

---

### Algorithm for Time Series Forecasting

---

**Input:**  $\mathbf{z}, \mathbf{x} \in \mathbb{R}^*$ , two sequences of IL-Ratio inputs (+ve reals).

**Output:**  $\mathbf{P} = \mathbb{R}^{length(X)}$  where  $p_t \in (0,1)$  is the  $t$ -th IL-Ratio.

**Hyperparameters:**  $l_{max} = 10, L_{enc} = 4, L_{dec} = 4, H, d_e = d_{model}, d_{mlp} \in \mathbb{N}$

**Parameters:**  $\theta$  includes all of the following parameters:

$\mathbf{W}_e \in \mathbb{R}^{d_e \times 1}, \mathbf{W}_p \in \mathbb{R}^{d_e \times l_{max}}$ , the token and positional embedding matrices.

For  $l \in [L_{enc}]$ :

|  $\mathbf{W}_l^{enc}$ , multi-head encoder attention parameters for layer  $l$

|  $\gamma_l^1, \beta_l^1, \gamma_l^2, \beta_l^2 \in \mathbb{R}^{d_e}$ , two sets of layer-norm parameters,

|  $\mathbf{W}_{mlp1}^l \in \mathbb{R}^{d_{mlp} \times d_e}, \mathbf{b}_{mlp}^l \in \mathbb{R}^{d_{mlp}}, \mathbf{W}_{mlp2}^l \in \mathbb{R}^{d_e \times d_{mlp}}, \mathbf{b}_{mlp2}^l \in \mathbb{R}^{d_e}$ , MLP parameters.

For  $l \in [L_{dec}]$ :

|  $\mathbf{W}_l^{dec}$ , multi-head decoder attention parameters for layer  $l$

|  $\mathbf{W}_l^{e/d}$ , multi-head cross-attention parameters for layer  $l$ .

|  $\gamma_l^3, \beta_l^3, \gamma_l^4, \beta_l^4 \in \mathbb{R}^{d_e}$ , two sets of layer-norm parameters,

|  $\mathbf{W}_{mlp3}^l \in \mathbb{R}^{d_{mlp} \times d_e}, \mathbf{b}_{mlp}^l \in \mathbb{R}^{d_{mlp}}, \mathbf{W}_{mlp4}^l \in \mathbb{R}^{d_e \times d_{mlp}}, \mathbf{b}_{mlp4}^l \in \mathbb{R}^{d_e}$ , MLP parameters.

$\mathbf{W}_u \in \mathbb{R}^{length(X) \times d_e}$ , the unembedding matrix.

/\* Encoder portion \*/

```

1   $l_z \leftarrow length(\mathbf{z})$ 
2  for  $t \in [l_z]$ :  $\mathbf{e}_t \leftarrow \mathbf{W}_e[:, z[t]] + \mathbf{W}_p[:, t]$ 
3   $\mathbf{Z} \leftarrow [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{l_z}]$ 
4  for  $l = 1, 2, \dots, L_{enc}$  do
5       $\mathbf{Z} \leftarrow \mathbf{Z} + MHAttention(\mathbf{Z} | \mathbf{W}_l^{enc}, Mask \equiv 1)$ 
6      for  $t \in [l_z]$ :  $\mathbf{Z}[:, t] \leftarrow layer\_norm(\mathbf{Z}[:, t] | \gamma_l^1, \beta_l^1)$ 
7       $\mathbf{Z} \leftarrow \mathbf{Z} + \mathbf{W}_{mlp2}^l ReLU(\mathbf{W}_{mlp1}^l \mathbf{Z} + \mathbf{b}_{mlp1}^l \mathbf{1}^T) + \mathbf{b}_{mlp2}^l \mathbf{1}^T$ 
8      for  $t \in [l_z]$ :  $\mathbf{Z}[:, t] \leftarrow layer\_norm(\mathbf{Z}[:, t] | \gamma_l^2, \beta_l^2)$ 
9  end

```

/\* Decoder portion \*/

```

10  $l_x \leftarrow length(\mathbf{x})$ 
11 for  $t \in [l_x]$ :  $\mathbf{e}_t \leftarrow \mathbf{W}_e[:, x[t]] + \mathbf{W}_p[:, t]$ 
12  $\mathbf{X} \leftarrow [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{l_x}]$ 
13 for  $l = 1, 2, \dots, L_{dec}$  do
14      $\mathbf{X} \leftarrow \mathbf{X} + MHAttention(\mathbf{X} | \mathbf{W}_l^{dec}, Mask[t, t'] \equiv [[t \leq t']])$ 
15     for  $t \in [l_x]$ :  $\mathbf{X}[:, t] \leftarrow layer\_norm(\mathbf{X}[:, t] | \gamma_l^3, \beta_l^3)$ 
16      $\mathbf{X} \leftarrow \mathbf{X} + MHAttention(\mathbf{X}, \mathbf{Z} | \mathbf{W}_l^{e/d}, Mask \equiv 1)$ 
17     for  $t \in [l_z]$ :  $\mathbf{X}[:, t] \leftarrow layer\_norm(\mathbf{X}[:, t] | \gamma_l^4, \beta_l^4)$ 
18      $\mathbf{X} \leftarrow \mathbf{X} + \mathbf{W}_{mlp4}^l ReLU(\mathbf{W}_{mlp3}^l \mathbf{X} + \mathbf{b}_{mlp3}^l \mathbf{1}^T) + \mathbf{b}_{mlp4}^l \mathbf{1}^T$ 
19     for  $t \in [l_x]$ :  $\mathbf{X}[:, t] \leftarrow layer\_norm(\mathbf{X}[:, t] | \gamma_l^5, \beta_l^5)$ 
20 end
21 return  $\mathbf{P} = \mathbf{W}_u \mathbf{X}$ 

```