

Можно ли извлечь новые знания
из результатов работы ИИ?

7 июня 2019

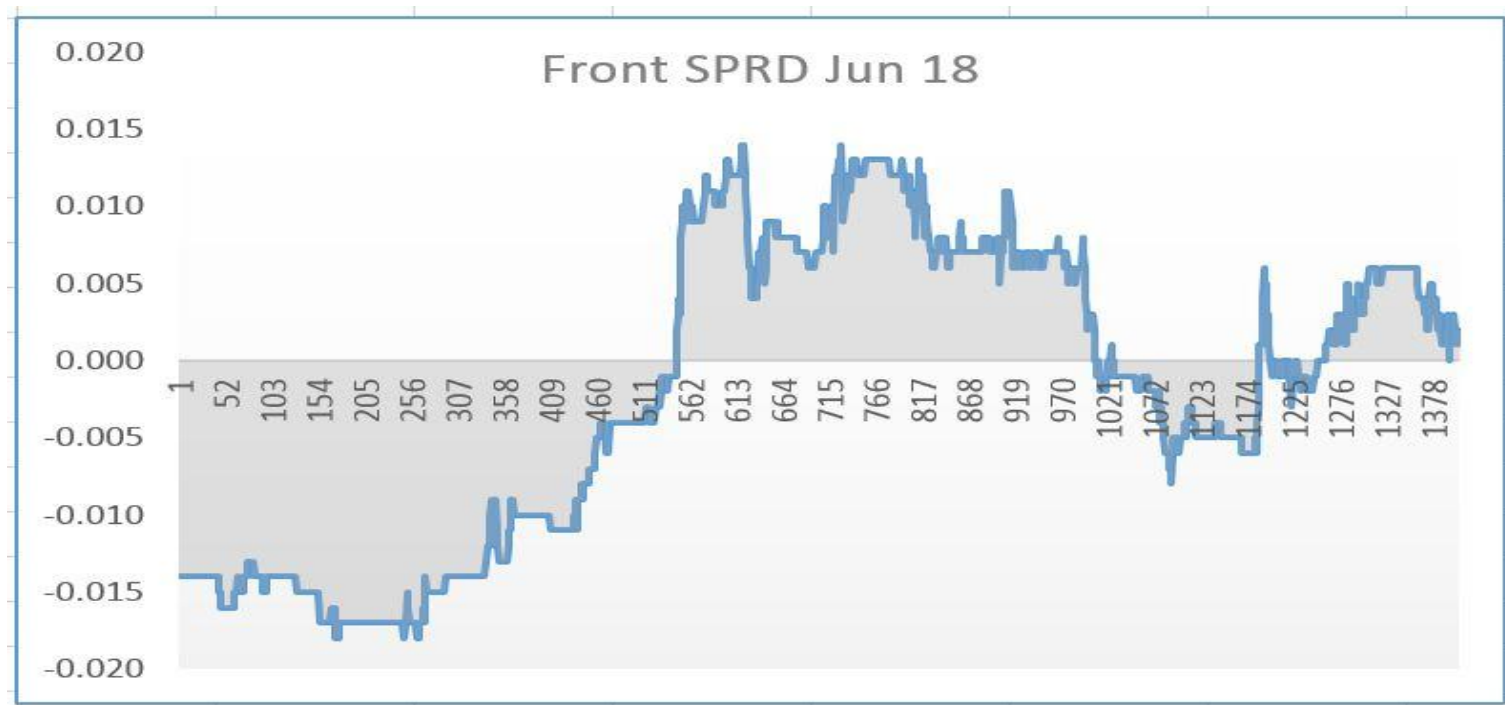
О чем мы не будем говорить...



Торговля commodity futures/derivates

Цель - предсказать тренд движения цены финансового инструмента в ближайшие 30 мин. Нейронная сеть

Тренд – {UP, DOWN, FLAT}



Classification and Clustering

Classification is a function that assigns items in a collection to target categories or classes.

The goal of **classification** is to accurately predict the target class for each case in the **data**.

Cluster:

A collection of data objects similar (or related) to one another within the same group dissimilar (or unrelated) to the objects in other groups

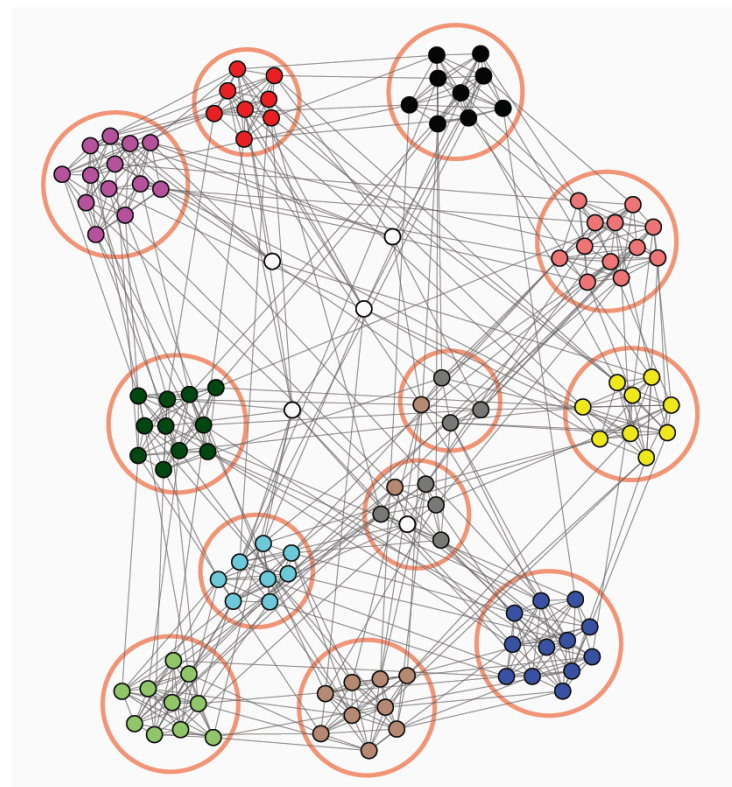
Cluster analysis

Finding similarities between data according to the characteristics found in the data and grouping similar data objects into clusters

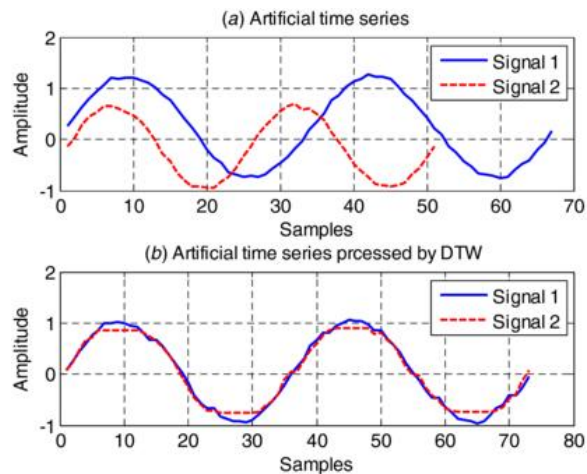
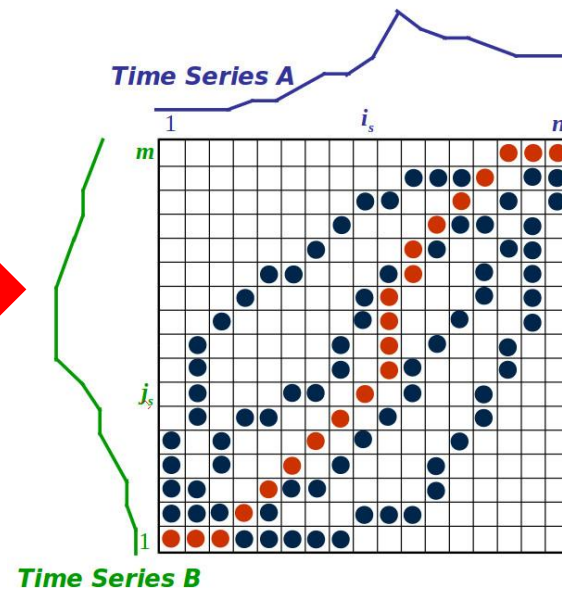
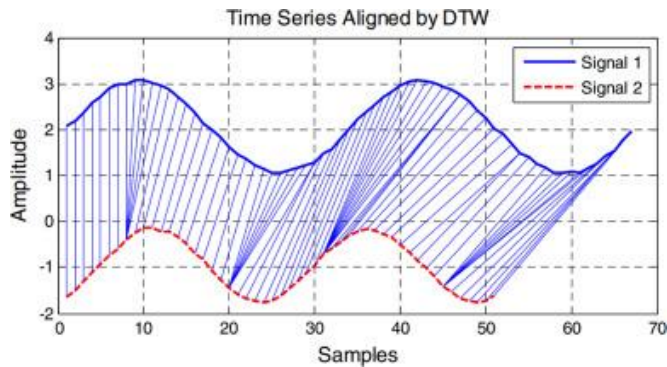
Данные проекта

Форма кривой (timeseries) предшествующей текущему моменту времени важный параметр.

Кривые распределяются по **кластерам**



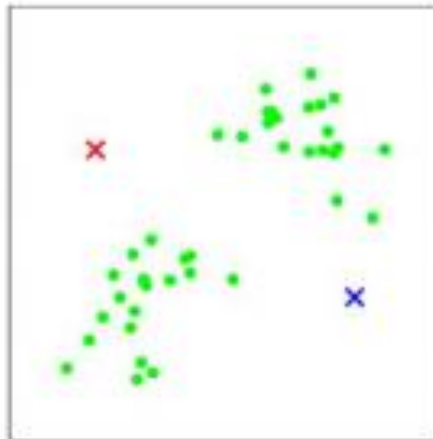
Метрика для кластера



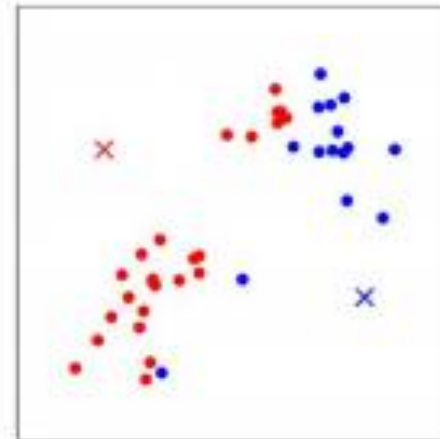
K-means метод



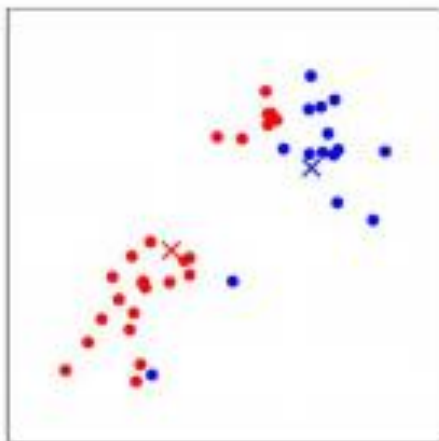
(a)



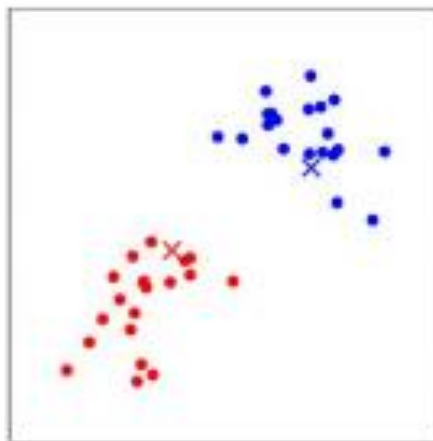
(b)



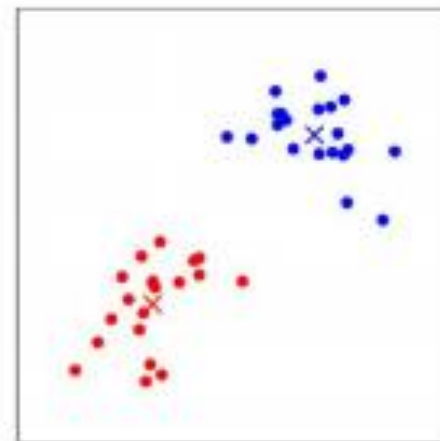
(c)



(d)



(e)



(f)

Результат clustering



Результат работы NN

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1																	
2																	
3						18	117	90%	98%	100%							
4						Loss:	0.27				1,828						
5						Accuracy:	73.00%				135						
6						Mae/Mse	0.46	0.819			71						
7						1,363	116	349									
8						59	66	10									
9						21	-	50									
10						56%	25%	92%									
11						44%	75%	8%									
12																	
13																	
14																	
15																	
16																	
17																	
18																	
19																	
20																	
21																	
22																	
23																	
24																	
25																	
26																	
27																	
28																	
29																	

True positive rate (TPR),
probability of detection =
 $\frac{\sum \text{True positive}}{\sum \text{Condition positive}}$

False negative rate (FNR), Miss rate =
 $\frac{\sum \text{False negative}}{\sum \text{Condition positive}}$

Specificity (SPC), Selectivity, True
negative rate (TNR) =
 $\frac{\sum \text{True negative}}{\sum \text{Condition negative}}$

False positive rate (FPR), Fall-out,
probability of false alarm =
 $\frac{\sum \text{False positive}}{\sum \text{Condition negative}}$

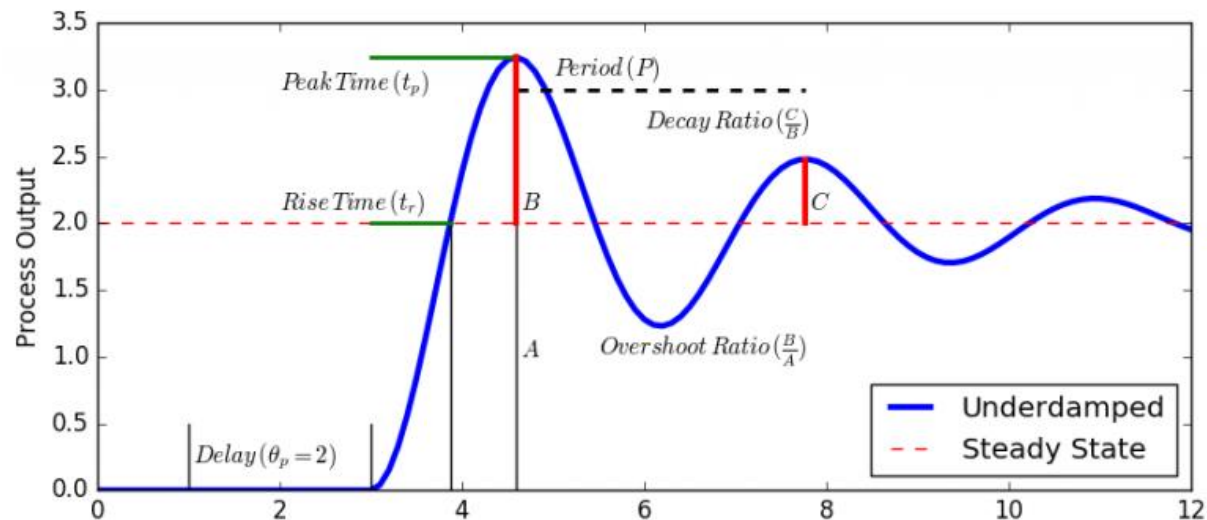
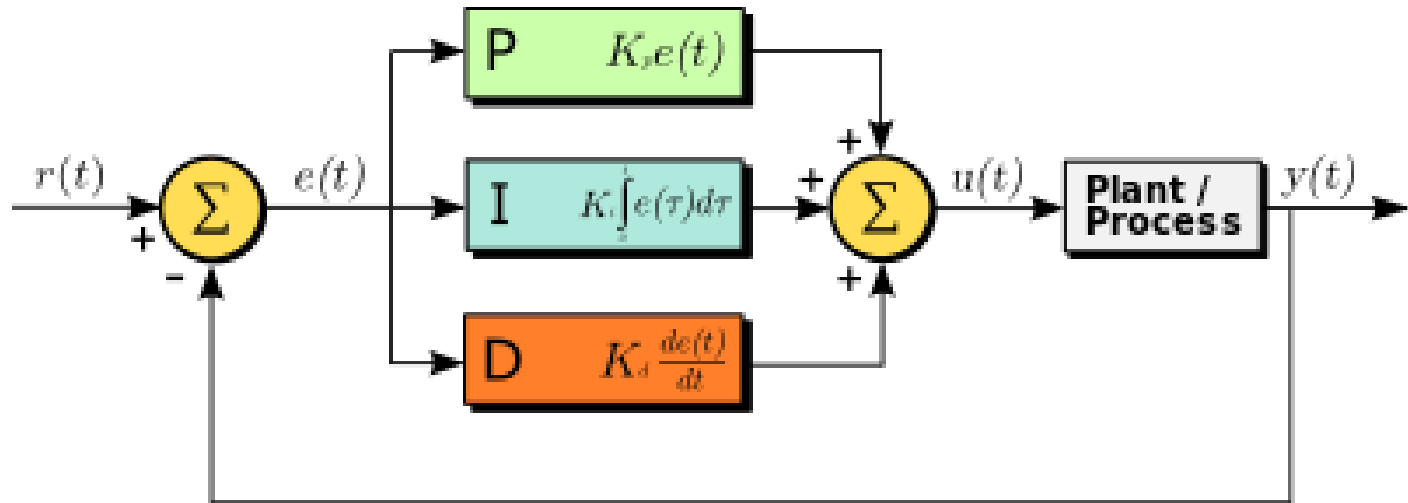
False discovery rate (FDR) =
 $\frac{\sum \text{False positive}}{\sum \text{Predicted condition positive}}$

Negative predictive value (NPV) =
 $\frac{\sum \text{True negative}}{\sum \text{Predicted condition negative}}$

Positive predictive value (PPV),
Precision =
 $\frac{\sum \text{True positive}}{\sum \text{Predicted condition positive}}$

False omission rate (FOR) =
 $\frac{\sum \text{False negative}}{\sum \text{Predicted condition negative}}$

Управление отоплением дома



Управление отоплением дома

PI Control:

$$G_c(s) = K_c \left(1 + \frac{1}{\tau_I s} \right)$$

$$Y(s) = \frac{8s}{s(s+2)^3 + \frac{8K_c}{\tau_I} + 8K_c s} \cdot \frac{1}{s} \quad \lim_{s \rightarrow 0} sY(s) = 0 \quad \text{no offset}$$

adjust K_c and τ_I to obtain satisfactory response (roots of equation which is 4th order)

PID Control: (pure PID)

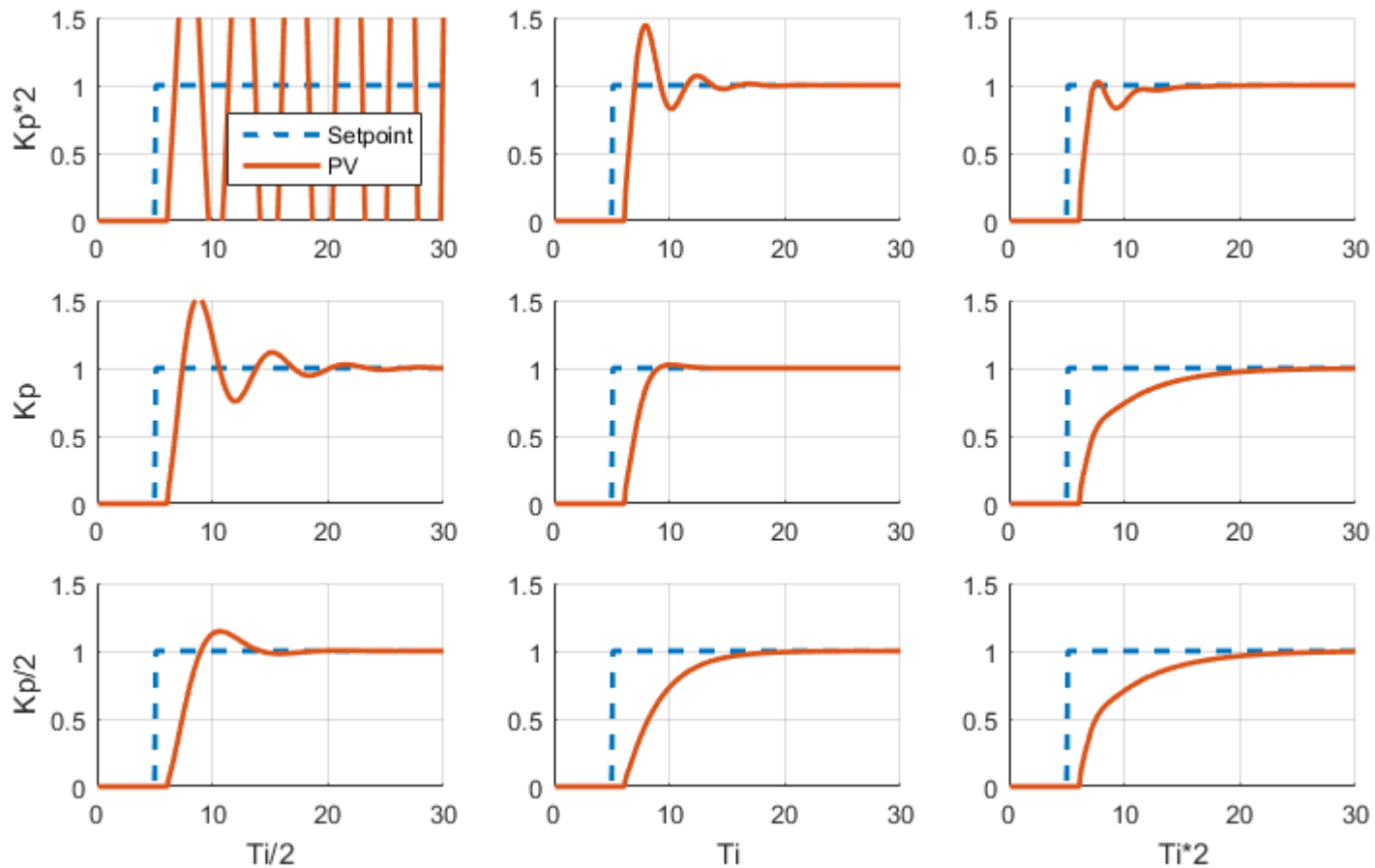
$$G_c(s) = K_c \left(1 + \frac{1}{\tau_I s} + \tau_D s \right)$$

No offset, adjust K_c , τ_I , τ_D to obtain satisfactory result (requires solving for roots of 4th order characteristic equation).

∴ Analysis of roots of characteristic equation is one way to analyze controller behavior

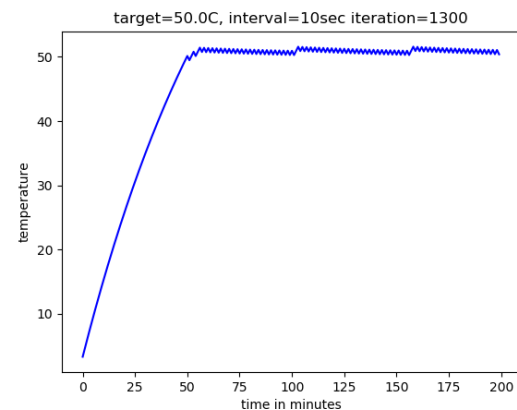
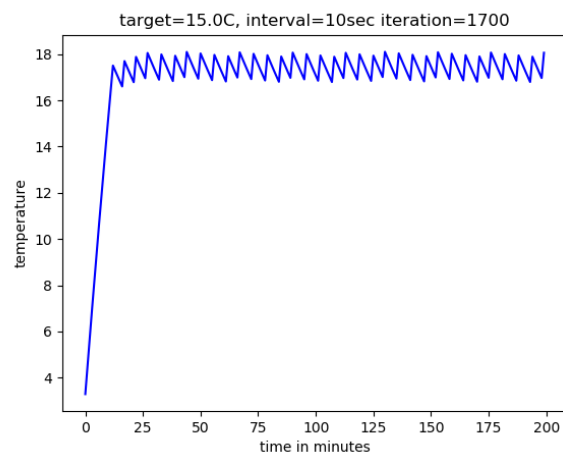
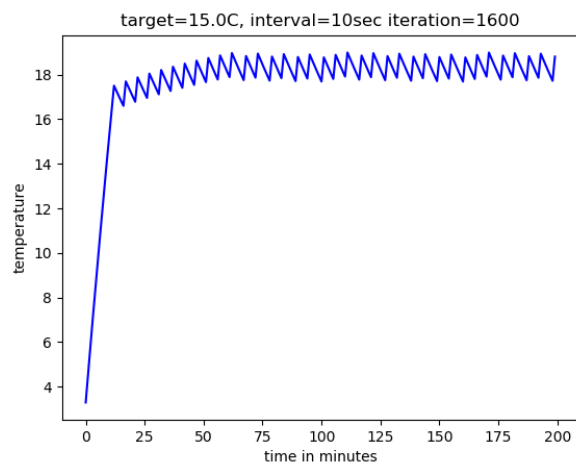
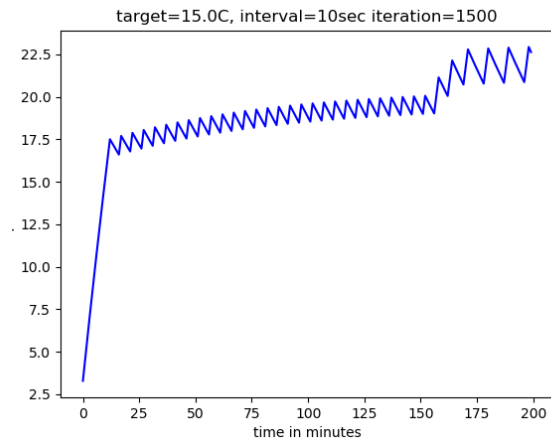
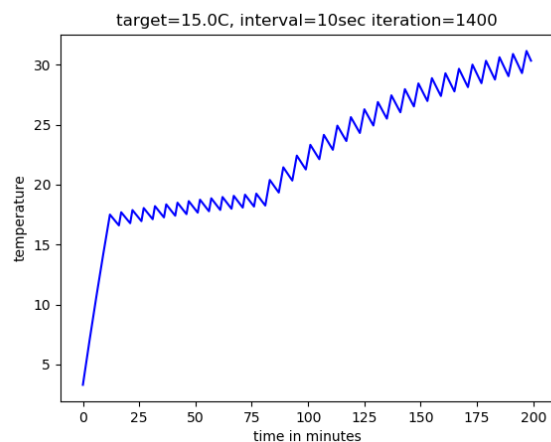
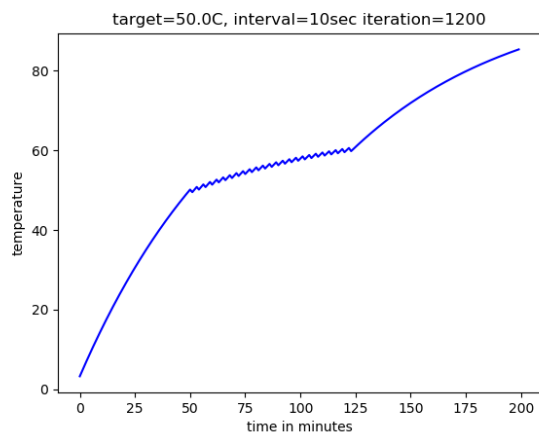
$$1 + G_c G_v G_p G_m = 0$$

Управление отоплением дома

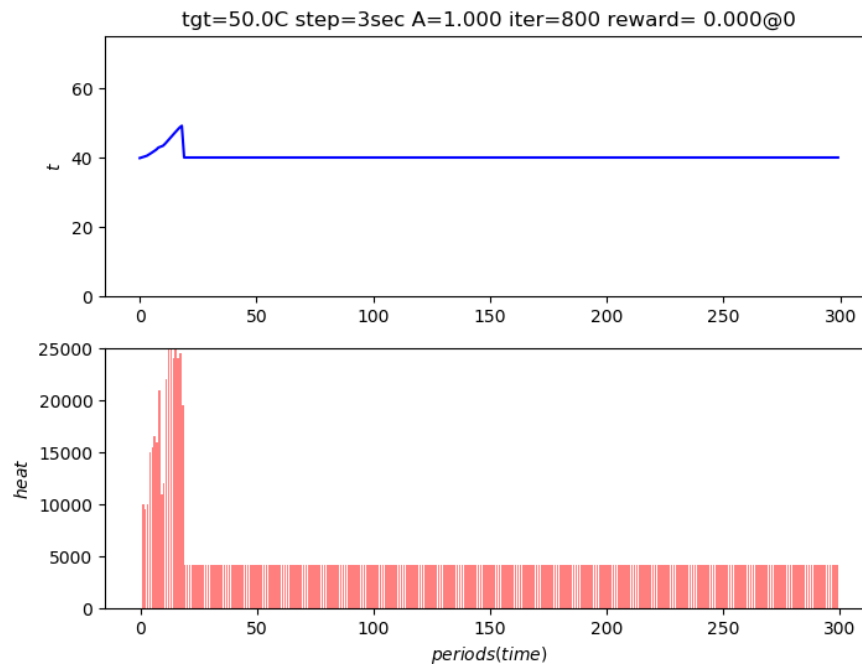
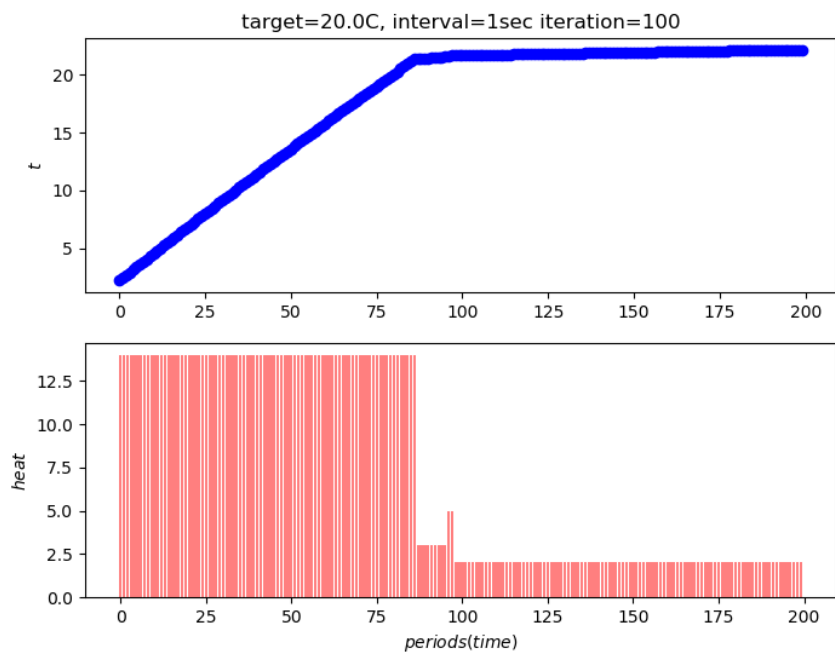


Управление отоплением дома

Средства RL – gradient policy agent. Самообучение



Управление отоплением дома



Медицинский проект на примере Титаника



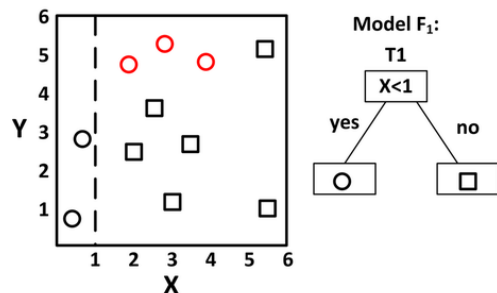
Титаник: Данные о пассажирах

Пассажир	Выжил	Класс	Имя	Пол	Возраст	Родств	Дети	Билет	Цена	Каюта	Порт
1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.25		S
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599	71.28	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26	0	0	O2. 3101282	7.93		S
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.10	C123	S
5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.05		S
6	0	3	Moran, Mr. James	male		0	0	330877	8.46		Q
7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.86	E46	S
8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909	21.08		S
9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742	11.13		S
10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736	30.07		C
11	1	3	Sandstrom, Miss. Marguerite Rut	female	4	1	1	PP 9549	16.70	G6	S
12	1	1	Bonnell, Miss. Elizabeth	female	58	0	0	113783	26.55	C103	S
13	0	3	Saunderscock, Mr. William Henry	male	20	0	0	A/5. 2151	8.05		S
14	0	3	Andersson, Mr. Anders Johan	male	39	1	5	347082	31.28		S
15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14	0	0	350406	7.85		S
16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55	0	0	248706	16.00		S
17	0	3	Rice, Master. Eugene	male	2	4	1	382652	29.13		Q
18	1	2	Williams, Mr. Charles Eugene	male		0	0	244373	13.00		S
19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vandemoortel)	female	31	1	0	345763	18.00		S
20	1	3	Masselmani, Mrs. Fatima	female		0	0	2649	7.23		C
21	0	2	Fynney, Mr. Joseph J	male	35	0	0	239865	26.00		S
22	1	2	Beesley, Mr. Lawrence	male	34	0	0	248698	13.00	D56	S
23	1	3	McGowan, Miss. Anna "Annie"	female	15	0	0	330923	8.03		Q
24	1	1	Sloper, Mr. William Thompson	male	28	0	0	113788	35.50	A6	S
25	0	3	Palsson, Miss. Torborg Danira	female	8	3	1	349909	21.08		S

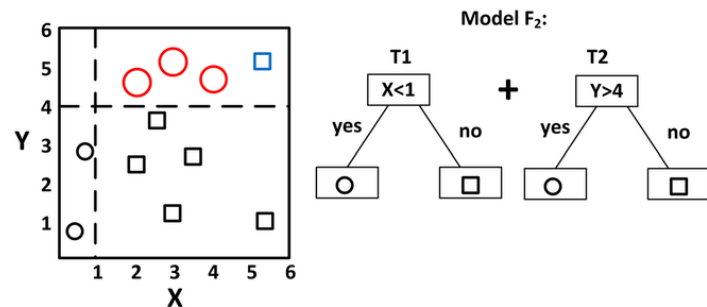
Титаник: метод ИИ

gradient boost классификатор

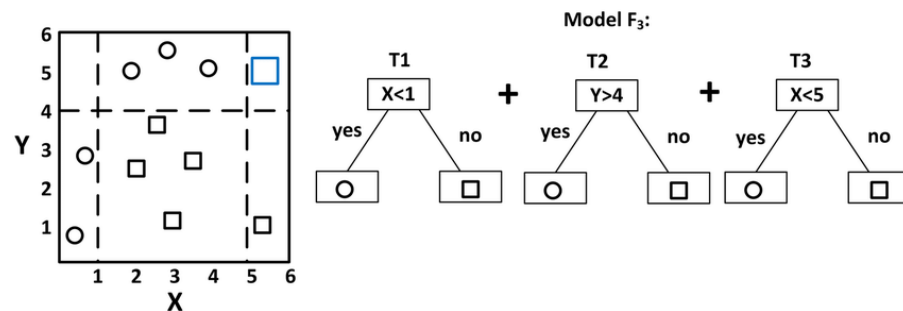
Iteration 1



Iteration 2



Iteration 3

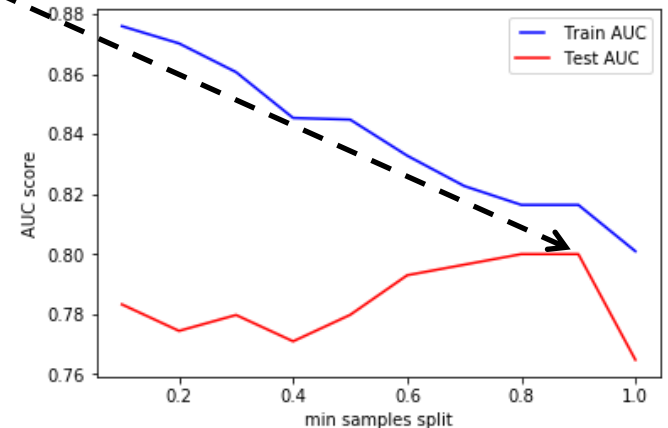
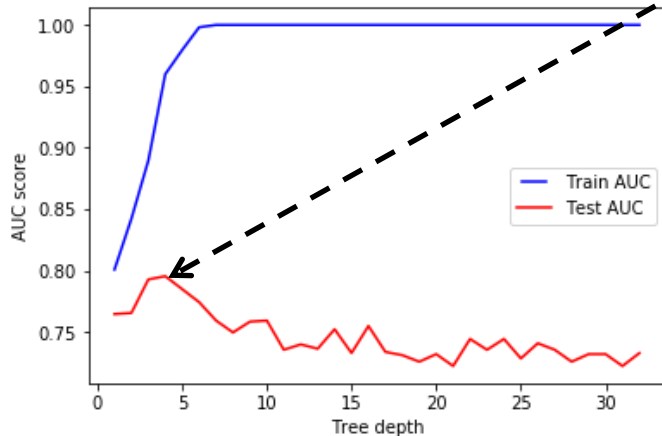
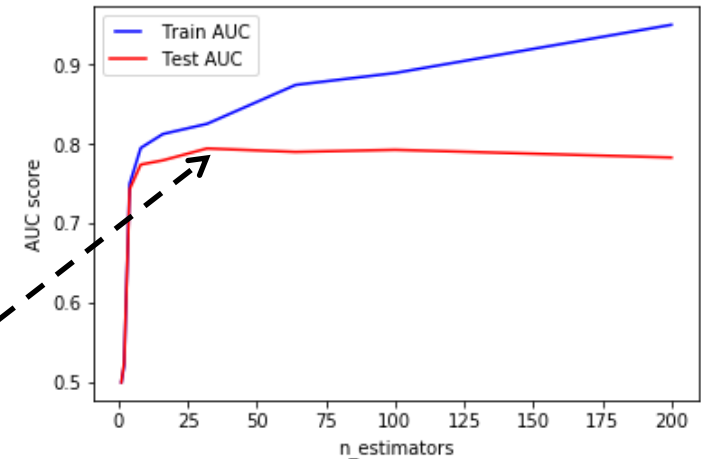
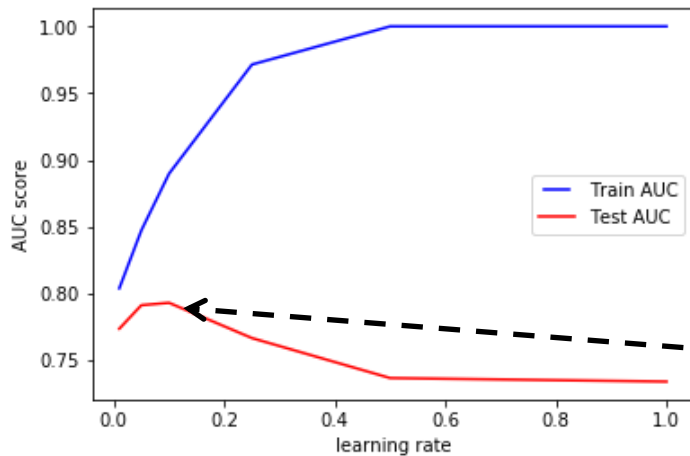


Gradient boosting - это один из методов machine learning. Используется в задачах регрессии и классификации. Состоит в построении предсказательной модели в форме подмножества «слабых» предсказательных моделей (обычно в форме деревьев решений)

Аккуратность предсказаний $\approx 80\%$

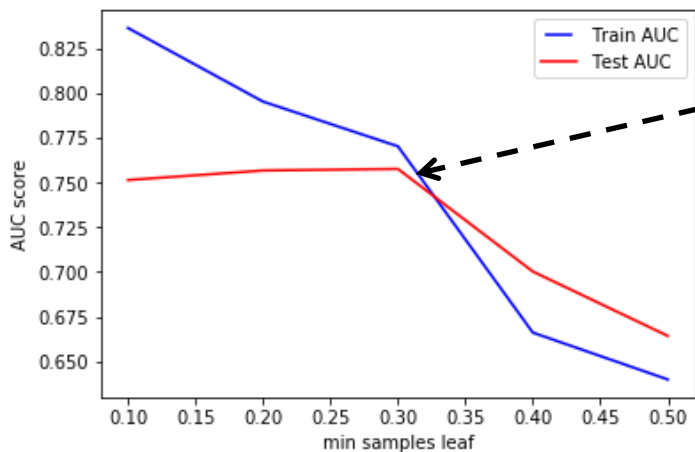
Титаник: hyper parameters

- ❑ **learning_rate** - шаг сдвига по градиенту
- ❑ **n_estimators** - количество деревьев решений в «лесу деревьев»
- ❑ **max_depth** - максимальная глубина дерева
- ❑ **min_samples_split** - min число образцов, необходимых для расщепления дерева

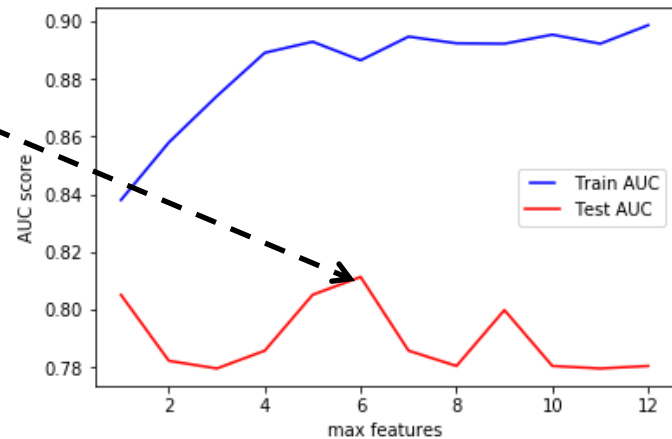


Титаник: hyper parameters

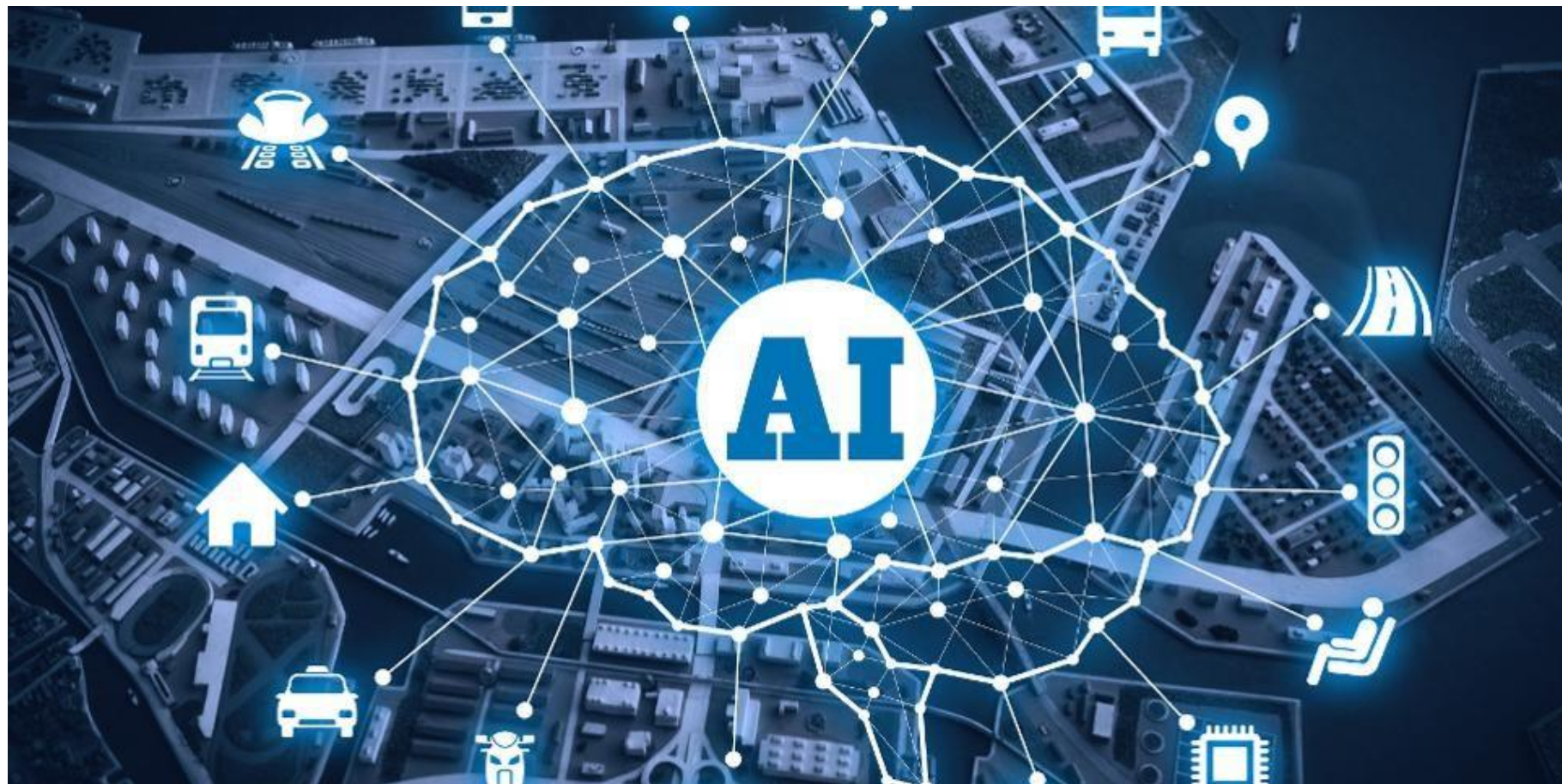
- ❑ **min_samples_leaf** - min число образцов на листе дерева
- ❑ **max_features** - max количество категорий при поиске лучшего расщепления дерева



best



Выводы



Контакты:

Геннадий Суворов

OSTC Ltd., www.ostc.com

трейдер-аналитик

тел: +7 925 800 7797

мейл: absfinance@gmail.com