

# Midterm 1 - Assignment 1

Autoregressive analysis

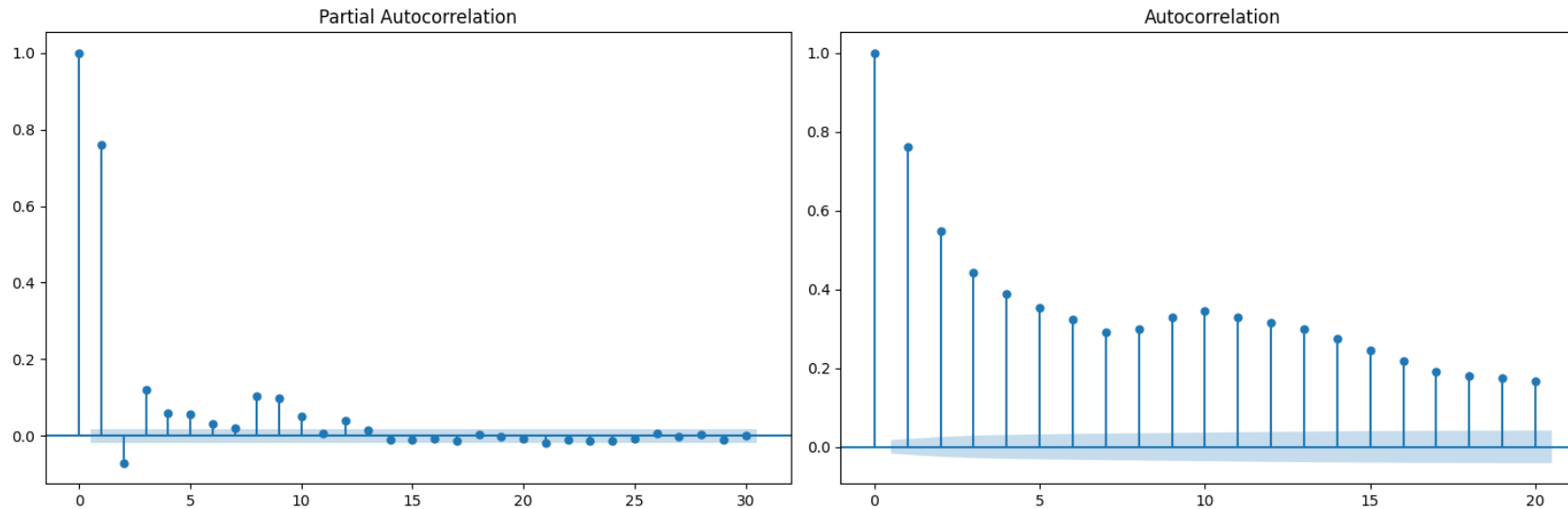
Alex Pasquali

- ▶ Check if the series is **stationary**: use *Augmented Dickey-Fuller test* (ADF)

```
adf = adfuller(whole_data)
adf, pvalue = adf[0], adf[1]
print(f"ADF: {adf}\np-value: {pvalue}")
```

ADF: -21.61637819803612  
p-value: 0.0

- ▶ Plot **autocorrelation** and **partial autocorrelation** to understand visually which lags have more influence on the current value  $y_t$

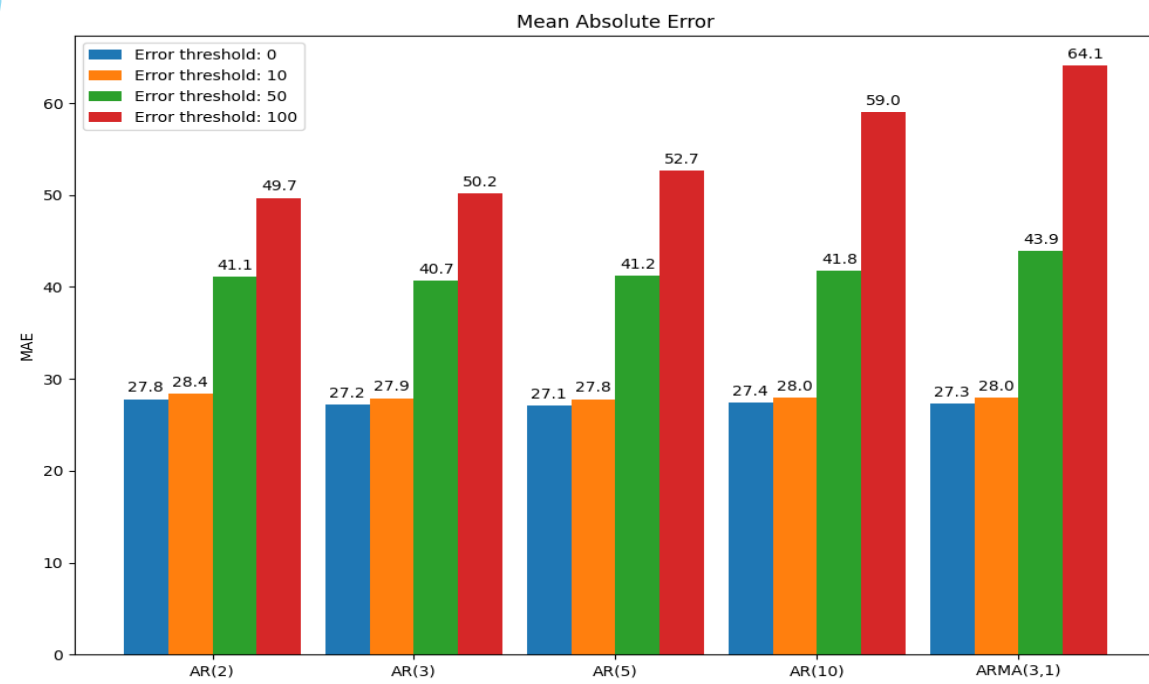


- ▶ It looks like an **AR** model is a good choice

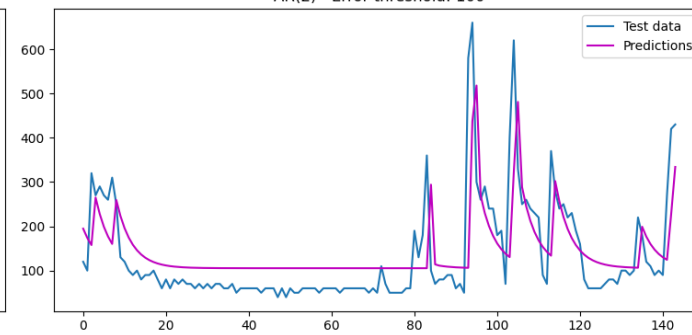
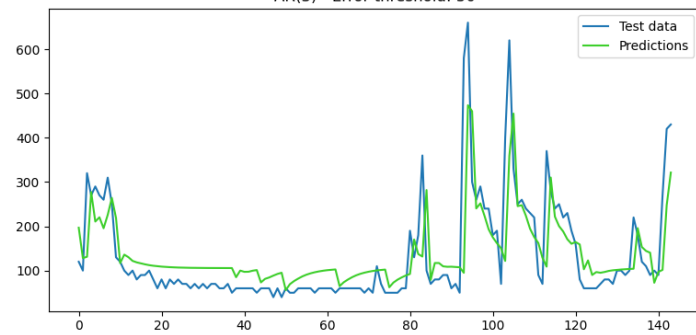
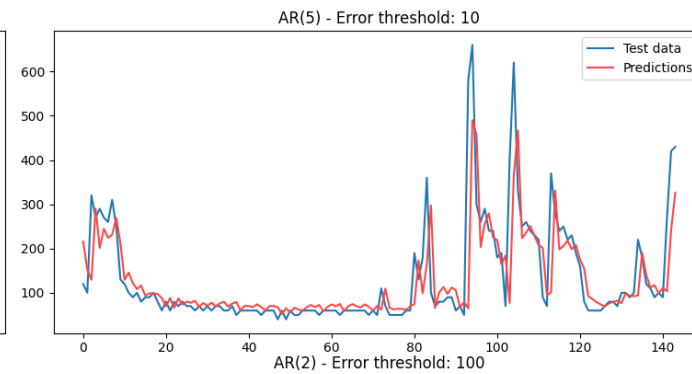
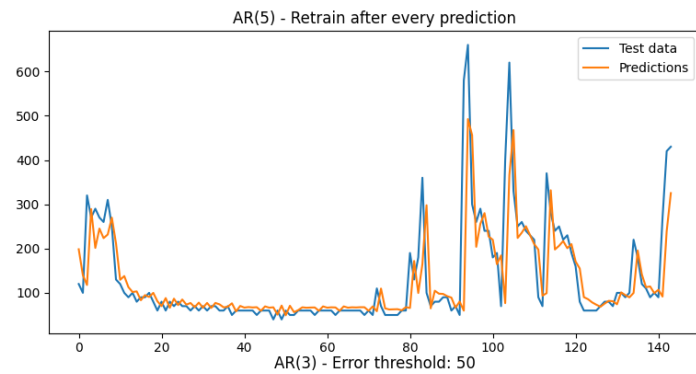
# Training and model selection:

```
def train(order_ar, order_ma, tr_data, ts_data, retrain, err_thresh):  
    # create and fit the model on the training set  
    model = ARIMA(endog=tr_data, order=(order_ar, 0, order_ma))  
    res = model.fit()  
  
    # start forecasting on the test set and retrain when needed  
    idx_last_retrain = 0  
    count_no_retrain = 1  
    predictions = []  
    for i in tqdm(range(len(ts_data))):  
        predictions.append(res.forecast(steps=count_no_retrain)[-1])  
        err = abs(ts_data[i] - predictions[-1])  
        if retrain and err > err_thresh:  
            idx_last_retrain = i  
            count_no_retrain = 1  
            tr_data = np.concatenate((tr_data, ts_data[idx_last_retrain: i + 1]))  
            model = ARIMA(endog=tr_data, order=(order_ar, 0, order_ma))  
            res = model.fit()  
        else:  
            count_no_retrain += 1
```

- ▶ **AR:** try the following orders: 2, 3, 5, 10
- ▶ **ARMA:** try AR(3) + MA(1)
- ▶ Error (E) measured in terms of *Mean Absolute Error* (MAE)
- ▶ retraining schedules:
  - ▶ Retrain after every prediction
  - ▶ Retrain when  $E > 10$
  - ▶ Retrain when  $E > 50$
  - ▶ Retrain when  $E > 100$



# Results



# Final considerations

- ▶ With a retraining after every prediction all the models provide similar results, even though the best one is AR(5)
- ▶ With less frequent retraining, simpler models tend to predominate
- ▶ The ARMA model was not the best one with any retraining policies
  - ▶ Still, it was good with frequent retraining
  - ▶ The training time was considerably longer than in the other models
  - ▶ With more rare retraining, it is noticeably worse than the others
  - ▶ Overall, it was not worth it for this specific time series

The background features abstract, overlapping geometric shapes in various shades of blue, ranging from light sky blue to deep navy blue. These shapes are primarily located on the left and right sides of the frame, creating a modern, dynamic feel. The central area is a plain, light grayish-white.

Thank you for your attention

Alex Pasquali