

02427 lecture 6: Exercises

Jan Kloppenborg Møller

DTU-Compute - Technical University of Denmark
jkmo@dtu.dk

Advanced time series analysis – 2026

Exercise 1

The the data file `exercise.csv`, contains 2 forecasts ($F1, F2$) given as the probability $P(Y = 1)$, and the binary outcome y . Calculate

- RMSE for the two forecasts
- MAE for the the two forecasts
- The log score for the two out comes
- Which forecaster would you prefer?

Exercise 2 - AR-processes: Simulation

Simulate 3 independent AR(2) processes, you may choose any set stationnary paramter values, e.g. $\phi_1 = 1.2$, $\phi_2 = -0.3$. The length of the simulation set should be $2 \cdot n$, you can choose n (e.g. $n = 100$). The 3 data sets will be referred to as

- 1 Training set, for estimation parameters
- 2 Validation set, used for hierarchical learning later
- 3 Test set, used for assesing model quality

Estimate

- 1 Parameters for an AR(1) process using the training set
- 2 Parameters for an AR(2) process using the training set

Exercise 2 - AR-processs forecasting

Organize the data in the test and validation set as

$$\mathbf{Y} = \begin{bmatrix} y_1 & y_2 \\ y_3 & y_4 \\ \vdots & \\ y_{2n-1} & y_{2n} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1^T \\ \mathbf{y}_2^T \\ \vdots \\ \mathbf{y}_n^T \end{bmatrix}$$

Using the estimated models from the training set, make the corresponding predictions

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{y}_2 = E[\mathbf{y}_2^T | \mathbf{y}_1^T] \\ \hat{y}_3 = E[\mathbf{y}_3^T | \mathbf{y}_2^T] \\ \vdots \\ \hat{y}_n = E[\mathbf{y}_n^T | \mathbf{y}_{n-1}^T] \end{bmatrix}$$

Notice that the number of rows in $\hat{\mathbf{Y}}$ is $n - 1$ (not n).

Exercise 2 - AR-processs forecasting

We assume the following model for the probabilistic forecasting

$$\mathbf{y}_i \sim N(\hat{\mathbf{y}}_i, \Sigma)$$

For each of the models find estimates of Σ , based on

- The estimated and assumed model
- The sum of squared error on the validation set

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{i=2}^n (\mathbf{y}_i - \hat{\mathbf{y}}_i)^T (\mathbf{y}_i - \hat{\mathbf{y}}_i)$$

You now have 4 models (2 mean value structures and 2 covariance structures).

Exercise 2 - AR-processs: Evaluation

For each of the 4 forecasting model calculate probabilistic scores on the test set, you should at least consider

- RMSE
- The multivariate log score

You may consider other score as well (or give the PIT histograms), e.g.

- univariate CRPS
- The variogram score

Conclude based on the best model.

Exercise 3 - AR-processs: Top level time series

Define a top level time-series (for all 3 data sets) as

$$y_{\text{top},t} = 2\bar{y}_t.$$

Estimate an AR(1) process on the top level times series on the training set and update the data matrices related to the validation and test set

$$Y = \begin{bmatrix} y_{\text{top},1} & y_1 & y_2 \\ y_{\text{top},2} & y_3 & y_4 \\ \vdots & & \\ y_{\text{top},n} & y_{2n-1} & y_{2n} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1^T \\ \mathbf{y}_2^T \\ \vdots \\ \mathbf{y}_n^T \end{bmatrix}.$$

Set up the corresponding matrix of forecasts for the validation and training set.

Exercise 3 - AR-processs: Hierarchical forecasting

Define the summation matrix S and make the reconciled forecast on the test set, you may use different versions of the covariance matrix, but at least consider the shrinkage estimator (in R you can use the library `corpcor` and `cor.shrink`) to find the optimal shrinkage.

Some questions to consider:

- Evaluate/compare the quality of the forecast using some probabilistic scores on the bottom level.
- What are the parameters in the weight matrix?
- What happens to the weights if you replace the bottom the level forecasts with forecasts based on the AR(2) model?
- What are the standard error of the (top level) weight matrix? (a bit complicated to find)