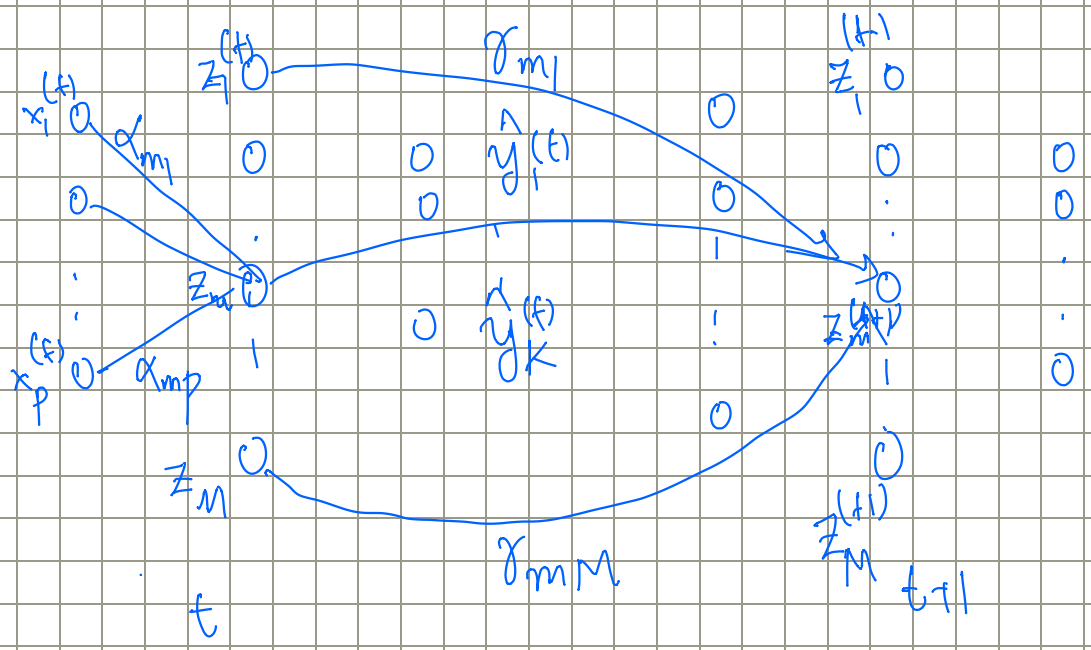
19/10/23

Deep learning

- Recap: Transformer
- Time Series models
  - Recurrent Neural Nets (RNN)
  - Gated Recurrent Units (GRU)
  - Long Short Term Memory (LSTM)
- Intro to RNN

ANN: $z_m^{(t)} = \sigma(\langle \alpha_m, x^{(t)} \rangle)$ (hidden layer output)

RNN: $z_m^{(t)} = \sigma(\langle \alpha_m, x^{(t)} \rangle + \langle \gamma_m, z^{(t-1)} \rangle)$



RNN (an unfolded ANN)

$$D = \{(x^{(1)}, y^{(1)}), \ldots (x^{(t)}, y^{(t)}), \ldots (x^{(T)}, y^{(T)})\}$$

ANN loss risk: $R(\theta) = \frac{1}{T} \sum_{t=1}^{T} d(y^{(t)}, \hat{y}^{(t)})$

$$R(\theta) = \frac{1}{T} \sum_{t=1}^{T} R_t(\theta)$$

$$\partial R / \partial \alpha_{m1} = \frac{1}{T} \sum_{t=1}^{T} \partial R_t(\theta) / \partial \alpha_{m1}$$

- Find $\frac{d}{d\alpha_{ml}} z_m^{(t)} = \frac{d}{d\alpha_{ml}} \sigma\left(\langle \underline{\alpha}_m, \underline{x}^{(t)} \rangle + \langle \underline{r}_m, \underline{z}^{(t-1)} \rangle\right)$

- Ans: Apply total derivative

$$\frac{d z_m^{(t)}}{d\alpha_{ml}} = \underbrace{\frac{\partial z_m^{(t)}}{\partial \alpha_{ml}}}_{\text{partial derivative}} + \sum_{i=1}^{t-1} \left(\prod_{j=i+1}^{t} \frac{\partial z_m^{(j)}}{\partial z_m^{(j-1)}} \right) \frac{\partial z_m^{(i)}}{\partial \alpha_{ml}} := (1)$$

- (1) goes into the back prop through time (BPTT) algorithm