Machine Learning techniques to Model Data Intensive Application Performance

A. Battistello P. Ferretti

Contents

1 Query comparison

1.1 R5 → R2 (Fast - Fast)

1.1.1 R5 → R2 — Datasize 250GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.2664	-3.8012	7131	0.0868
Linear SVR	1.4192	0.4500	41150	0.4928
Polynomial SVR (2)	2.0017	0.3318	57963	0.6936
Polynomial SVR (3)	1.7351	0.3201	50336	0.6032
Polynomial SVR (4)	2.0052	0.2591	58051	0.6946
Polynomial SVR (6)	2.0177	0.3148	58314	0.6974
Gaussian SVR	1.9531	0.0953	56646	0.6783

Table 1: Results for R5 → R2-250

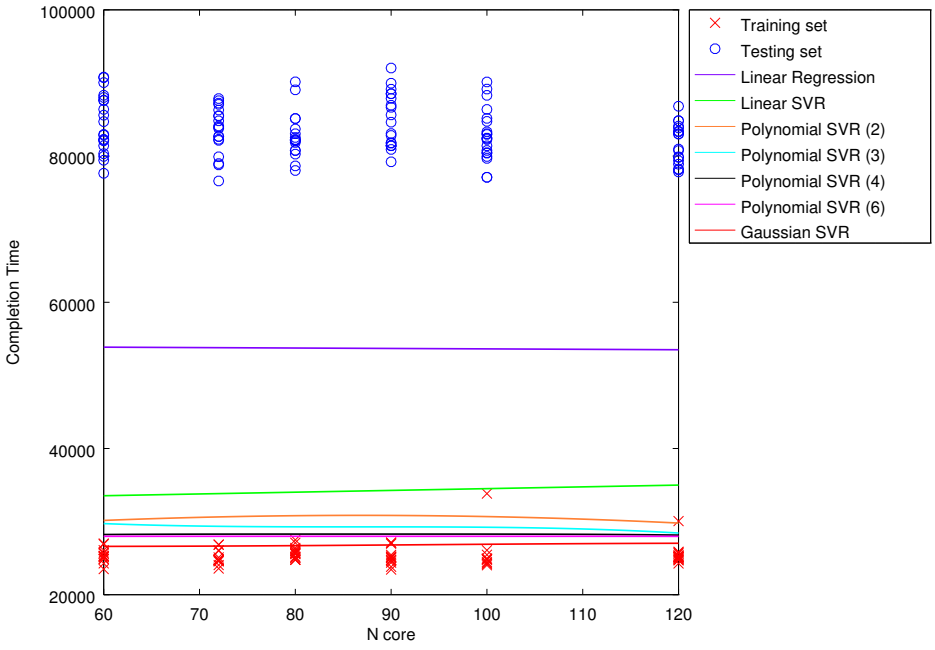

Figure 1: Completion time vs ncores for: R5 → R2 with datasize 250GB

1.1.2 R5 → R2 — Datasize 500GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.0927	0.4825	1627	0.0219
Linear SVR	0.8177	0.8804	20156	0.2760
Polynomial SVR (2)	1.9565	0.8888	47958	0.6544
Polynomial SVR (3)	1.3110	0.8506	32346	0.4437
Polynomial SVR (4)	1.9813	0.8130	48399	0.6598
Polynomial SVR (6)	1.9723	0.7576	48394	0.6606
Gaussian SVR	1.8458	0.4395	45448	0.6213

Table 2: Results for R5 → R2-500

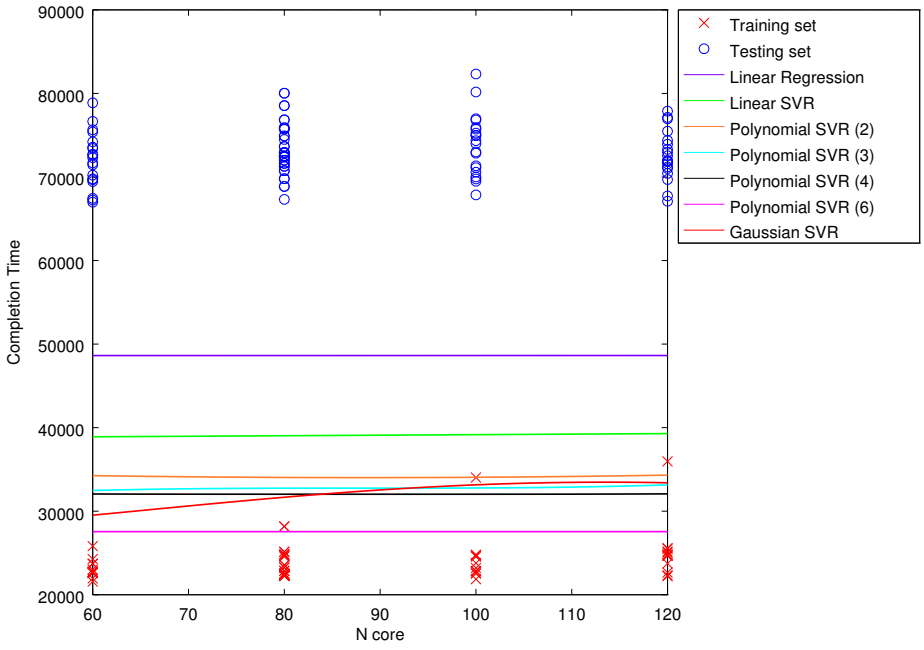

Figure 2: Completion time vs ncores for: R5 → R2 with datasize 500GB

1.1.3 R5 → R2 — Datasize 750GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.7284	-36.9573	19886	0.2523
Linear SVR	1.2200	0.7629	33371	0.4226
Polynomial SVR (2)	1.9610	0.7081	53638	0.6792
Polynomial SVR (3)	1.7155	0.7155	46976	0.5953
Polynomial SVR (4)	1.9544	0.6962	53405	0.6760
Polynomial SVR (6)	1.9545	0.6360	53347	0.6749
Gaussian SVR	1.8982	0.4132	51944	0.6579

Table 3: Results for R5 → R2-750

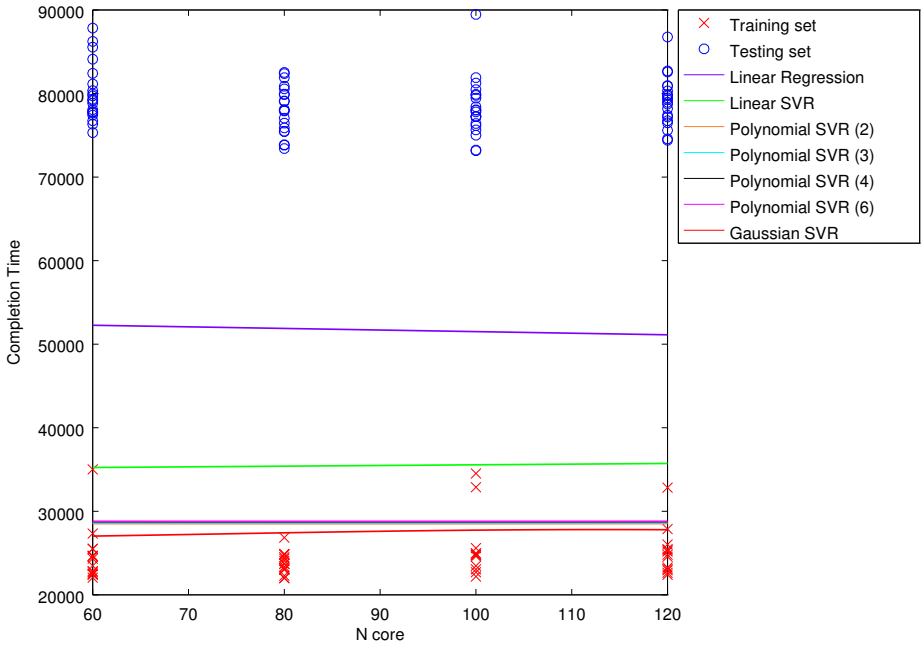

Figure 3: Completion time vs ncores for: R5 → R2 with datasize 750GB

1.1.4 R5 → R2 — Datasize 1000GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.2791	0.8182	124323	0.1468
Linear SVR	1.9131	-0.0000	865805	0.9413
Polynomial SVR (2)	1.9131	-0.0000	865805	0.9413
Polynomial SVR (3)	1.9131	-0.0000	865805	0.9413
Polynomial SVR (4)	1.9131	-0.0000	865805	0.9413
Polynomial SVR (6)	1.9131	-0.0000	865805	0.9413
Gaussian SVR	1.9131	-0.0000	865805	0.9413

Table 4: Results for R5 → R2-1000

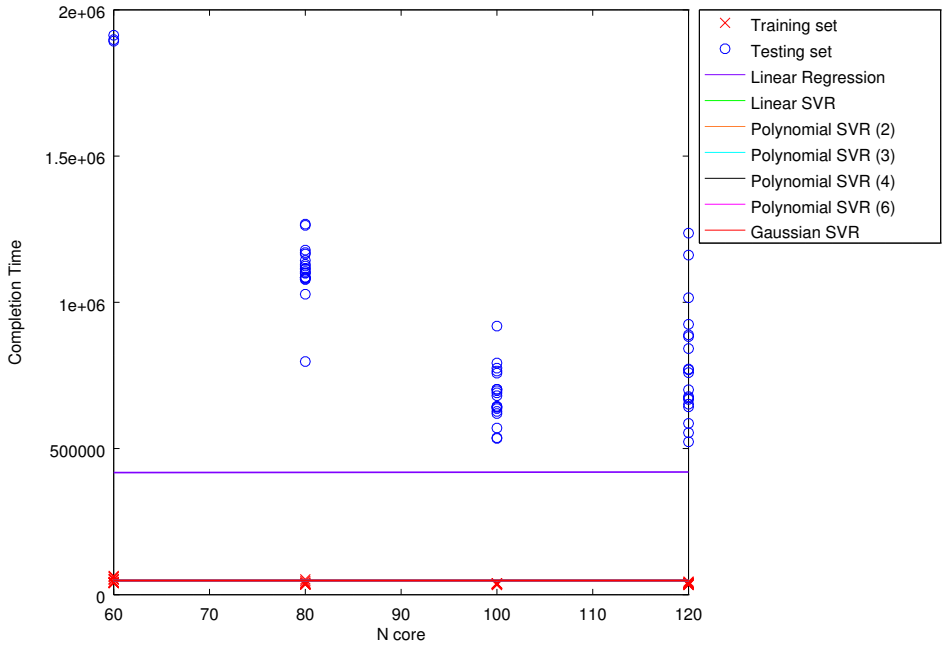

Figure 4: Completion time vs ncores for: R5 → R2 with datasize 1000GB

1.2 R2 → R5 (Fast - Fast)

1.2.1 R2 → R5 — Datasize 250GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.5076	-54.2010	14618	0.5690
Linear SVR	0.0742	0.5354	1906	0.0748
Polynomial SVR (2)	1.9582	0.5005	56793	2.2284
Polynomial SVR (3)	1.4334	0.4575	41627	1.6299
Polynomial SVR (4)	1.9578	0.4334	56818	2.2282
Polynomial SVR (6)	1.9529	0.3733	56692	2.2226
Gaussian SVR	1.8646	0.4311	54134	2.1221

Table 5: Results for R2 → R5-250

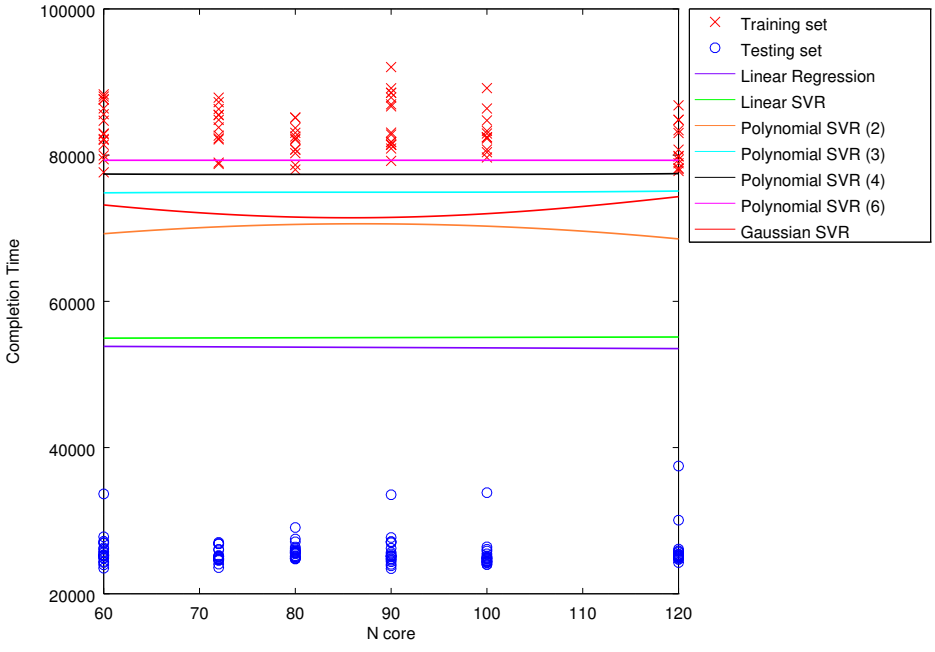

Figure 5: Completion time vs ncores for: R2 → R5 with datasize 250GB

1.2.2 R2 → R5 — Datasize 500GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.2910	-3.7899	6672	0.2718
Linear SVR	0.7805	0.3478	19144	0.8083
Polynomial SVR (2)	1.9764	0.3328	48648	2.0492
Polynomial SVR (3)	1.5362	0.3112	37827	1.5916
Polynomial SVR (4)	1.9546	0.2841	48117	2.0265
Polynomial SVR (6)	1.9390	0.2654	47733	2.0103
Gaussian SVR	1.9761	0.1983	48681	2.0483

Table 6: Results for R2 → R5-500

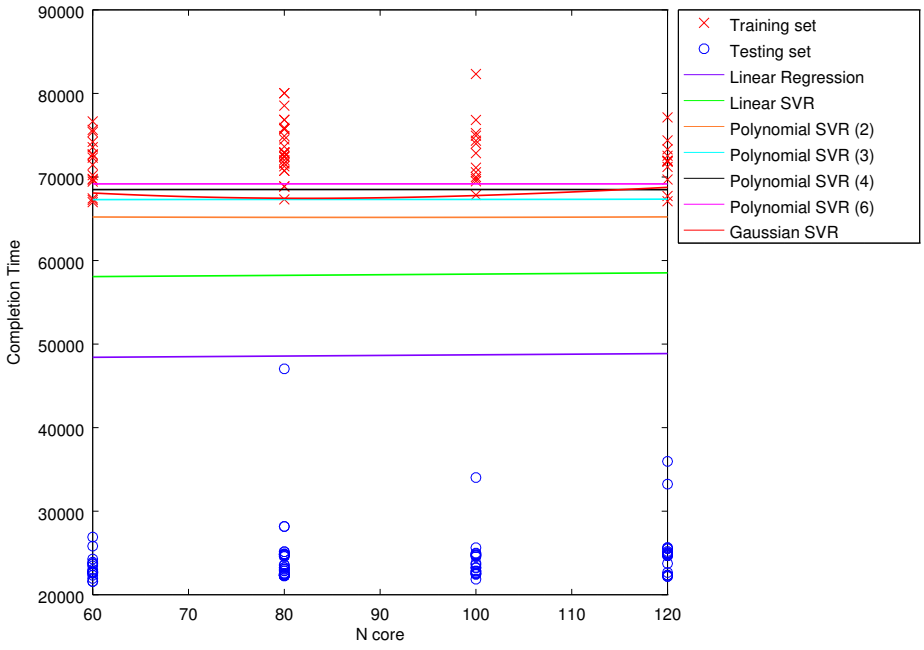

Figure 6: Completion time vs ncores for: R2 → R5 with datasize 500GB

1.2.3 R2 → R5 — Datasize 750GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.7050	-58.7255	19273	0.7906
Linear SVR	0.5814	0.1630	15781	0.6549
Polynomial SVR (2)	1.9918	0.3443	54498	2.2536
Polynomial SVR (3)	1.5014	0.3534	41121	1.6974
Polynomial SVR (4)	1.9754	0.2597	54059	2.2349
Polynomial SVR (6)	1.9710	0.2735	53943	2.2300
Gaussian SVR	2.0285	0.4331	55559	2.2948

Table 7: Results for R2 → R5-750

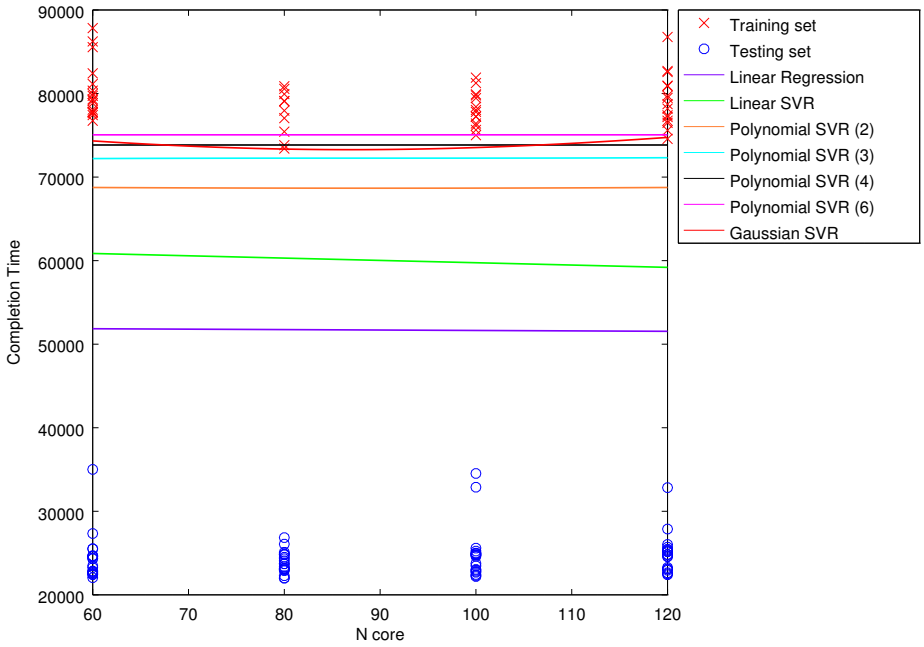

Figure 7: Completion time vs ncores for: R2 → R5 with datasize 750GB

1.2.4 R2 → R5 — Datasize 1000GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.2457	-258.0712	117597	2.9684
Linear SVR	0.0746	0.0791	31327	0.7786
Polynomial SVR (2)	1.5254	0.2338	734175	18.4731
Polynomial SVR (3)	0.8871	0.1001	426009	10.6683
Polynomial SVR (4)	1.4660	0.0268	705234	17.7261
Polynomial SVR (6)	1.4552	0.0225	700727	17.5862
Gaussian SVR	2.1204	0.0853	1020795	25.5869

Table 8: Results for R2 → R5-1000

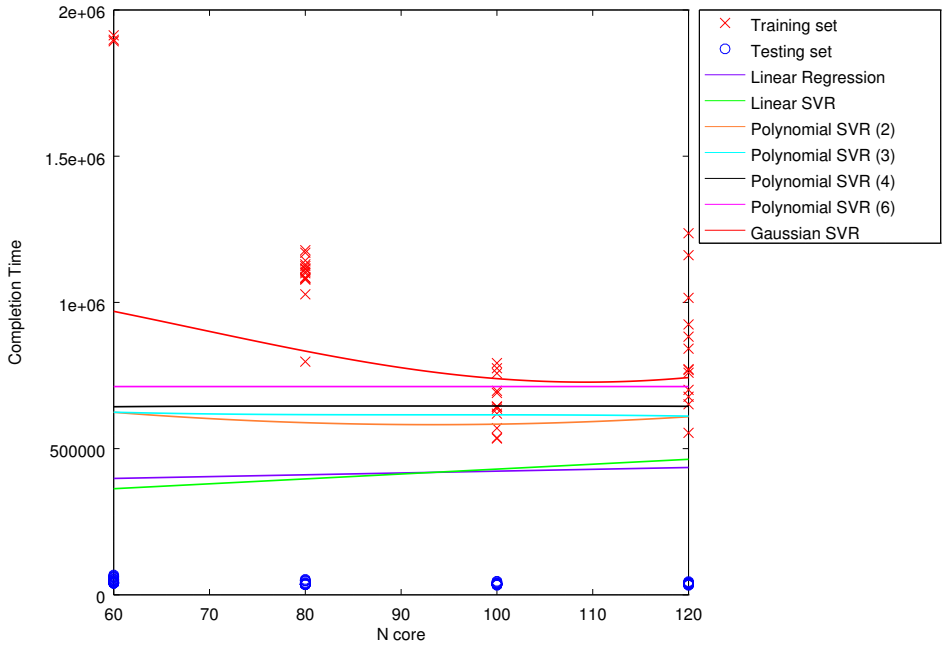

Figure 8: Completion time vs ncores for: R2 → R5 with datasize 1000GB

1.3 R3 → R4 (Slow - Slow)

1.3.1 R3 → R4 — Datasize 250GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.0999	0.9905	7743	0.0556
Linear SVR	0.1017	0.9936	7894	0.0559
Polynomial SVR (2)	1.2864	0.0000	109952	0.8321
Polynomial SVR (3)	0.5276	0.8127	40426	0.2156
Polynomial SVR (4)	0.9298	0.2785	77573	0.5374
Polynomial SVR (6)	0.8746	0.2982	69520	0.4722
Gaussian SVR	0.5125	0.8738	39988	0.2815

Table 9: Results for R3 → R4-250

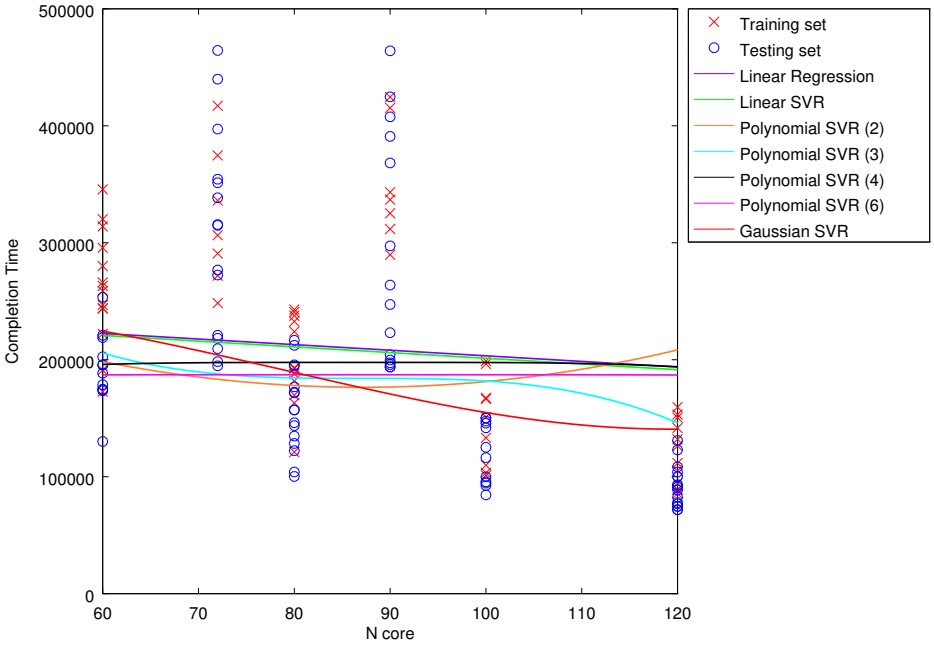

Figure 9: Completion time vs ncores for: R3 → R4 with datasize 250GB

1.3.2 R3 → R4 — Datasize 500GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.5575	0.7917	79305	0.1806
Linear SVR	0.6248	0.8895	87857	0.1990
Polynomial SVR (2)	1.4064	0.3856	175966	0.4039
Polynomial SVR (3)	0.9288	0.7713	129311	0.3215
Polynomial SVR (4)	1.3185	0.1435	177714	0.4267
Polynomial SVR (6)	1.4713	0.1476	177441	0.4040
Gaussian SVR	1.0797	0.3126	134221	0.2887

Table 10: Results for R3 → R4-500

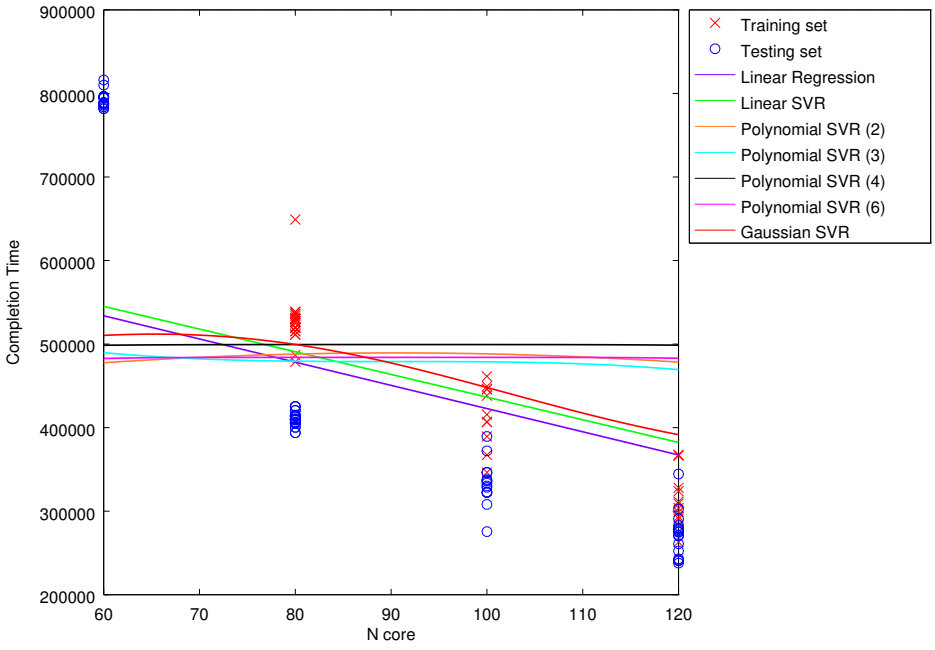

Figure 10: Completion time vs ncores for: R3 → R4 with datasize 500GB

1.3.3 R3 → R4 — Datasize 750GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.0933	0.9899	12026	0.0195
Linear SVR	0.2629	0.9912	33564	0.0561
Polynomial SVR (2)	1.8337	0.7689	217984	0.3828
Polynomial SVR (3)	0.7352	0.7039	87799	0.1548
Polynomial SVR (4)	1.8539	0.4641	210264	0.3621
Polynomial SVR (6)	1.1873	0.4401	148303	0.2645
Gaussian SVR	1.0774	0.8421	126509	0.2349

Table 11: Results for R3 → R4-750

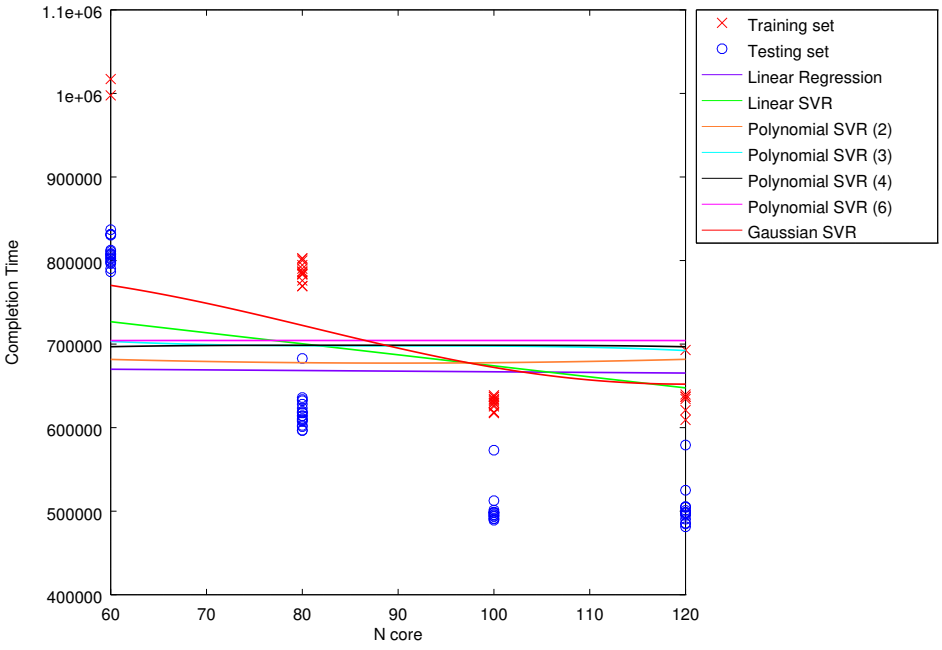

Figure 11: Completion time vs ncores for: R3 → R4 with datasize 750GB

1.3.4 R3 → R4 — Datasize 1000GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	1.7878	-1.2231	784437	0.3797
Linear SVR	1.6361	0.0281	660983	0.2902
Polynomial SVR (2)	1.6549	0.0157	689481	0.3172
Polynomial SVR (3)	1.6757	0.0650	705047	0.3218
Polynomial SVR (4)	1.9237	0.1209	764929	0.3431
Polynomial SVR (6)	1.8627	0.0254	735220	0.3294
Gaussian SVR	1.5644	0.0772	640303	0.2875

Table 12: Results for R3 → R4-1000

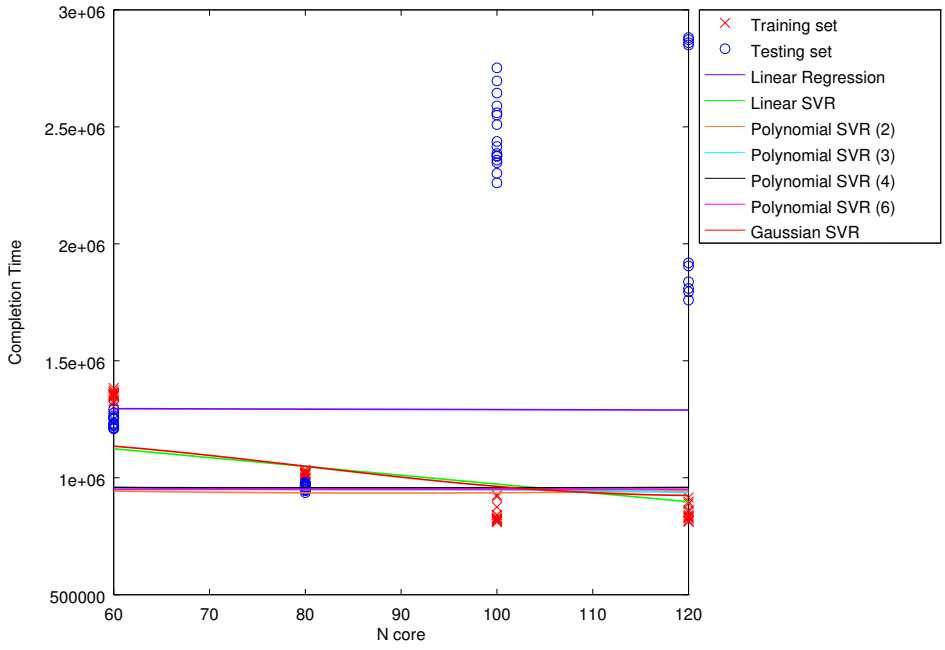

Figure 12: Completion time vs ncores for: R3 → R4 with datasize 1000GB

1.4 R4 → R3 (Slow - Slow)

1.4.1 R4 → R3 — Datasize 250GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.1117	0.9859	6402	0.0272
Linear SVR	0.1241	0.9882	8184	0.0379
Polynomial SVR (2)	1.0187	0.0102	84602	0.4395
Polynomial SVR (3)	0.7609	0.7115	53158	0.2639
Polynomial SVR (4)	1.0989	0.0178	85066	0.3771
Polynomial SVR (6)	1.0632	0.0244	82125	0.3725
Gaussian SVR	0.5420	0.8101	41963	0.2180

Table 13: Results for R4 → R3-250

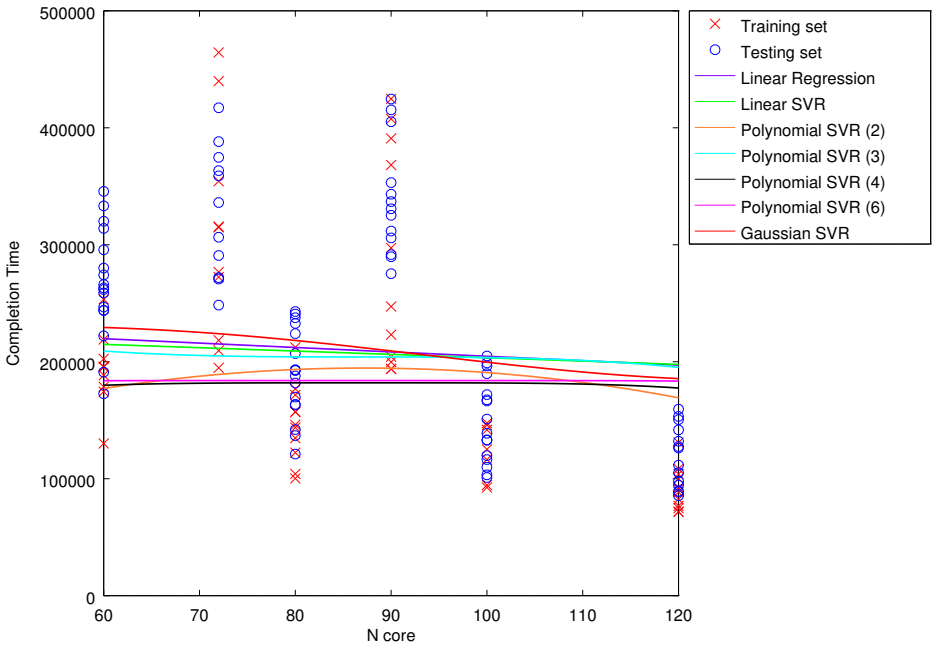

Figure 13: Completion time vs ncores for: R4 → R3 with datasize 250GB

1.4.2 R4 → R3 — Datasize 500GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.1181	0.9592	13349	0.0297
Linear SVR	0.1281	0.9719	13961	0.0304
Polynomial SVR (2)	1.0739	0.0989	142216	0.3246
Polynomial SVR (3)	0.5917	0.7504	77226	0.1606
Polynomial SVR (4)	1.0211	0.2288	134494	0.3151
Polynomial SVR (6)	0.9462	0.0853	117567	0.2718
Gaussian SVR	0.3507	0.7951	46118	0.1032

Table 14: Results for R4 → R3-500

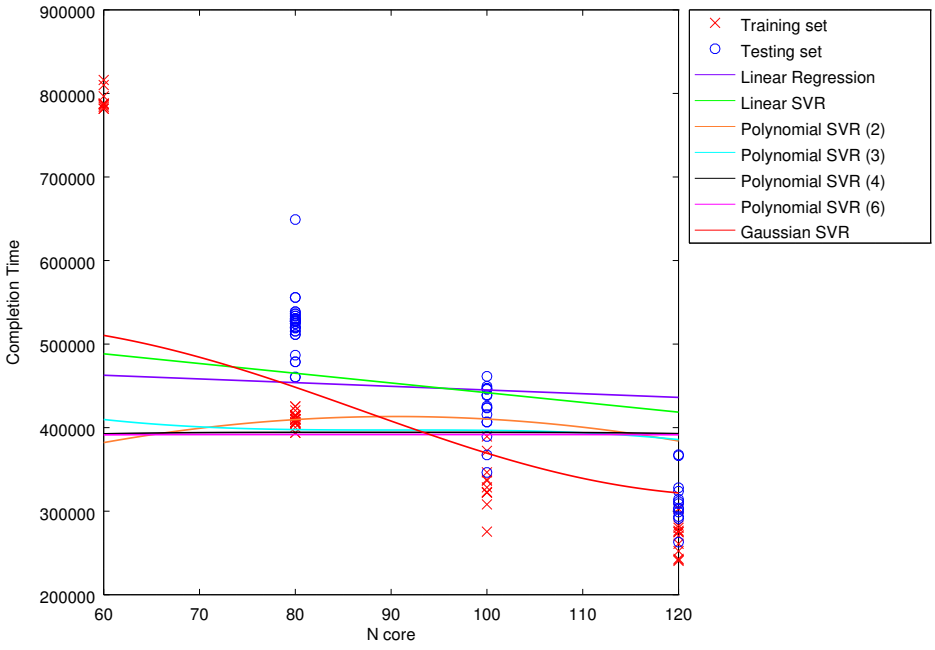

Figure 14: Completion time vs ncores for: R4 → R3 with datasize 500GB

1.4.3 R4 → R3 — Datasize 750GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.1065	0.9853	13983	0.0192
Linear SVR	0.1572	0.9937	18494	0.0250
Polynomial SVR (2)	1.7890	0.8077	185642	0.2313
Polynomial SVR (3)	1.0593	0.8125	126124	0.1704
Polynomial SVR (4)	2.8642	0.6965	233765	0.2775
Polynomial SVR (6)	2.9392	0.5268	213611	0.2468
Gaussian SVR	0.8538	0.7386	85514	0.1055

Table 15: Results for R4 → R3-750

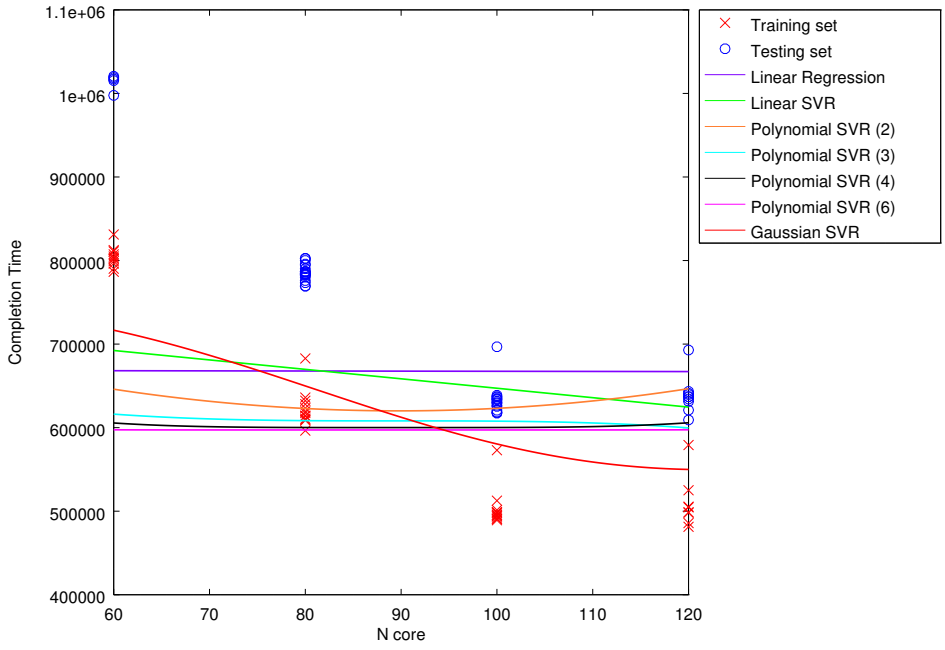

Figure 15: Completion time vs ncores for: R4 → R3 with datasize 750GB

1.4.4 R4 → R3 — Datasize 1000GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.2979	0.2999	163252	0.1736
Linear SVR	0.2150	0.9597	109503	0.1219
Polynomial SVR (2)	0.7033	0.6852	382105	0.4195
Polynomial SVR (3)	0.5805	0.7986	324151	0.3440
Polynomial SVR (4)	0.3638	0.7740	189564	0.2076
Polynomial SVR (6)	0.4766	0.6324	252222	0.2741
Gaussian SVR	1.0142	0.5173	574326	0.6115

Table 16: Results for R4 → R3-1000

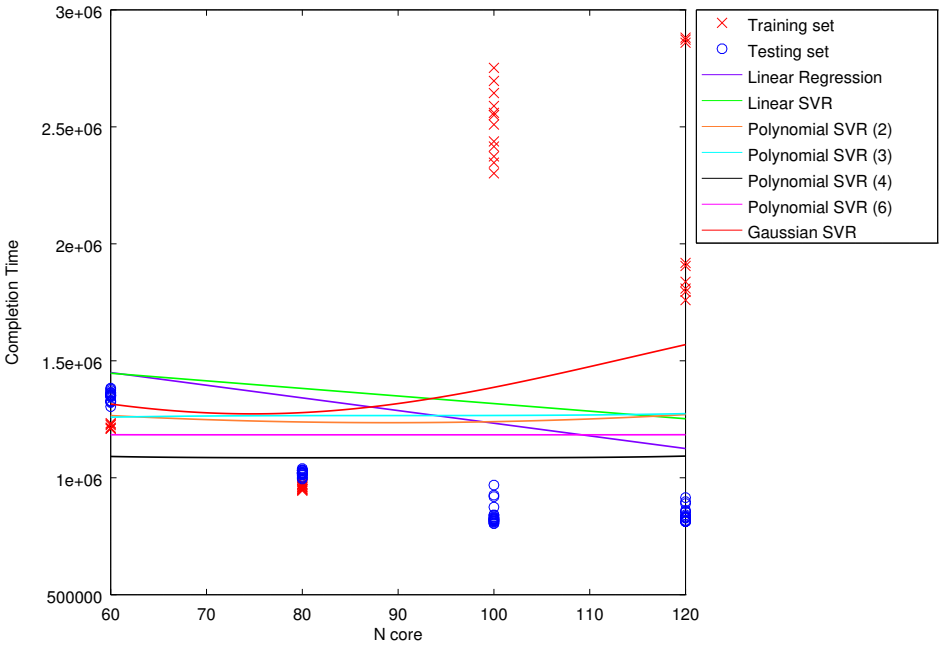

Figure 16: Completion time vs ncores for: R4 → R3 with datasize 1000GB

1.5 R1, R2, R4 → R3 (Fast and Slow - Average)

1.5.1 R1,R2,R4 → R3 — Datasize 250GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.1484	0.9772	9501	0.0466
Linear SVR	0.1560	0.9845	10574	0.0553
Polynomial SVR (2)	0.8548	0.7060	57023	0.2555
Polynomial SVR (3)	0.4469	0.8616	28093	0.1225
Polynomial SVR (4)	0.7404	0.7503	42222	0.1783
Polynomial SVR (6)	1.1648	0.5504	61124	0.2455
Gaussian SVR	0.3926	0.9405	24779	0.1084

Table 17: Results for R1,R2,R4 → R3-250

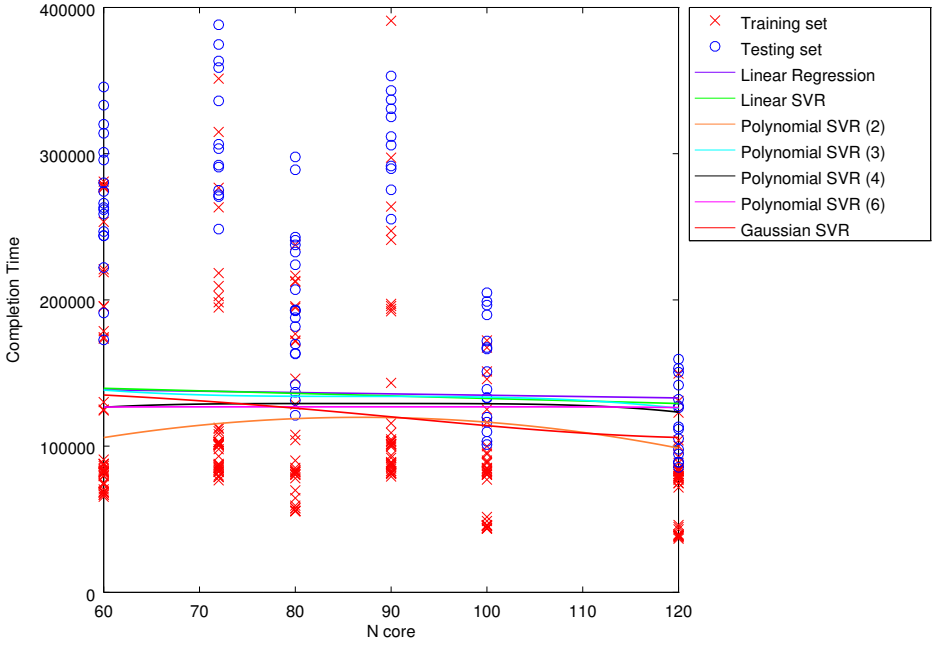

Figure 17: Completion time vs ncores for: R1,R2,R4 → R3 with datasize 250GB

1.5.2 R1,R2,R4 → R3 — Datasize 500GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.0914	0.9687	13457	0.0317
Linear SVR	0.0833	0.9741	13357	0.0308
Polynomial SVR (2)	0.6833	0.6864	118761	0.2933
Polynomial SVR (3)	0.2660	0.8449	42425	0.0947
Polynomial SVR (4)	0.7291	0.6105	129234	0.3055
Polynomial SVR (6)	0.7959	0.4006	136789	0.3104
Gaussian SVR	0.2844	0.8990	43387	0.1147

Table 18: Results for R1,R2,R4 → R3-500

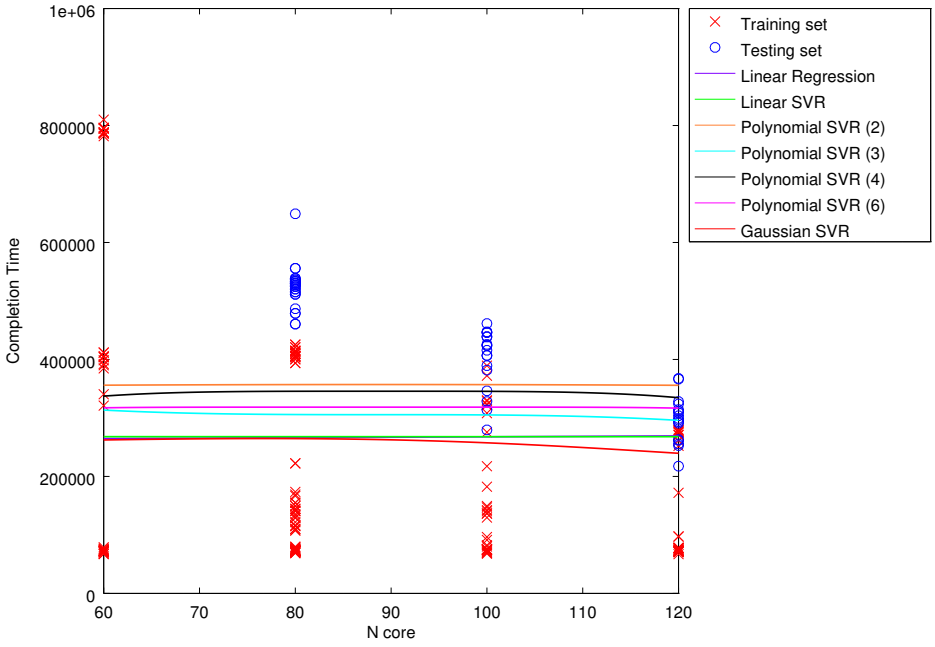

Figure 18: Completion time vs ncores for: R1,R2,R4 → R3 with datasize 500GB

1.5.3 R1,R2,R4 → R3 — Datasize 750GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.0581	0.9694	13358	0.0186
Linear SVR	0.0432	0.9924	10296	0.0152
Polynomial SVR (2)	0.4219	0.7036	106082	0.1502
Polynomial SVR (3)	0.2763	0.7323	53615	0.0742
Polynomial SVR (4)	0.6688	0.4095	168830	0.2349
Polynomial SVR (6)	0.6903	0.2073	166482	0.2269
Gaussian SVR	0.3779	0.5523	82501	0.1133

Table 19: Results for R1,R2,R4 → R3-750

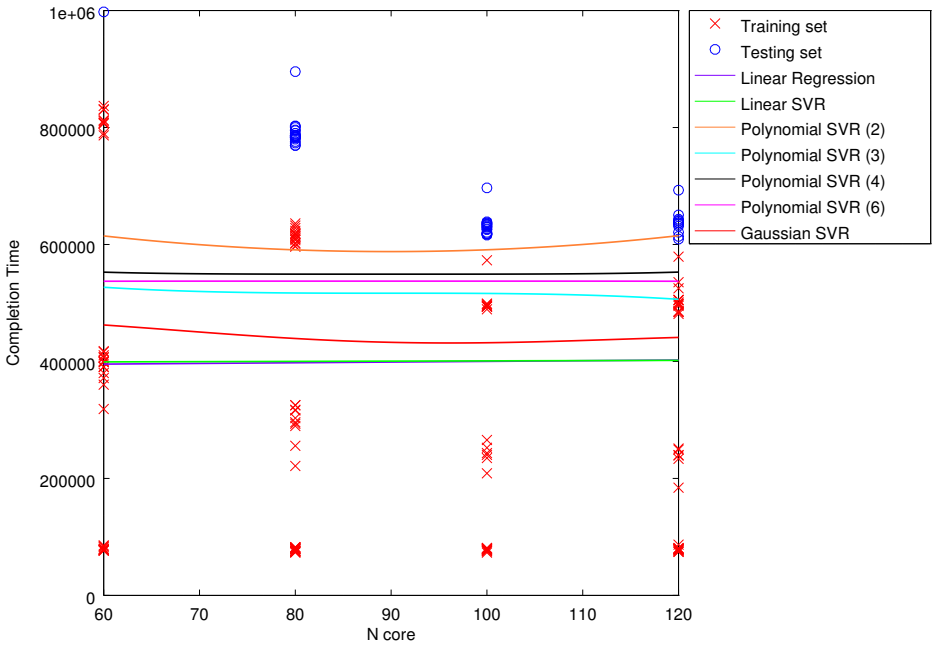

Figure 19: Completion time vs ncores for: R1,R2,R4 → R3 with datasize 750GB

1.5.4 R1,R2,R4 → R3 — Datasize 1000GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.5423	-0.7126	260530	0.2553
Linear SVR	0.5775	0.9500	275655	0.2718
Polynomial SVR (2)	0.5045	0.7467	202083	0.1809
Polynomial SVR (3)	0.4542	0.0345	148622	0.1290
Polynomial SVR (4)	0.5079	0.0024	181230	0.1545
Polynomial SVR (6)	0.5769	0.0175	203255	0.1711
Gaussian SVR	0.3920	0.7004	168171	0.1877

Table 20: Results for R1,R2,R4 → R3-1000

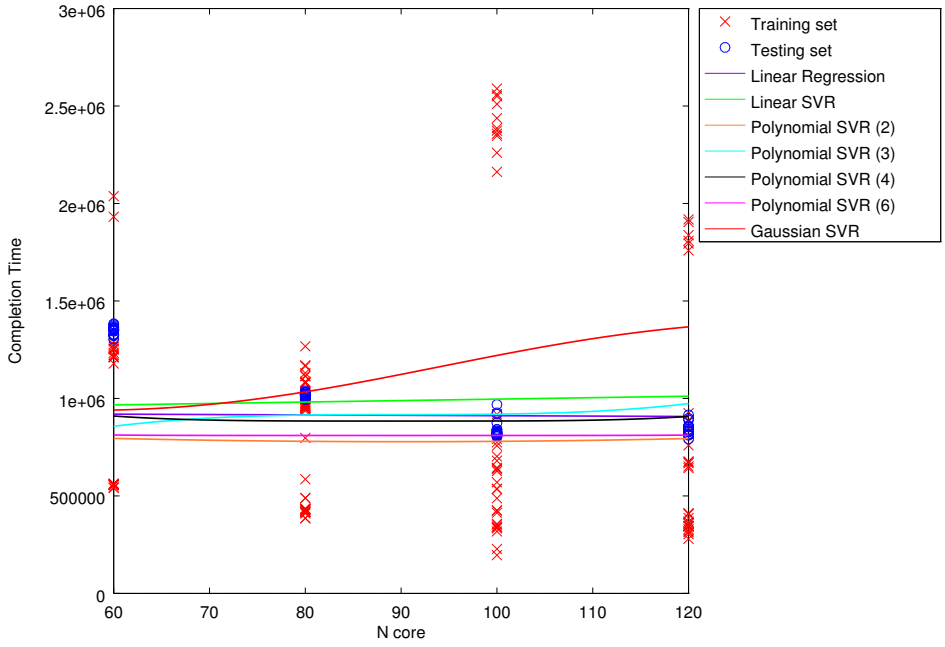

Figure 20: Completion time vs ncores for: R1,R2,R4 → R3 with datasize 1000GB

1.6 R2, R5 → R3 (Fast - Slow)

1.6.1 R2,R5 → R3 — Datasize 250GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	2.9358	-8.9645	225067	1.0067
Linear SVR	2.0189	0.1000	154238	0.6860
Polynomial SVR (2)	2.3413	0.0143	179635	0.8056
Polynomial SVR (3)	2.0566	0.0575	157367	0.6993
Polynomial SVR (4)	2.1957	0.0474	167202	0.7392
Polynomial SVR (6)	2.1578	0.0358	163842	0.7215
Gaussian SVR	2.0866	0.2880	160725	0.7205

Table 21: Results for R2,R5 → R3-250

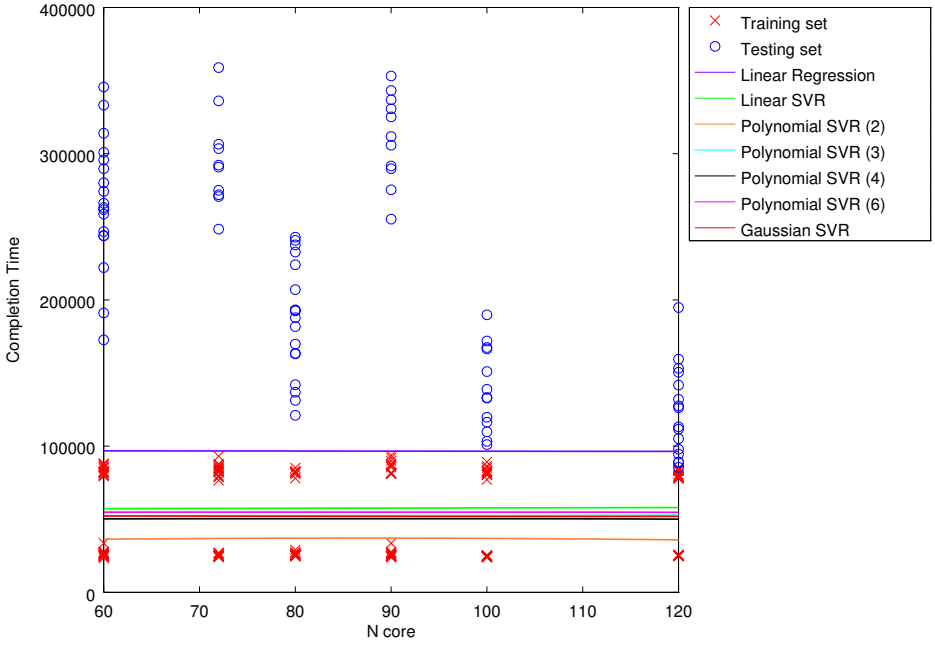

Figure 21: Completion time vs ncores for: R2,R5 → R3 with datasize 250GB

1.6.2 R2,R5 → R3 — Datasize 500GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	24.4317	-1732.1625	4082755	9.5054
Linear SVR	1.3901	0.0020	217391	0.4766
Polynomial SVR (2)	2.3279	0.8694	385016	0.8855
Polynomial SVR (3)	2.1472	0.6028	358017	0.8305
Polynomial SVR (4)	2.2733	0.4990	377968	0.8744
Polynomial SVR (6)	2.2643	0.2486	376874	0.8730
Gaussian SVR	2.2414	0.1919	373980	0.8683

Table 22: Results for R2,R5 → R3-500

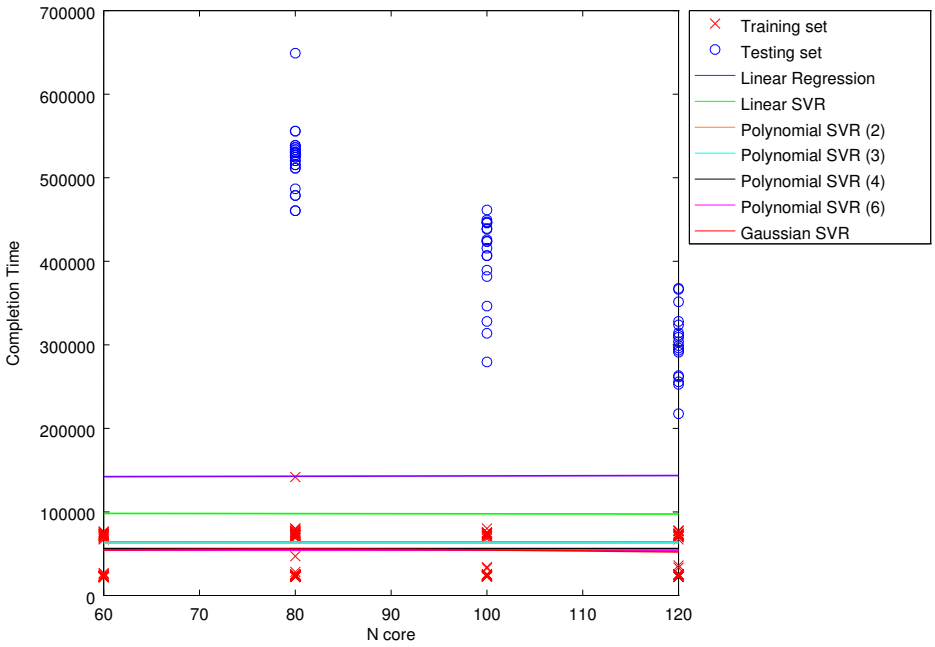

Figure 22: Completion time vs ncores for: R2,R5 → R3 with datasize 500GB

1.6.3 R2,R5 → R3 — Datasize 750GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	3.1762	-108.9756	905865	1.2847
Linear SVR	2.2593	0.0458	645146	0.9158
Polynomial SVR (2)	2.2682	0.1040	647657	0.9193
Polynomial SVR (3)	2.2586	0.3000	644988	0.9156
Polynomial SVR (4)	2.2552	0.0930	643893	0.9139
Polynomial SVR (6)	2.2549	0.0726	643840	0.9139
Gaussian SVR	2.2664	0.6737	647187	0.9187

Table 23: Results for R2,R5 → R3-750

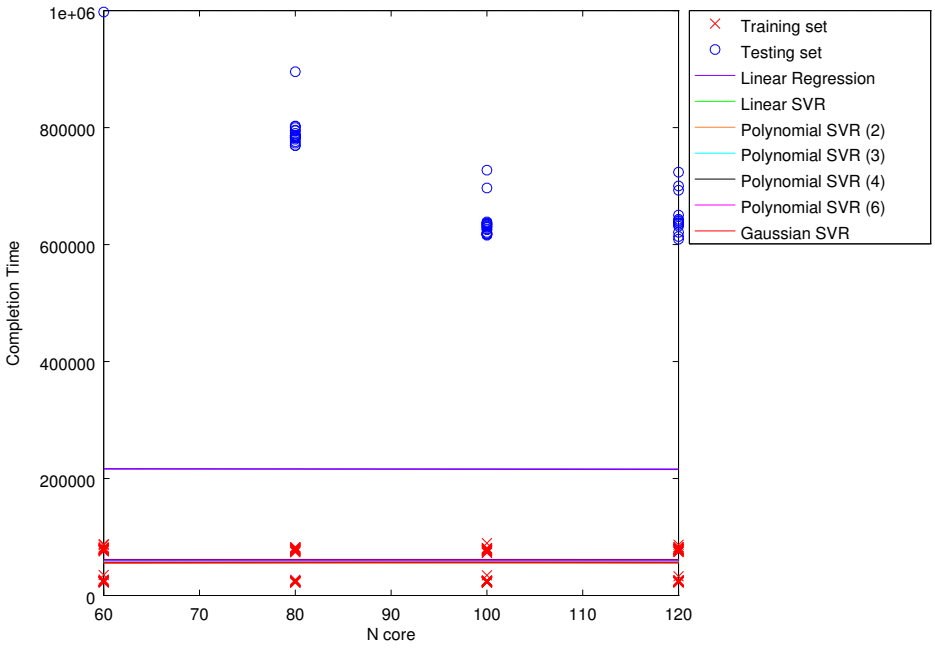

Figure 23: Completion time vs ncores for: R2,R5 → R3 with datasize 750GB

1.6.4 R2,R5 → R3 — Datasize 1000GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.4825	-0.2162	225269	0.2231
Linear SVR	1.5082	0.8905	700319	0.6897
Polynomial SVR (2)	1.8532	0.0474	856486	0.8417
Polynomial SVR (3)	1.7507	0.0017	810643	0.7966
Polynomial SVR (4)	1.8048	0.1072	838182	0.8256
Polynomial SVR (6)	1.8737	0.3832	871983	0.8603
Gaussian SVR	1.5367	0.8611	723592	0.7227

Table 24: Results for R2,R5 → R3-1000

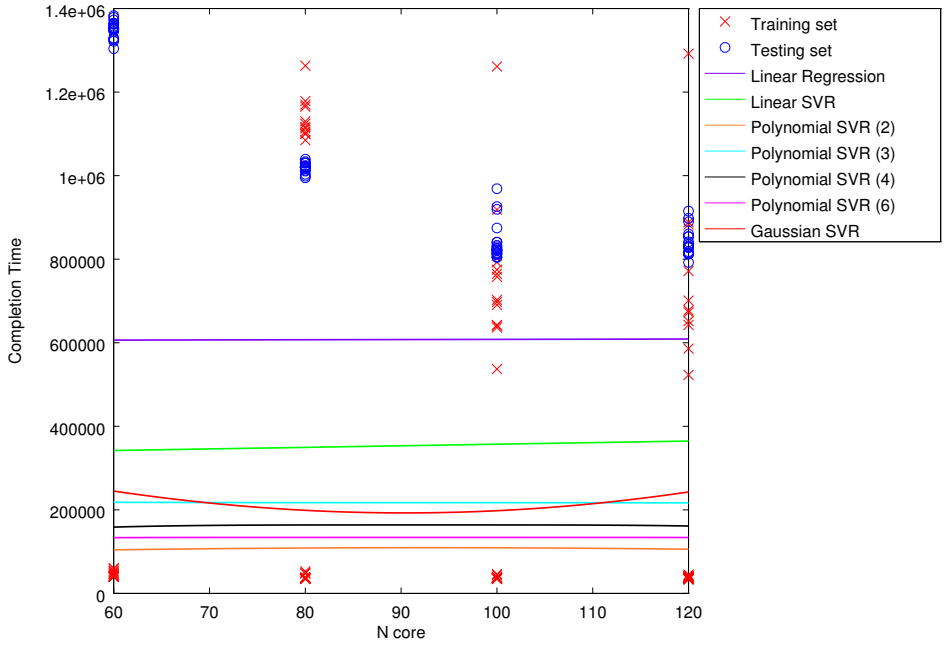

Figure 24: Completion time vs ncores for: R2,R5 → R3 with datasize 1000GB

1.7 R3, R4 → R2 (Fast - Slow)

1.7.1 R3,R4 → R2 — Datasize 250GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.7320	-324.7047	67111	0.8046
Linear SVR	0.6184	0.3343	56625	0.6788
Polynomial SVR (2)	2.2958	0.1994	209908	2.5228
Polynomial SVR (3)	0.6067	0.0155	47847	0.5734
Polynomial SVR (4)	2.0229	0.0101	180533	2.1680
Polynomial SVR (6)	1.7103	0.0153	146100	1.7568
Gaussian SVR	1.0720	0.2857	98273	1.1766

Table 25: Results for R3,R4 → R2-250

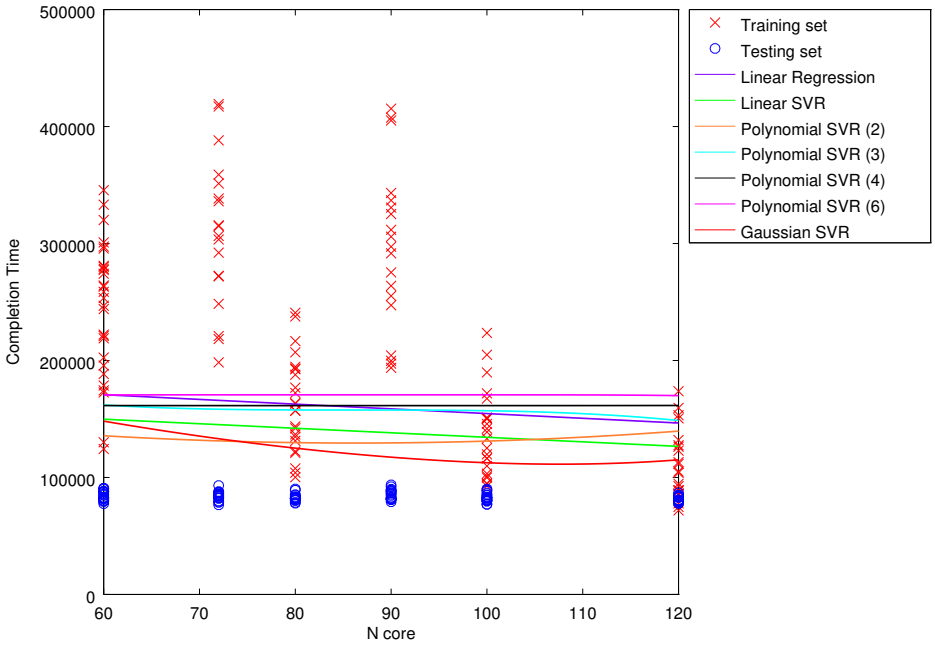

Figure 25: Completion time vs ncores for: R3,R4 → R2 with datasize 250GB

1.7.2 R3,R4 → R2 — Datasize 500GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.2274	-43.9565	50511	0.6952
Linear SVR	0.4418	0.0006	96855	1.3309
Polynomial SVR (2)	1.8749	0.0345	420366	5.7616
Polynomial SVR (3)	0.9866	0.0290	213001	2.9014
Polynomial SVR (4)	1.8453	0.0260	410413	5.6274
Polynomial SVR (6)	1.7365	0.0248	385605	5.2880
Gaussian SVR	1.4964	0.0442	337994	4.6265

Table 26: Results for R3,R4 → R2-500

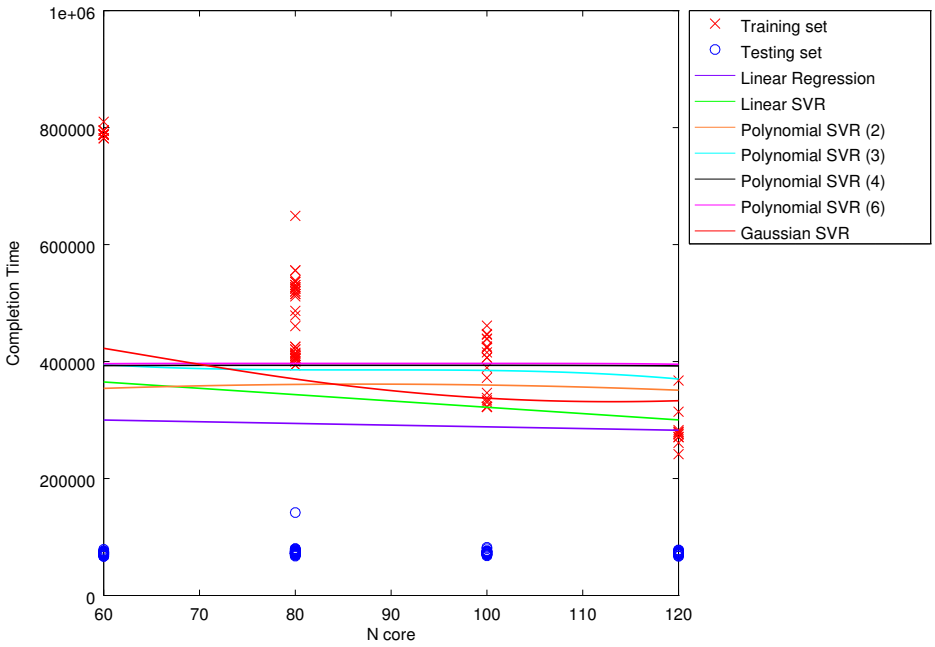

Figure 26: Completion time vs ncores for: R3,R4 → R2 with datasize 500GB

1.7.3 R3,R4 → R2 — Datasize 750GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.4540	-3592.1168	146163	1.8730
Linear SVR	0.4483	0.0421	137083	1.7535
Polynomial SVR (2)	2.4760	0.1503	788379	10.1194
Polynomial SVR (3)	1.1909	0.3274	304007	3.8427
Polynomial SVR (4)	2.6951	0.2837	803028	10.3423
Polynomial SVR (6)	3.3215	0.2411	930987	11.9433
Gaussian SVR	1.9111	0.1845	615634	7.8865

Table 27: Results for R3,R4 → R2-750

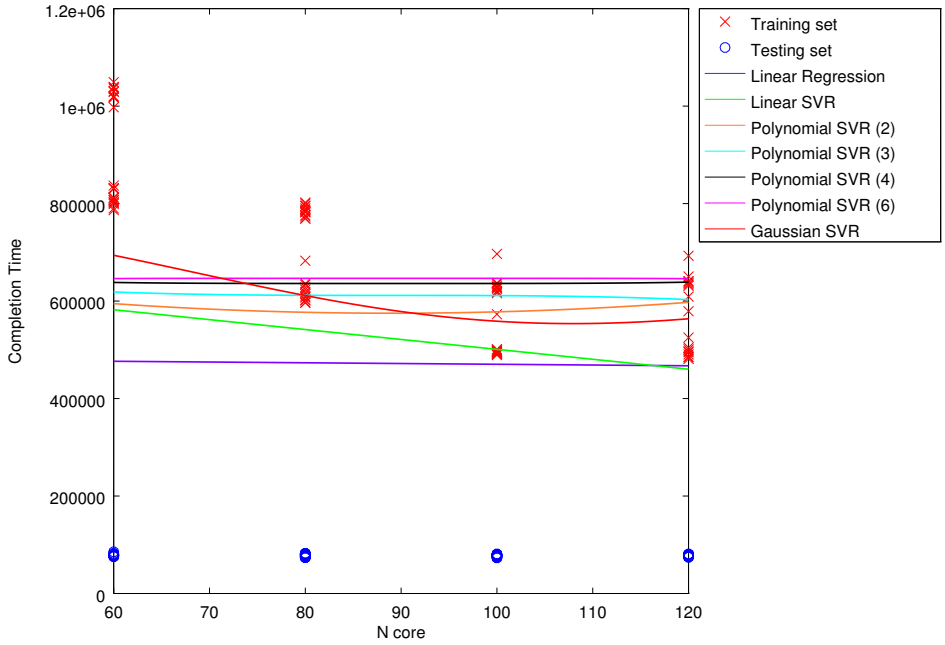

Figure 27: Completion time vs ncores for: R3,R4 → R2 with datasize 750GB

1.7.4 R3,R4 → R2 — Datasize 1000GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	11.2761	-583.0303	3410161	3.2357
Linear SVR	7.7162	0.9822	3657876	4.2191
Polynomial SVR (2)	8.0373	0.9441	2766694	2.6339
Polynomial SVR (3)	10.5398	0.8977	3750682	3.6335
Polynomial SVR (4)	10.3551	0.8329	3230811	2.9786
Polynomial SVR (6)	7.2376	0.6977	2298685	2.2241
Gaussian SVR	2.4222	0.5863	1124917	1.5115

Table 28: Results for R3,R4 → R2-1000

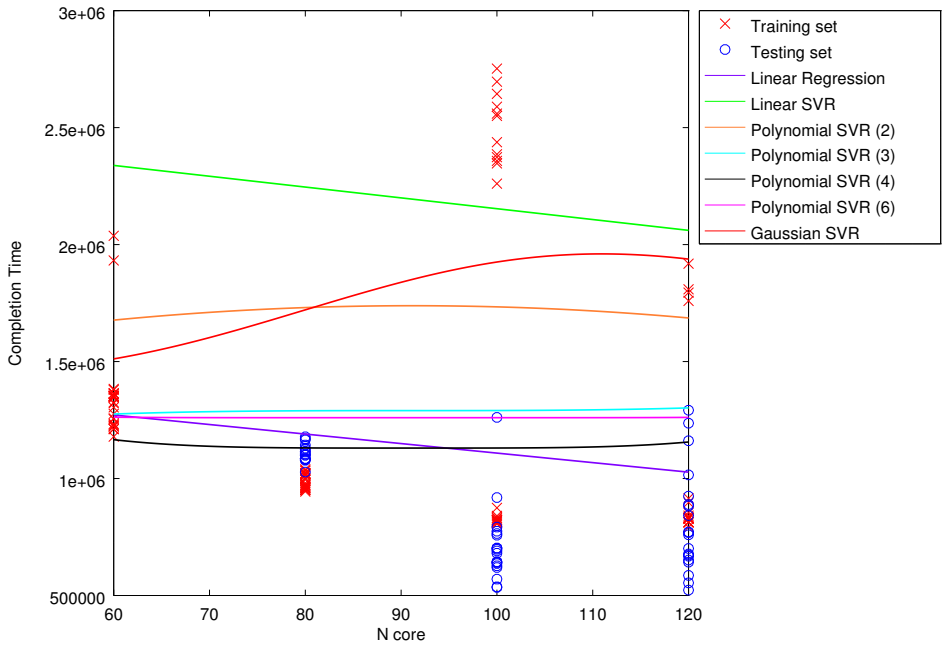

Figure 28: Completion time vs ncores for: R3,R4 → R2 with datasize 1000GB