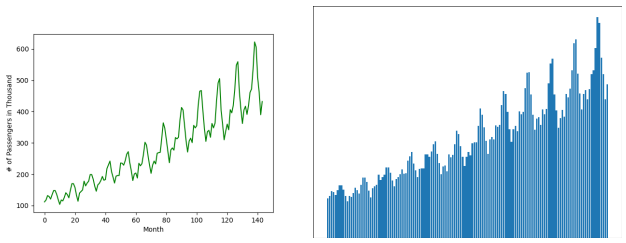
Data Analytics
HW 5

Ahmed Shahabaz

QUESTION 1(A)



(a) Line graph (left) and Bar graph (right)

Fig. 1: Pattern in the Airline Passenger data

QUESTION 1(B)

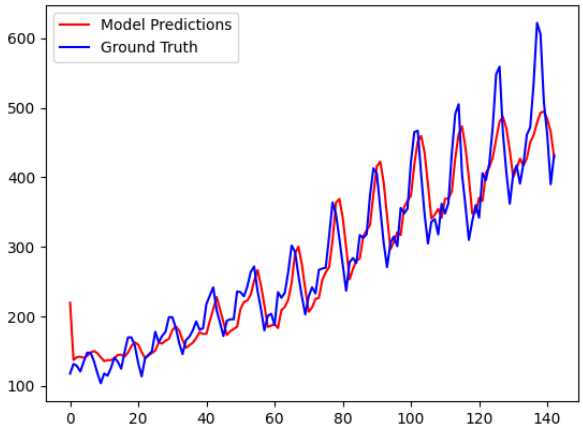
The long term pattern that we can see in the data is that the number of Passenger keeps increasing every year. The increase of the number of passengers in 10-12 months is kind of exponential.

But as for the short term patterns we can find multiple Gaussian distributions. The spread of those smaller Gaussian distributions are for 12 months. So each year the number of passenger is highest in the middle of the year and by the end of the year it is on the fall. But the number of passenger by the end of the year is always higher than that of the beginning of the year.

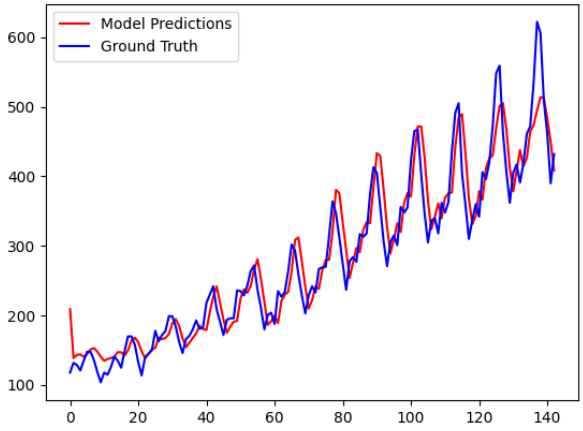
QUESTION 1(C) & 1(D)

Figure [2] shows the model prediction vs ground truth. Our model was used to make 143 predictions, where as there were 144 data samples. As the LSTM model makes prediction about the next month given the input of current month and previous two months, so there was no prediction made for the first month. Predictions were made from the second month until the last month. So for making the first prediction (second month), three inputs that were given/fed to the LSTM model are : $0, 0, first\_months\_value$.

For initializing the hidden state (h_0) and the cell state (c_0) we used the pytorch default initialization which is zero.



(a) c_0 and h_0 initialized with zero (default initialization)



(b) c_0 and h_0 initialized with a dense layer

Fig. 2: Prediction vs Ground Truth