i .
- The output z is the weighted arithmetic mean of all the y_i , where the weight represent the normalized relevance for each variable according the context

Attention model1: dot product relevance

