Questions 1

Question 1-a

$$L = \frac{1}{2}(h_T - y)^2$$

$$\frac{\partial L}{\partial h_T} = (h_T - y)$$

Question 1-b

$$\frac{\partial L}{\partial h_t} = \frac{\partial L}{\partial h_t} \frac{\partial h_t}{\partial h_{t-1}}$$

$$\frac{\partial h_t}{\partial h_{t-1}} = a$$

$$\frac{\partial L}{\partial h_t} = \frac{\partial L}{\partial h_{t-1}} a$$

$$\frac{\partial L}{\partial h_t} = \frac{\partial L}{\partial h_T} a^{T-t}$$

Question 1-c

$$\frac{\partial L}{\partial a} = \frac{\partial L}{\partial h_T} \cdot \frac{\partial h_t}{\partial a}$$

$$\frac{\partial h_t}{\partial a} = h_{t-1}$$

$$\frac{\partial L}{\partial a} = \sum_{t=1}^{T} \frac{\partial L}{\partial h_t} h_{t-1}$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial h_T} \cdot \frac{\partial h_t}{\partial b}$$

$$\frac{\partial h_t}{\partial b} = x_t$$

$$\frac{\partial L}{\partial b} = \sum_{t=1}^{T} \frac{\partial L}{\partial h_t} x_t$$

Question 1-d

If a is less than one as $T - t$ becomes large a^{T-t} becomes 0

If a is greater than one as $T - t$ becomes large a^{T-t} becomes large

Question 1-e

If the average of A or B are much less than 1 it becomes 0 as T is much larger than t

If the average of A or B are much more than 1 it becomes infiity as T is much larger than t

Question 1-f

- Gradient Clipping
- Weight Regularization