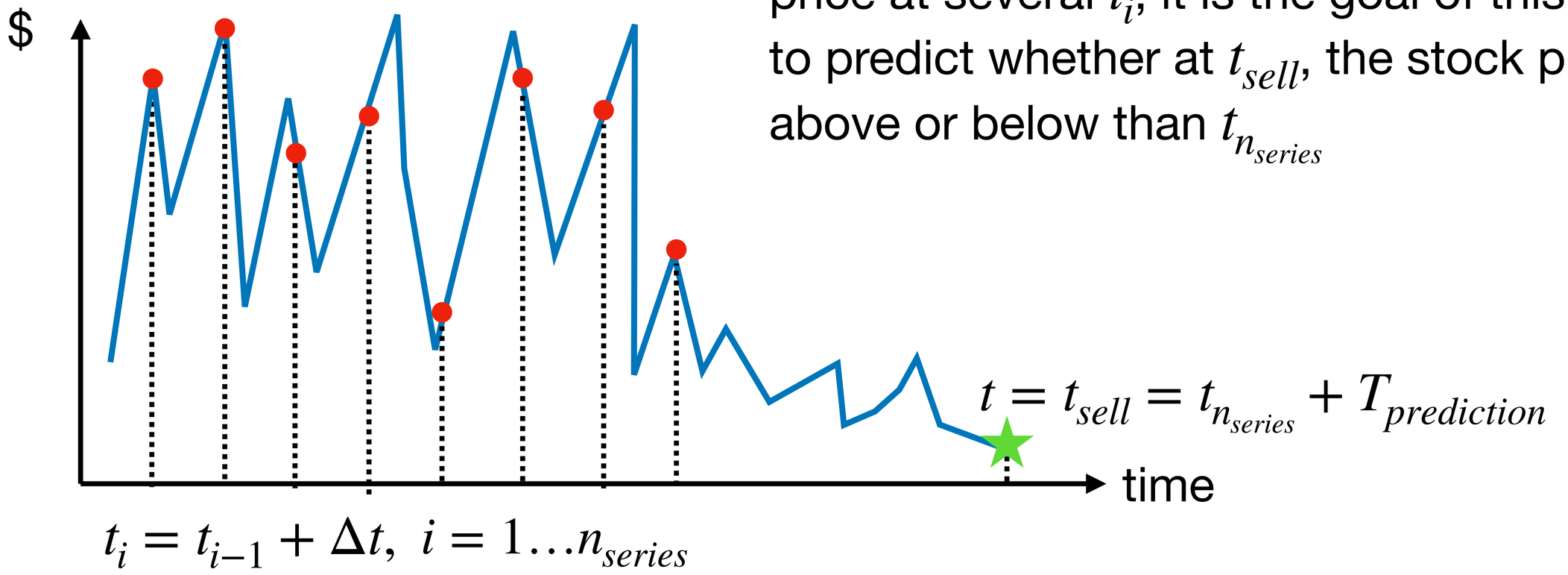
StocConv

Alvaro Tomas Gil - September 2020

Objectives/Goals

- Can Convolutional NN be applied to predict whether a stock value is to rise or fall in the future?
- How can the time series of a particular stock value be processed by a Convolutional NN?
- Can several values be taken into account for the prediction of a single value?

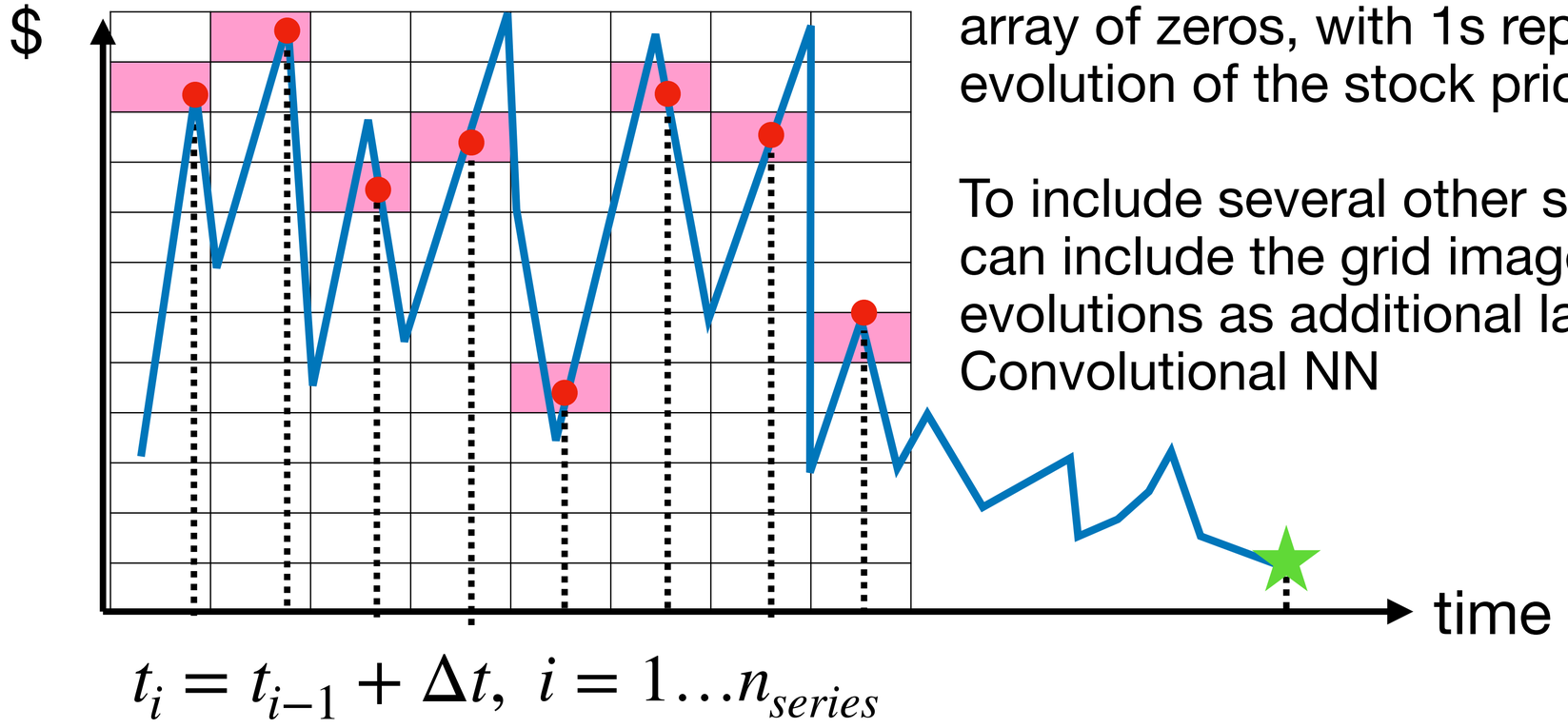
The Essence of the Prediction



Given a set of values of a particular stock price at several t_i, it is the goal of this model to predict whether at t_{sell}, the stock price is above or below than $t_{n_{series}}$

How to process the time series?

For the input



$$t_i = t_{i-1} + \Delta t,\ i = 1 \ldots n_{series}$$

The time series can be converted into a grid image. More simply, instead of a vector of stock prices with the same length as the time series under study, define a 2D array of zeros, with 1s representing the line tracing the evolution of the stock price.

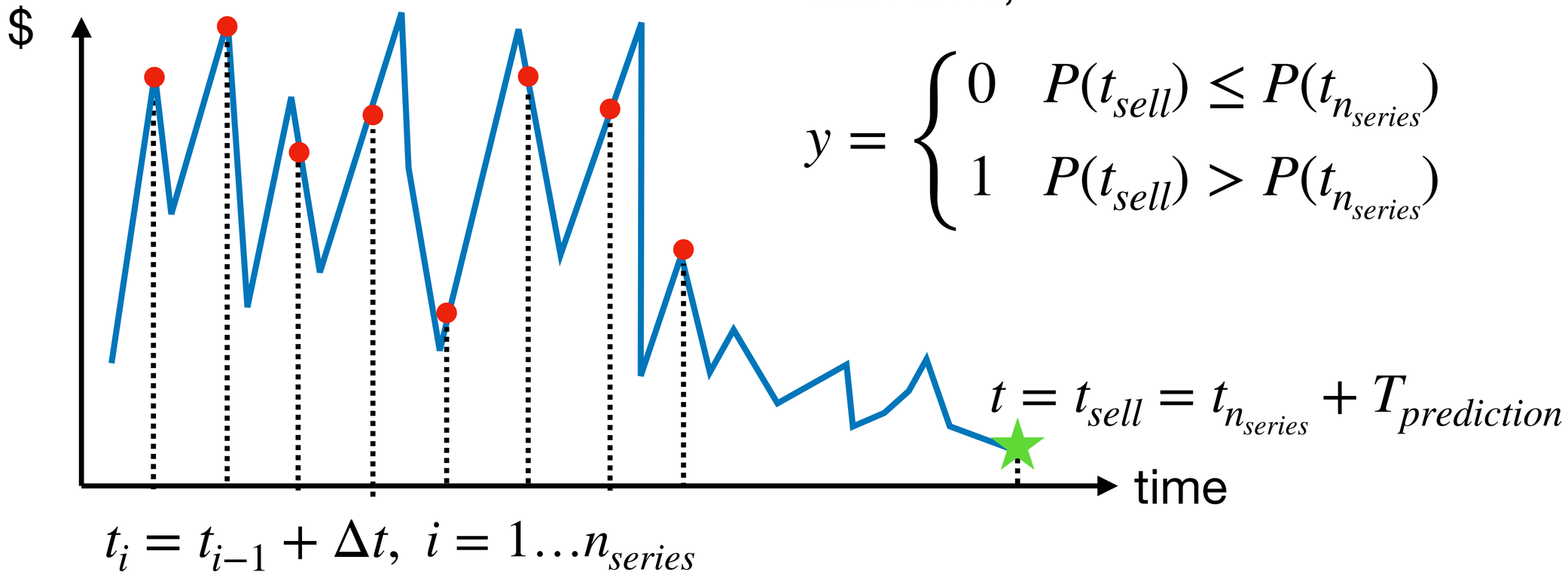
To include several other stocks into the prediction, one can include the grid images of these other stock price evolutions as additional layers to be inputted to the Convolutional NN

How to process the time series?

For the output

If y is the output value/label and $P(t)$ represents the evolution of the stock price with time,

$$y = \begin{cases} 0 & P(t_{sell}) \leq P(t_{n_{series}}) \\ 1 & P(t_{sell}) > P(t_{n_{series}}) \end{cases}$$

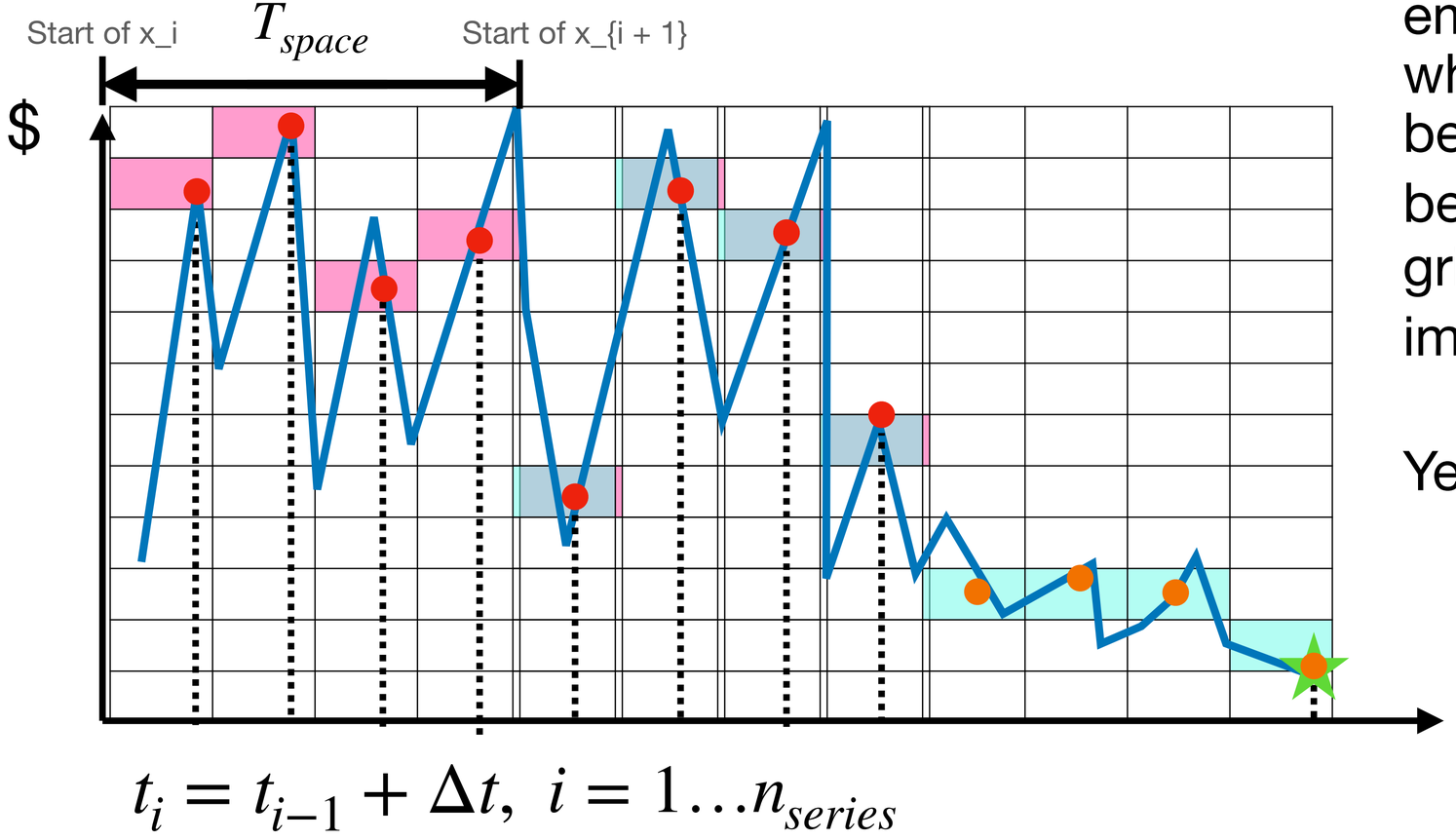


$$t = t_{sell} = t_{n_{series}} + T_{prediction}$$

$$t_i = t_{i-1} + \Delta t, \; i = 1 \ldots n_{series}$$

How to process the time series?

Spacing between data samples



$$t_i = t_{i-1} + \Delta t,\ i = 1 \ldots n_{series}$$

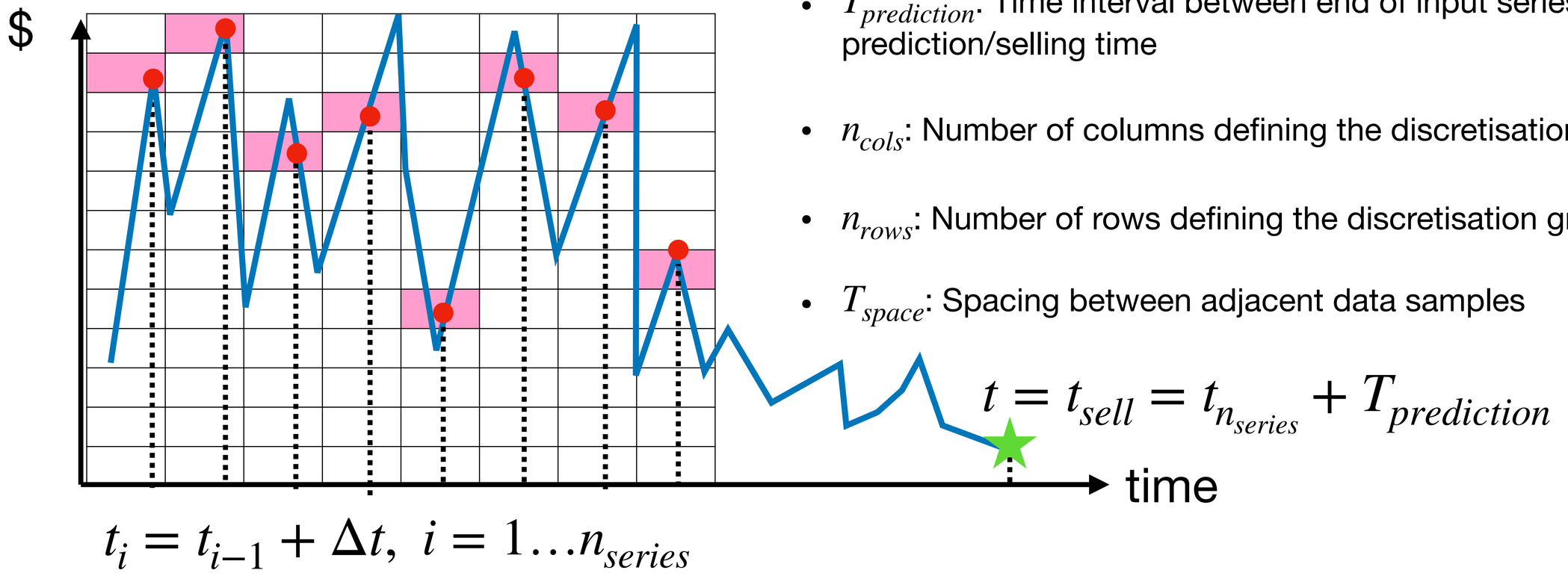
Since the same global time series of a single stock value is employed for the whole dataset, what should the temporal spacing between adjacent data samples be? In other words if x_i is an input grid image, when should the grid image x_{i+1} begin?

Yet another parameter: T_{space}

How to process the time series?

Resulting parameters

- Δt: Time interval between input prices
- n_{series}: Number of time instants in the input series
- $T_{prediction}$: Time interval between end of input series and prediction/selling time
- n_{cols}: Number of columns defining the discretisation grid
- n_{rows}: Number of rows defining the discretisation grid
- T_{space}: Spacing between adjacent data samples



$$t_i = t_{i-1} + \Delta t,\ i = 1 \ldots n_{series}$$

$$t = t_{sell} = t_{n_{series}} + T_{prediction}$$

Initial Parameter Selection

Resulting parameters

- Δt: 1 day
- n_{series}: 20
- $T_{prediction}$: 10 days
- n_{cols}: 2 * n_series
- n_{rows}: 20, to be studied further
- T_{space}: 10 days

Processing the Time Series

Results for AAPL

- Δt: 1 day
- n_{series}: 20
- $T_{prediction}$: 10 days
- n_{cols}: 2 * n_series
- n_{rows}: 20, to be studied further
- T_{space}: 10 days