ISTIC - IT & Electronics Department
Master EIT Digital in Data Science

UNIVERSITÉ DE RENNES 1

ARBITRATED ENSEMBLING

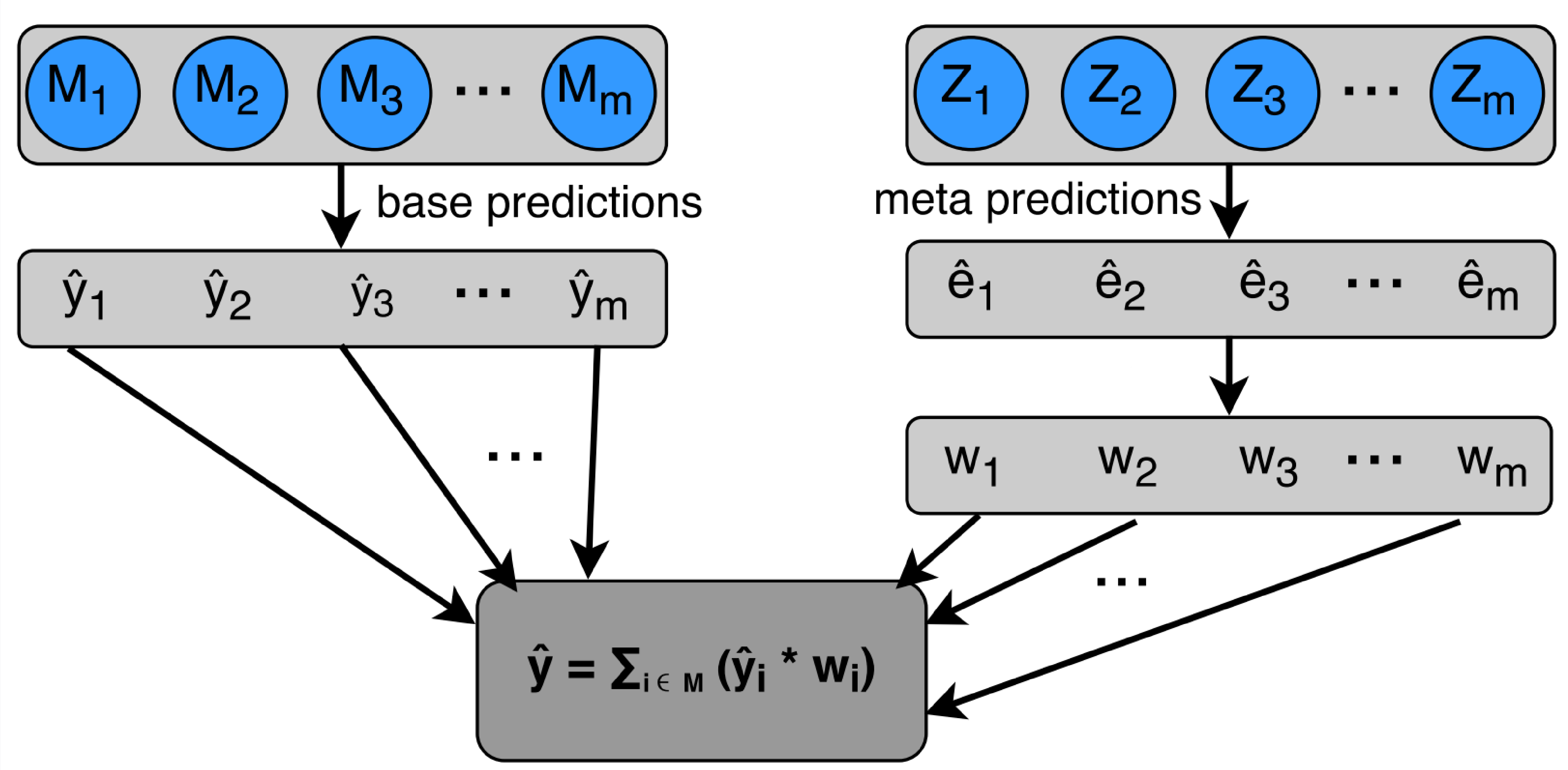
Supervisor: Simon Malinowski

Ostap Kharysh: ostap.kharysh@etudiant.univ-rennes1.fr
Ignacio Regana: ignacio.regana@etudiant.univ-rennes1.fr

SCOPE

- WHAT IS ARBITRATED DYNAMIC ENSEMBLE?
- REGRESSION:
  - WEIGHTING: EXPONENTIAL VS INVERSE
- CLASSIFICATION:
  - AVERAGING

ARCHITECTURE OF ADE



HOW WE SPLIT THE DATASET IN CHUNKS?

BETA = 3

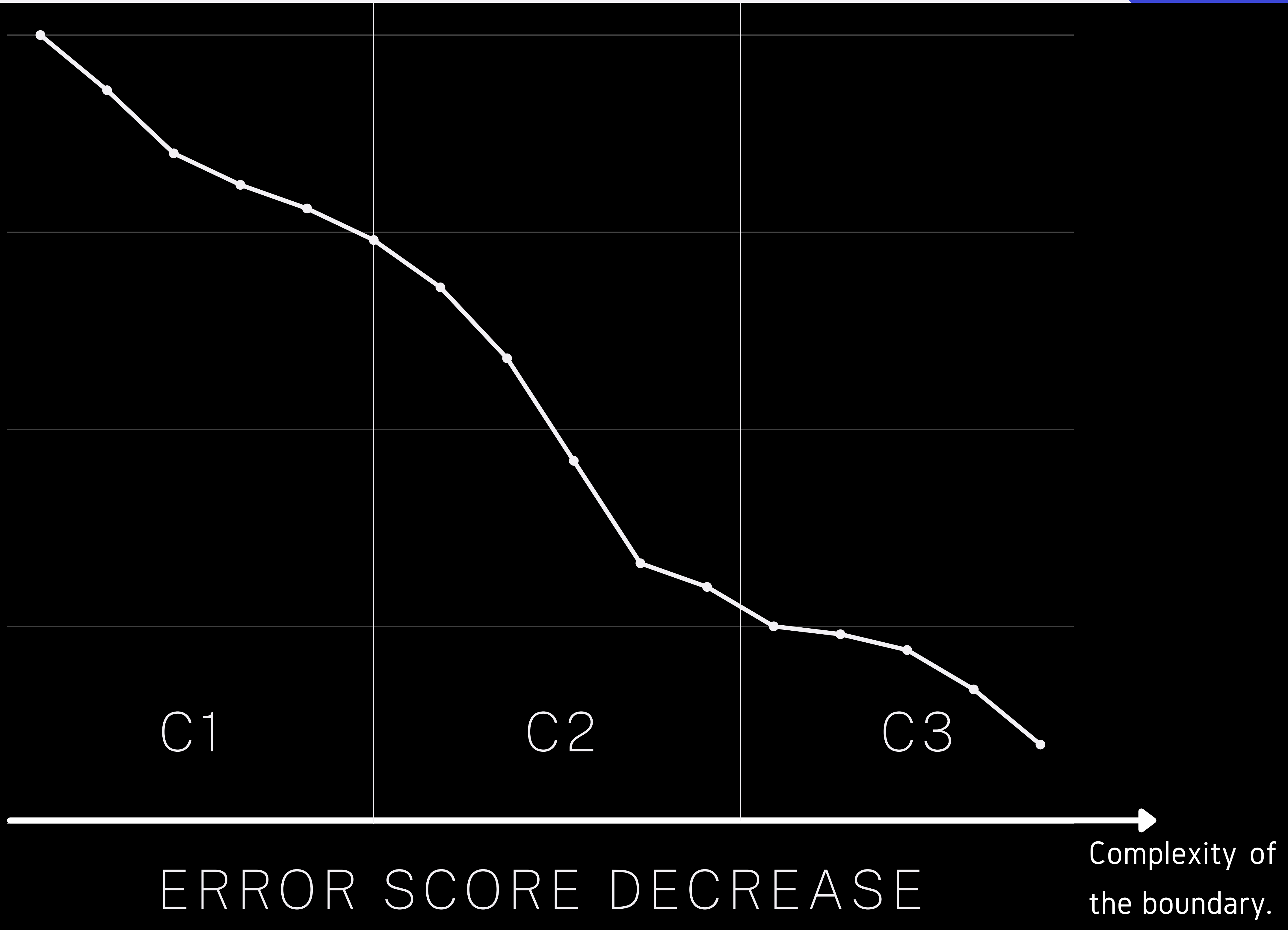
#2	#1	Chunk					
TRAIN	TRAIN	C1	1	X11	X12	X13	y1
			2	X21	X22	X23	y2
	TEST	C2	3	X31	X32	X33	y3
			4	X41	X42	X43	y4
TEST		C3	5	X51	X52	X53	y5
			6	X61	X62	X63	y6

M trained on C1 and predict C2

M trained on C1 + C2 and predict C3

M trained on C1 + ... + Cn-1 and predict Cn

WHAT DOES IT GIVE US?



ERROR SCORE DECREASE

REGRESSION

$$C_n \left\{ \begin{array}{cccc|c} \mathbf{1} & X11 & X12 & X13 & y1 \\ \mathbf{2} & X21 & X22 & X23 & y2 \\ \mathbf{3} & X31 & X32 & X33 & y3 \\ \mathbf{4} & X41 & X42 & X43 & y4 \end{array} \right.$$

j_{i1}

Base learner is: $j \in M$,

j - base - learner model

$\boxed{(y1 - \hat{y}1)^2 = \text{squared error for jn at observation t.}}$

$$\left| \begin{array}{c} ei1 \\ ei2 \\ ei3 \\ ei4 \end{array} \right|$$

REGRESSION

Cn					
1	X11	X12	X13	y1	$\hat{y}1$
2	X21	X22	X23	y2	$\hat{y}2$
3	X31	X32	X33	y3	$\hat{y}3$
4	X41	X42	X43	y4	$\hat{y}4$

$$\boxed{(y1 - \hat{y}1)^2 = e}$$

What do we need for metalearner?

e1	e2	e4 ...	en
0.5	0.3	0.2	0.4

This is fitted to our metalearner.

RFR

WHY WE CHOSE RANDOM FOREST REGRESOR (RFR)?

- GOOD PREDICTIVE PERFORMANCE OVER NOISY DATA.
- GOOD GENERALIZATION POTENTIAL.
- CONSISTENT FOR AVOIDING OVERFITTING.
- SUITABLE FOR DATASETS WITH HIGH NUMBER OF FEATURES.

REGRESSION PREDICTION

Meta-learner uses observation to predict an error for each base learner.

Meta-learner

Cn + 1

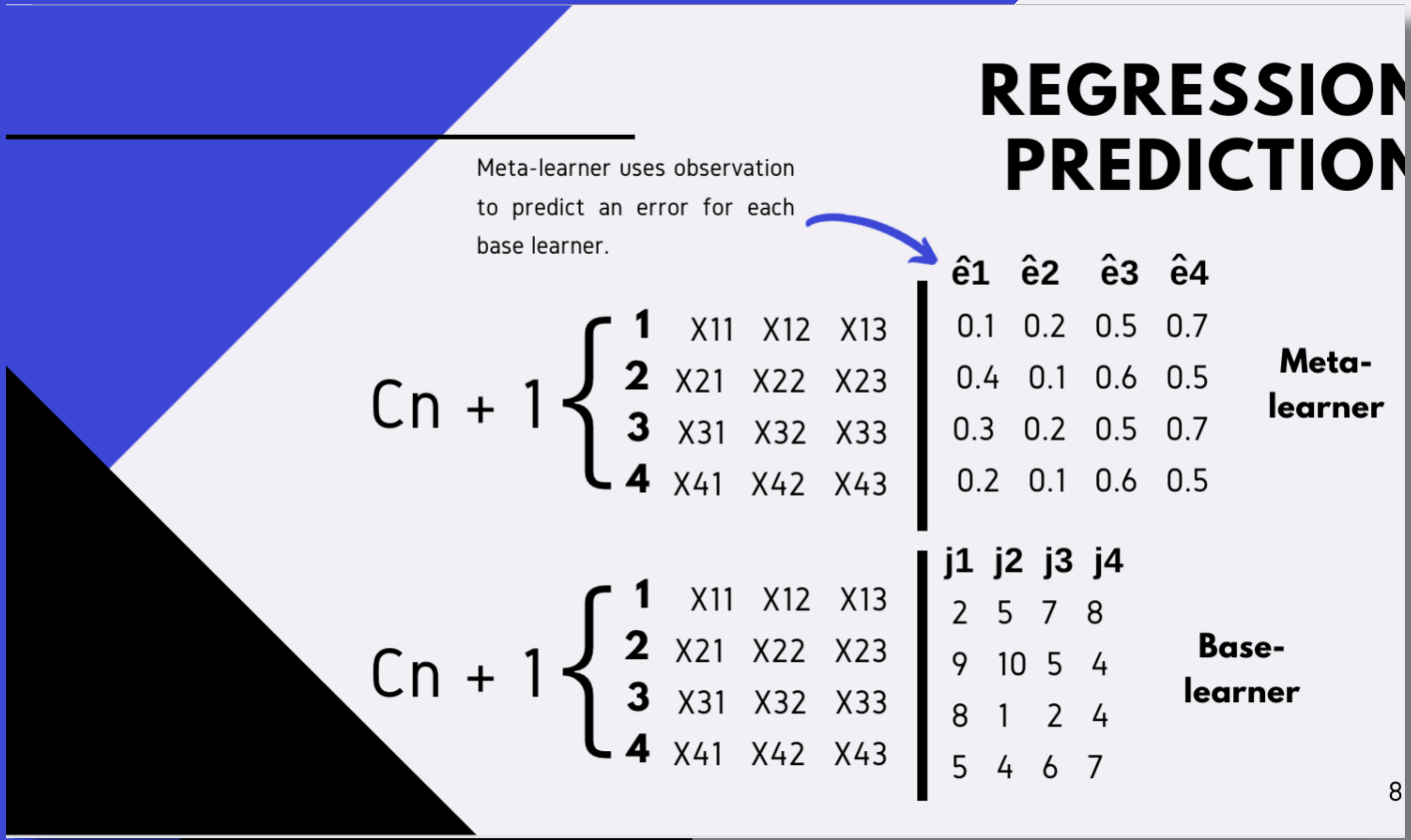
				ê1	ê2	ê3	ê4
1	X11	X12	X13	0.1	0.2	0.5	0.7
2	X21	X22	X23	0.4	0.1	0.6	0.5
3	X31	X32	X33	0.3	0.2	0.5	0.7
4	X41	X42	X43	0.2	0.1	0.6	0.5

Base-learner

Cn + 1

				j1	j2	j3	j4
1	X11	X12	X13	2	5	7	8
2	X21	X22	X23	9	10	5	4
3	X31	X32	X33	8	1	2	4
4	X41	X42	X43	5	4	6	7

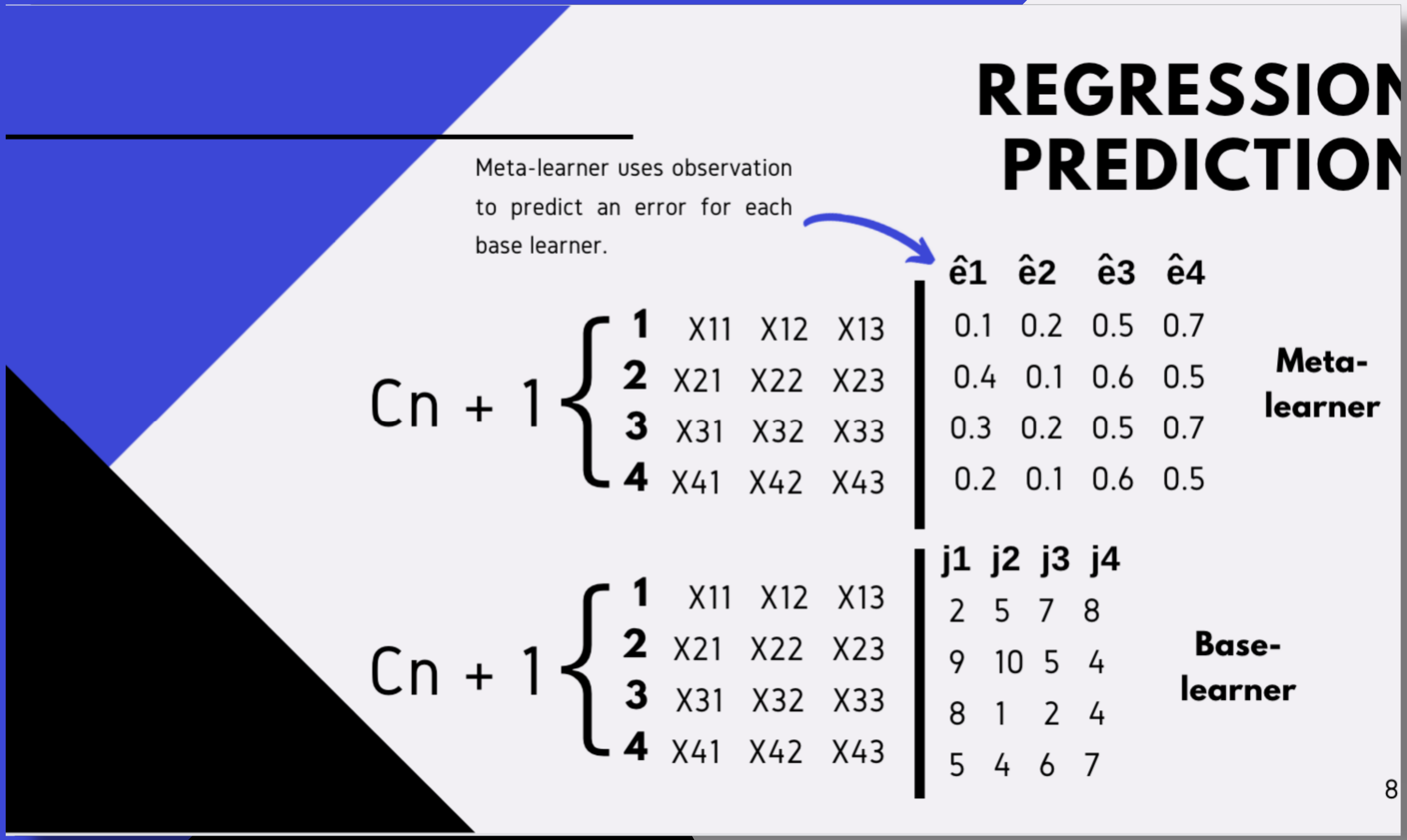
REGRESSION PREDICTION: EXPONENTIAL WEIGHTING



$$w^{j}{}_{t} = \frac{exp(-\hat{e_t}^{j})}{\sum_{j\epsilon M} exp(-\hat{e_t}^{j})}$$

$$W11 = \frac{\exp(-0.1)}{\exp(-0.1) + \exp(-0.2) + \exp(-0.5) + \exp(-0.7)}$$

W11 = 0.32



REGRESSION PREDICTION: EXPONENTIAL WEIGHTING

$$\hat{y}_t^{ADE} = \sum_{j \; \epsilon_M} \hat{y}_t^j w_t^j$$

Wjt = (0.32, 0.24, 0.27, 0.28)

$\hat{y}1^{ADE}$ = 2*0.32 + 5*0.24 + 7*0.27 + 8*0.28 = 6.09

$\hat{y}1...4^{ADE}$ = (6.09, 7.86, 4.5, 6.24)

REGRESSION PREDICTION: INVERSE WEIGHTING

PROPORTIONING

$$p_{jt} = \frac{\hat{e}_t}{\sum_{j\epsilon M} \hat{e}_t^{\,j}}$$

$$p1n = \frac{0.1}{0.1 + 0.2 + 0.5 + 0.7} = 0.066$$

p1...4n = (0.066, 0.25, 0.17, 0.14)

REGRESSION PREDICTION: INVERSE WEIGHTING

REGRESSIO
PREDICTIO

Meta-learner uses observation to predict an error for each base learner.

Meta-learner

					ê1	ê2	ê3	ê4
Cn + 1	1	X11	X12	X13	0.1	0.2	0.5	0.7
	2	X21	X22	X23	0.4	0.1	0.6	0.5
	3	X31	X32	X33	0.3	0.2	0.5	0.7
	4	X41	X42	X43	0.2	0.1	0.6	0.5

Base-learner

					j1	j2	j3	j4
Cn + 1	1	X11	X12	X13	2	5	7	8
	2	X21	X22	X23	9	10	5	4
	3	X31	X32	X33	8	1	2	4
	4	X41	X42	X43	5	4	6	7

8

$$p_{1j}{}^{-1}x_t + p_{2j}{}^{-1}x_t + p_{3j}{}^{-1}x_t + ... + p_{nj}{}^{-1}x_t = 1$$

p1...4n = (0.066, 0.25, 0.17, 0.14)

1

p1...4n^{-1} = (15, 4, 5.66, 7)

BALANCING

$$w_{jt} = p_{jt}^{-1}x_t$$

15*X +4*X + 5.66*X + 7*X = 1

X = 0.031

0.47 +0.13 + 0.18 + 0.22 = 1

REGRESSION PREDICTION: INVERSE WEIGHTING

$$\hat{y}_t^{ADE} = \sum_{j \, \epsilon_M} \hat{y}_t^j w_t^j$$

0.47 +0.13 + 0.18 + 0.22 = 1

$\hat{y}1^{ADE}$ = 0.47*2 +0.13*2 + 0.18*7 + 0.22*8 = 4.6

$\hat{y}1...4^{ADE}$ = (4.6, 7.3, 5.15, 5.49)

DATASETS

Regression Datasets		
Name	Attributes	# instances
Cancer	14	3047
Bike Sharing	15	17379
Insurance (Charge)	6	1338
Real Estate	6	415

Time series Datasets	
Name	# instances
Bike Sharing	17379
IBM stock price	1008
Ozon	518
Temperature	1461

EVALUATION

PRIORIZATION

As we evaluate the mean of the predicted error the lower value achieved the better.

Regression: Base-learners								
Dataset	SVR	RFR	DTR	LR	LAS	ELAS	BAY	ARD
Cancer	817.30	**477.36**	586.67	520.72	519.38	719.79	528.71	509.76
Bike Sharing	23306	**13118**	13881	22150	22159	26643	22152	22154
Insurance	1.64	1.33	1.35	**1.2713**	**1.2713**	1.30	1.2765	1.28
Real Estate	60.10	**58.51**	60.02	71.88	71.74	74.40	71.26	105.82

Regression: ADE								
Dataset	Ens3	Ens3P	ExEns3	ExEns3P	Ens5	Ens5P	ExEns5	ExEns5P
Cancer	582.04	519.29	604.34	**503.87**	626.60	708.31	584.01	708.31
Bike Sharing	22344	23719	26655	23719	**21292**	23508	26654	23508
Insurance	1.32	1.63	3.24	1.63	**1.29**	1.6484	3.24	1.6484
Real Estate	63.65	70.45	64.36	70.45	**60.71**	72.25	63.37	72.25

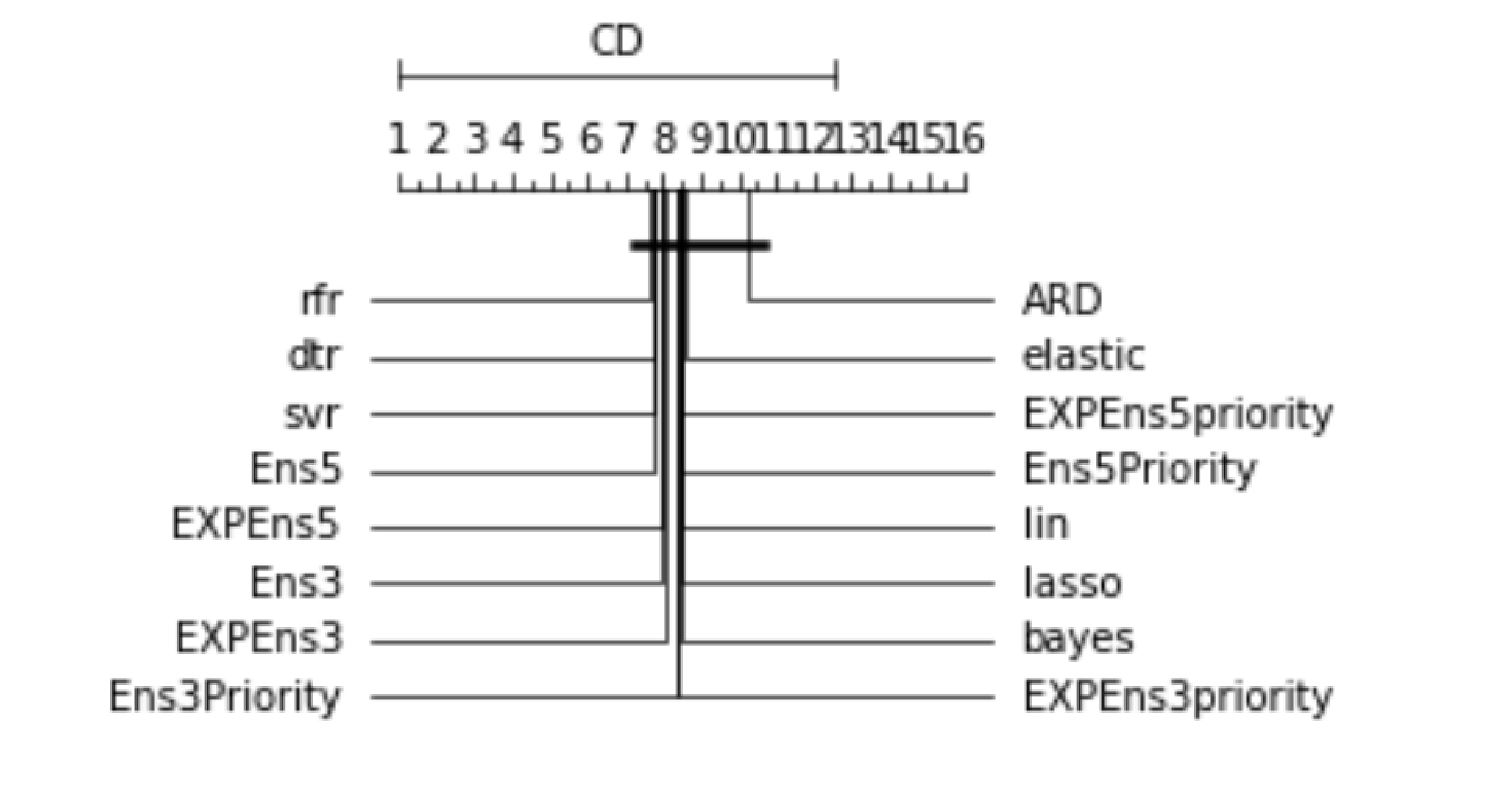
RESULTS

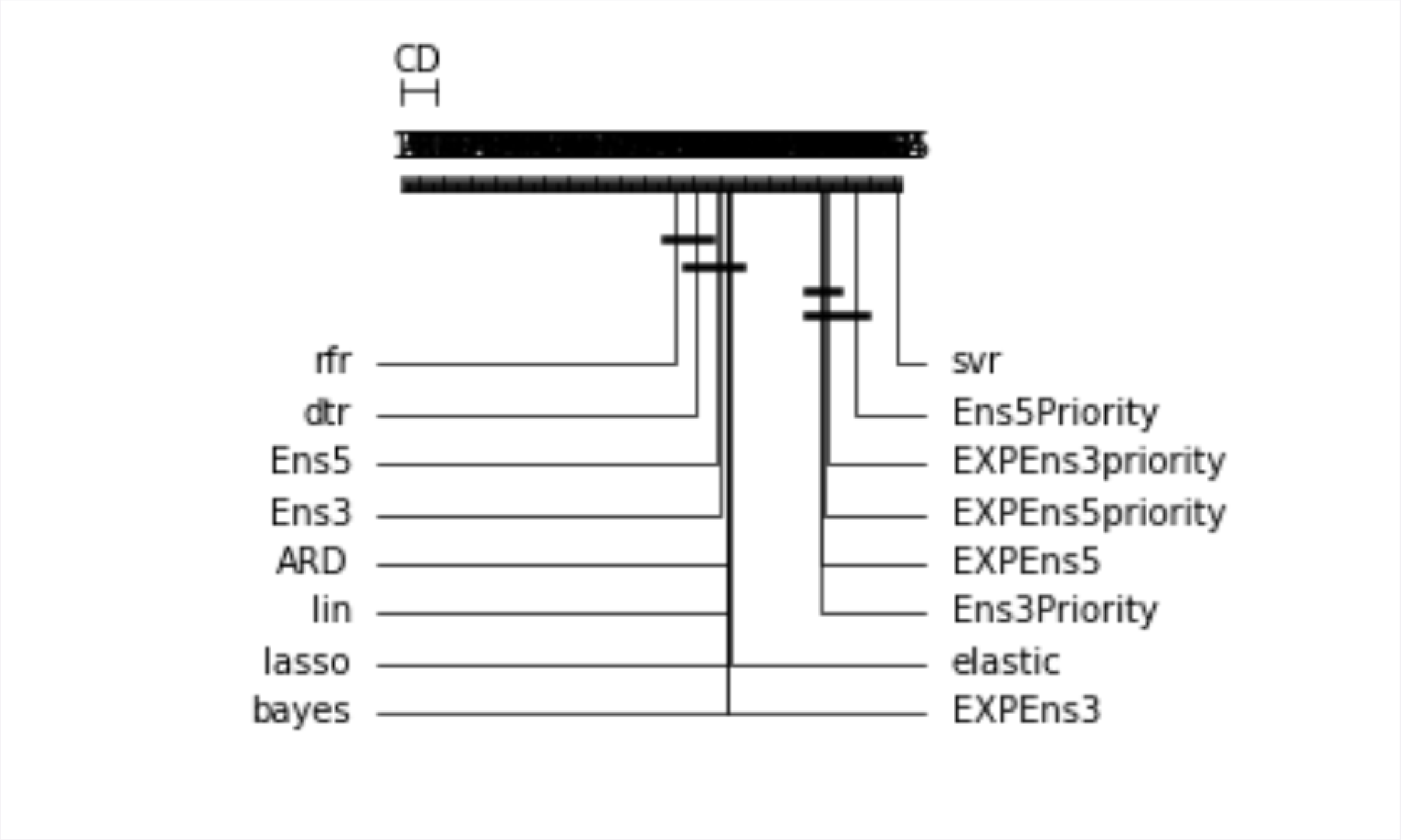
Results of the datasets used in the report are shown for regression and for time series.

Regression for Time series: Base-learners								
Dataset	SVR	RFR	DTR	LR	LAS	ELAS	BAY	ARD
Bike Sharing	26904	**8302**	9589	11711	11711	11792	11712	11710
IBM	44.38	27.46	58.88	**16.30**	16.31	16.56	16.35	16.35
Ozone	543.5	**358.5**	569.8	373.4	373.5	386.3	375.0	389.1
Temperature	22.74	17.86	21.06	**16.81**	16.85	16.97	16.85	16.87

Regression for Time series: ADE								
Dataset	Ens3	Ens3P	ExEns3	ExEns3P	Ens5	Ens5P	ExEns5	ExEns5P
Bike Sharing	11250	19070	11734	19731	**10869**	22414	19145	19378
IBM	17.11	119.50	16.94	119.50	**16.35**	44.29	16.33	44.29
Ozone	413.0	703.4	**396.3**	703.4	562.8	610.2	608.1	610.2
Temperature	20.92	26.03	24.45	26.03	**20.37**	23.07	22.56	23.07

NEMENYI TEST





REGRESSION

Mean errors with Real Estate dataset.

TIME SERIES

Mean errors with Bike Sharing dataset.

CLASSIFICATION

Cn+1

				j1	j2	j3	j4	z1	z2	z3	z4
1	X11	X12	X13	$\hat{y}1$	$\hat{y}1$	$\hat{y}1$	$\hat{y}1$	$\hat{e}i1$	$\hat{e}i1$	$\hat{e}i1$	$\hat{e}i1$
2	X21	X22	X23	$\hat{y}2$	$\hat{y}2$	$\hat{y}2$	$\hat{y}2$	$\hat{e}i2$	$\hat{e}i2$	$\hat{e}i2$	$\hat{e}i2$
3	X31	X32	X33	$\hat{y}3$	$\hat{y}3$	$\hat{y}3$	$\hat{y}3$	$\hat{e}i3$	$\hat{e}i3$	$\hat{e}i3$	$\hat{e}i3$
4	X41	X42	X43	$\hat{y}4$	$\hat{y}4$	$\hat{y}4$	$\hat{y}4$	$\hat{e}i4$	$\hat{e}i4$	$\hat{e}i4$	$\hat{e}i4$

Instances

base-learners predictions

meta-learners predictions

HOW TO CALCULATE THE ERROR?

CLASSIFICATION: ERROR CALCUALTION

$$\hat{y}_t^{ADE} = c,\ \ where\ \hat{e}_{tc} = \min\left(\frac{\sum_{j \epsilon MC} \hat{e}_{jc}}{N_{MC}},\ for\ each\ c\ \ where\ c\ \epsilon\ C\right)$$

			y
X11	X12	X13	**Dog**

	Cat	**Dog**	**Wolf**
j1	0.2	0.5	0.3
j2	0.4	0.3	0.3
j3	0.1	0.8	0.1
j4	0.2	0.3	0.5

$(1 - yj_{Dog})$ = error for jn at observation t.

Errors

j1	0.5
j2	0.7
j3	0.2
j4	0.3

			y	**e1**	**e2**	**e3**	**e4**
X11	X12	X13	**Dog**	0.5	0.7	0.2	0.3

CLASSIFICATION PREDICTION

$X11 \quad X12 \quad X13 \mid ?$

	$\hat{y}$	$\hat{e}$
j1	wolf	0.4
j2	dog	0.2
j3	cat	0.6
j4	dog	0.3
j5	wolf	0.4
j6	wolf	0.5

cat = [0.6] / Ncat = 0.6

dog = [0.2, 0.3] / Ndog = $\boxed{0.25}$

wolf = [0.4, 0.4, 0.5] / Nwolf = 0.43

rabbit, fox, squirrel = $\varnothing$

$\hat{y}1^{ADE}$ = dog

Dataset

Possible classes:

{cat, dog, wolf, rabbit, fox, squirrel}

EVALUATION

DATASETS

2 datasets for binary classification and 2 for multinomial classification.

Classification Datasets			
Name	Attributes	Classes	# Instances
Car	6	3	1728
Obesity	16	4	2211
Chess	36	2	3196
Tic-tac-toe	9	2	958

RESULTS

Results of the datasets used in the report are shown for classification. As we evaluate the accuracy the higher value achieved the better.

Classification: Base-learners								
Dataset	MLR	SVM	SGD	RFC	multNB	bernNB	KNN	ADA
Car	0.917	**0.973**	0.905	0.967	0.855	0.863	0.946	0.795
Obesity	0.831	0.534	0.556	**0.958**	0.611	0.556	0.910	0.317
Chess	**0.959**	0.946	0.956	0.970	0.849	0.850	0.924	0.957
Tic-tac-toe	**0.982**	0.975	0.979	0.972	0.677	0.666	0.968	0.857

Classification: ADE				
Dataset	Ens3	Ens3P	Ens5	Ens5P
Car	0.761	0.797	0.855	**0.963**
Obesity	0.446	0.902	0.798	**0.906**
Chess	0.871	0.965	0.912	**0.967**
Tic-tac-toe	**0.982**	**0.982**	0.975	0.947

istic Informatique Électronique