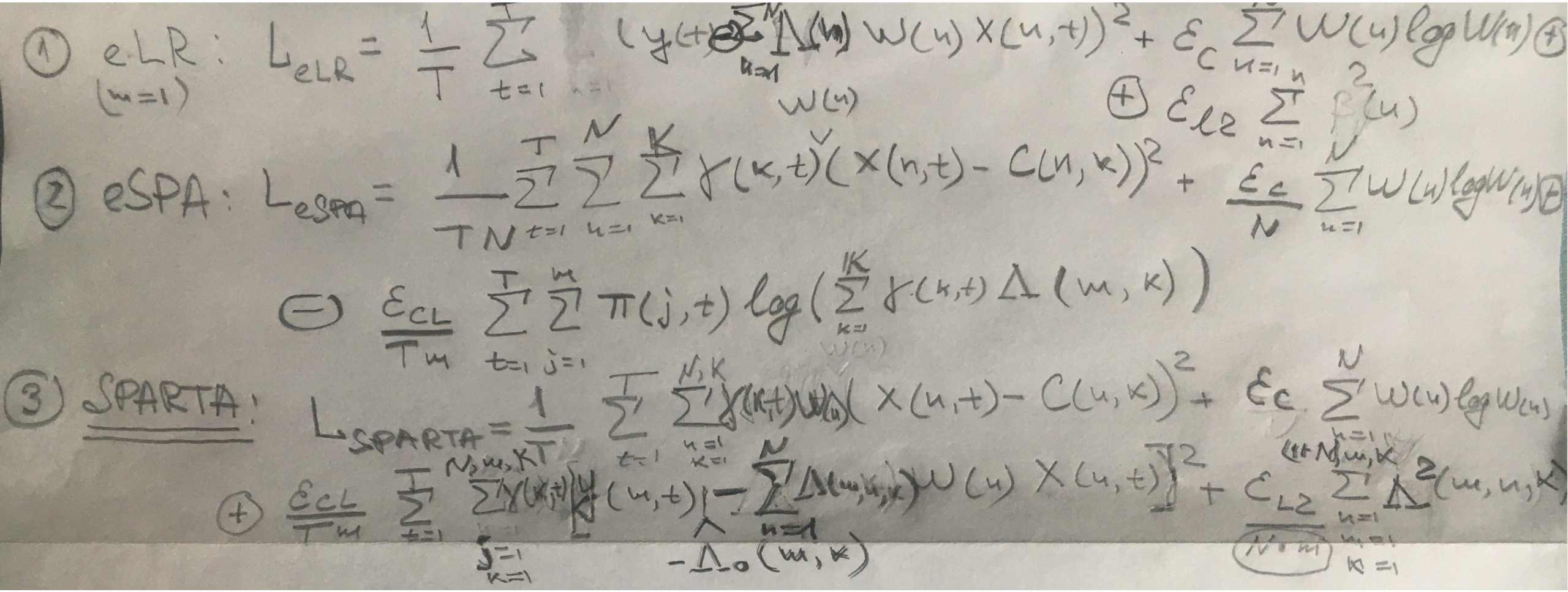
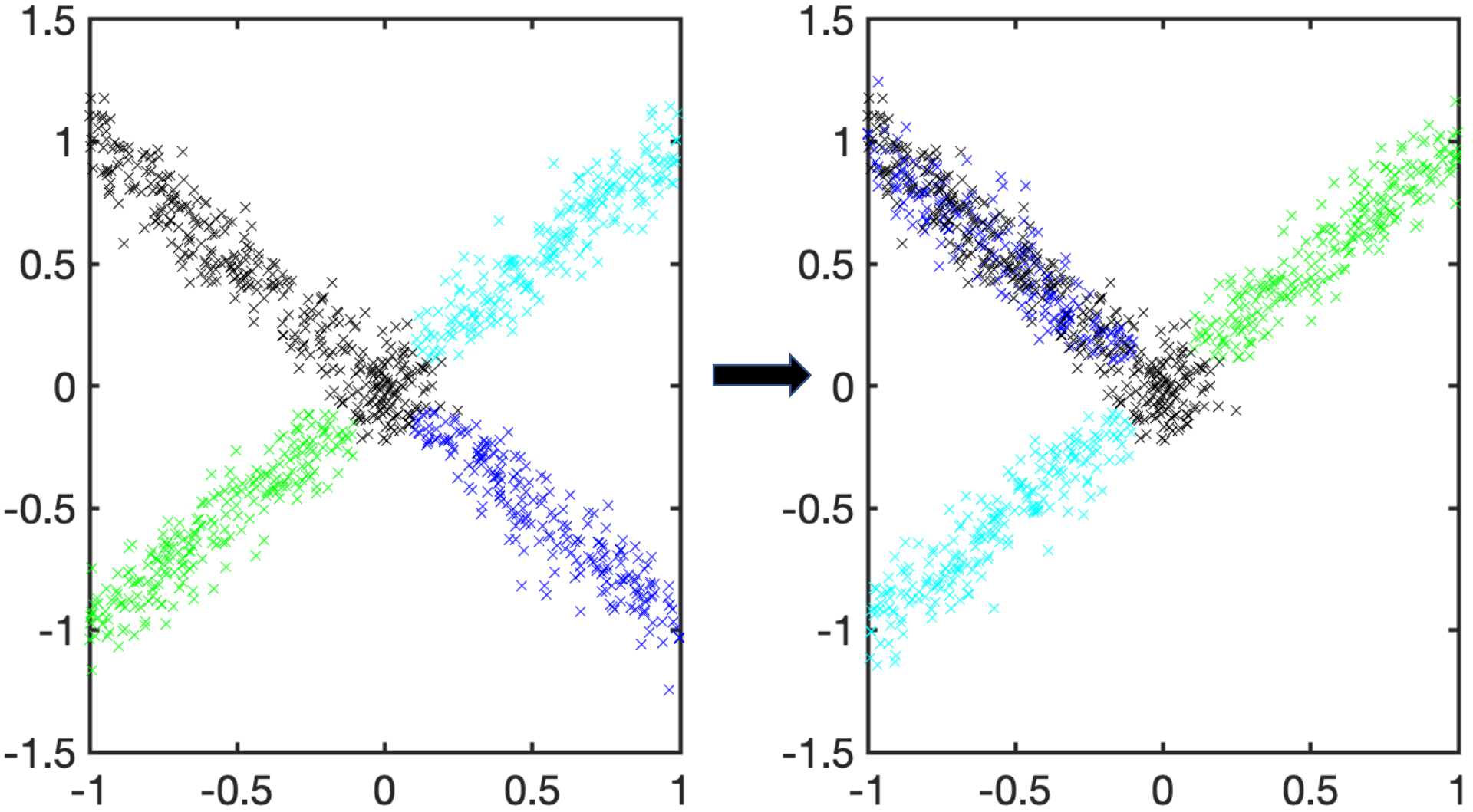
From eSPA to entropically-sparsified Linear Regression (eLR)
and to Sparse Probabilistic Approximation for Regression Task Analysis (**SPARTA**)



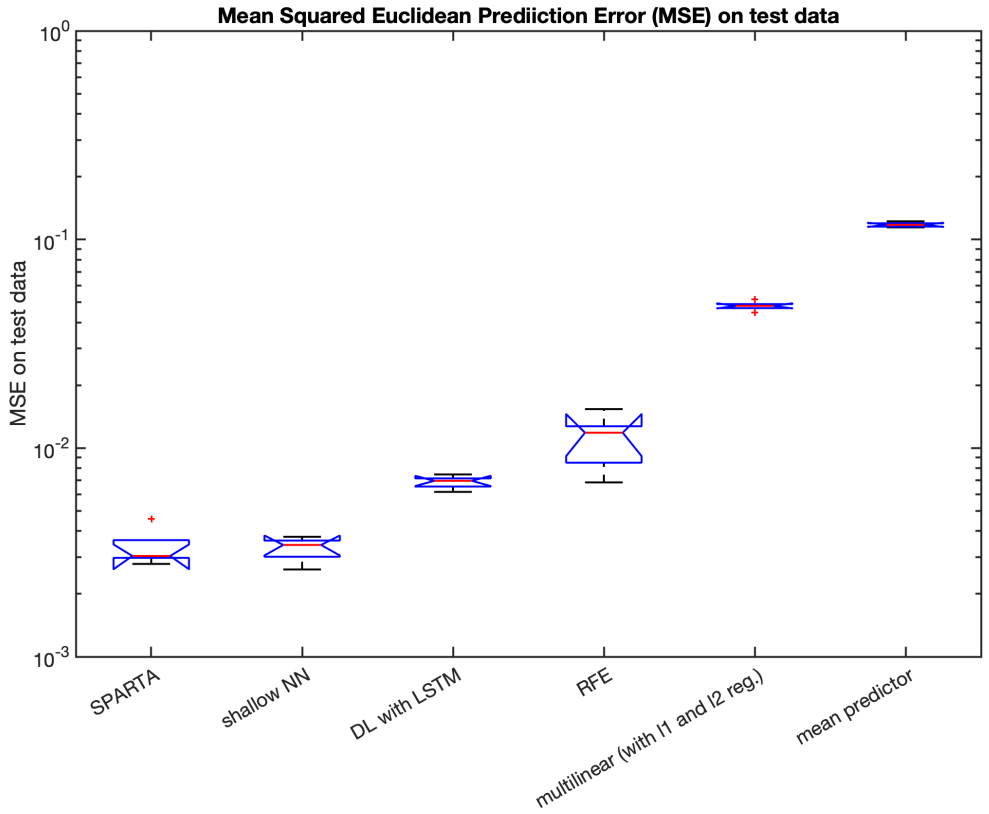
Mathematical idea behind SPARTA: joint numerically-scalable (linear in T and K, loglinear in N) solution of feature X discretization (first term), entropic feature sparsification (second term) and regularized **piecewise-linear** regression problems (third and fourth terms)

Synthetic example: piecewise-linear transformation $f:R^N \rightarrow R^2$ of a 2D image of letter **X** (hidden N dimensions without any image information) into a 2D image of letter **Y**

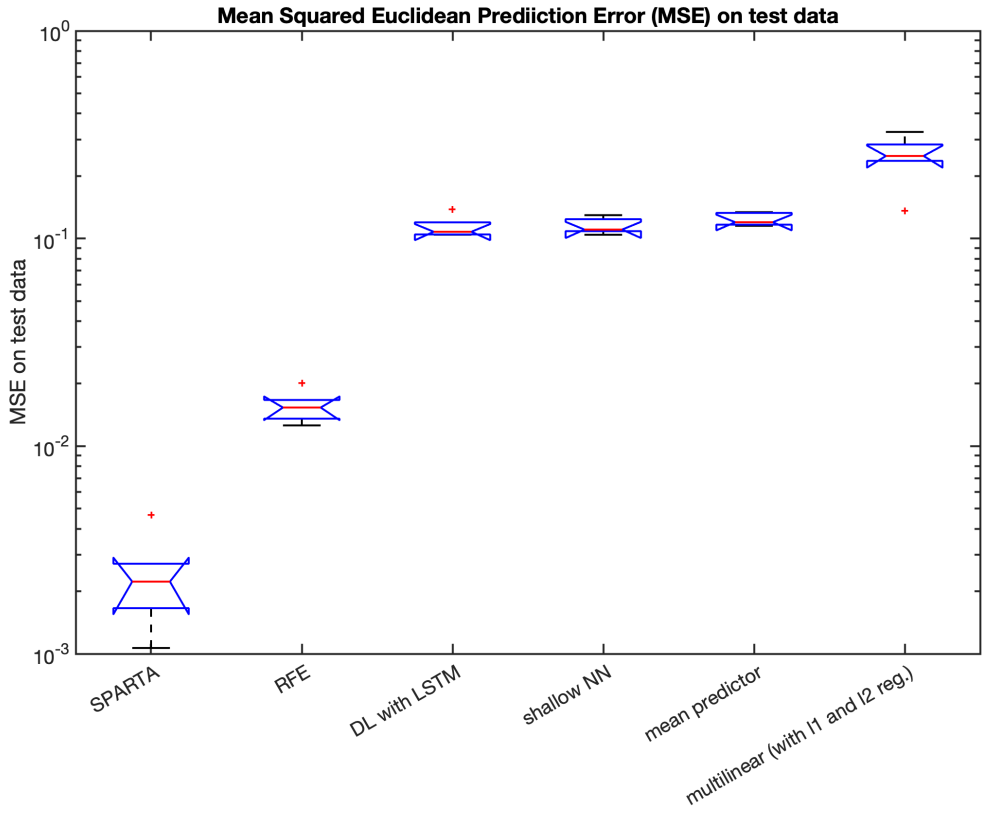


Synthetic example: piecewise-linear transformation $f:R^N \rightarrow R^2$ of a 2D image of letter **X** (hidden N dimensions without any image information) into a 2D image of letter **Y**

N=10, T=1000

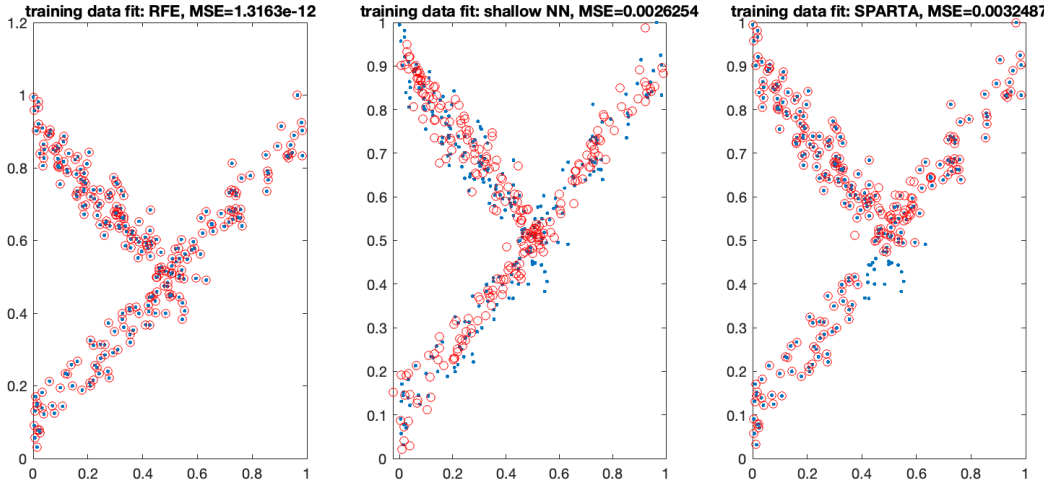


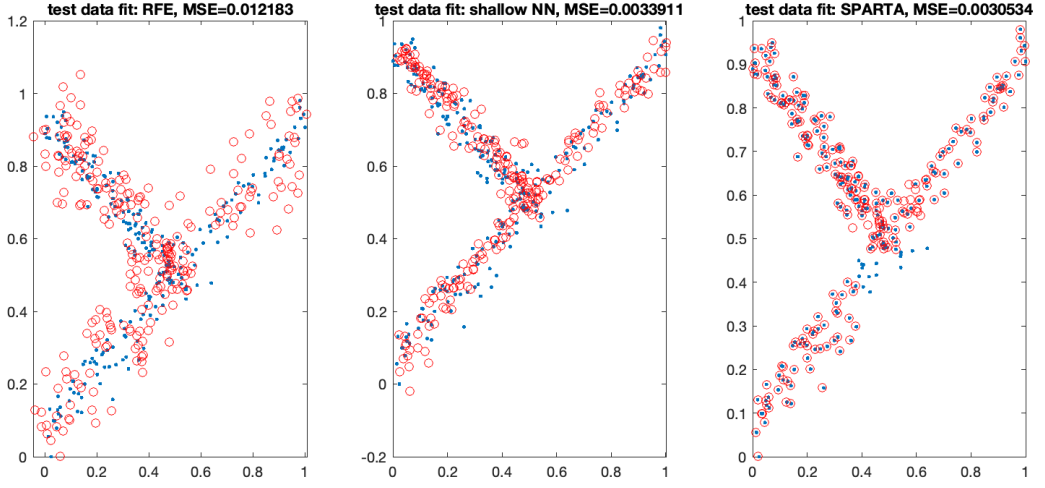
N=200, T=1000



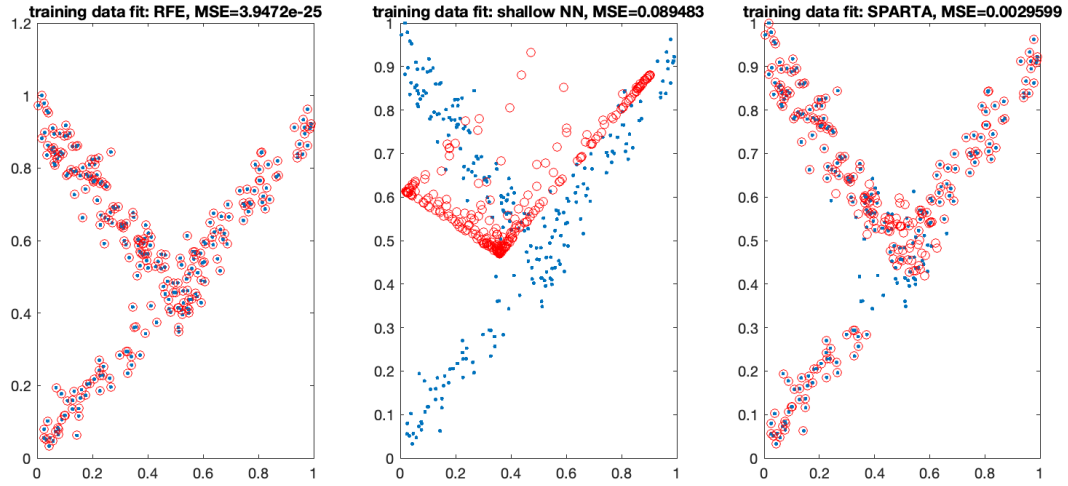
Synthetic example: piecewise-linear transformation $f:R^N \rightarrow R^2$ of a 2D image of letter **X** (hidden N dimensions without any image information) into a 2D image of letter **Y**

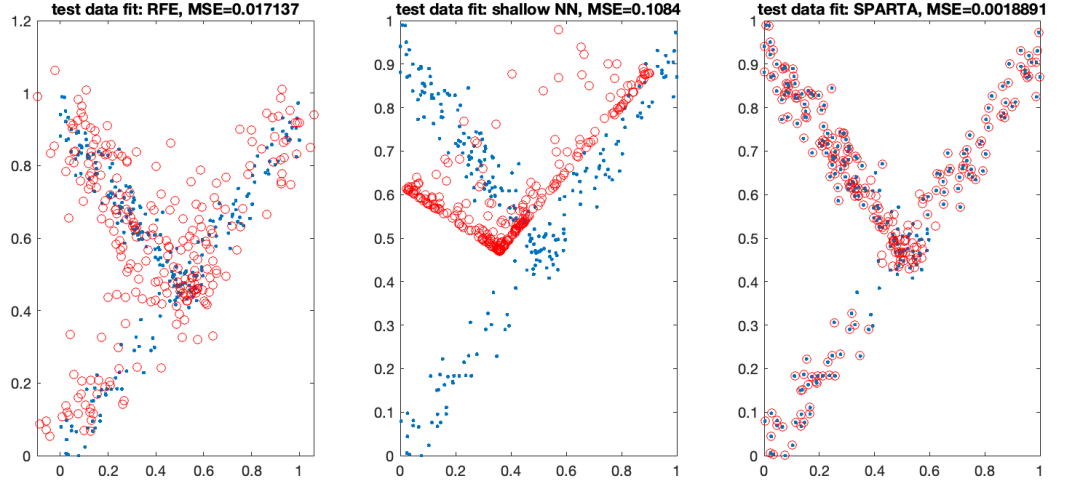
N=10, T=1000





N=200, T=1000

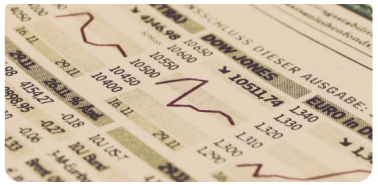




Application example 1: predicting dayly change of price (S_{t+1}-S_t) and its dependence from previous days price changes (open, low, high, close) AND news proportion sentiments (positive, neutral, negative) for dayly Apple Stock between 2006 and 2016. Data from Kaggle.com.

Impact of News on the Share closing value

Stock prices and the News related to the Apple and Microsoft



https://www.kaggle.com/datasets/BidecInnovations/stock-price-and-news-realted-to-it

Data | Code (2) | Discussion (2) | Metadata

About Dataset

The Dataset here consists of Stock Value of Apple(AAPL) and Microsoft(MSFT) from 2006 to 2016 and News summary, abstract and snippets on News featuring these two tech giants during the same period. We are trying to understand and depict the impact of News stories on the stock prices. The News snippets, summary and abstracted were retrieved from The New York Times API. The stock values were obtained from Yahoo Finance.
The final data set is created by applying sentiment analysis on those News string and converting them into a score. For this purpose, Natural Language Toolkit(NLTK) was used. The final dataset were further used for a Regression Model.

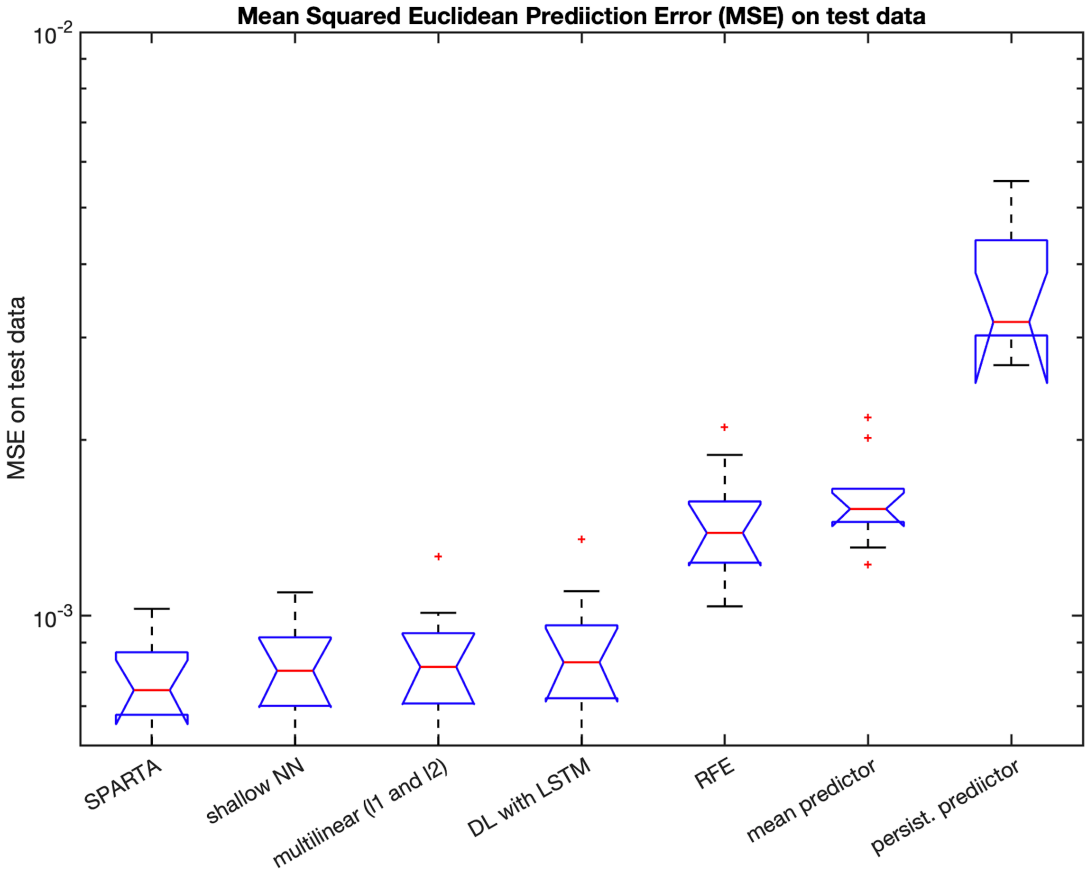
Usability ⓘ
7.35

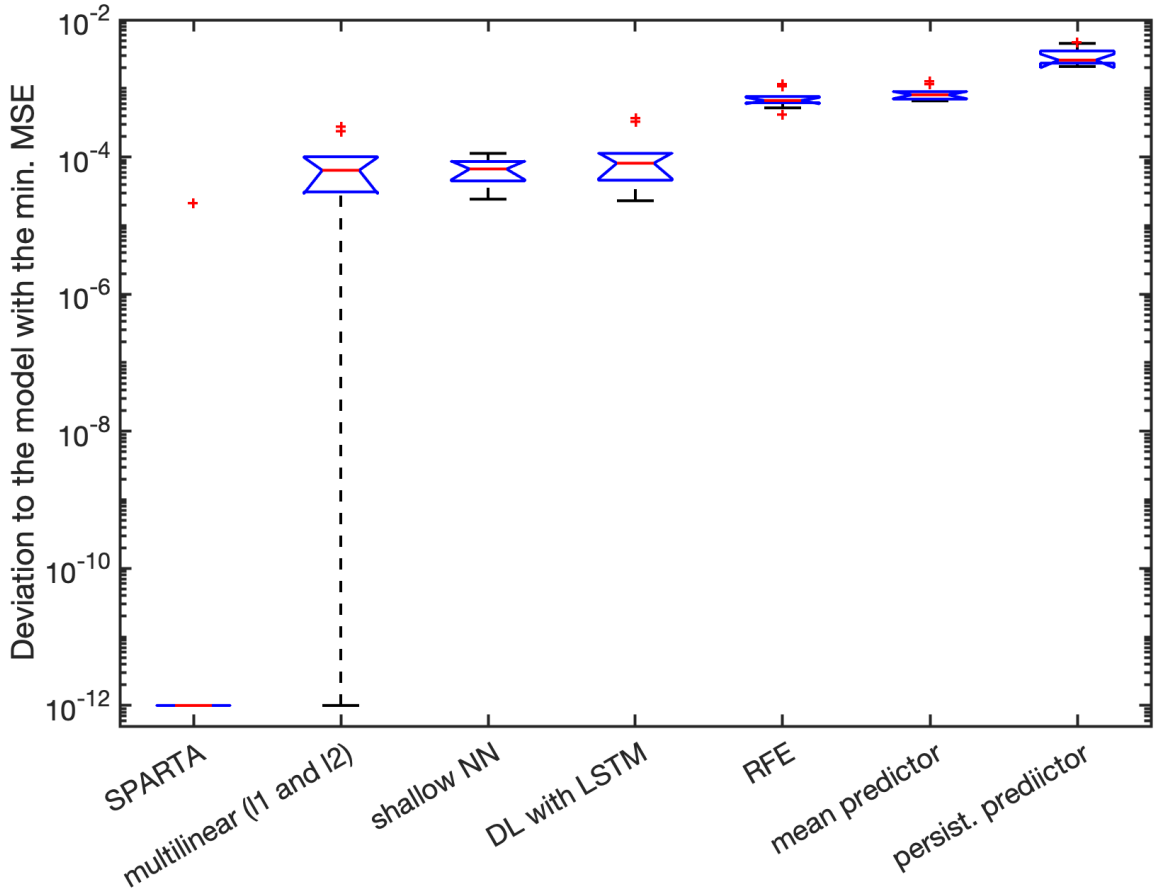
License
Unknown

Expected update frequency
Not specified

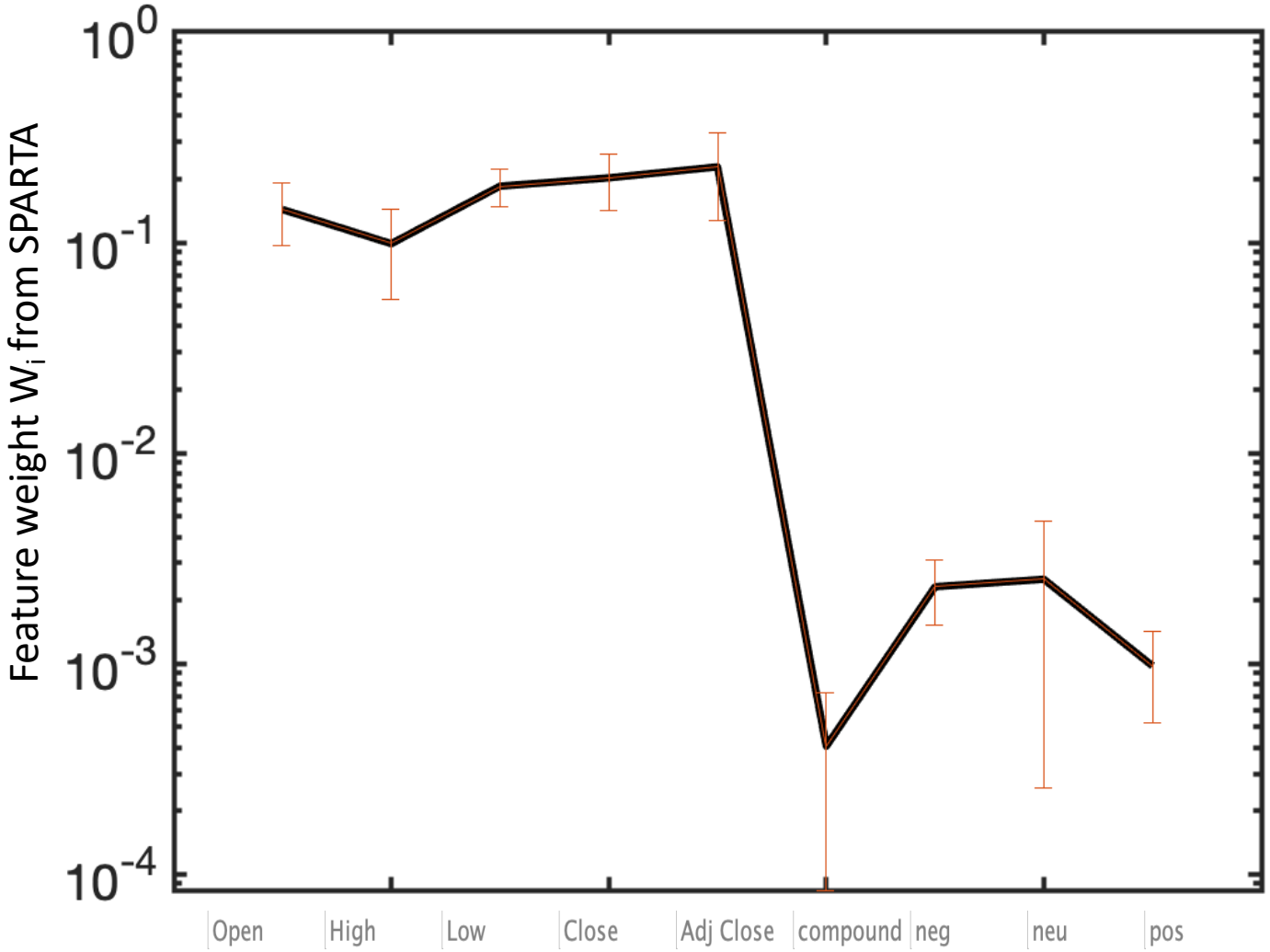
1	Date	Open	High	Low	Close	Adj Close	compound	neg	neu	pos
2	2006-12...	13.1143	13.19	12.8714	91.32	13.0457	0.7707	0.032	0.905	0.063
3	2006-12...	13.1257	13.15	12.9286	91.12	13.0171	0.872	0.011	0.904	0.085
4	2006-12...	13.0929	13.19	12.9814	91.27	13.0386	0.0	0.0	0.0	0.0
5	2006-12...	12.9486	13.0557	12.81	89.83	12.8329	0.6858	0.029	0.878	0.093
6	2006-12...	12.8614	12.9286	12.4143	87.04	12.4343	-0.6712	0.091	0.869	0.04
7	2006-12...	12.4614	12.77	12.4286	88.26	12.6086	-0.1796	0.084	0.848	0.069
8	2006-12...	12.7	12.7571	12.5786	88.75	12.6786	-0.8743	0.105	0.852	0.042
9	2006-12...	12.6586	12.6914	12.2186	86.14	12.3057	0.0	0.0	0.0	0.0
10	2006-12...	12.5643	12.7243	12.45	89.05	12.7214	0.936	0.018	0.9	0.082
11	2006-12...	12.7214	12.8571	12.6086	88.55	12.65	0.962	0.026	0.88	0.094
12	2006-12...	12.7171	12.7457	12.4757	87.72	12.5314	-0.5228	0.051	0.905	0.044
13	2006-12...	12.5186	12.5714	12.0843	85.47	12.21	0.7059	0.03	0.887	0.082
14	2006-12...	12.1043	12.3829	11.9457	86.31	12.33	-0.6705	0.132	0.78	0.088
15	2006-12...	12.3529	12.3814	12.1057	84.76	12.1086	-0.802	0.131	0.781	0.088
16	2006-12...	12.1	12.2114	11.7429	82.9	11.8429	0.0	0.0	0.0	0.0

Application example 1: predicting dayly change of price (S_{t+1}-S_t) and its dependence
from previous days price changes (open, low, high, close) AND news proportion sentiments (positive, neutral, negative)
for dayly Apple Stock between 2006 and 2016. Data from Kaggle.com.

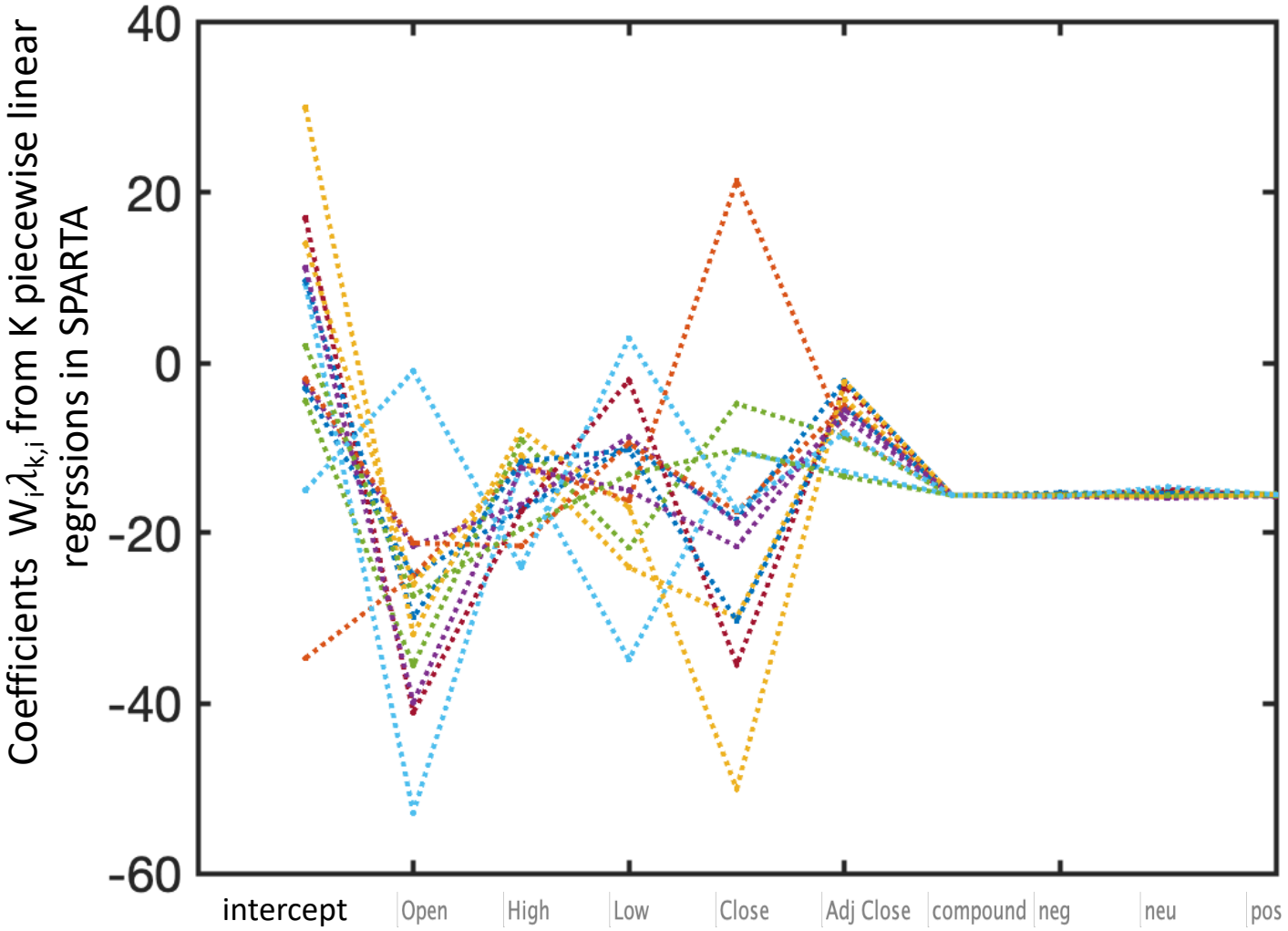




Application example 1: predicting dayly change of price (S_{t+1}-S_t) and its dependence
from previous days price changes (open, low, high, close) AND news proportion sentiments (positive, neutral, negative)
for dayly Apple Stock between 2006 and 2016. Data from Kaggle.com.



Application example 1: predicting dayly change of price (S_{t+1}-S_t) and its dependence from previous days price changes (open, low, high, close) AND news proportion sentiments (positive, neutral, negative) for dayly Apple Stock between 2006 and 2016. Data from Kaggle.com.



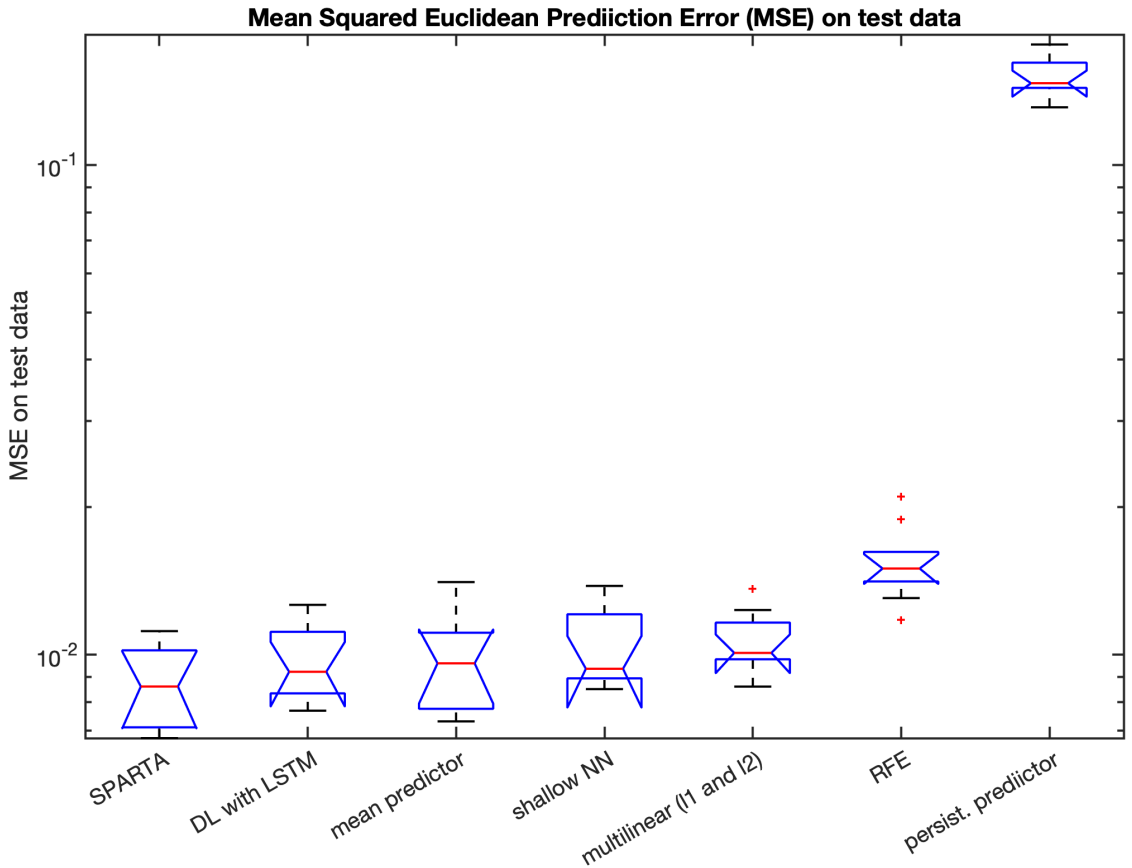
Application example 2: predicting interday absolut change of price of S&P $|S_t^{close} - S_t^{open}|$. Data from Kaggle.com.

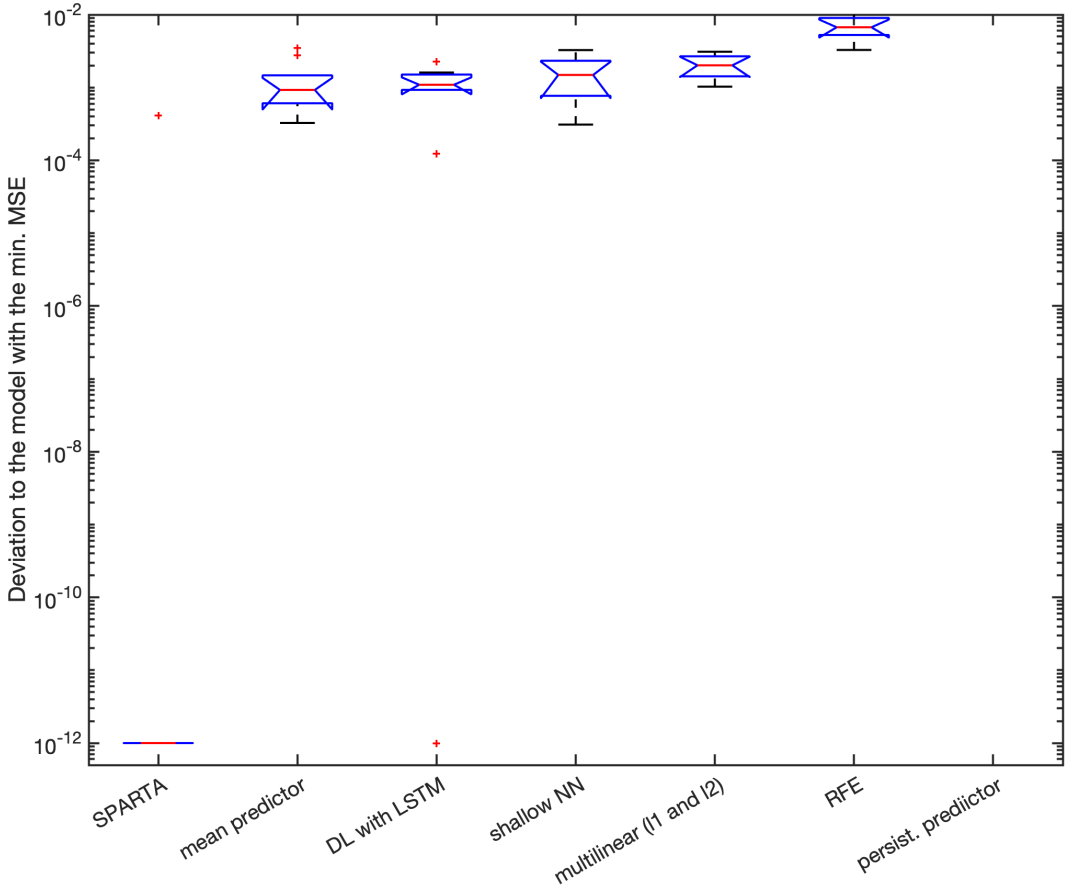
https://www.kaggle.com/datasets/sid321axn/gold-price-prediction-dataset

The dataset has 1718 rows in total and 80 columns in total. Data for attributes, such as Oil Price, Standard and Poor's (S&P) 500 index, Dow Jones Index US Bond rates (10 years), Euro USD exchange rates, prices of precious metals Silver and Platinum and other metals such as Palladium and Rhodium, prices of US Dollar Index, Eldorado Gold Corporation and Gold Miners ETF were gathered.

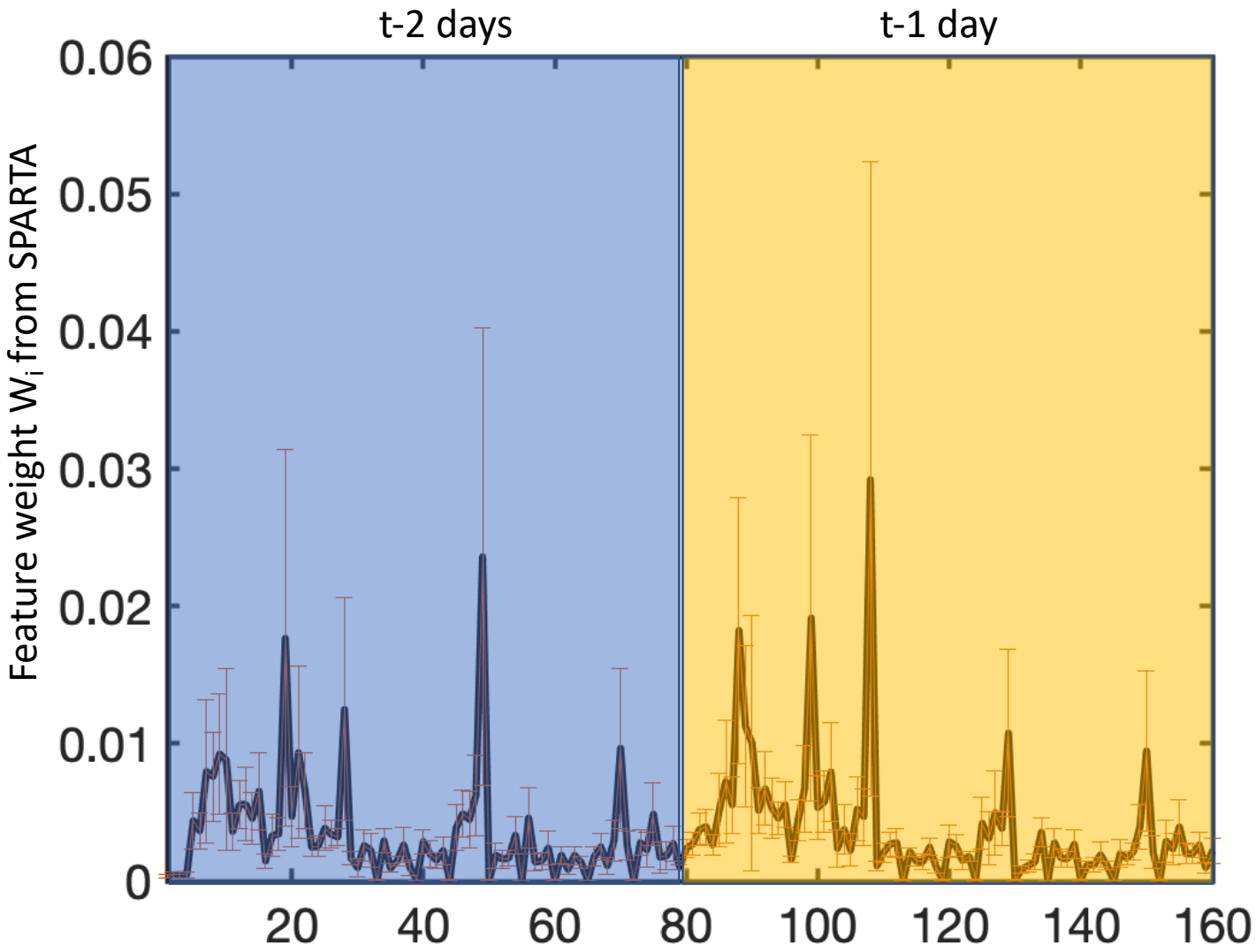
The historical data of Gold ETF fetched from Yahoo finance has 7 columns, Date, Open, High, Low, Close, Adjusted Close, and Volume, the difference between Adjusted Close and Close is that the closing price of a stock is the price of that stock at the close of the trading day. Whereas the adjusted closing price takes into account factors such as dividends, stock splits, and new stock offerings to determine a value. So, Adjusted Close is the outcome variable which is the value you have to predict.

Application example 2: predicting interday absolut change of price of S&P $|S_t^{close} - S_t^{open}|$. Data from Kaggle.com.



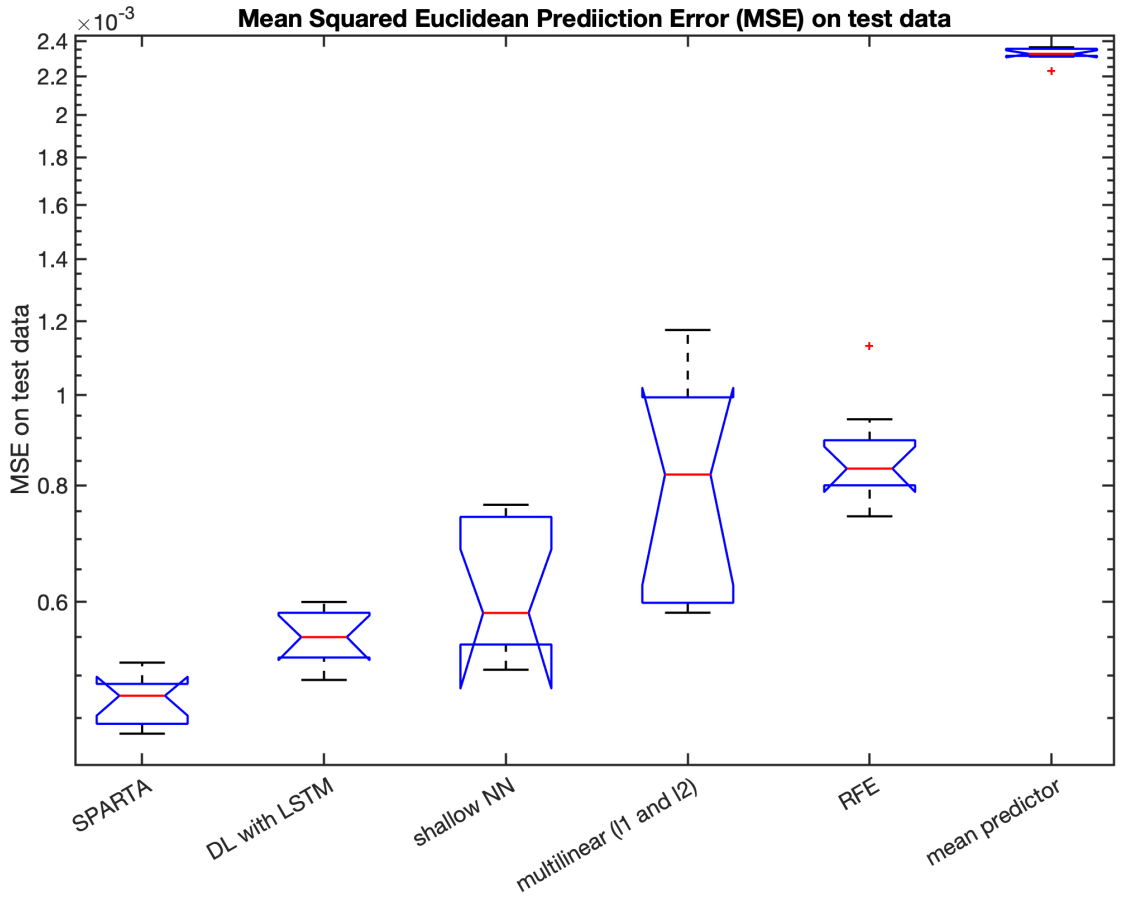


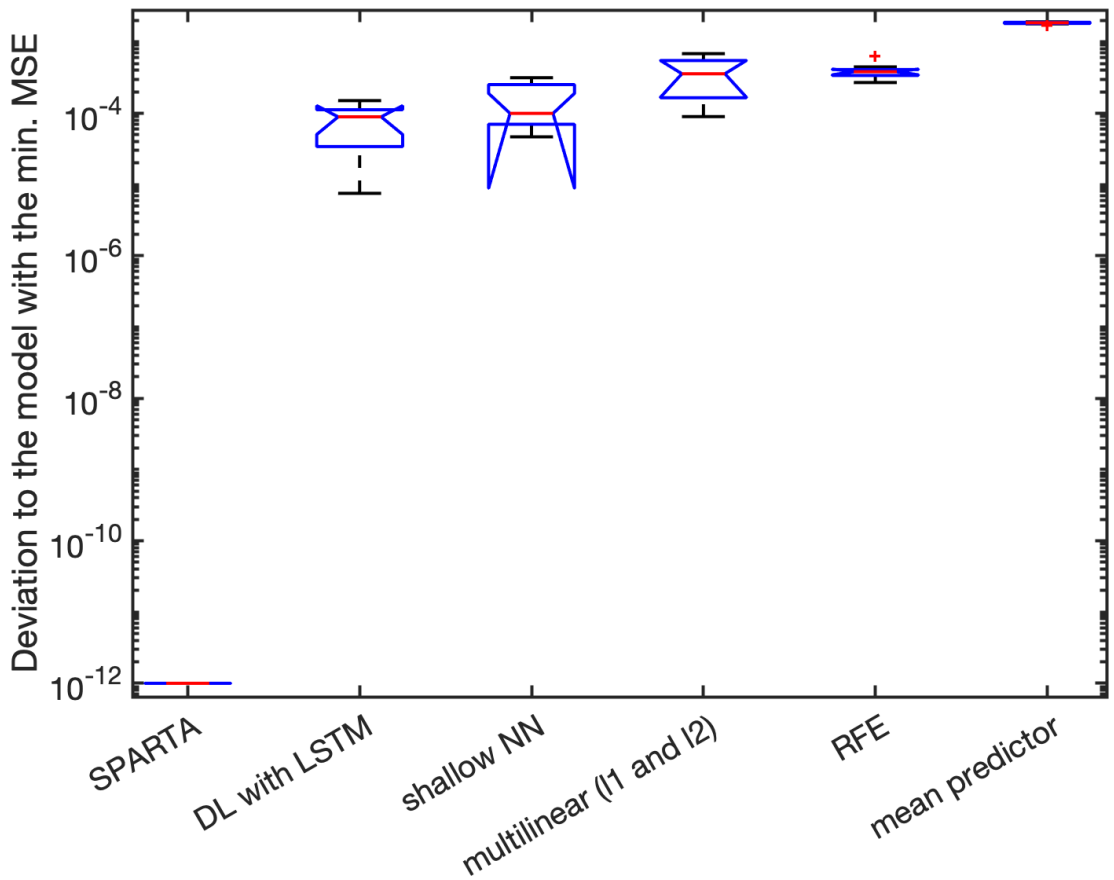
Application example 2: predicting interday absolut change of price of S&P $|S_t^{close} - S_t^{open}|$. Data from Kaggle.com.



Application example 3: predicting house prices. Data from 23'000 houses in King County in US between 2014 and 2015from Kaggle.com.

https://www.kaggle.com/datasets/harlfoxem/housesalesprediction





Application example 4: predicting El Nino Southern Oscillation (ENSO) with a 12 months of lead time.
ENSO is one of the dominant climate indicators and was shown to have a major impact on commodity prices.

Data from Kitsios et- al. " Forecasting commodity returns by exploiting climate model forecasts of the El Niño Southern Oscillation" Environmental Data Science (2022), 1: e7, 1–16doi:10.1017/eds.2022.6