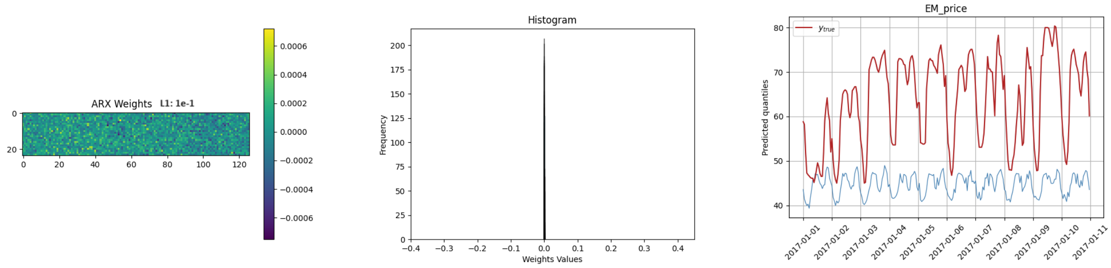
Assignment1PEF_AY23_24
Group 14

Idda, Raffaeli, Riondato, Stillo

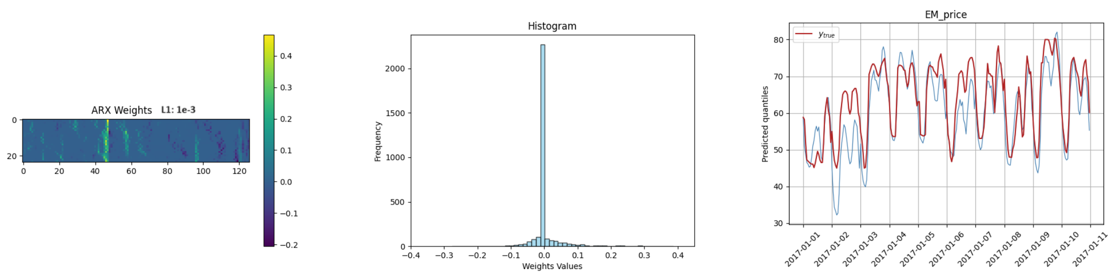
1 LASSO

for l1=1e-01:

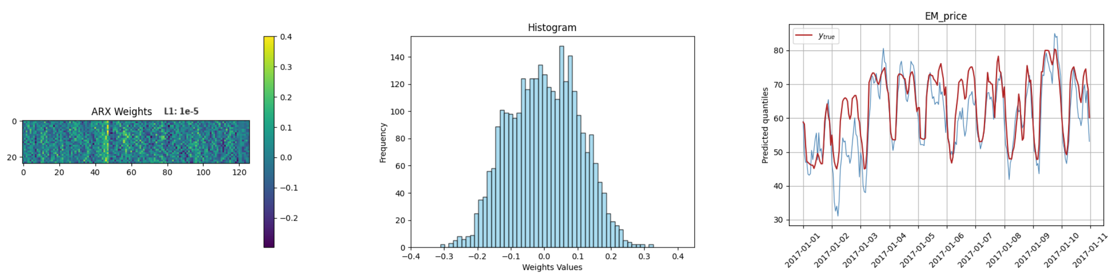
(a) Weight (b) Histogram (c) EM price



for l1=1e-03:



for l1=1e-05:

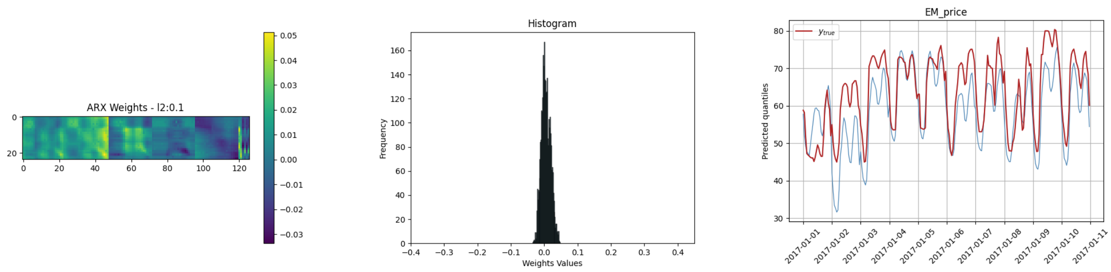


L1-01	431,17
L1-03	34,95
L1-05	41,11

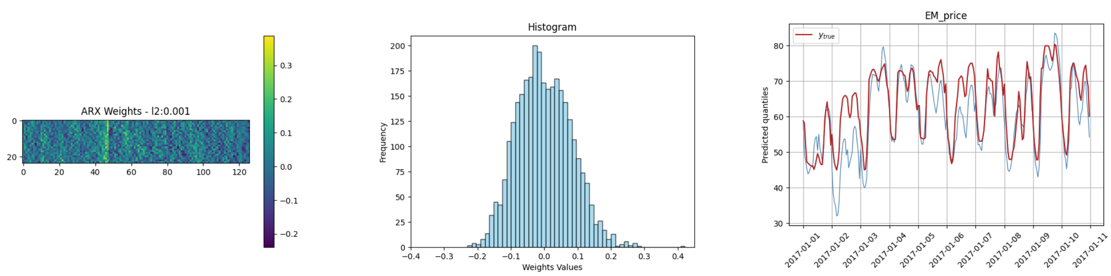
(a) l1 MSE

2 Ridge

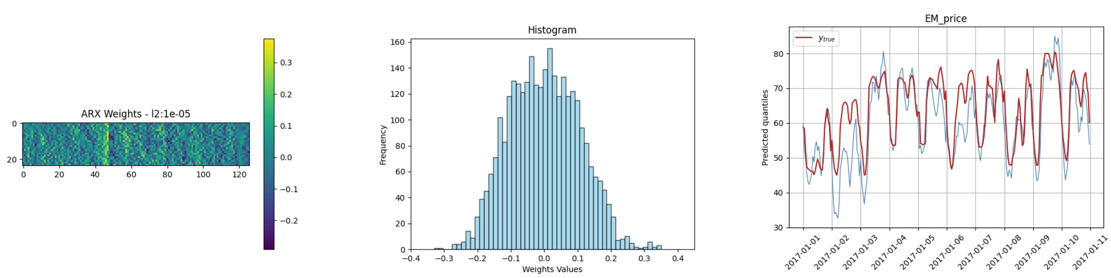
for l2=1e-01:



for l2=1e-03:



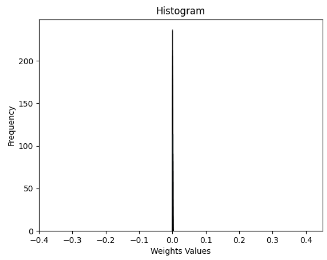
for l2=1e-05:



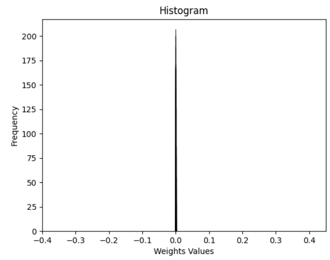
L2-0.1	51,25
L2-0.3	36,7
L2-0.5	40,46

(a) l2 MSE

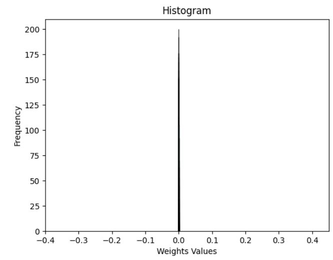
3 Elastic Net



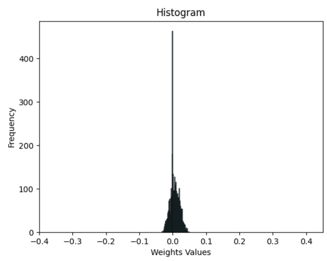
(a) l1=1e-01, l2=1e-01



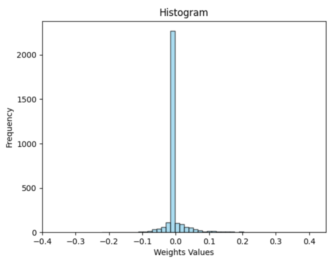
(b) l1=1e-01, l2=1e-03



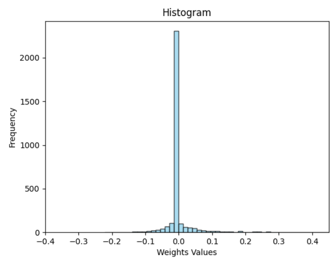
(c) l1=1e-01, l2=1e-05



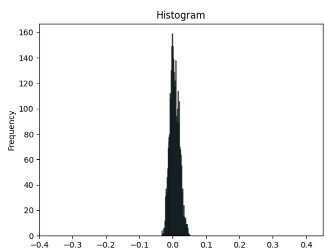
(a) l1=1e-03, l2=1e-01



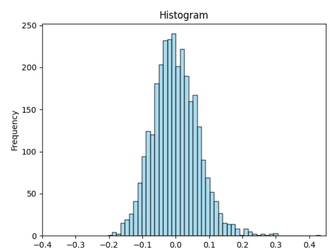
(b) l1=1e-03, l2=1e-03



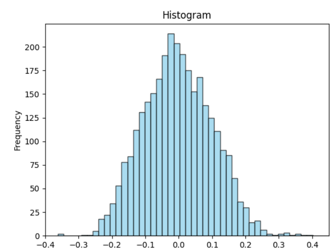
(c) l1=1e-03, l2=1e-05



(a) l1=1e-05, l2=1e-01



(b) l1=1e-05, l2=1e-03



(c) l1=1e-05, l2=1e-05

L1 \ L2	1E-01	1E-03	1E-05
1E-01	430,08	432,07	430
1E-03	54,35	35,4	35,02
1E-05	51,02	37,69	40,83

(a) l1-l2 MSE

4 Comments

In this laboratory we were tasked with the application of probabilistic time series forecasting systems to day-ahead electricity price forecasting. This study examines state-of-the-art methods, including ARX with l1 regularization (LASSO), l2 regularization (Ridge), and their combination (Elastic Net), in forecasting day-ahead electricity prices. The goal is to assess the effectiveness of these regularization approaches in multi-step settings to enhance model precision.

1. LASSO (l1 Regularization): LASSO regularization penalizes the absolute value of coefficients, promoting sparsity and reducing complexity in the model. The LASSO objective function is:

$$\min_{\Omega} \sum_{n=1}^{N} (f_{\Omega}(x_n) - y_n)^2 + \lambda \sum_{j=1}^{\Omega} |w_j|$$

2. Ridge (l2 Regularization): Ridge regularization penalizes the sum of the squared coefficients, helping to prevent over-fitting and stabilize the model. The Ridge objective function is:

$$\min_{\Omega} \sum_{n=1}^{N} (f_{\Omega}(x_n) - y_n)^2 + \gamma \sum_{j=1}^{\#\Omega} (w_j)^2$$

3. Elastic Net (l1 + l2 Regularization): This approach combines LASSO and Ridge regularization, providing a balance between promoting sparsity and reducing model complexity. The Elastic Net objective function is:

$$\min_{\Omega} \sum_{n=1}^{N} (f_{\Omega}(x_n) - y_n)^2 + \lambda \sum_{j=1}^{\#\Omega} |w_j| + \gamma \sum_{j=1}^{\#\Omega} (w_j)^2$$

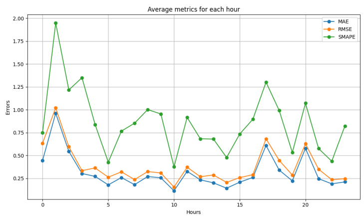
To assess the performance of the methods, the mean squared error (MSE) was computed for each method across various regularization parameters $[10^{-1}, 10^{-3}, 10^{-5}]$. This computation has been integrated with the graphical analysis of the distribution of weights for every combination, to ascertain the optimal one.

LASSO: As observed from the table, the highest regularization parameter 10^{-1} yields a notably high MSE, indicating potential over-regularization and consequent loss of model performance. The optimal MSE is achieved at $\lambda = 10^{-3}$, although this choice is not optimal in terms of data fitting. Specifically, a prominent blue-colored area in the weight plot suggests non-uniform prediction. The optimal choice appears to be $\lambda = 10^{-5}$, exhibiting relatively low MSE and uniform weight distribution.

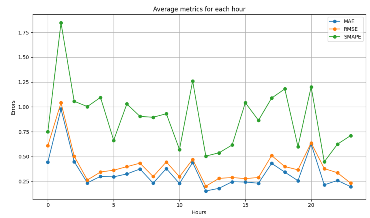
RIDGE: It can be noticed that in this scenario, the MSE maintains the same magnitude across all parameter choices, suggesting less sensitivity of the prediction to the gamma parameter selection. Once again, the optimal compromise appears to be $\gamma = 10^{-5}$.

ELASTIC NET: Conducting a similar analysis as in the preceding cases, it is evident that the best results are obtained when setting $\lambda = 10^{-5}$ and $\gamma = 10^{-5}$. Indeed, the MSE demonstrates relatively low MSE and fairly homogeneous weight distribution (when compared to the other combinations).

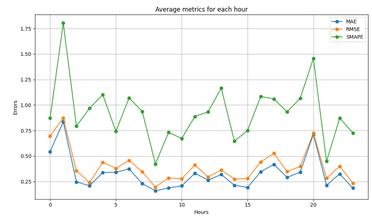
5 Extra point



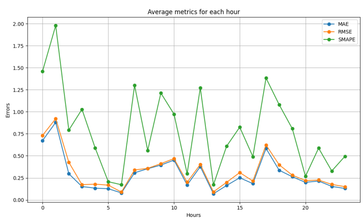
(a) l1=1e-01, l2=1e-01



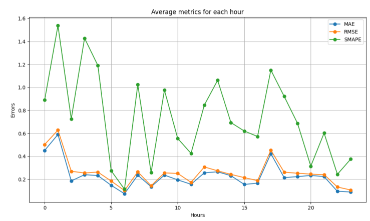
(b) l1=1e-01, l2=1e-03



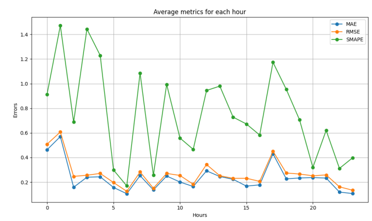
(c) l1=1e-01, l2=1e-05



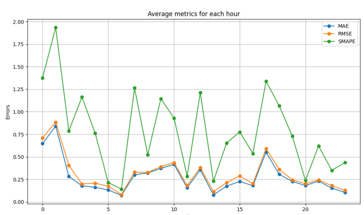
(a) l1=1e-03, l2=1e-01



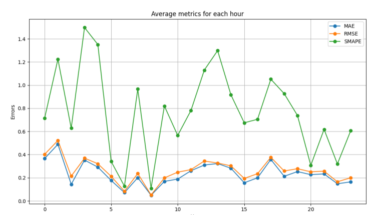
(b) l1=1e-03, l2=1e-03



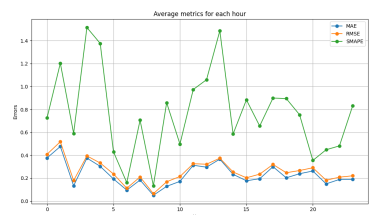
(c) l1=1e-03, l2=1e-05



(a) l1=1e-05, l2=1e-01



(b) l1=1e-05, l2=1e-03



(c) l1=1e-05, l2=1e-05

L1 \ L2	1E-01	1E-03	1E-05
1E-01	18,89	18,96	18,97
1E-03	6,15	4,66	4,64
1E-05	5,97	4,71	4,99

(a) MAE

L1 \ L2	1E-01	1E-03	1E-05
1E-01	20,74	20,79	20,74
1E-03	7,37	5,95	5,92
1E-05	7,14	6,14	6,39

(b) RMSE

L1 \ L2	1E-01	1E-03	1E-05
1E-01	0,34	0,34	0,34
1E-03	0,10	0,08	0,08
1E-05	0,10	0,08	0,08

(c) sMAPE

To evaluate the point prediction accuracy, several indicators are commonly considered, including Mean Absolute Error

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_n - \hat{y}_n|$$

Root Mean Squared Error

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_n - \hat{y}_n)^2}$$

Symmetric Mean Absolute Percentage Error

$$\text{sMAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_n - \hat{y}_n|}{0.5(|y_n| - |\hat{y}_n|)}$$

We implemented these evaluation metrics in two ways (specifically for the Elastic Net method applied to several hyper-parameters combinations):

HOURLY: we considered the prediction of each hour of the day over the 10-days time horizon and we computed the metrics. We plotted the 24 hours averages on the same figure (re-scaling all the data to better understand the metrics differences). it can be noticed that the method with $\lambda = 10^{-5}$ and $\gamma = 10^{-5}$ is the best.

AVERAGE: we computed metrics for all data and entered the results into tables. Looking at the tables we have further confirmation that the method with $\lambda = 10^{-5}$ and $\gamma = 10^{-5}$ is the best, because it has the bast value of evaluation metrics and fairly homogeneous weights distribution.