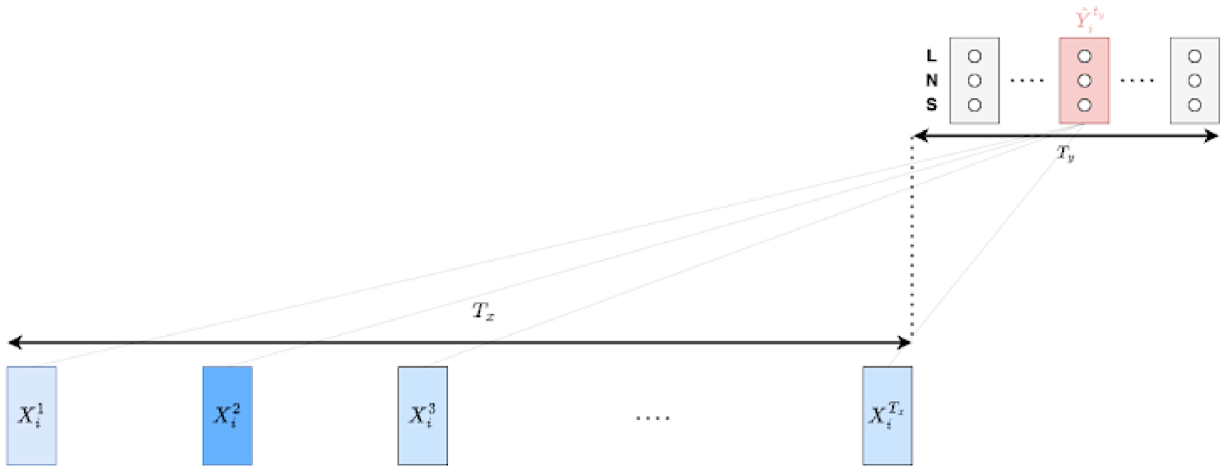
Interactive Session 3 - Seq2Seq with Attention

In financial markets, we often need to predict future trading positions based on historical market features. Our objective is to transform a sequence of input features (such as price movements, volume indicators, technical indicators) into a sequence of future trading positions. Given an input sequence of market features over Tx time periods, we want to predict the optimal trading positions (Long, Short, or Neutral) for the next Ty time periods. This sequential prediction problem requires understanding both the temporal dependencies in the input features and the ability to focus on the most relevant historical information when making each future position decision. The sequence-to-sequence framework with attention mechanism allows us to capture these complex relationships and generate informed trading strategies.

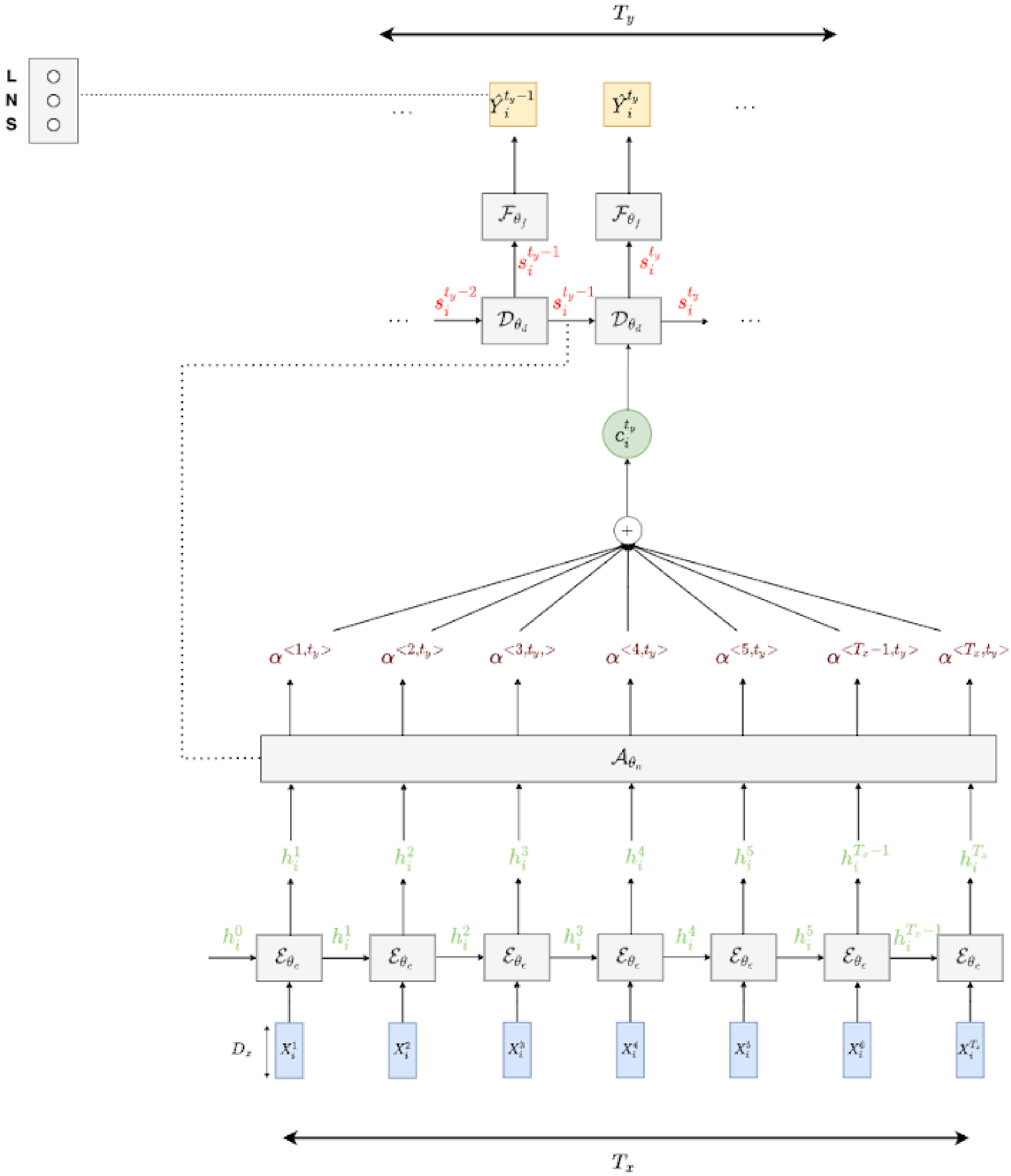
We have a dataset $(X_i, y_i)_{1 \leq i \leq N}$ where each sample consists of:

- Input sequence: $X_i = (X_i^1, X_i^2, \ldots, X_i^T)$
- Each $X_i^t \in \mathbb{R}^D$ represents D financial features at time step t
- Target label: $y_i \in \{L, S, N\}$ representing Long/Short/Neutral trading positions

In financial markets, we often need to predict future trading positions based on historical market features. Our objective is to transform a sequence of input features (such as price movements, volume indicators, technical indicators) into a sequence of future trading positions. Given an input sequence of market features over Tx time periods, we want to predict the optimal trading positions (Long, Short, or Neutral) for the next Ty time periods. This sequential prediction problem requires understanding both the temporal dependencies in the input features and the ability to focus on the most relevant historical information when making each future position decision. The sequence-to-sequence framework with attention mechanism allows us to capture these complex relationships and generate informed trading strategies.



The Architecture



1. Which of the following correctly represents the categorical cross-entropy loss function for our sequence-to-sequence model predicting Long/Short/Neutral positions? 2 points

Mark only one oval.

$$L(\theta) = -\frac{1}{N \cdot T_y} \sum_{i=1}^{N} \sum_{t_y=1}^{T_y} \sum_{c=1}^{3} Y_i^{t_y,c} \log(\hat{Y}_i^{t_y,c})$$

- [ ] A

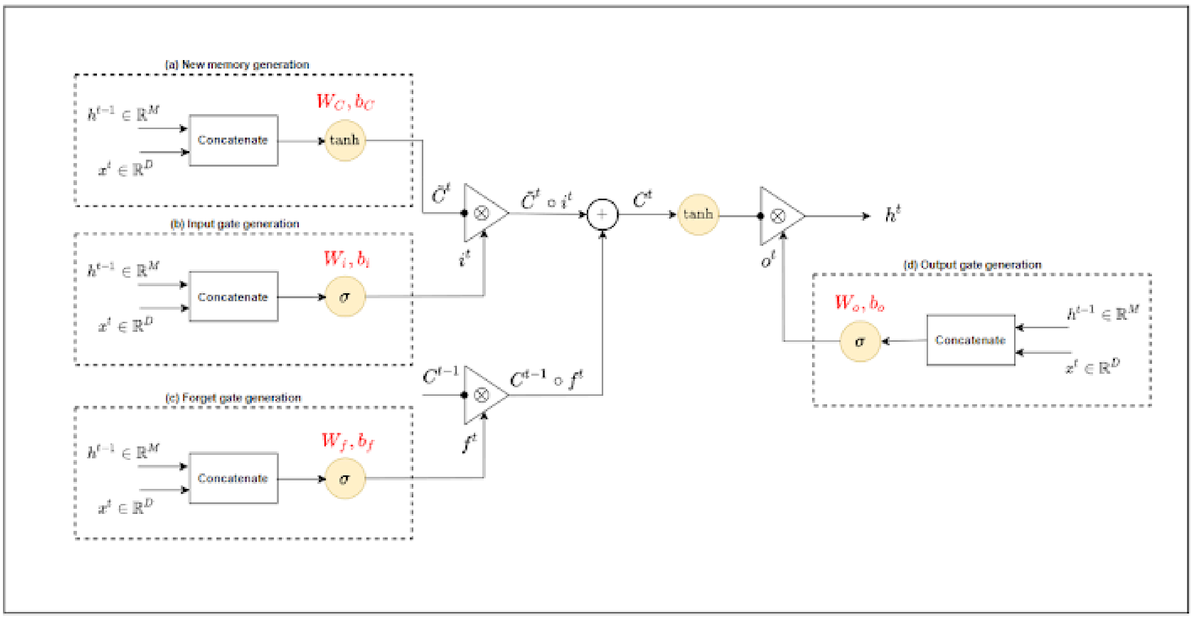
$$L(\theta) = \frac{1}{N \cdot T_y} \sum_{i=1}^{N} \sum_{t_y=1}^{T_y} (Y_i^{t_y} - \hat{Y}_i^{t_y})^2$$

- [ ] B

$$L(\theta) = -\frac{1}{N} \sum_{i=1}^{N} \log(\hat{Y}_i^{T_y})$$

- [ ] C

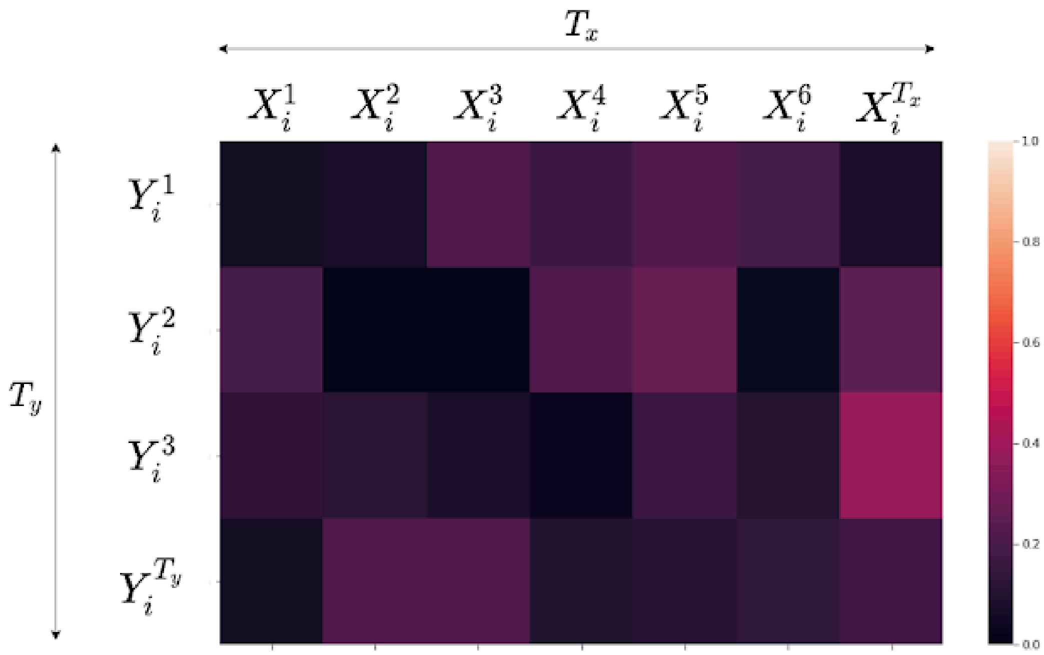
2. For an LSTM encoder with input dimension D and hidden dimension M, what is the total number of parameters? 2 points



Mark only one oval.

- 4 × M × (D + M + 1)
- M × (D + M + 1)
- 4 x M × D

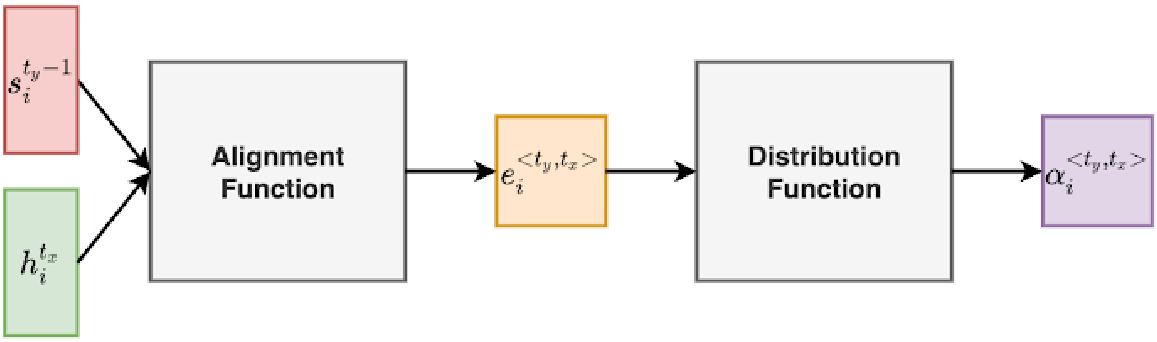
3. What is the shape of the attention weight matrix α in our sequence-to-sequence model? 2 points



Mark only one oval.

- (T_x, T_y)
- (T_y, T_x)
- (T_x, T_x)
- (T_y, T_y)

4. What does row t_y of the attention weight matrix α represent? 2 points



Mark only one oval.

- The importance of output position t_y relative to other output positions
- The attention distribution over all input time steps when predicting position at time t_y
- The probability distribution of trading positions at time t_y

5. How is the context vector computed in the attention mechanism? 2 points

Mark only one oval.

$$c_i^{t_y} = \sum_{t_x=1}^{T_x} \alpha_i^{\langle t_y, t_x \rangle} \times h_i^{t_x}$$

A

$$c_i^{t_y} = \sum_{t_x=1}^{T_x} \alpha_i^{\langle t_y, t_x \rangle} \times h_i^{t_x}$$

B

$$c_i^{t_y} = \max_{t_x}(\alpha_i^{\langle t_y, t_x \rangle} \times h_i^{t_x})$$

C