

Machine Learning techniques to Model Data Intensive Application Performance

A. Battistello

P. Ferretti

Contents

1	Fixed Datasize, all features with 1/nCores,	2
1.1	Query Q2	2
1.1.1	Query Q2 — Datasize 40GB	2
1.1.2	Query Q2 — Datasize 50GB	3
1.2	Query Q3	4
1.2.1	Query Q3 — Datasize 50GB	4
1.3	Query Q4	5
1.3.1	Query Q4 — Datasize 50GB	5

1 Fixed Datasize, all features with 1/nCores,

1.1 Query Q2

1.1.1 Query Q2 — Datasize 40GB

Model	RMSE	R ²	Mean absolute error	Mean relative error
Linear regression	1.1724	-0.0377	24035	0.1012
Linear SVR	0.3713	0.9310	7170	0.0293
Polynomial SVR (2)	0.7597	0.7219	15618	0.0647
Polynomial SVR (3)	0.5592	0.8204	10138	0.0414
Polynomial SVR (4)	0.9967	0.4517	18992	0.0756
Polynomial SVR (6)	1.1150	0.0733	20847	0.0824
Gaussian SVR	0.5455	0.9096	10367	0.0410

Table 1: Results for Q2-40GB

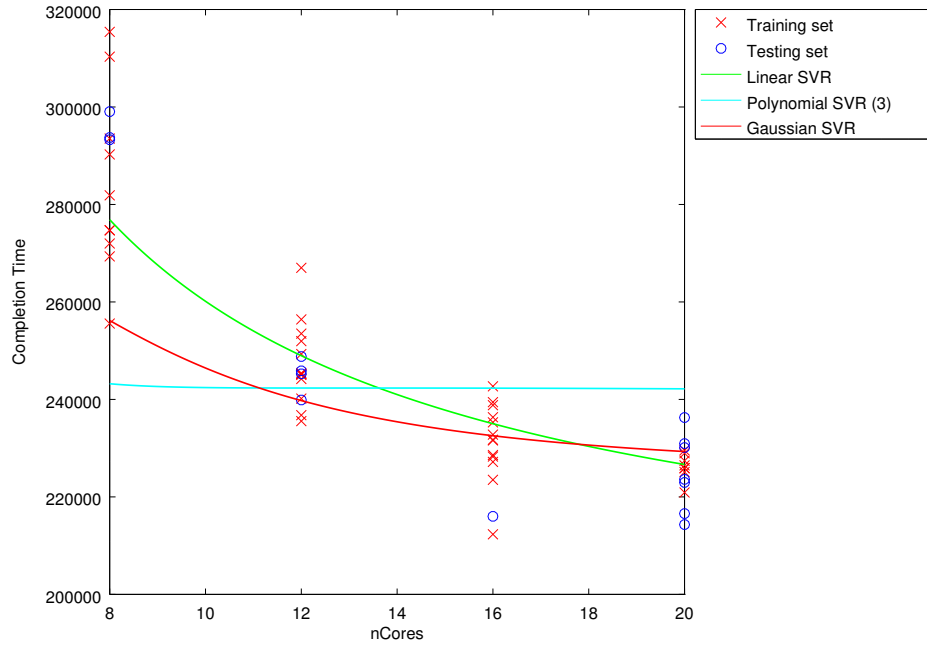


Figure 1: Completion time vs ncores for query Q2 with datasize 40

1.1.2 Query Q2 — Datasize 50GB

Model	RMSE	R ²	Mean absolute error	Mean relative error
Linear regression	0.9401	0.1721	35276	0.1336
Linear SVR	0.4863	0.7826	15991	0.0631
Polynomial SVR (2)	0.7756	0.4662	28112	0.1025
Polynomial SVR (3)	0.7518	0.4994	26075	0.0987
Polynomial SVR (4)	1.1967	0.2245	36195	0.1382
Polynomial SVR (6)	1.1080	0.1845	37971	0.1412
Gaussian SVR	0.4857	0.8544	16546	0.0638

Table 2: Results for Q2-50GB

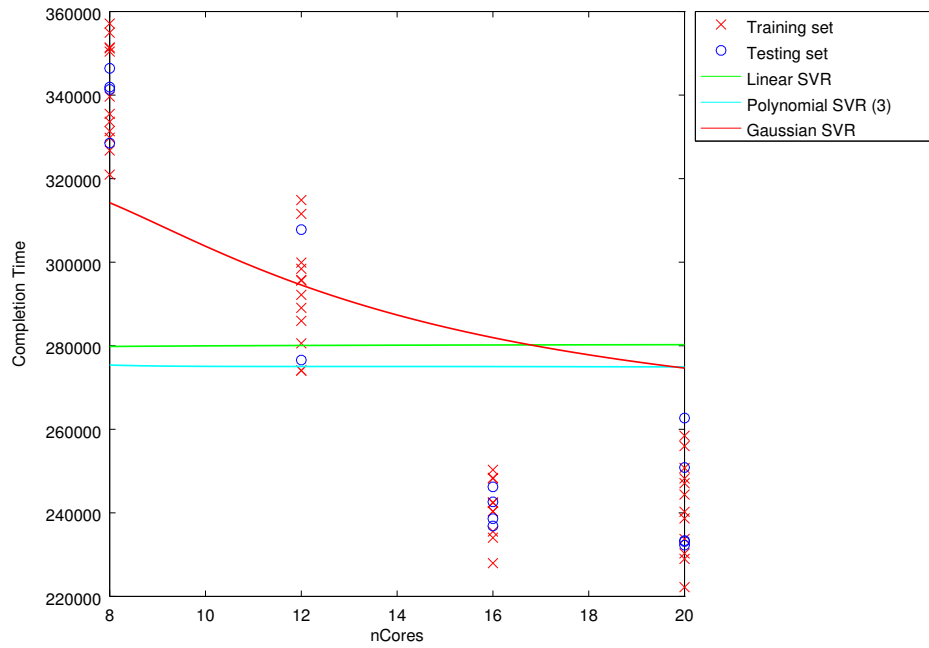


Figure 2: Completion time vs ncores for query Q2 with datasize 50

1.2 Query Q3

1.2.1 Query Q3 — Datasize 50GB

Model	RMSE	R ²	Mean absolute error	Mean relative error
Linear regression	0.1821	0.9385	5147	0.0238
Linear SVR	0.3698	0.7939	9957	0.0472
Polynomial SVR (2)	0.6657	0.2370	22068	0.1086
Polynomial SVR (3)	0.4115	0.7344	10516	0.0487
Polynomial SVR (4)	0.6275	0.4275	16290	0.0778
Polynomial SVR (6)	0.6407	0.4513	16807	0.0790
Gaussian SVR	0.2121	0.9383	6227	0.0295

Table 3: Results for Q3-50GB

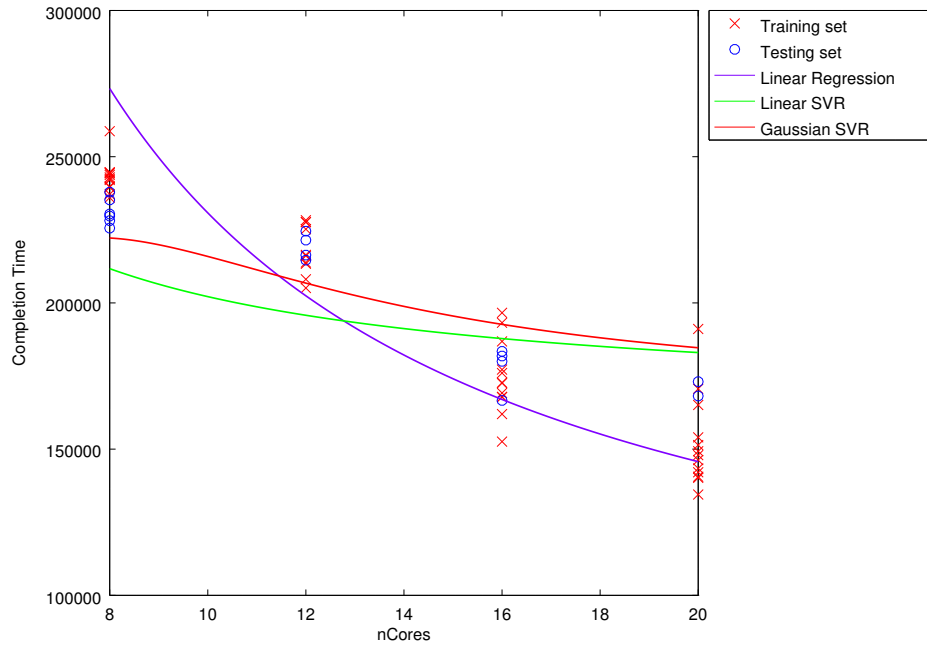


Figure 3: Completion time vs ncores for query Q3 with datasize 50

1.3 Query Q4

1.3.1 Query Q4 — Datasize 50GB

Model	RMSE	R ²	Mean absolute error	Mean relative error
Linear regression	0.0990	0.9881	4688	0.0154
Linear SVR	0.1853	0.9602	11231	0.0364
Polynomial SVR (2)	0.6609	0.5640	41416	0.1327
Polynomial SVR (3)	0.3972	0.8135	23263	0.0776
Polynomial SVR (4)	0.4891	0.7383	28158	0.0915
Polynomial SVR (6)	0.6421	0.6013	39270	0.1209
Gaussian SVR	0.2128	0.9463	12910	0.0451

Table 4: Results for Q4-50GB

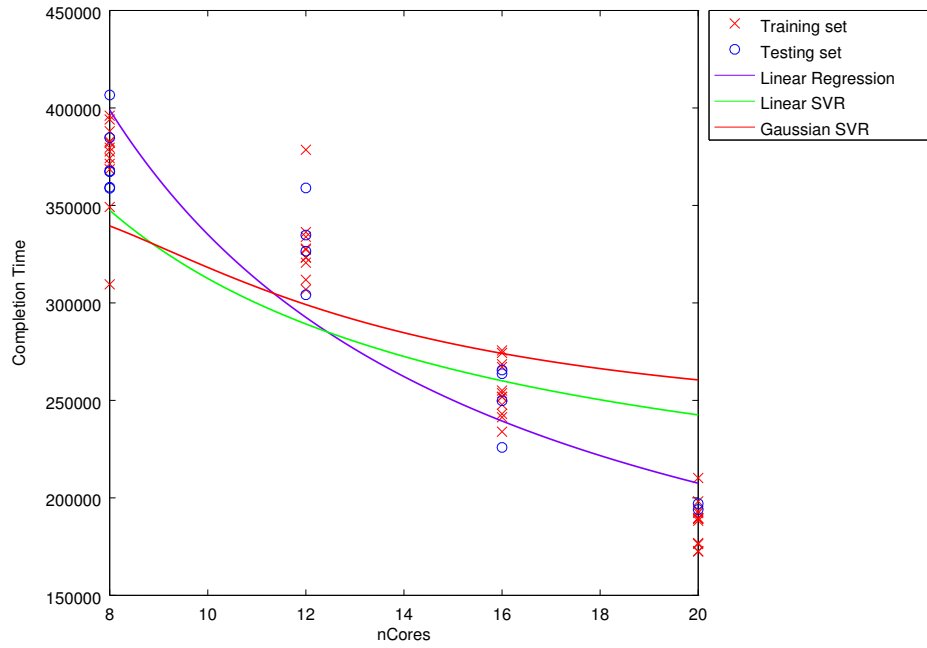


Figure 4: Completion time vs ncores for query Q4 with datasize 50