03/11/23

Deep learning

- Recap: GRU, RNN

- LSTM

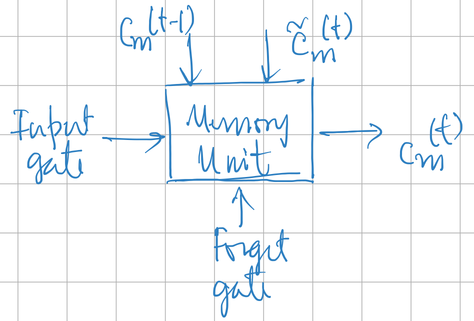
- RNN: $z_m^{(t)} = \sigma\left(\underline{\alpha}_m^T \cdot \underline{x}^{(t)} + \underline{\gamma}_m^T \cdot \underline{z}^{(t-1)}\right)$

- GRU: $z_m^{(t)} = s_m^{(t)} \cdot z_m^{(t-1)} + (1-s_m^{(t)}) \cdot \tilde{z}_m^{(t)}$

  $\tilde{z}_m^{(t)} = \tanh\left(\underline{\alpha}_m^T \cdot \underline{x}^{(t)} + r_m^{(t)} \cdot \underline{\gamma}_m^T \cdot \underline{z}^{(t-1)}\right)$

  $s_m^{(t)}$: update gate (long term)
  $r_m^{(t)}$: reset gate (short term)

- LSTM:



$c_m^{(t)} = f_m^{(t)} \cdot \underline{c_m^{(t-1)}} + i_m^{(t)} \cdot \underline{\tilde{c}_m^{(t)}}$

$\tilde{c}_m^{(t)} = \tanh\left(\underline{\alpha}_m^T \cdot \underline{x}^{(t)} + \underline{\gamma}_m^T \cdot \underline{z}^{(t-1)}\right)$

$z_m^{(t)} = o_m^{(t)} \cdot \tanh\left(c_m^{(t)}\right)$

Gates in an LSTM:
$$\left\{\begin{aligned} f_m^{(t)} &= \sigma\left(\underline{u}_m^T \underline{x}^{(t)} + \underline{v}_m^T \cdot \underline{z}^{(t-1)}\right) \\ i_m^{(t)} &= \sigma\left(\underline{w}_m^T \underline{x}^{(t)} + \underline{p}_m^T \cdot \underline{z}^{(t-1)}\right) \\ o_m^{(t)} &= \sigma\left(\underline{s}_m^T \underline{x}^{(t)} + \underline{d}_m^T \cdot \underline{z}^{(t-1)}\right) \end{aligned}\right.$$

$c_m^{(t)}$: m^{th} memory unit contents at time t
$\tilde{c}_m^{(t)}$: m^{th} candidate mem. unit contents at t
$f_m^{(t)}$: forget gate unit m at time t
$i_m^{(t)}$: input gate unit m at time t
$o_m^{(t)}$: output gate unit m time t

- $\Theta = \left\{ \underline{\alpha}^{\{1..M\}}, \underline{\beta}^{\{1..M\}}, \underline{\gamma}^{\{1..M\}}, \underline{u}^{\{1..M\}}, \underline{v}^{\{1..M\}}, \underline{w}^{\{1..M\}}, \underline{p}^{\{1..M\}}, \underline{s}^{\{1..M\}}, \underline{d}^{\{1..M\}} \right\}$

- Again, LSTM models have the ability to "remember" inputs in the long term and in the near term.

- Generative Models.

Ex 1: Let X be a continuous RV with CDF $F_X(x)$. What is the distribution of the RV Y which is defined as $Y = F_X(X)$? ("Encoder" of a VAE)

Ex 2: Let $X \sim U[0,1]$. Let $Y = g(X)$. Find $g(\cdot)$ such that $Y \sim f_d(y)$ i.e. Y follows a desired distribution. ("Decoder" of a VAE)

Ans 1: $P(Y \leq y) = P(F_X(X) \leq y)$

Case $0 \leq y \leq 1$ $= P(X \leq F_X^{-1}(y))$

$$= \int_{-\infty}^{F_X^{-1}(y)} f_X(x)\, dx$$

$$= F_X(F_X^{-1}(y)) - \overset{\to 0}{\cancel{F_X(-\infty)}}$$

$$F_Y(y) = y$$