Machine Learning techniques to Model Data Intensive Application Performance

A. Battistello   P. Ferretti

Contents

1 Fixed Datasize, only nCores,

1.1 Query Q2

1.1.1 Query Q2 — Datasize 40GB

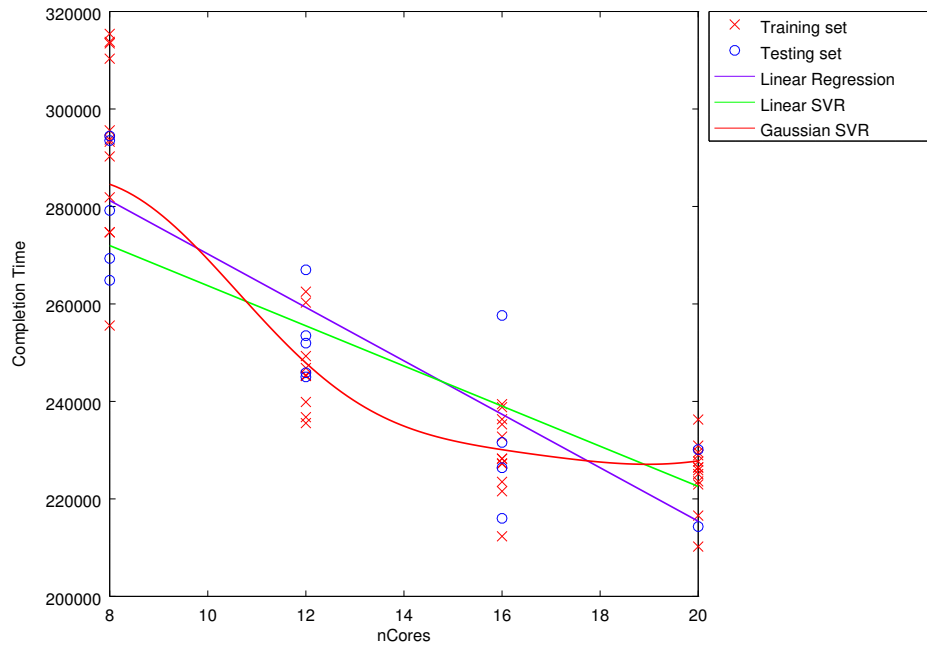| Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.4732	0.7213	11516	0.0462
Linear SVR	0.4773	0.7465	10981	0.0436
Polynomial SVR (2)	0.9249	0.1892	21428	0.0832
Polynomial SVR (3)	0.5122	0.6857	11435	0.0460
Polynomial SVR (4)	0.9083	0.2025	21353	0.0831
Polynomial SVR (6)	0.9293	0.2194	21908	0.0848
Gaussian SVR	0.4556	0.7493	9715	0.0382

Table 1: Results for Q2-40GB

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

2

1.1.2 Query Q2 — Datasize 50GB

Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.4165	0.8333	13809	0.0505
Linear SVR	0.4199	0.8411	14497	0.0514
Polynomial SVR (2)	1.0202	0.0000	38508	0.1365
Polynomial SVR (3)	0.5115	0.7521	16040	0.0601
Polynomial SVR (4)	1.0202	0.0000	38508	0.1365
Polynomial SVR (6)	1.0202	0.0000	38508	0.1365
Gaussian SVR	0.2887	0.9208	10435	0.0374

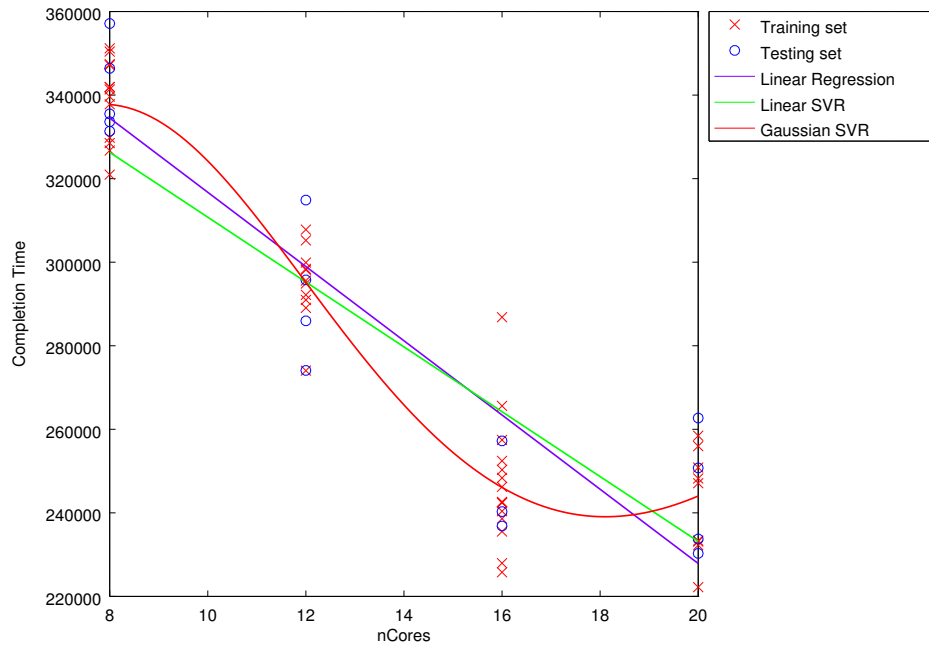**Table 2:** Results for Q2-50GB

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

1.2 Query Q3

1.2.1 Query Q3 — Datasize 50GB

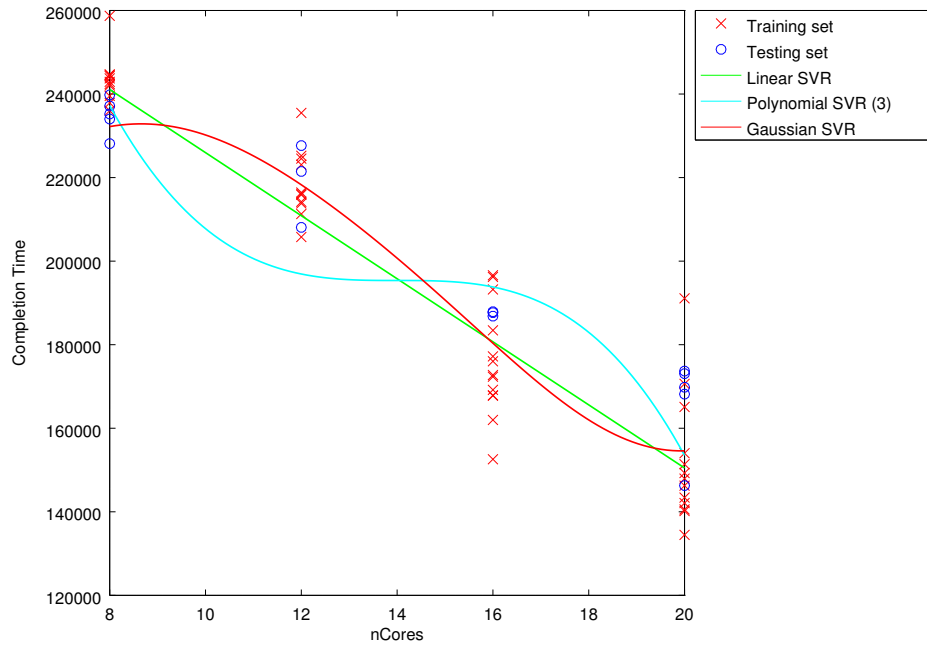| Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.3813	0.7992	11506	0.0597
Linear SVR	0.3626	0.9342	10534	0.0556
Polynomial SVR (2)	0.8629	-0.0000	27382	0.1376
Polynomial SVR (3)	0.4068	0.8520	11280	0.0585
Polynomial SVR (4)	0.8629	-0.0000	27382	0.1376
Polynomial SVR (6)	0.8629	-0.0000	27382	0.1376
Gaussian SVR	0.2917	0.9418	8764	0.0471

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

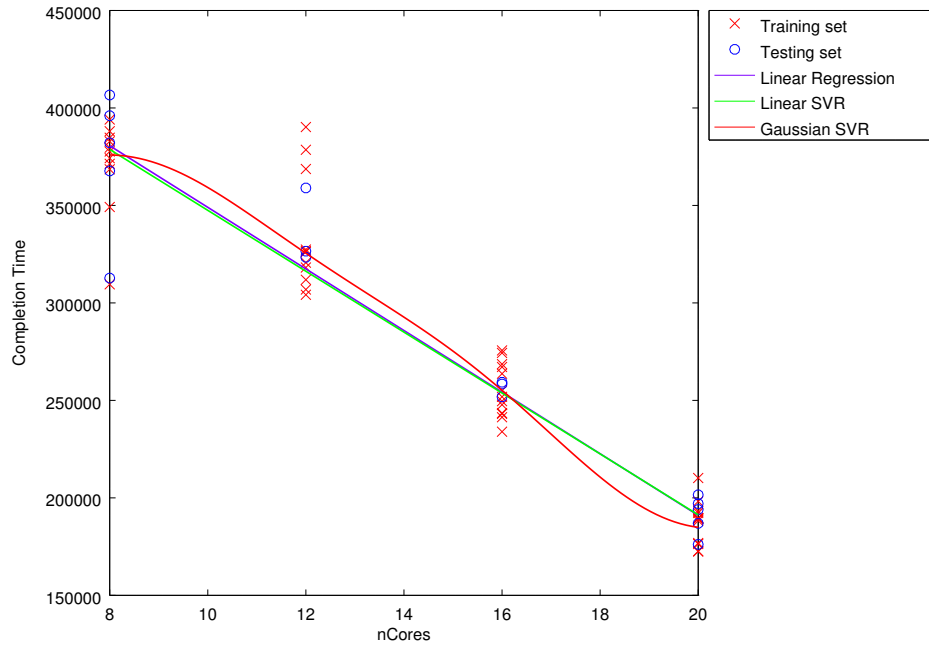| Model	RMSE	R^2	Mean absolute error	Mean relative error
Linear regression	0.3026	0.9203	14054	0.0471
Linear SVR	0.3036	0.9208	14764	0.0491
Polynomial SVR (2)	1.0691	0.0077	71807	0.2703
Polynomial SVR (3)	0.4319	0.8457	25075	0.0836
Polynomial SVR (4)	1.0698	0.0077	71807	0.2722
Polynomial SVR (6)	1.0698	0.0077	71807	0.2722
Gaussian SVR	0.2880	0.9306	14033	0.0477

Table 4: Results for Q4-50GB

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