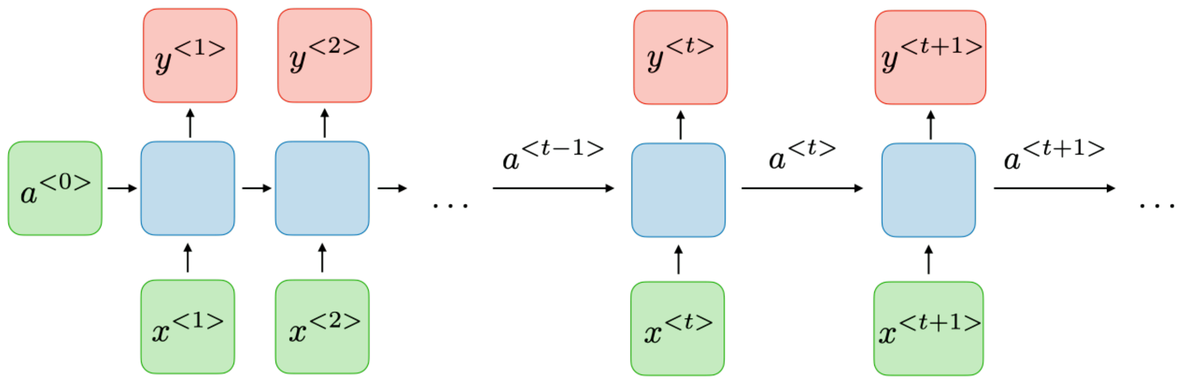
Alberto Andrés Valdés González.
Degree: Mathematical Engineer.
Work position: ML-Engineer.
Mail: anvaldes@uc.cl/alberto.valdes.gonzalez.96@gmail.com
Location: Santiago, Chile.

RNN

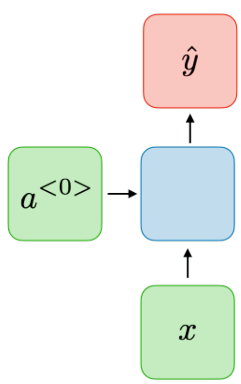


$$\boxed{a_{(t)} = g_1\left(W_{a,a} \cdot a_{(t-1)} + W_{a,x} \cdot x_{(t)} + b_a\right)}$$

$$\boxed{y_{(t)} = g_2\left(W_{y,a} \cdot a_{(t)} + b_y\right)}$$

Types of RNN

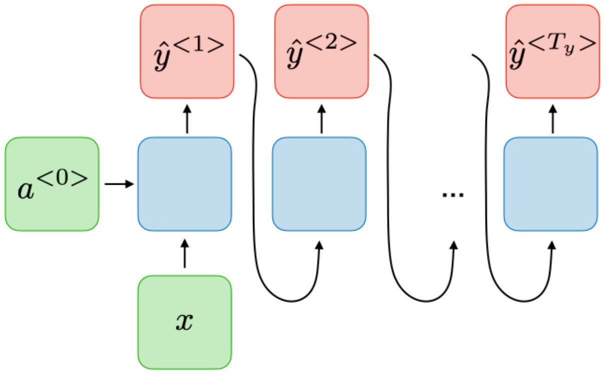
i. One to one $(T_x = 1, T_y = 1)$**:**



<u>Example:</u> Traditional neural network.

---

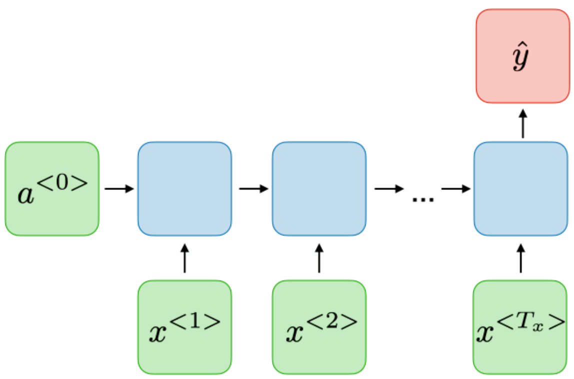
ii. One to many $(T_x = 1, T_y > 1)$**:**



Example: Music generation.

---

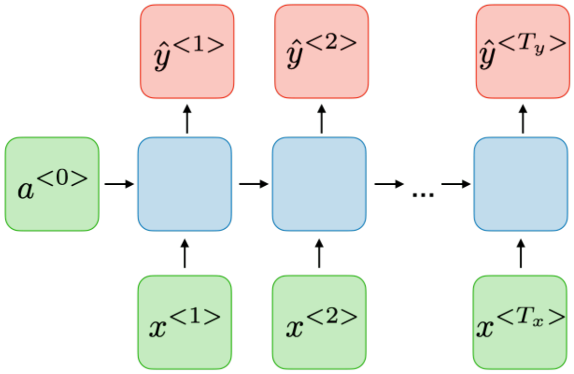
iii. Many to one $(T_x > 1, T_y = 1)$**:**



Example: Sentiment classification.

---

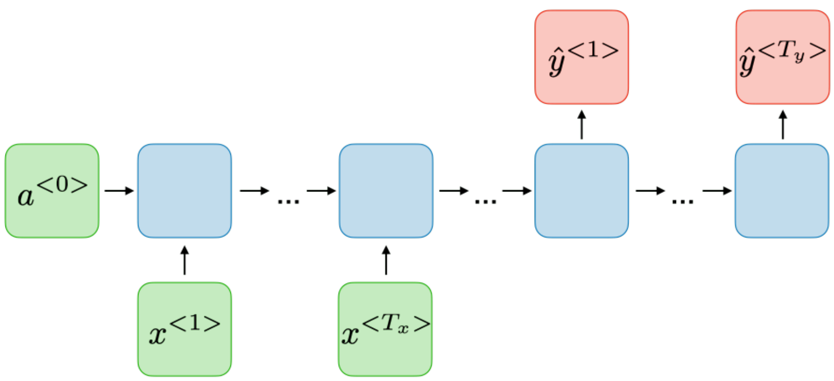
iv. Many to many $(T_x = T_y)$**:**



Example: Name entity recognition.

---

v. Many to many $(T_x \neq T_y)$**:**

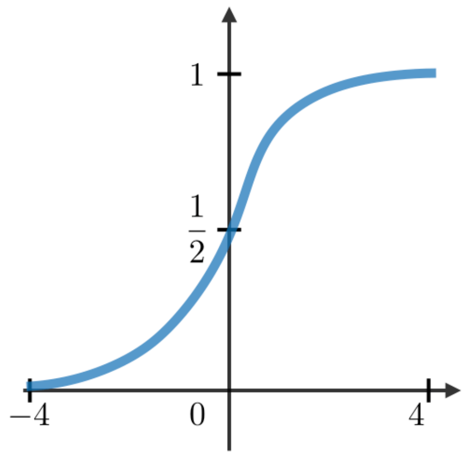


Example: Machine Translation.

Commonly used activation functions

i. Sigmoid:

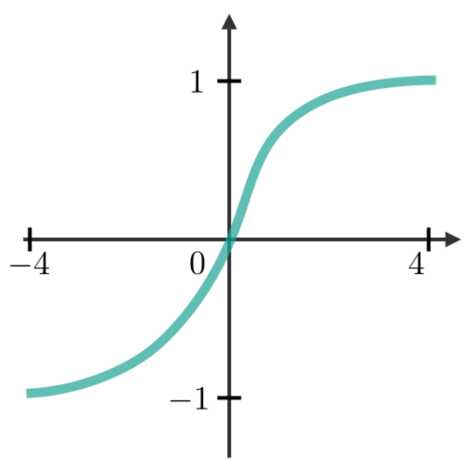
$$\boxed{g(z) = \frac{1}{1 + e^{-z}}}$$



---

ii. Tanh:

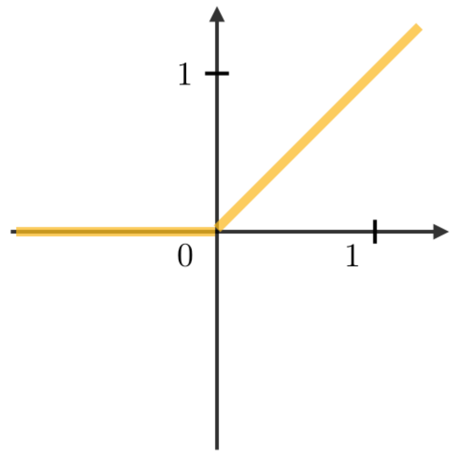
$$\boxed{g(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}}$$



---

iii. ReLU:

$$\boxed{g(z) = max(0, x)}$$

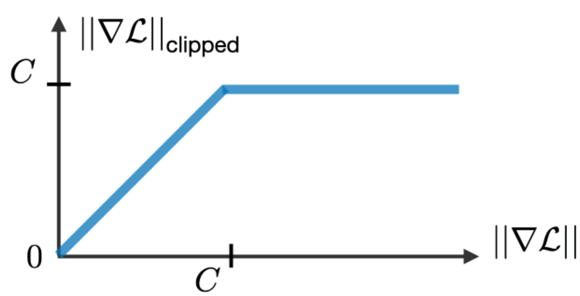


Vanishing/exploding gradient

The vanishing and exploding gradient phenomena are often encountered in the context of RNNs. The reason why they happen is that it is difficult to capture long term dependencies because of multiplicative gradient that can be exponentially decreasing/increasing with respect to the number of layers.

Gradient clipping

It is a technique used to cope with the exploding gradient problem sometimes encountered when performing backpropagation. By capping the maximum value for the gradient, this phenomenon is controlled in practice.



Types of gates

$$\boxed{\Gamma = \sigma\left(W \cdot x_{(t)} + U \cdot a_{(t-1)} + b\right)}$$

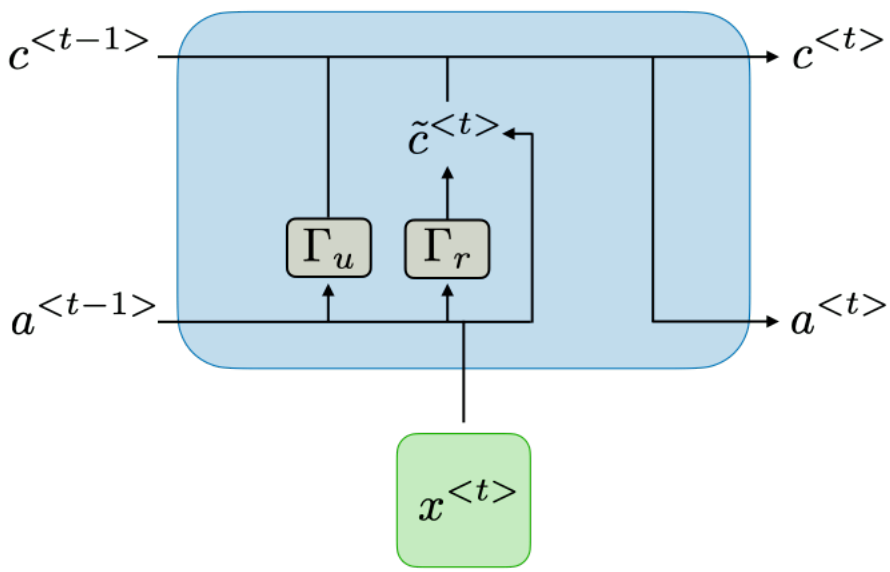
GRU

$$\boxed{\tilde{c}_{(t)} = tanh\left(W_c \cdot [\Gamma_r \star a_{(t-1)}, x_{(t)}] + b_c\right)}$$

$$\boxed{c_{(t)} = \Gamma_u \star \tilde{c}_{(t)} + (1 - \Gamma_u) \star c_{(t-1)}}$$

$$\boxed{a_{(t)} = c_{(t)}}$$

$\star$ is element-wise multiplication.



LSTM

$$\boxed{\tilde{c}_{(t)} = tanh\left(W_c \cdot [\Gamma_r \star a_{(t-1)}, x_{(t)}] + b_c\right)}$$

$$\boxed{c_{(t)} = \Gamma_u \star \tilde{c}_{(t)} + \Gamma_f \star c_{(t-1)}}$$

$$\boxed{a_{(t)} = \Gamma_o \star c_{(t)}}$$

$\star$ is element-wise multiplication.