

Исходные данные:

№пп	y	x1	x2	x3	x4
1	0,904	75,5	25,2	3343	77
2	0,922	78,5	21,8	3001	78,2
3	0,763	78,4	25,7	3101	68
4	0,923	77,7	17,8	3543	77,2
5	0,918	84,4	15,9	3237	77,2
6	0,906	75,9	22,4	3330	77,2
7	0,905	76	20,6	3808	75,7
8	0,545	67,5	25,2	2415	62,6
9	0,894	78,2	20,7	3295	78
10	0,9	78,1	17,5	3504	78,2
11	0,932	78,6	19,7	30565	79
12	0,74	84	18,5	3007	67,6
13	0,701	59,2	54,4	2844	69,8
14	0,744	90,2	23	2861	68,4
15	0,921	72,8	20,2	3259	77,9
16	0,927	67,7	25,2	3350	78,1
17	0,802	82,6	22,4	3344	72,5
18	0,747	74,4	22,7	2704	66,6
19	0,927	83,3	18,1	3642	76,7
20	0,721	83,7	20,1	2753	68,8
21	0,913	73,8	17,3	2916	76,8
22	0,918	79,2	16,8	3551	78,1
23	0,833	71,5	29,9	3177	73,9
24	0,914	75,3	20,3	3280	78,6
25	0,923	79	14,1	3160	78,5

1-12 – значения для тренировки модели

13-20 – значения для отбора по МГУА

21-25 – тестовая выборка

Для подбора коэффициентов опорной функции решим систему:

$$\begin{cases} na_0 + a_1 \sum x_1 + a_2 \sum x_2 = \sum y \\ \sum x_1 + \sum x_1^2 + \sum x_1 x_2 = \sum x_1 y \\ \sum x_2 + \sum x_1 x_2 + \sum x_2^2 = \sum x_2 y \end{cases}$$

Для основных математических расчётов я воспользовался программой Microsoft Office Excel.

x1^2	x2^2	x3^2	x4^2	x1x2	x1x3	x1x4	x2x3	x2x4	x3x4
5700,25	635,04	11175649	5929	1902,6	252396,5	5813,5	84243,6	1940,4	257411
6162,25	475,24	9006001	6115,24	1711,3	235578,5	6138,7	65421,8	1704,76	234678,2
6146,56	660,49	9616201	4624	2014,88	243118,4	5331,2	79695,7	1747,6	210868
6037,29	316,84	12552849	5959,84	1383,06	275291,1	5998,44	63065,4	1374,16	273519,6
7123,36	252,81	10478169	5959,84	1341,96	273202,8	6515,68	51468,3	1227,48	249896,4
5760,81	501,76	11088900	5959,84	1700,16	252747	5859,48	74592	1729,28	257076
5776	424,36	14500864	5730,49	1565,6	289408	5753,2	78444,8	1559,42	288265,6
4556,25	635,04	5832225	3918,76	1701	163012,5	4225,5	60858	1577,52	151179
6115,24	428,49	10857025	6084	1618,74	257669	6099,6	68206,5	1614,6	257010
6099,61	306,25	12278016	6115,24	1366,75	273662,4	6107,42	61320	1368,5	274012,8
6177,96	388,09	9,34E+08	6241	1548,42	2402409	6209,4	602130,5	1556,3	2414635
7056	342,25	9042049	4569,76	1554	252588	5678,4	55629,5	1250,6	203273,2
72711,58	5366,66	1,05E+09	67207,01	19408,47	5171083	69730,52	1345076	18650,62	5071825

yx1	yx2	yx3	yx4
68,252	22,7808	3022,072	69,608
72,377	20,0996	2766,922	72,1004
59,8192	19,6091	2366,063	51,884
71,7171	16,4294	3270,189	71,2556
77,4792	14,5962	2971,566	70,8696
68,7654	20,2944	3016,98	69,9432
68,78	18,643	3446,24	68,5085
36,7875	13,734	1316,175	34,117
69,9108	18,5058	2945,73	69,732
70,29	15,75	3153,6	70,38
73,2552	18,3604	28486,58	73,628
62,16	13,69	2225,18	50,024
799,5934	212,492	58987,3	772,0503
7			

Системы уравнений для каждой комбинации будем решать с помощью матричного метода.

Как известно:

$$AX=B \rightarrow X=A^{-1} \times B.$$

Решения систем уравнений:

1	12	932,8	251	10,252	86,63596	-0,86123	-0,93737	a0	0,376925
	932,8	72711,58	19408,47	799,5934	-0,86123	0,008958	0,007884	a1	0,008595
	251	19408,47	5366,66	212,4927	-0,93737	0,007884	0,015516	a2	-0,00912
2	12	932,8	66149	10,252	30,05491	-0,38616	8,35E-06	a0	-0,15656

	932,8	72711,58	5171083,2		799,5934		-0,38616	0,004983	-2,1E-07		a1	0,012787
	66149	5171083,2	1050647173		58987,3		8,35E-06	-2,1E-07	1,47E-09		a2	3,06E-06
3	12	932,8	895,9		10,252		35,78617	-0,32161	-0,14336		a0	-0,95855
	932,8	72711,58	69730,52		799,5934		-0,32161	0,005647	-0,00157		a1	0,004629
	895,9	69730,52	67207,01		772,0503		-0,14336	-0,00157	0,003556		a2	0,019463
4	12	251	66149		10,252		4,065682	-0,18553	-1,8E-05		a0	1,16951
	251	5366,66	1345076,1		212,4927		-0,18553	0,00874	4,91E-07		a1	-0,01578
	66149	1345076,1	1050647173		58987,3		-1,8E-05	4,91E-07	1,49E-09		a2	2,72E-06
5	12	251	895,9		10,252		36,2206	-0,45127	-0,3576		a0	-0,64711
	251	5366,66	18650,62		212,4927		-0,45127	0,01086	0,003002		a1	-0,00115
	895,9	18650,62	67207,01		772,0503		-0,3576	0,003002	0,003949		a2	0,020433
6	12	66149	895,9		10,252		18,50106	4,05E-05	-0,24968		a0	-0,70671
	66149	1050647173	5071824,8		58987,3		4,05E-05	1,59E-09	-6,6E-07		a1	-4,6E-07
	895,9	5071824,8	67207,01		772,0503		-0,24968	-6,6E-07	0,003393		a2	0,020943

Предсказания модели после первой эпохи обучения.

y1	y2	y3	y4	y5	y6
0,389685	0,609166	0,606129	0,311	0,166469	0,822089
0,942492	1,005625	1,021154	0,806643	0,8115	0,822335
0,818465	0,784345	0,787166	0,850866	0,858761	0,82367
0,729031	0,719408	0,718891	0,771942	0,756073	0,823954
0,882638	0,909921	0,918958	0,816125	0,819427	0,823985
0,809419	0,803104	0,809924	0,811373	0,819567	0,821747
0,927868	0,919785	0,927825	0,884011	0,903079	0,825031
0,913067	0,922176	0,93412	0,85242	0,870509	0,821952

Значения среднеквадратичной ошибки после первой эпохи:

	(y- y1^)^2	(y- y2^)^2	(y- y3^)^2	(y- y4^)^2	(y- y5^)^2	(y- y6^)^2
--	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

	0,096917	0,008434	0,000729	0,1521	0,000241	0,00279
	0,039399	0,068448	0,002137	0,003924	0,000398	0,000381
	0,010514	0,018675	0,000697	0,004919	1,57E-07	5,13E-06
	0,039192	0,043094	0,002717	0,024043	5,29E-05	1,69E-07
	0,006502	0,011647	0,001079	0,0002	4,26E-05	6,61E-05
	0,003896	0,003148	0,004217	0,004144	0,003526	0,003617
	7,53E-07	5,21E-05	5,13E-05	0,001848	0,000768	0,000844
	0,03689	0,040472	0,002202	0,017271	0,000212	0,000142
СР. КВ. ОШИБКА	0,029164	0,024246	0,001729	0,026056	0,000655	0,00098
	Лучший результат			0,000655021		

Модель	СР. КВ. ОШИБКА
y5	0,000655021
y6	0,000980467
y3	0,001728752
y2	0,024246098
y4	0,026056044
y1	0,029163756

Оставим y6, y2, y4 и y3 для следующей эпохи обучения.

2 эпоха обучения:

№пп	y	y5	y6	y3	y2	y5^2	y6^2	y3^2	y2^2
1	0,904	0,897253	0,904377	0,889572	0,819128	0,805063	0,817897	0,791339	0,670971
2	0,922	0,925685	0,929667	0,926814	0,856442	0,856892	0,86428	0,858984	0,733493
3	0,763	0,712778	0,716	0,727827	0,85547	0,508052	0,512656	0,529732	0,731829
4	0,923	0,909854	0,908473	0,903648	0,847873	0,827834	0,825323	0,816580	0,718888
5	0,918	0,91204	0,908614	0,934659	0,93261	0,831817	0,82558	0,873588	0,869762
6	0,906	0,904561	0,908571	0,895317	0,824203	0,818231	0,825502	0,801592	0,679311
7	0,905	0,875982	0,876936	0,866585	0,826946	0,767345	0,769017	0,750969	0,68384
8	0,545	0,603014	0,603223	0,572274	0,713986	0,363625	0,363878	0,327498	0,509776
9	0,894	0,922864	0,925342	0,921533	0,853507	0,851678	0,856258	0,849223	0,728474
10	0,90	0,930632	0,929434	0,924963	0,852868	0,866077	0,863848	0,855556	0,727384
11	0,932	0,944448	0,933696	0,942847	0,942175	0,891982	0,871789	0,888961	0,887694
12	0,74				0,926791	0,508211			0,858941
			<u>0,707666</u>	<u>0,745961</u>		<u>0,500791</u>	<u>0,556458</u>		
СУММА	10,252	10,252	10,252	10,252	10,252	8,896807	8,89682	8,900479	8,800364

y5y6	y5y3	y5y2	y6y3	y6y2	y3y2	yy5	yy6	yy3	yy2
------	------	------	------	------	------	-----	-----	-----	-----

0,81145 5	0,79817 1	0,73496 5	0,80450 9	0,7408	0,72867 4	0,81111 7	0,81755 7	0,80417 4	0,74049 2
0,86057 8	0,85793 8	0,79279 5	0,86162 8	0,79620 6	0,79376 3	0,85348 1	0,85715 3	0,85452 3	0,78964
0,51034 9	0,51877 9	0,60976	0,52112 4	0,61251 6	0,62263 4	0,54385	0,54630 8	0,55533 2	0,65272 3
0,82657 8	0,82218 8	0,77144 1	0,82094	0,77027	0,76617 9	0,83979 5	0,83852 1	0,83406 7	0,78258 7
0,82869 3	0,85244 7	0,85057 8	0,84924 5	0,84738 3	0,87167 3	0,83725 3	0,83410 8	0,85801 7	0,85613 6
0,82185 8	0,80986 9	0,74554 2	0,81345 9	0,74884 7	0,73792 3	0,81953 2	0,82316 6	0,81115 7	0,74672 8
0,76818	0,75911 3	0,72439	0,75993 9	0,72517 9	0,71661 9	0,79276 4	0,79362 7	0,78425 9	0,74838 6
0,36375 2	0,34508 9	0,43054 3	0,34520 9	0,43069 3	0,40859 6	0,32864 2	0,32875 7	0,31188 9	0,38912 2
0,85396 5	0,85044 9	0,78767 1	0,85273 3	0,78978 6	0,78653 4	0,82504	0,82725 6	0,82385	0,76303 5
0,86496 2	0,8608	0,79370 7	0,85969 2	0,79268 5	0,78887 1	0,83756 9	0,83649 1	0,83246 6	0,76758 2
0,88182 7	0,89047	0,88983 5	0,88033 3	0,87970 5	0,88832 7	0,88022 5	0,87020 5	0,87873 4	0,87810 7
0,50448 7	0,53178 8	0,66069 9	0,52789 1	0,65585 8	0,69135	0,52753 8	0,52367 3	0,55201 1	0,68582 5
8,896684	8,8971	8,791927	8,896702	8,789929	8,801143	8,896807	8,89682	8,900479	8,800364

[illegible]

4	12	10,252	10,252		10,252		5,365054	-5,99599	-0,18628		a0	-0,0007
	10,252	8,89682	8,896702		8,89682		-5,99599	263,5926	-256,574		a1	0,03093
	10,252	8,896702	8,900479		8,900479		-0,18628	-256,574	256,7923		a2	0,969893
5	12	10,252	10,252		10,252		17,96807	-1,86183	-19,0723		a0	-0,19901
	10,252	8,89682	8,789929		8,89682		-1,86183	8,717152	-6,53788		a1	0,93178
	10,252	8,789929	8,800364		8,800364		-19,0723	-6,53788	28,86208		a2	0,301164
6	12	10,252	10,252		10,252		17,57299	0,161763	-20,6335		a0	0,016069
	10,252	8,900479	8,801143		8,900479		0,161763	10,14782	-10,3372		a1	1,00805
	10,252	8,801143	8,800364		8,800364		-20,6335	-10,3372	34,48871		a2	-0,02686

Предсказания после 2 эпохи:

z1	z2	z3	z4	z5	z6
0,736022	0,670261	0,66349	0,676312	0,686838	0,679125
0,724223	0,796832	0,769778	0,788143	0,778909	0,785649
0,922407	0,891884	0,899836	0,895512	0,897484	0,896796
0,923788	0,870457	0,882019	0,876521	0,881791	0,878669
0,809346	0,837464	0,825199	0,83407	0,829888	0,833201
0,687147	0,681922	0,684857	0,682071	0,682859	0,682053
0,898612	0,921688	0,912961	0,919214	0,914692	0,918606
0,734121	0,771286	0,75971	0,766775	0,76163	0,765411

	(y- z1^)^2	(y- z2^)^2	(y- z3^)^2	(y- z4^)^2	(y- z5^)^2	(y- z6^)^2
	0,001227	0,000945	0,001407	0,000609	0,000201	0,000479
	0,000391	0,002791	0,000664	0,001949	0,001219	0,001735
	1,98E-06	0,000848	0,000448	0,00065	0,000553	0,000586
	1,03E-05	0,003197	0,002023	0,002548	0,002044	0,002336
	5,4E-05	0,001258	0,000538	0,001029	0,000778	0,000974

	0,003582	0,004235	0,003862	0,004216	0,004114	0,004218
	0,000806	2,82E-05	0,000197	6,06E-05	0,000151	7,05E-05
	0,000172	0,002529	0,001498	0,002095	0,001651	0,001972
СР. КВ. ОШИБКА	0,000781	0,001979	0,00133	0,001645	0,001339	0,001546
	Лучший результат			0,000780549		

Лучшие модели на данный момент:

Модель	СР. КВ. ОШИБКА
y5	0,000655021
z1	0,000780549
y6	0,000980467
z3	0,00132977

Как мы можем заметить, уже на второй эпохи алгоритмы показали результат хуже, чем на первой, а это значит, что уже получена максимально эффективная модель.

$$y_5 = -0,647114185 - 0,001150619x_2 + 0,02043328x_4$$

Проверим модель на тестовой выборке:

№	y	x1	x2	x3	x4	y5	(y-y5)^2
21	0,913	73,8	17,3	2916	76,8	0,902256014	0,000115
22	0,918	79,2	16,8	3551	78,1	0,929394587	0,00013
23	0,833	71,5	29,9	3177	73,9	0,828501702	2,02E-05
24	0,914	75,3	20,3	3280	78,6	0,935584061	0,000466
25	0,923	79	14,1	3160	78,5	0,940674571	0,000312
						Ср. кв. ошибка	0,000209

Итоговый средний квадрат ошибки: 0,000209.

Вывод:

Метод группового учёта аргументов успешно справляется с задачей регрессии. Причём необязательно, что с каждым этапом селекции моделей результат предсказания будет улучшаться.