Q.1. Fashion MNST is a dataset consist of 28 x 28 pixel grayscale images of 70,000 fashion products. It has 10 categories with 7000 images per category. The training set has 60,000 images and the test set has 10,000 images.

Normally Convolutional Neural Network (CNN) is considered to be the first choice on image classification, because it learns to recognize components of images firstly and then learn the larger structure, which can meet the needs of image classification. Unlikely to CNN, RNN learns to recognize image features across time.

Traditional RNN is sometimes unstable in practice, especially when backpropagating gradients through long time windows, which may cause gradient explosion or vanishing. Therefore we can try LSTM model to improve the model.

Unlike feedforward neural networks, RNNs can use their internal state (memory) to process sequences of i/p. This makes them applicables to tasks such as unsegmented, connected handwritting recognition or speech recognition or speech recognition. In other neural networks, all the i/ps are independent of each other. But in RNN, all the i/ps are related to each other.

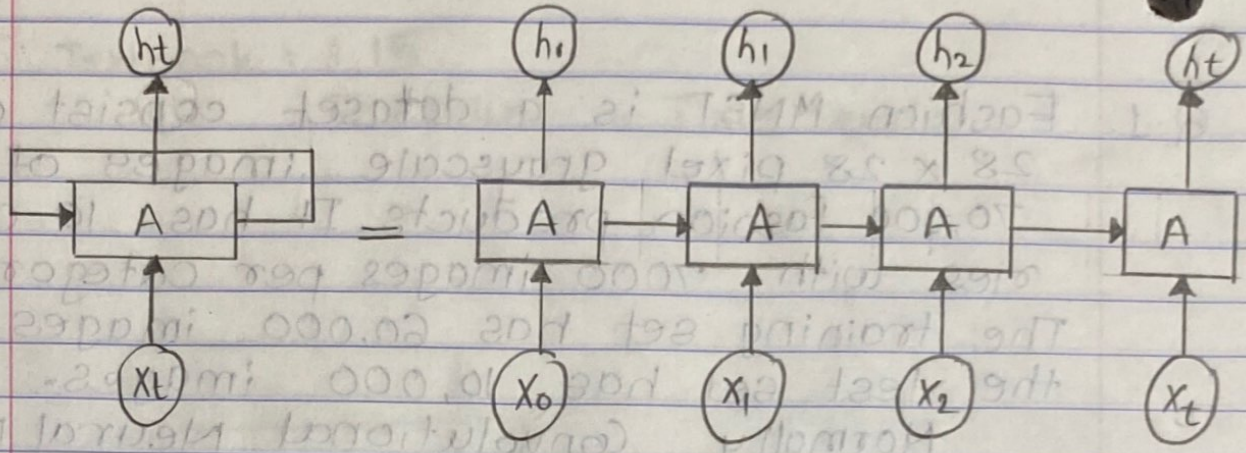

fig. unrolled recurrent neural Network.

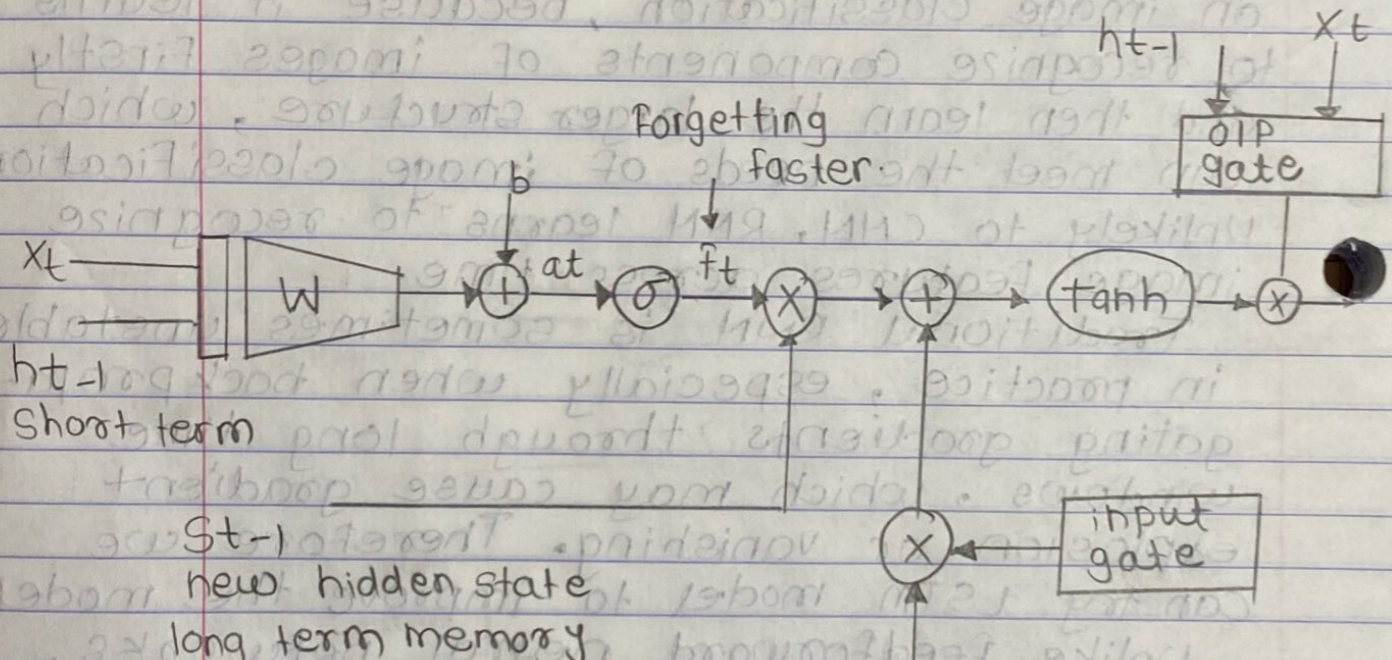

fig. LSTM

Long short term memory (LSTM) networks are special kind of RNN, capable of learning long term dependencies.
LSTM are explicitly designed to avoid the long term dependency problem.
LSTM does have the ability to remove or add information to the cell state, regulated by structures called gates.
Gates are a way to optionally let information through. They are composed out of a sigmoid neural net layer and pointwise multiplication operation.
Sigmoid layer outputs number between zero & one, describing how much of each component should be let through. A value of zero means "let nothing through" while a value of one means "let everything through".
Tanh function gives weightage to the values which are passed deciding their level of importance ranging from -1 to 1 and multiplied with output of sigmoid.
In the Fashion MNST example the first LSTM layer encodes every column of pixel of shape (28,28,1) to a column vector (128). The second LSTM layer encodes then these 28 column vectors of shape (28,128) to image vector representing the whole image.
The dense layer is added for prediction.