

Machine Learning techniques to Model Data Intensive Application Performance

A. Battistello P. Ferretti

Contents

1	Fixed Datasize, only 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, only 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	0.4755	0.7186	9942	0.0394
Linear SVR	0.4356	0.7670	9580	0.0378
Polynomial SVR (2)	0.7545	0.3885	17726	0.0705
Polynomial SVR (3)	0.5410	0.6751	12134	0.0488
Polynomial SVR (4)	0.6630	0.5285	15720	0.0630
Polynomial SVR (6)	0.6016	0.5722	14553	0.0587
Gaussian SVR	0.4536	0.7500	9672	0.0381

Table 1: Results for Q2-40GB

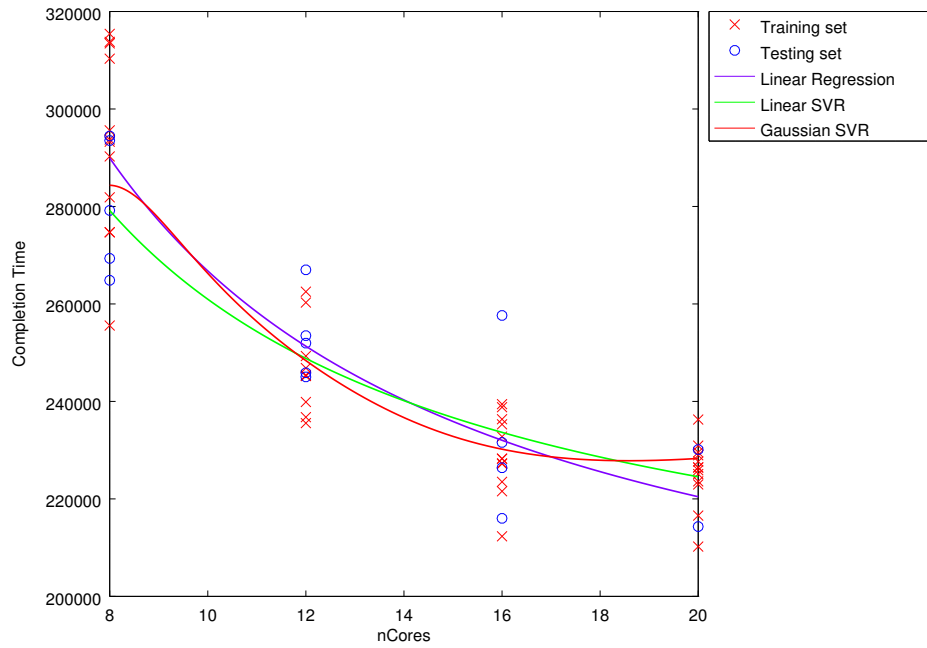


Figure 1: Completion time vs ncores for query Q2 with datasize 40

1.1.2 Query Q2 — Datasize 50GB

Model	RMSE	R ²	Mean absolute error	Mean relative error
Linear regression	0.3294	0.8958	11482	0.0409
Linear SVR	0.3301	0.8973	11520	0.0410
Polynomial SVR (2)	0.8382	0.4016	32740	0.1153
Polynomial SVR (3)	0.4515	0.8074	15713	0.0568
Polynomial SVR (4)	0.6385	0.6123	21200	0.0775
Polynomial SVR (6)	0.5805	0.6807	18767	0.0683
Gaussian SVR	0.2887	0.9208	10434	0.0374

Table 2: Results for Q2-50GB

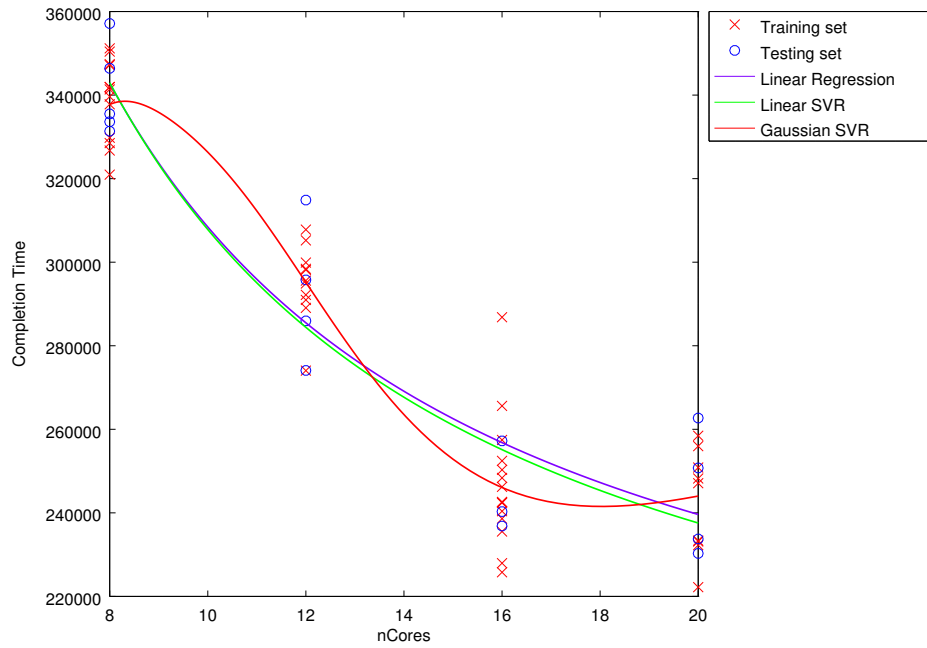


Figure 2: Completion time vs ncores for query Q2 with datasize 50

1.2 Query Q3

1.2.1 Query Q3 — Datasize 50GB

Model	RMSE	R ²	Mean absolute error	Mean relative error
Linear regression	0.4364	0.7371	14301	0.0707
Linear SVR	0.3581	0.8593	10636	0.0527
Polynomial SVR (2)	0.7692	0.1922	22882	0.1221
Polynomial SVR (3)	0.4719	0.7356	12895	0.0635
Polynomial SVR (4)	0.6625	0.4085	18813	0.0980
Polynomial SVR (6)	0.6111	0.5019	16636	0.0864
Gaussian SVR	0.2815	0.9430	8413	0.0456

Table 3: Results for Q3-50GB

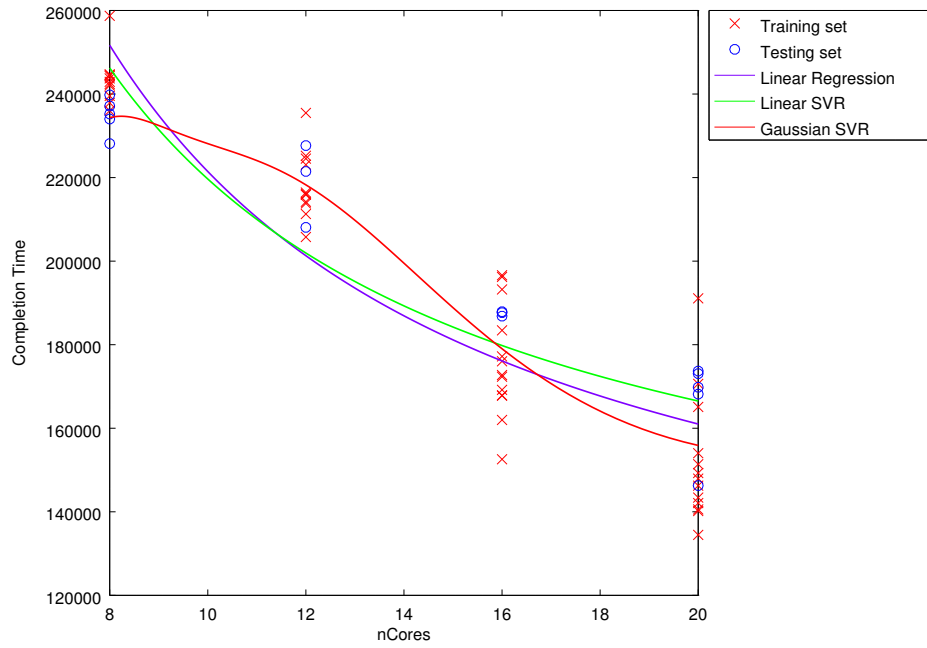


Figure 3: Completion time vs ncores for query Q3 with datasize 50

1.3 Query Q4

1.3.1 Query Q4 — Datasize 50GB

Model	RMSE	R ²	Mean absolute error	Mean relative error
Linear regression	0.4439	0.8284	26027	0.0969
Linear SVR	0.4490	0.8300	27044	0.1029
Polynomial SVR (2)	1.0085	0.1628	66229	0.2716
Polynomial SVR (3)	0.6029	0.7055	37412	0.1473
Polynomial SVR (4)	0.8773	0.3698	55542	0.2300
Polynomial SVR (6)	0.7938	0.4631	47441	0.1942
Gaussian SVR	0.2888	0.9308	14254	0.0484

Table 4: Results for Q4-50GB

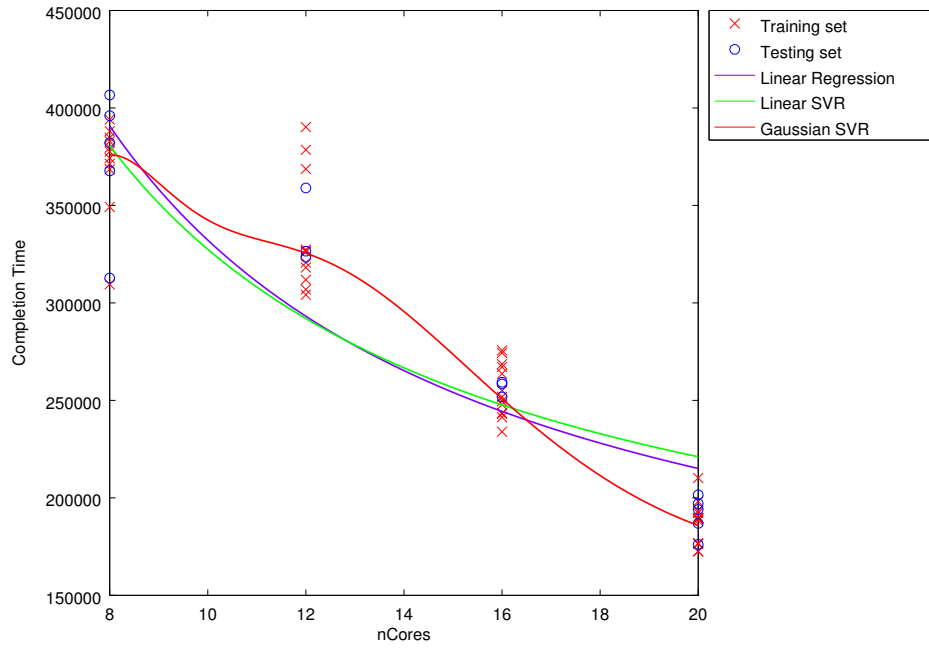


Figure 4: Completion time vs ncores for query Q4 with datasize 50