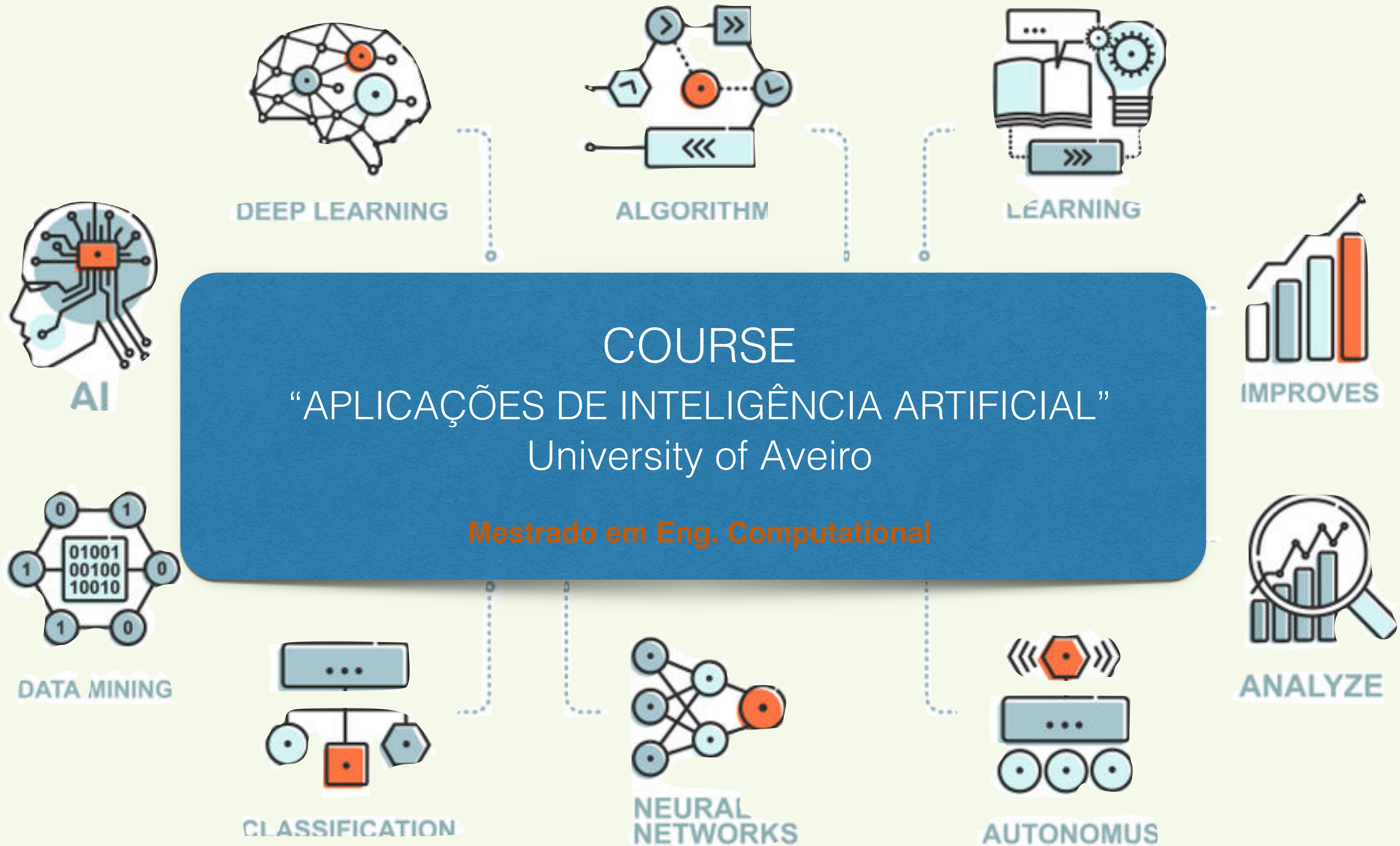

COURSE

“APLICAÇÕES DE INTELIGÊNCIA ARTIFICIAL”

University of Aveiro

Mestrado em Eng. Computational

EXTRA INFO - this label will appear in slides with complementary information

SUMMARY: Time series methods of auto-regression type:

[T] SARIMAX variant models;

[P] Apply the auto-regression methods.

EXTRA INFO - this label will appear in slides with complementary information

SARIMAX Variant Models

Autoregressive (AR) Models

Suppose we have a time series given by y_t. An $AR(p)$ model can be specified by

$$y_t = \beta + \epsilon_t + \sum_{i=1}^{p} \theta_i y_{t-i}$$

Where p is the number of time lags to regress on, ϵ_t is the noise at time t and β is a constant.

This equation can be made more concise through the use of the lag operator, L.

$$L^n y_t = y_{t-n}$$

Taking $\Theta(L)^p$ to be an order p polynomial function of L, we can instead define an autoregressive model by

$$y_t = \Theta(L)^p y_t + \epsilon_t$$

Taking note that the constant has been absorbed into the polynomial Θ.

SARIMAX Variant Models

Moving average (MA) Models.

Whereas autoregressive models regress on prior values of y_t, moving average models regress on prior values of error. An $MA(q)$ model can be specified by

$$y_t = \Phi(L)^q \epsilon_t + \epsilon_t$$

Where q is the number of time lags of the error term to regress on and Φ is defined analagously to Θ.

Autoregressive Moving Average (ARMA) Models

$ARMA(p, q)$ models are simply a sum of $AR(p)$ and $MA(q)$ models.

$$y_t = \Theta(L)^p y_t + \Phi(L)^q \epsilon_t + \epsilon_t$$

SARIMAX Variant Models

Autoregressive Integrated Moving Average (ARIMA) Models

To help tackle non-stationary data, we introduce an integration operator Δ^d, defined as follows

$$y_t^{[d]} = \Delta^d y_t = y_t^{[d-1]} - y_{t-1}^{[d-1]}$$

where $y_t^{[0]} = y_t$ and d is the order of differencing used.

We can now fit an $ARMA(p,q)$ model to $y_t^{[d]}$ rather than y_t.

$$y_t^{[d]} = \Theta(L)^p y_t^{[d]} + \Phi(L)^q \epsilon_t^{[d]} + \epsilon_t^{[d]}$$

This is equivalent to an $ARIMA(p,d,q)$ model on y_t

$$\Delta^d y_t = \Theta(L)^p \Delta^d y_t + \Phi(L)^q \Delta^d \epsilon_t + \Delta^d \epsilon_t$$

With some algebra, we can re-arrange the equation and absorb constants into the polynomials Θ and Φ.

$$\Theta(L)^p \Delta^d y_t = \Phi(L)^q \Delta^d \epsilon_t$$

SARIMAX Variant Models

SARIMA

SARIMA models take seasonality into account by essentially applying an ARIMA model to lags that are integer multiples of seasonality. Once the seasonality is modelled, an ARIMA model is applied to the leftover to capture non-seasonal structure.

To see this more clearly, suppose we have a time series y_t with seasonality s. We can try to eliminate the seasonality with differencing, by applying the differencing operator Δ_s^D to take the seasonal differences of the time series. Here s is the number of time lags comprising one full period of seasonality. D takes on a similar meaning to d in ARIMA models, but instead applies to *seasonal* lags.

We can then capture any remaining structure by applying an $ARMA(P, Q)$ model, to the differenced values, but using seasonal lags. i.e. instead of using a regular lag operator L, we use L^s. P and Q are again seasonal time lags

$$\Delta_s^D y_t = \theta(L^s)^P \Delta_s^D y_t + \phi(L^s)^Q \Delta_s^D \epsilon_t + \Delta_s^D \epsilon_t$$

SARIMAX Variant Models

As with ARIMA, massaging the equation and absorbing constants into polynomials yields the following concise form

$$\theta(L^s)^P \Delta_s^D y_t = \phi(L^s)^Q \Delta_s^D \epsilon_t$$

With any seasonality now removed, we can apply another $ARIMA(p, d, q)$ model to $\Delta_s^D y_t$ by multiplying the seasonal model by the new ARIMA model.

$$\Theta(L)^p \theta(L^s)^P \Delta^d \Delta_s^D y_t = \Phi(L)^q \phi(L^s)^Q \Delta^d \Delta_s^D \epsilon_t$$

This is the general form of a $SARIMA(p, d, q)(P, D, Q, s)$ model.

ARIMAX and SARIMAX

ARIMAX and SARIMAX models simply take exogenous variables into account - ie variables measured at time t that influences the value of our time series at time t, but that are not autoregressed on. To do this, we simply add the terms in on the right hand side of our ARIMA and SARIMA equations.

SARIMAX Variant Models

For n exogenous variables defined at each time step t, denoted by x_t^i for $i \leq n$, with coefficients β_i, the $ARIMAX(p, d, q)$ model is defined by

$$\Theta(L)^p \Delta^d y_t = \Phi(L)^q \Delta^d \epsilon_t + \sum_{i=1}^{n} \beta_i x_t^i$$

and the $SARIMAX(p, d, q)(P, D, Q, s)$ model by

$$\Theta(L)^p \theta(L^s)^P \Delta^d \Delta_s^D y_t = \Phi(L)^q \phi(L^s)^Q \Delta^d \Delta_s^D \epsilon_t + \sum_{i=1}^{n} \beta_i x_t^i$$

Further References

Chatfield, C 2004, *The Analysis of Time Series : An Introduction*, 6th ed., Chapman & Hall/CRC, Boca Raton, Fla.

(Partial) AutoCorrelation Function on SARIMA

In the case of a SARIMA model with only a seasonal moving average process of order 1 and period of 12, denoted as:

SARIMA(0,0,0)(0,0,1;12)

- A spike is observed at lag 12
- Exponential decay in the seasonal lags of the PACF (lag 12, 24, 36, …)

Similarly, for a model with only a seasonal autoregressive process of order 1 and period of 12:

SARIMA(0,0,0)(1,0,0;12)

- Exponential decay in the seasonal lags of the ACF (lag 12, 24, 36, …)
- A spike is observed at lag 12 in the PACF

SARIMA APPLICATION

Explore the notebooks (S)ARIMA.ipynb and answer the several teacher questions. Apply to the dataset chosen in the last session.

```
simulated_ARMA_data = ArmaProcess(ar1, ma1).generate_sample(nsample=10000)
```

✓ 0.2s

+ Code + Markdown

```
plt.figure(figsize=[15, 7.5]) # Set dimensions for figure
plt.plot(simulated_ARMA_data)
plt.title("Simulated ARMA(1,1) Process")
plt.xlim([0, 200])
plt.show()
```

✓ 0.4s