

The NTU logo is a large, light gray watermark in the background. It features a stylized 'NTU' acronym and a circular emblem containing a traditional Chinese building with a curved roof.

Working with Time Series Data

MUKESH KUMAR

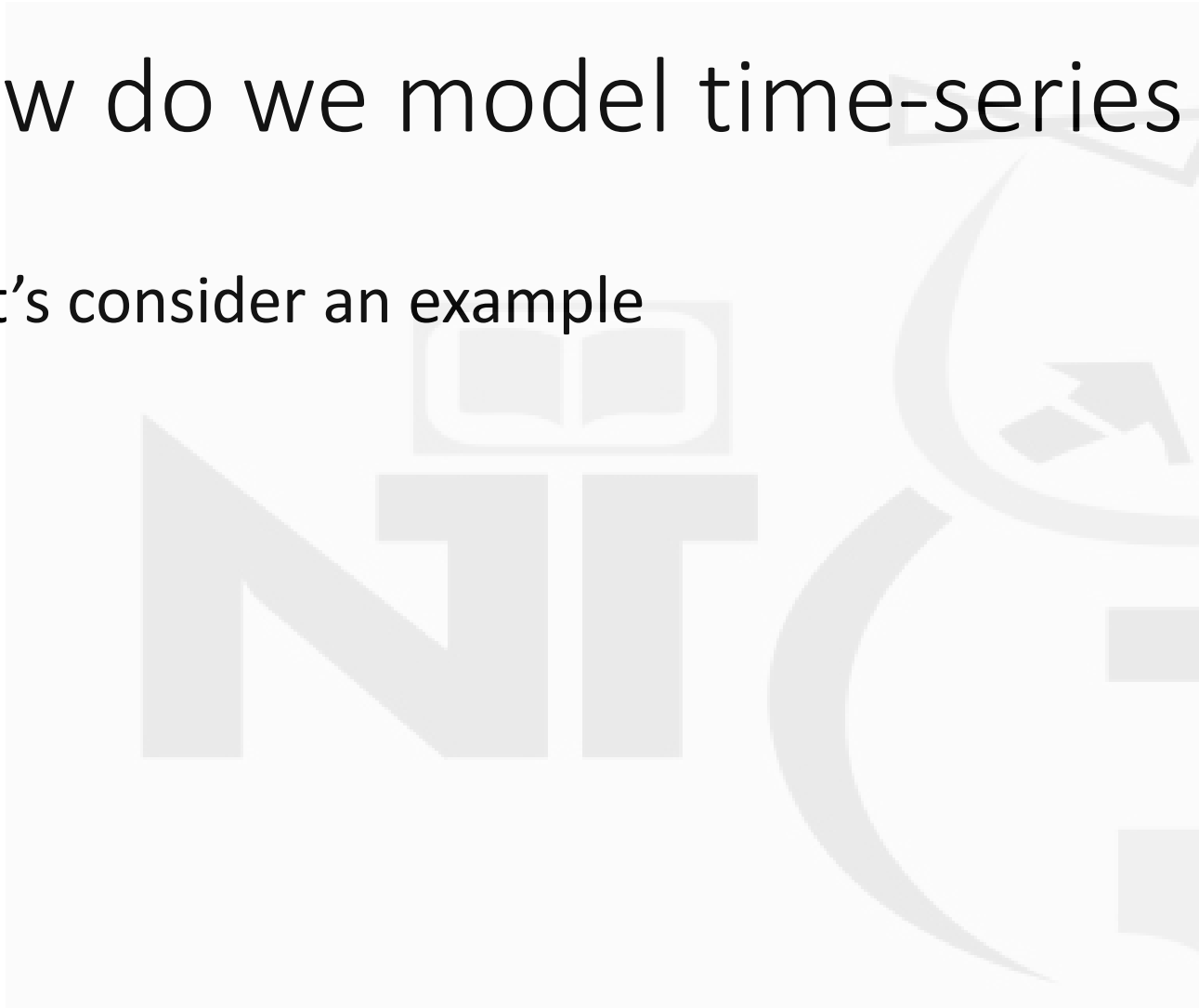
Examples of Time-Series data

1. Stock Prices at the end of each **day**
2. Temperature reading taken **hourly**
3. Number of patients in a hospital every **day**
4. Electricity demand for city on **daily** basis
5. No of births in a community every **year**
6. Organization's revenue at the end of each **month**

- Time-Series data
 - Is sequential data where data points are ordered in time

How do we model time-series data?

- Let's consider an example



Patients visiting a hospital on daily basis

Date	# of Patients
Jan-01	237
Jan-02	250
Jan-03	251
Jan-04	242
Jan-05	179
Jan-06	193
Jan-07	230
Jan-08	224

GOAL:

- Forecast the # of Patients that will visit hospital based on previous days visits

How do we model such data?

Patients visiting a hospital on daily basis

Date	# of Patients
Jan-01	237
Jan-02	250
Jan-03	251
Jan-04	242
Jan-05	179
Jan-06	193
Jan-07	230
Jan-08	224

Time-series data is sequential data and LSTM can be useful to understand the sequence!

To model the data, we need two things:

1. Input features or sequence (x)
2. Ground truth or Actual label (y)

Patients visiting a hospital on daily basis

Date	# of Patients
Jan-01	237
Jan-02	250
Jan-03	251
Jan-04	242
Jan-05	179
Jan-06	193
Jan-07	230
Jan-08	224

Preparing x and y for time-series

Let's assume # of patients on a day depend on # of patients in last 3 days

Record#	Input Sequence (x)	Label (Y)
1	237, 250, 251	242
2	250, 251, 242	179
3	251, 242, 179	193
4	242, 179, 193	230
5	179, 193, 230	224

Once we have the input sequence and actual label, it becomes fairly straightforward to model the data using LSTM.

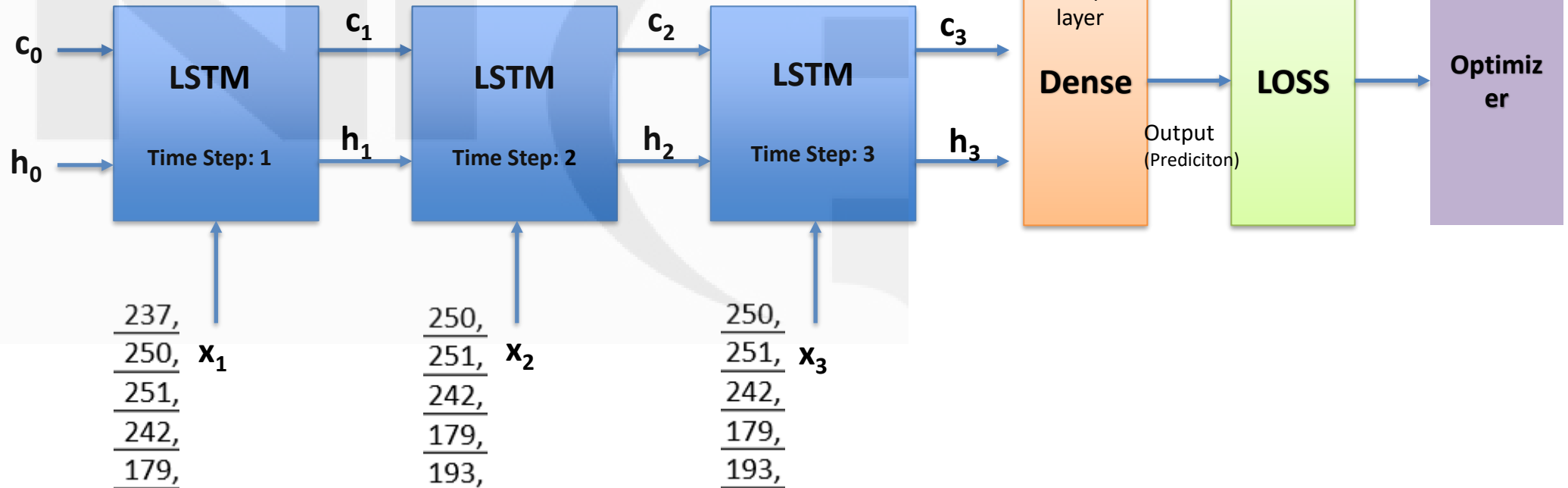
- Total records- window size = final record in new x and y



How do we model this data with LSTM?

Modeling time-series data with LSTM

Input Sequence (x)
237, 250, 251
250, 251, 242
251, 242, 179
242, 179, 193
179, 193, 230



The background of the slide features a large, light gray watermark of the Nanyang Technological University (NTU) logo. The logo consists of the letters 'NTU' in a stylized font, with a square icon above the 'T' containing an open book. To the right of the letters is a circular emblem with a crescent moon and a five-pointed star.

Exercise

Weather forecast using LSTM



What if we had both types of data

1. Time-Series data
2. Non time-series data

Date	# of Patients	Holiday?
Jan-01	237	No
Jan-02	250	No
Jan-03	251	No
Jan-04	342	Yes
Jan-05	352	Yes
Jan-06	193	No
Jan-07	230	No
Jan-08	224	No

Goal

Forecast **# of patients** that will visit hospital based on...

1. Number of patients visited on each of last 3 days (**time-series**)
2. Whether a day was holiday or not? (**Non time-series**)

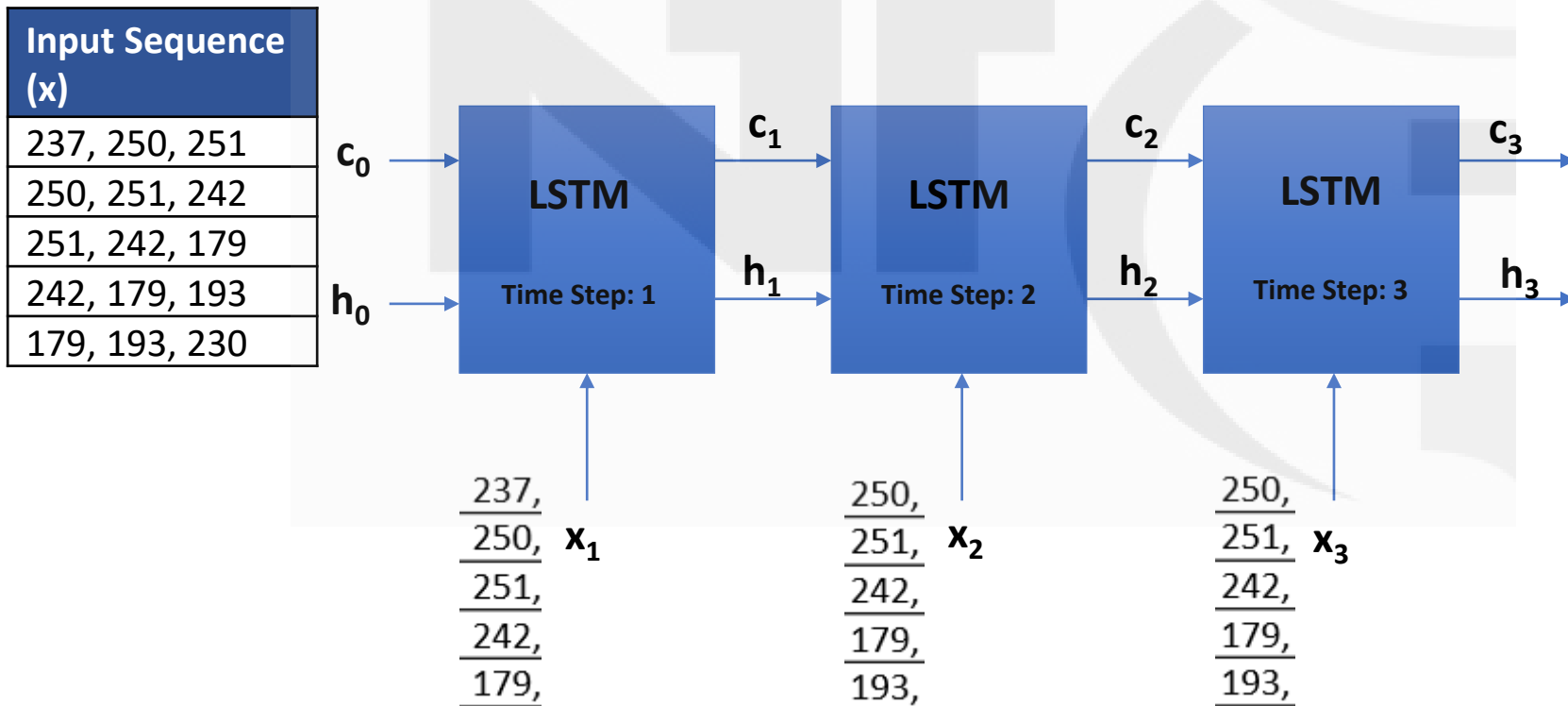
How do we model such data?

Date	# of Patients	Holiday?
Jan-01	237	No
Jan-02	250	No
Jan-03	251	No
Jan-04	342	Yes
Jan-05	352	Yes
Jan-06	193	No
Jan-07	230	No
Jan-08	224	No

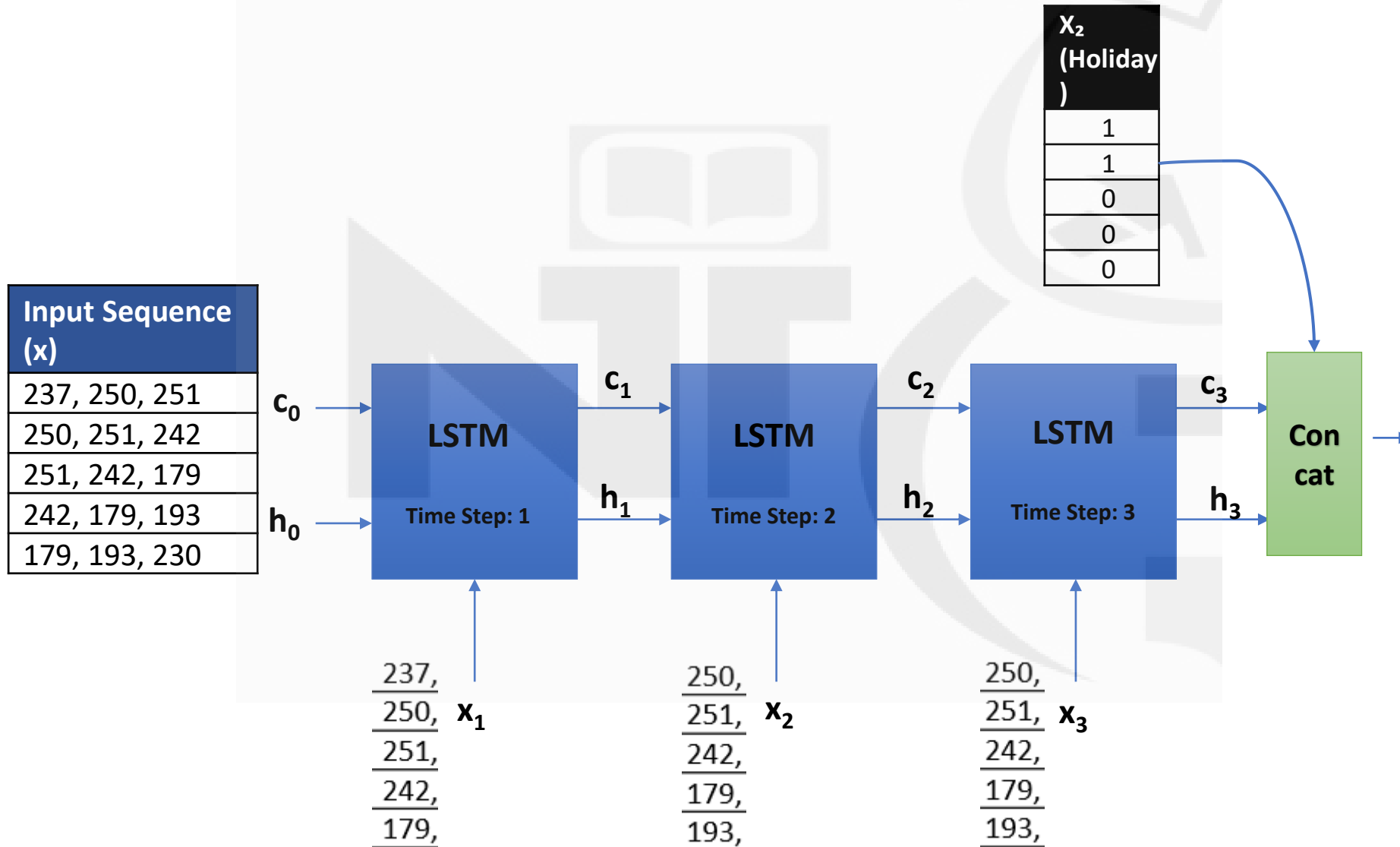
Preparing x and y

#	Input Features		Label (Y)
	X ₁ (patients in last 3 days)	X ₂ (Holiday)	
1	237, 250, 251	1	342
2	250, 251, 342	1	352
3	251, 342, 352	0	193
4	342, 352, 193	0	230
5	352, 193, 230	0	224

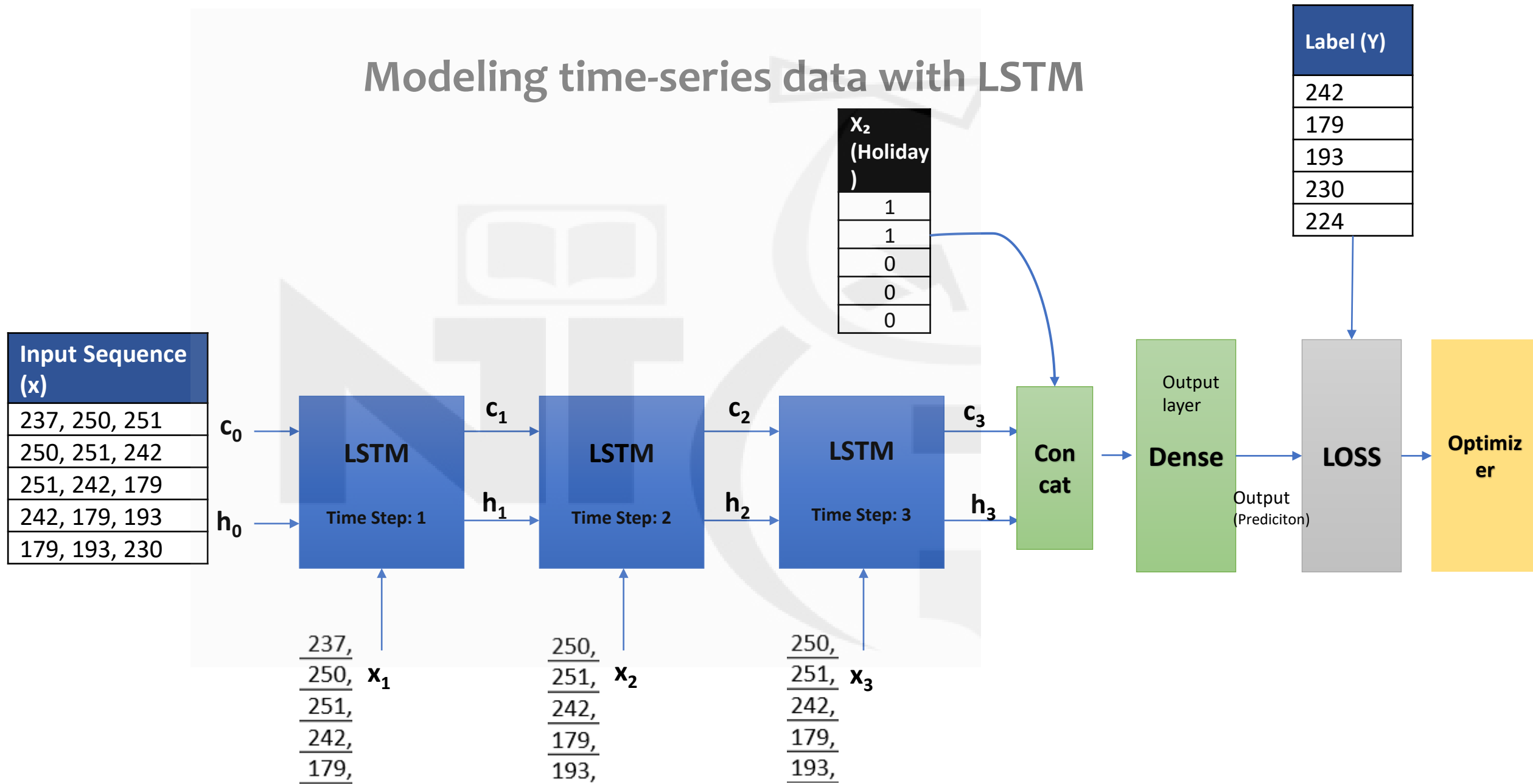
Modeling time-series data with LSTM



Modeling time-series data with LSTM



Modeling time-series data with LSTM





What if we have more than one time-series data?

Patients visiting a hospital on daily basis

Date	# of patients Tested for Covid in city	# of patients found + for covid	# of Patients visiting hospital
36892	3000	1200	237
37257	2900	1250	250
37622	2500	995	251
37987	3000	778	342
38353	3500	1020	352
38718	3200	950	193
39083	1800	800	230
39448	2500	775	224

Patients visiting a hospital on daily basis

Date	# of patients Tested for Covid in city	# of patients found + for covid	# of Patients visiting hospital
36892	3000	1200	237
37257	2900	1250	250
37622	2500	995	251
37987	3000	778	342
38353	3500	1020	352
38718	3200	950	193
39083	1800	800	230
39448	2500	775	224

Goal

Forecast # of patients that will visit hospital based...

1. Number of patients tested for last 3 days (*time-series*)
2. Number of patients found + for last 4 days (*time-series*)

Patients visiting a hospital on daily basis

Date	# of patients Tested for Covid in city	# of patients found + for covid	# of Patients visiting hospital
36892	3000	1200	237
37257	2900	1250	250
37622	2500	995	251
37987	3000	778	342
38353	3500	1020	352
38718	3200	950	193
39083	1800	800	230
39448	2500	775	224

#	Input Features			Label (Y)
	X_1 (patients tested in last 3 days)	X_2 (patients + in last 4 days)		
1	2900, 2500, 3000	1200, 1250, 995, 778		1310
2	2500, 3000, 3500	1250, 995, 778, 1020		1193
3	3000, 3500, 3200	995, 778, 1020, 950		1230
4	3500, 3200, 1800	778, 1020, 950, 800		1224

X_2 (patients + in last 4 days)
1200, 1250, 995, 778
1250, 995, 778, 1020
995, 778, 1020, 950
778, 1020, 950, 800

X_1 (patients tested in last 3 days)
2900, 2500, 3000
2500, 3000, 3500
3000, 3500, 3200
3500, 3200, 1800

