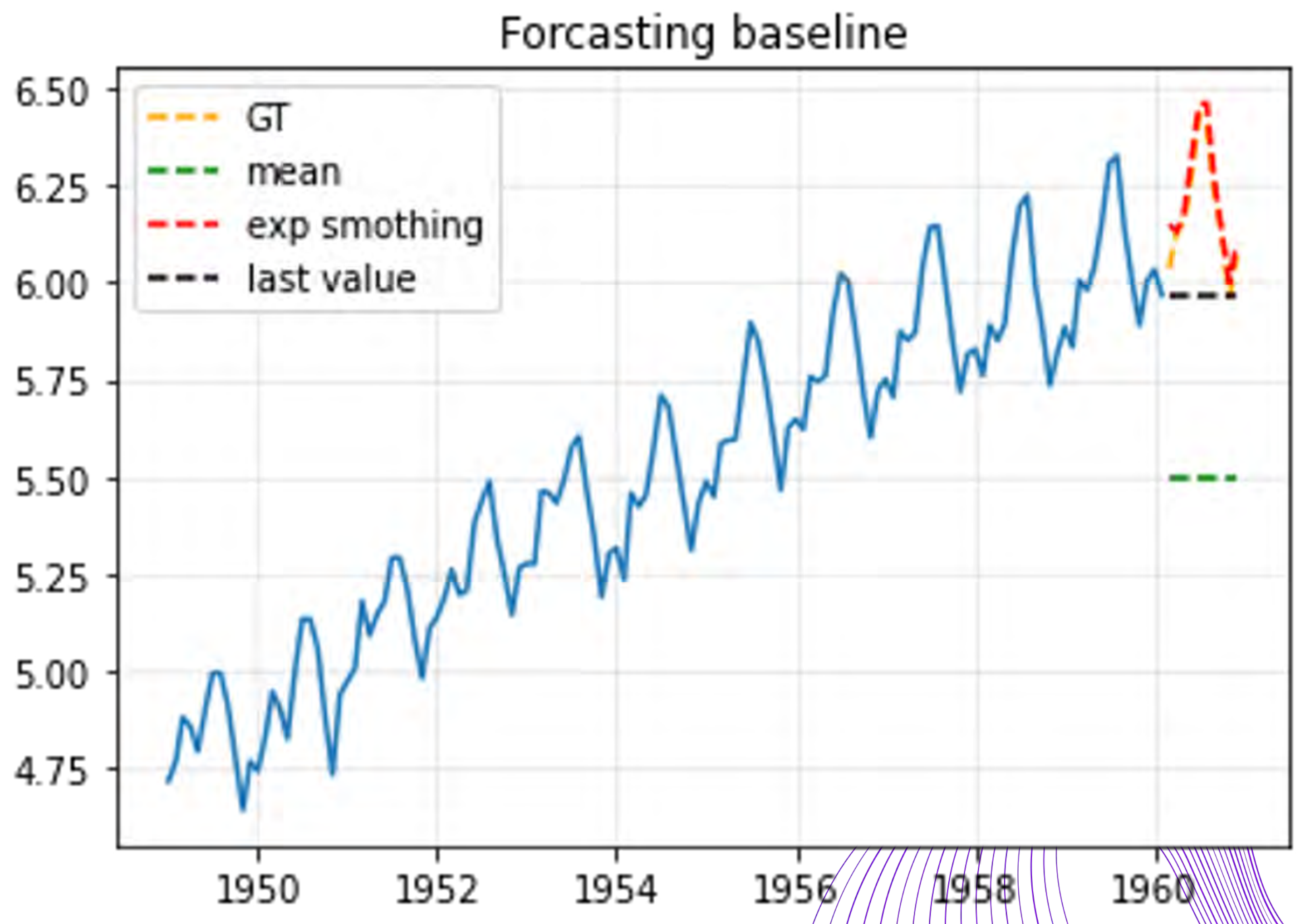
Modeling time series

Time series forecasting

- Fitting the model based on historical data and use them to predict the future
- Cover wide range of applications
  - Weather forecast
  - Stock price prediction
  - Inventory management

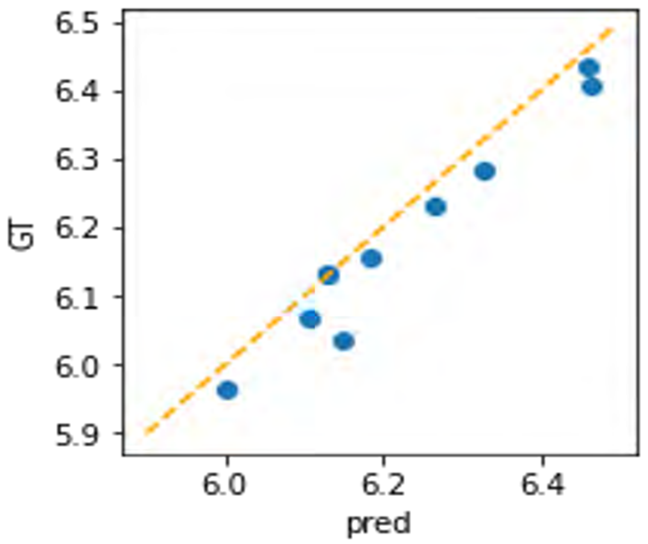
Forecasting baseline

- Simple average (quite strong baseline if stationary)
- Naïve approach (Last value)
- Moving average (smoothing)
    - Equal weighted
    - Exponential weighted



Model benchmarking

- There are many choices when creating a model.
- Which one is the best? Compare them using the same metric.
- Common metrics are:
  - Mean square error (MSE)
  - Mean absolute error (MAE)
  - Mean absolute percentage error (MAPE)
  - Correlation + R2



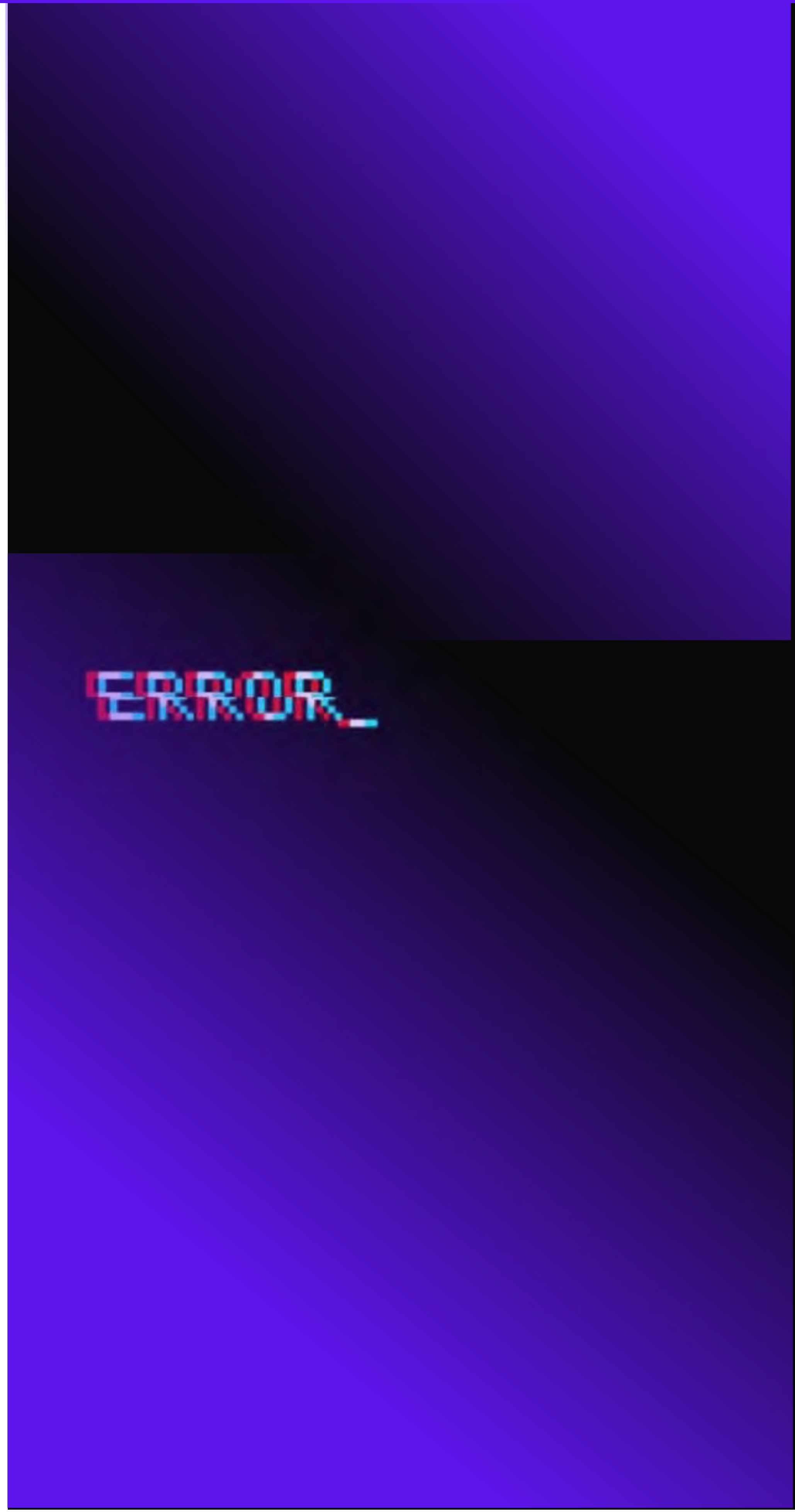
$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y_i})^2$$

$$\mathrm{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y_i}|$$

$$\mathrm{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y_i}}{y_i} \right|$$

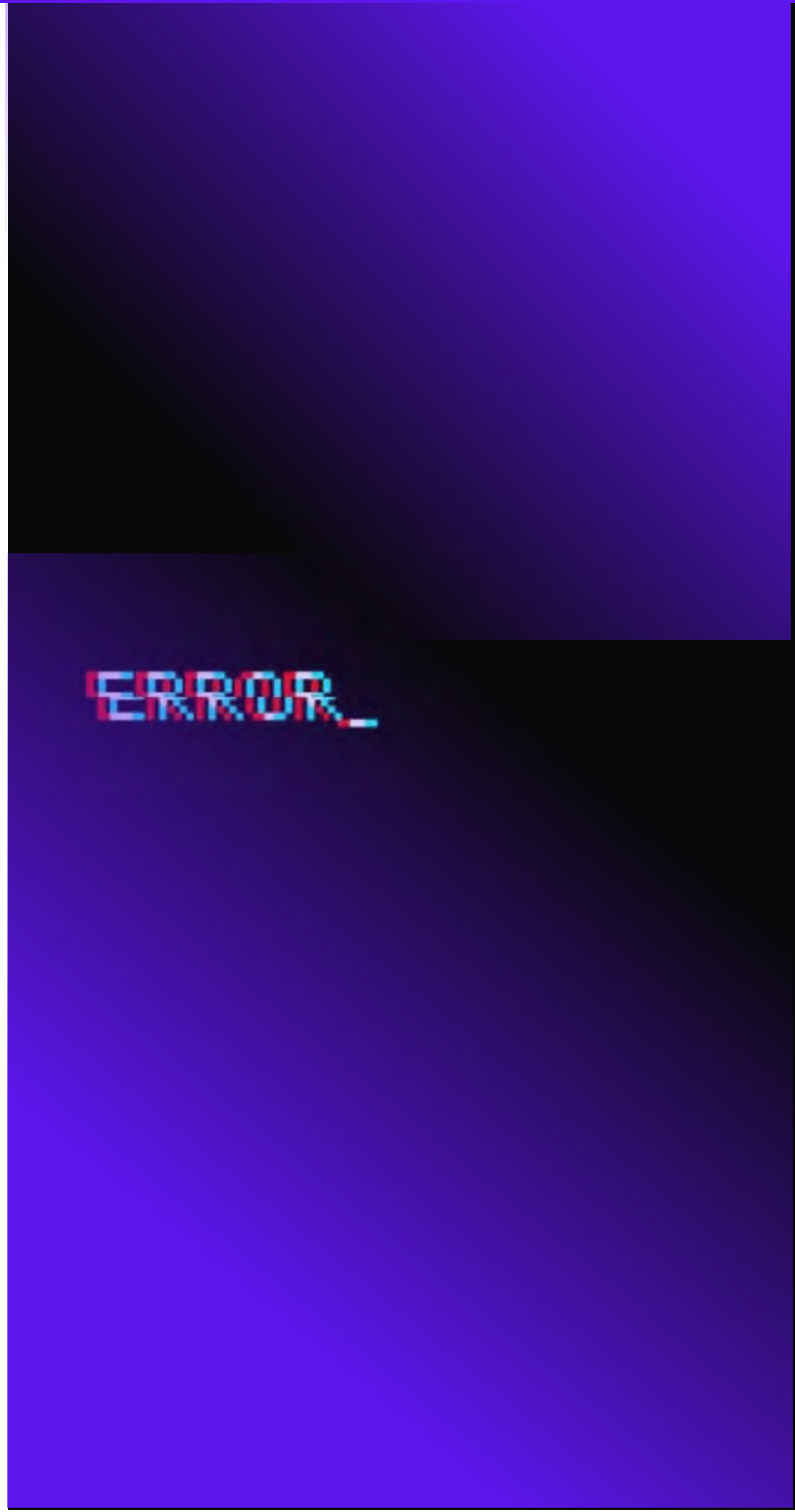
Common mistakes

- Split according to time, ~~random split,~~ sliding window



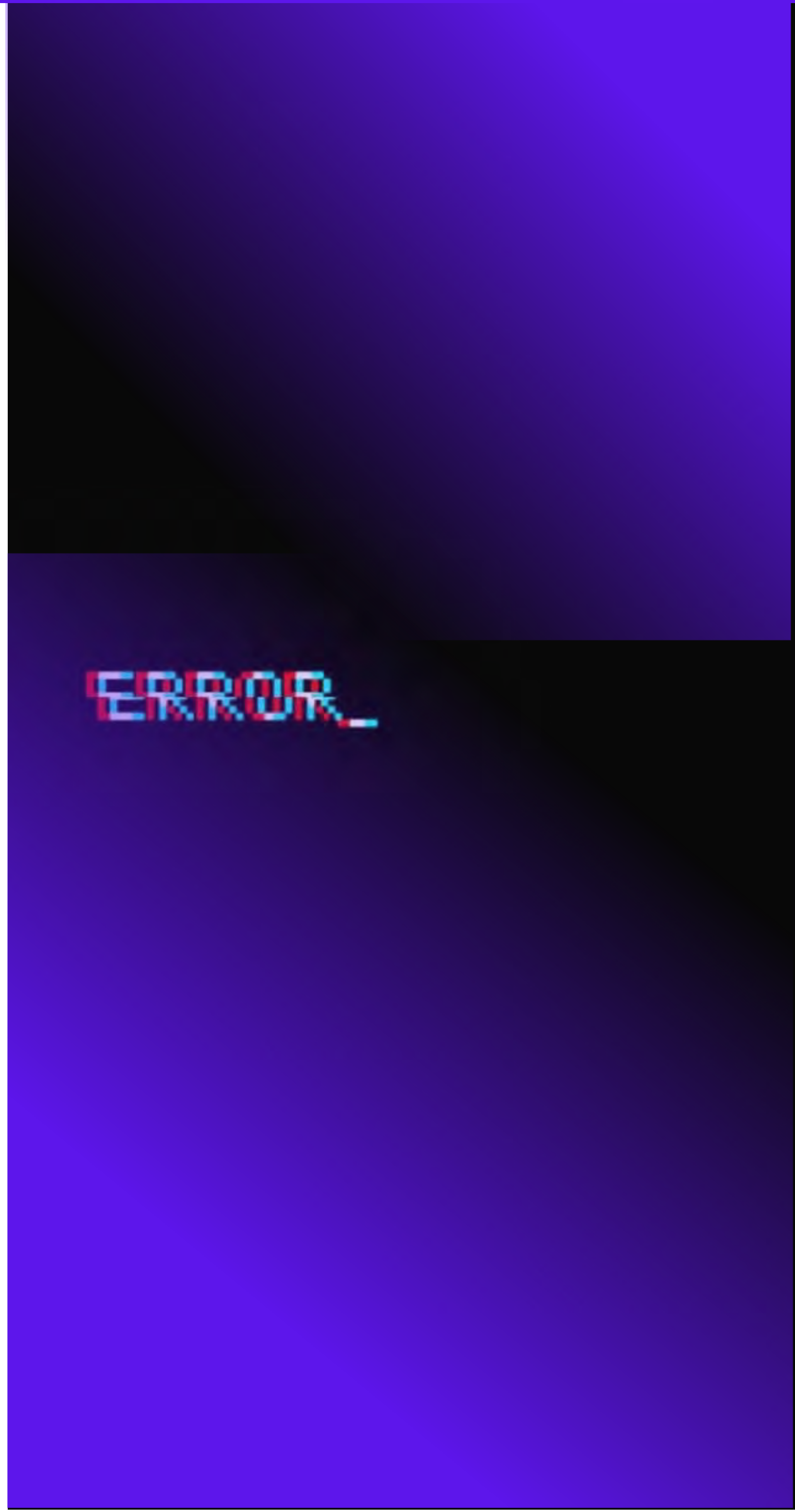
Common mistakes

- Split according to time, ~~random split,~~ sliding window
- The assumption is inappropriate for the data
  - Stationary
  - Linear



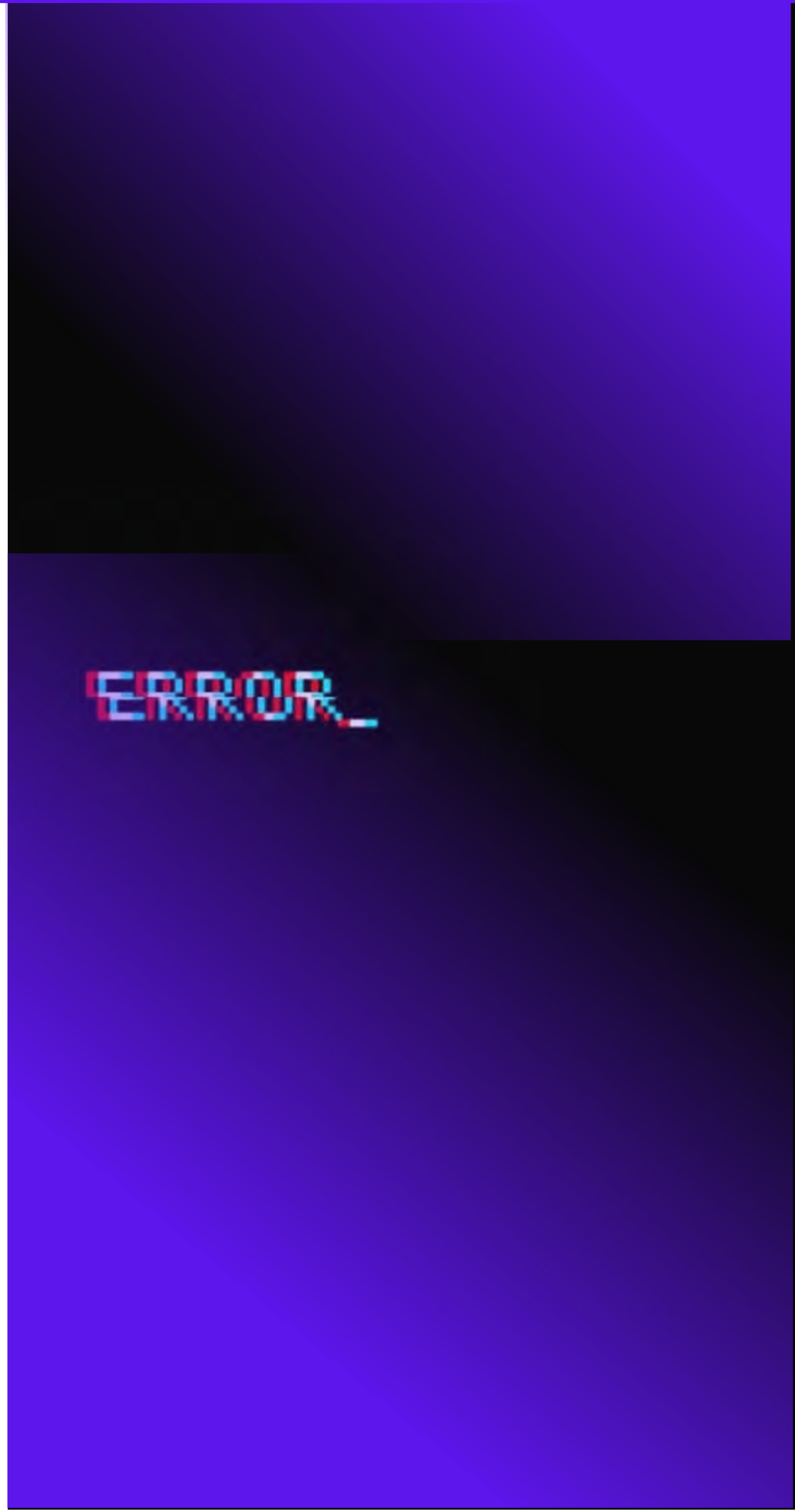
Common mistakes

- Split according to time, ~~random split,~~ sliding window
- The assumption is inappropriate for the data
  - Stationary
  - Linear
- The model can fail to capture some extra detail
  - Special holidays, 7/7, 11/11, 12/12
  - Stock price after dividend
  - New regulation
  - Model can only predict based on previous statistics



Common mistakes

- Split according to time, ~~random split,~~ sliding window
- The assumption is inappropriate for the data
  - Stationary
  - Linear
- The model can fail to capture some extra detail
  - Special holidays, 7/7, 11/11, 12/12
  - Stock price after dividend
  - New regulation
  - Model can only predict based on previous statistics
- Overfitting
  - Overly complex model
    - simple baselines can be very strong
  - Inappropriate evaluation (val /test does not represent the data)



ARMA

- Family of time series model widely used to describe time series data or predict future values
- Using information from autocorrelation
- Assumption:
  - Stationary
  - Univariate
- Consists of two main components: AR, MA

Autoregressive model (AR)

- Assumption : output variable depends on its own previous values
- Ex. Stock price prediction is based on the past stock price
- Observe last **p** values of the time series
- The model with the order of **p** (AR(**p**)) is modeled as:

$$y_t = w_1 y_{t-1} + w_2 y_{t-2} + \ldots + w_p y_{t-p} + c + \epsilon_t$$

  - w_i and c are model parameters
  - y_t is the value at the time t
  - ϵ_t is noise
- This is similar to regression in supervised learning with features as previous values

AR examples (p = 1, p = 2)

- We have the models $\hat{y}_t = w_1 y_{t-1}, \bar{y}_t = w_2 y_{t-1} + w_3 y_{t-2}$
  where w1 = 0.9, w2 = 0.8, w3 = 0.2
- y_t is the actual value at time t
- $\hat{y}_t, \bar{y}_t$ is the predicted value at time t

t	y_t	$\hat{y}_t$	$\bar{y}_t$
1	10	-	-
2	20	0.9×10	-
3	30	0.9×20	$0.8 \times 20 + 0.2 \times 10$
4	40	0.9×30	$0.8 \times 30 + 0.2 \times 20$
5	50	0.9×40	$0.8 \times 40 + 0.2 \times 30$

Moving average model (MA)

- Similar to AR model but based on past forecast error instead
- Intuition : The predicted value is obtained by adjusting the mean (mu) based on the past errors -> learn from past mistakes
- Observe last **q** errors
- The model with the order of **q** (MA(**q**)) is modeled as:

$$y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \ldots + \theta_q\epsilon_{t-q}$$

- θ is model parameter

MA example (q = 1)

- We have the model : $\hat{y}_t = \mu + \theta_1 \epsilon_{t-1}$ where mu = 5, theta1 = 0.5
- y_t is the ground truth value at time t
- $\hat{y}_t$ is the predicted value at time t

= 3 - 5

= 5 + 0.5 x (-2)

t	$\hat{y}_t$	ϵ_t	y_t
1	5	-2	3
2	4	2	6
3	5	0	5
4	5	-1	4
5	4.5	0	4.5

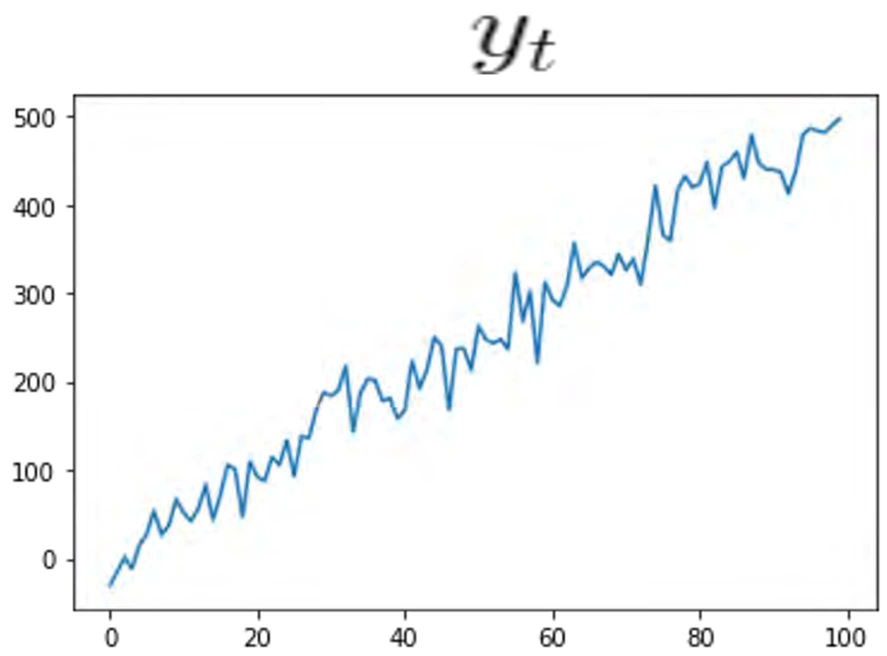
Autoregressive Moving Average model (ARMA)

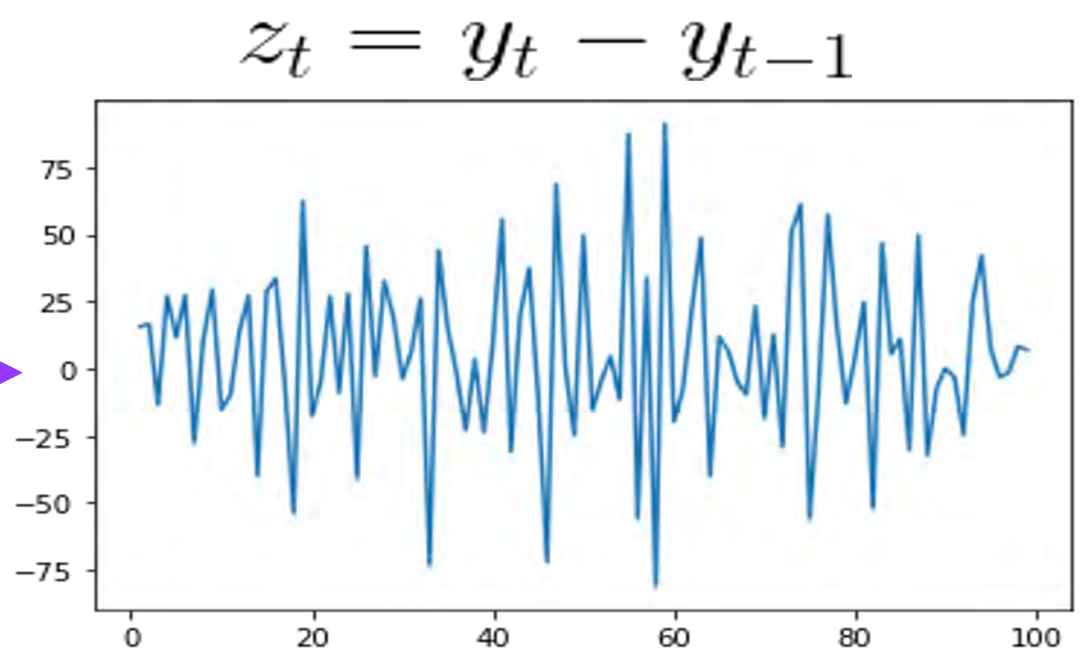
- ARMA(p, q) = AR(p) + MA(q)
- Accounting information of the past values and past mistakes

$$y_t = c + \epsilon_t + \sum_{i=1}^{p} w_i y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

Autoregressive Integrated Moving Average model (ARIMA)

- AR, MA, ARMA is inappropriate for non-stationarity time series
- ARIMA could relax the constant mean assumption
  - Add I component which adds previous terms (integration)
- Intuition : we predict the differenced series instead of the original series as it is more stationary





ARIMA (2)

- The original time series is differenced for **d** times to make the series stationary

$$d = 0, z_t = y_t$$
$$d = 1, z_t = y_t - y_{t-1}$$
$$d = 2, z_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$$

and so on

- Thus, ARIMA(p, d, q) is modeled as:

$$z_t = c + \epsilon_t + \sum_{i=1}^{p} w_i z_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

Notebook

Seasonal ARIMA (SARIMA)

- Further relax ARIMA by handling seasonal component
- Similar to ARIMA but have extra P, D, Q, m term
- At D = 1, $z_m = z_t - z_{t-m}$
- It considers m time step behind
- The SARIMA is modeled as SARIMA(p, d, q) (P, D, Q, m)

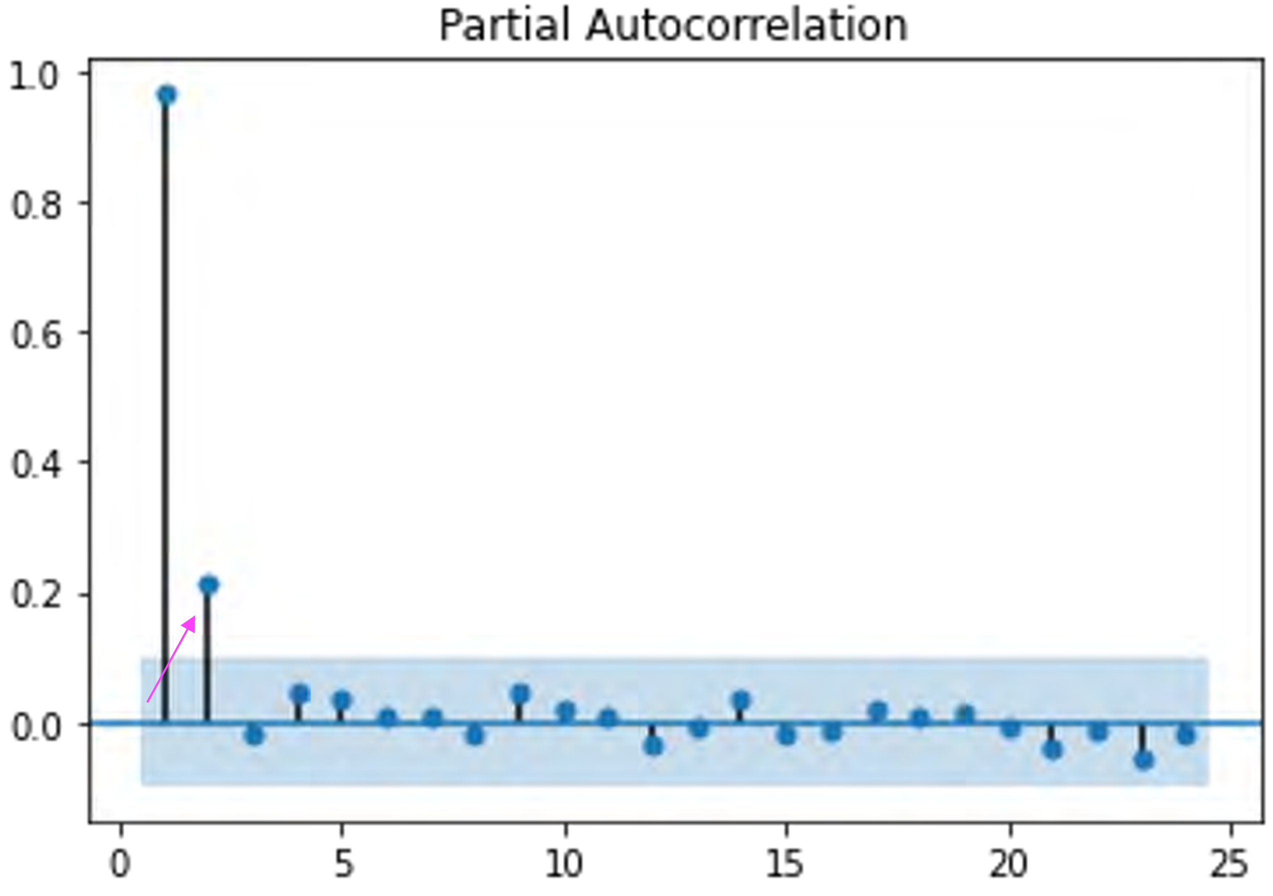
How to choose p, d, q

- Intuition
- p, q can be estimated by observing ACF and PACF plot
- p, q, d could be treated as parameters and tuned in a typical machine learning manner (train, val, test / cross validation)
- Akaike Information Criteria (AIC) and Bayesian Information Criterion (BIC)
  - Measure how well the model fits the data by discounting model size
  - Can be used to pick p and q

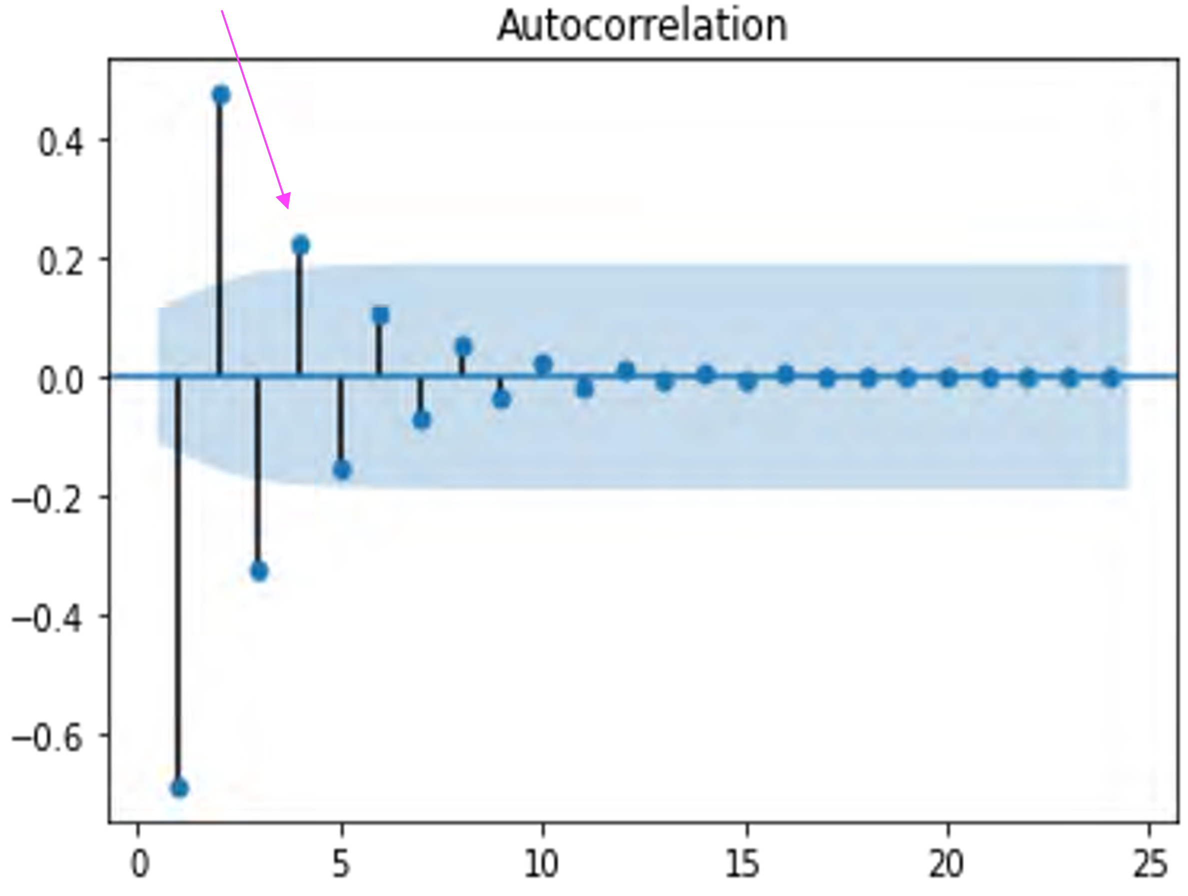
ACF, PACF

- Appropriate p is often chosen from PACF significant lag
- Appropriate q is often chosen from ACF significant lag

p=2



q=4

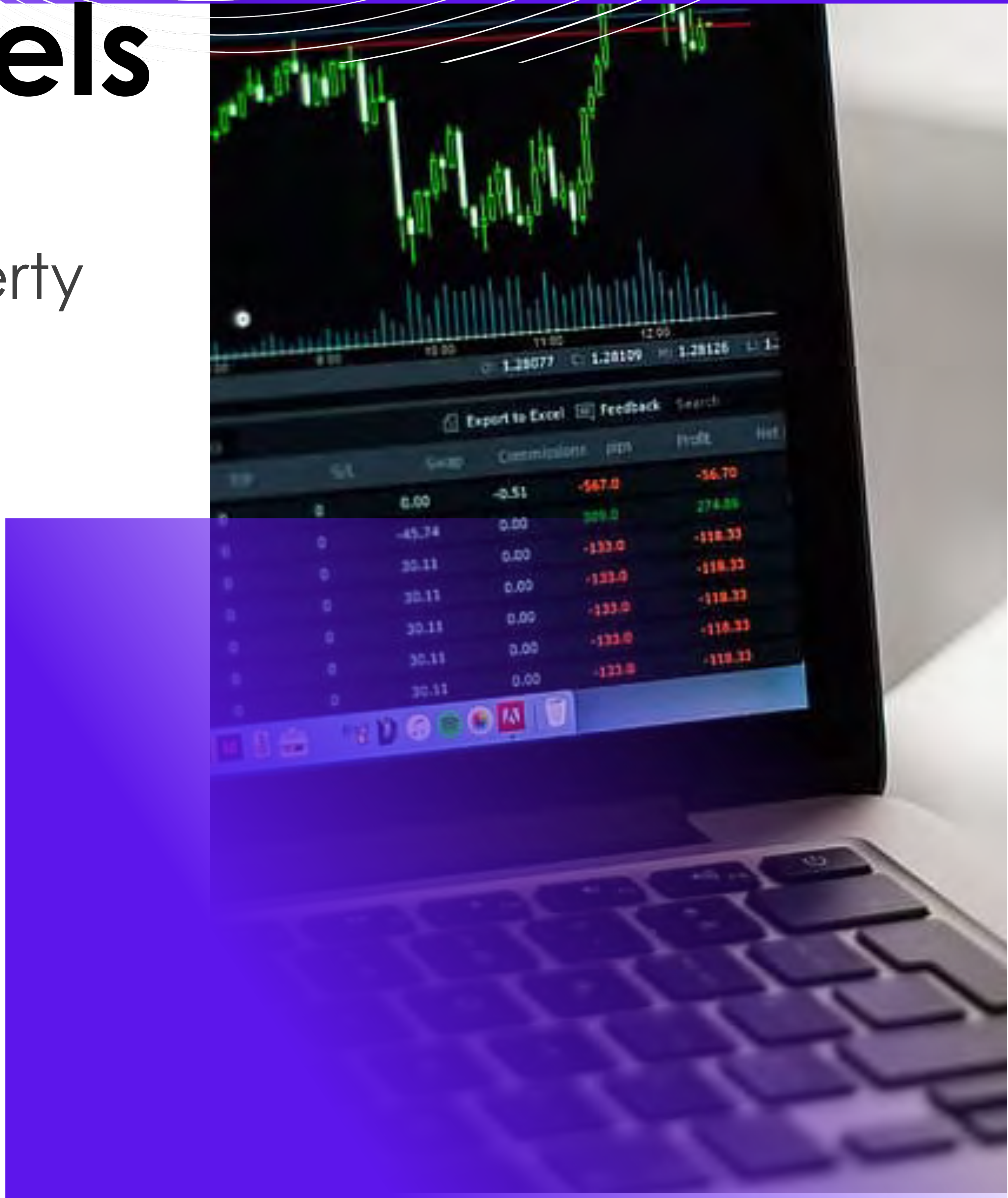


Parameters estimation

- After model order (p, d, q) was chosen the parameters were then estimated
- Typical estimation is done by using MLE

Remark on ARIMA models

- Stationary time series is still a desirable property
- Parameter tuning is important
- Number of datapoints



Conclusion

Notebook