

# Machine Learning techniques to Model Data Intensive Application Performance

A. Battistello      P. Ferretti

11 maggio 2016

## Indice

<b>1</b>	<b>Prime analisi</b>	<b>2</b>
1.1	SVR vs Linear Regression . . . . .	2
<b>2</b>	<b>Tests</b>	<b>4</b>
2.1	Query R1 . . . . .	4
2.1.1	R1 – Datasize 250 . . . . .	4
2.1.2	R1 – Datasize 500 . . . . .	5
2.1.3	R1 – Datasize 750 . . . . .	6
2.1.4	R1 – Datasize 1000 . . . . .	7
2.2	Query R2 . . . . .	8
2.2.1	R2 – Datasize 250 . . . . .	8
2.2.2	R2 – Datasize 500 . . . . .	9
2.2.3	R2 – Datasize 750 . . . . .	10
2.2.4	R2 – Datasize 1000 . . . . .	11
2.3	Query R3 . . . . .	12
2.3.1	R3 – Datasize 250 . . . . .	12
2.3.2	R3 – Datasize 500 . . . . .	13
2.3.3	R3 – Datasize 750 . . . . .	14
2.3.4	R3 – Datasize 1000 . . . . .	15
2.4	Query R4 . . . . .	16
2.4.1	R4 – Datasize 250 . . . . .	16
2.4.2	R4 – Datasize 500 . . . . .	17
2.4.3	R4 – Datasize 750 . . . . .	18
2.4.4	R4 – Datasize 1000 . . . . .	19
2.5	Query R5 . . . . .	20
2.5.1	R5 – Datasize 250 . . . . .	20
2.5.2	R5 – Datasize 500 . . . . .	21
2.5.3	R5 – Datasize 750 . . . . .	22
2.5.4	R5 – Datasize 1000 . . . . .	23

# 1 Prime analisi

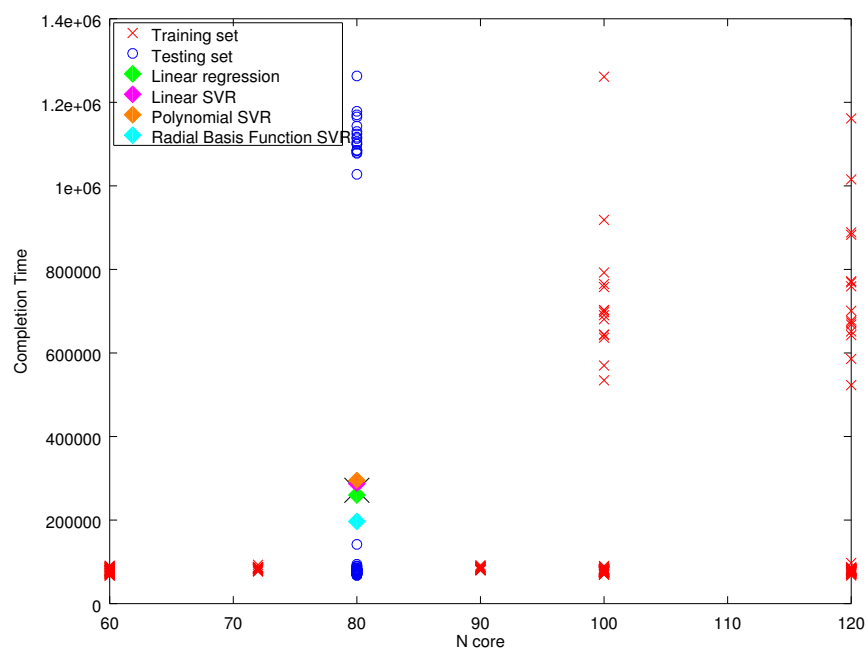
## 1.1 SVR vs Linear Regression

Presa in considerazione la query R2, cerchiamo di prevedere il tempo di esecuzione della query con 80 cores. Creeremo i nostri modelli facendo training su numeri di cores diversi da quello di test: 60, 72, 90, 100, 120. Dai risultati potremo confrontare la performance della regressione lineare rispetto a vari modelli di Support Vector Regression (lineare, polinomiale, sigmoidale).

Come si può vedere dalla Tabella 1 i risultati migliori si hanno dalla SVR lineare, mentre gli altri due tipi di SVR sono addirittura peggiori della semplice regressione lineare, probabilmente per problemi di *overfit*.

Modello	RMSE	R <sup>2</sup>	Errore assoluto medio	Errore relativo medio	Differenza medie
Regressione lineare	0.0940	0.9952	213397	0.0295	-0.0378
SVR lineare	0.0722	0.9991	220018	0.1730	0.0526
SVR polinomiale	0.1050	0.9976	226093	0.1831	0.0780
SVR sigmoidale	0.5862	0.9802	279777	0.2286	-0.2487

**Tabella 1:** Risultati per il primo test



**Figura 1:** Test su numero di cores. La croce nera indica la media originale dei valori di test.

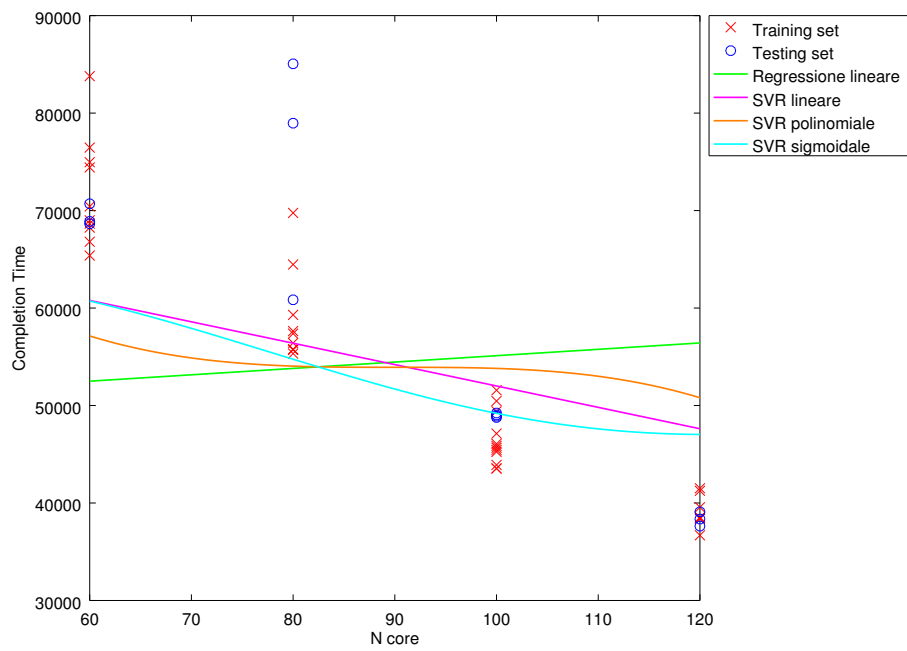
## 2 Fixed Datasize

### 2.1 Query R1

#### 2.1.1 R1 – Datasize 250

Modello	RMSE	R <sup>2</sup>	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.1795	0.9757	56367	0.2186
SVR lineare	0.1224	0.9927	55443	0.0743
SVR polinomiale	1.1146	0.8420	62456	1.2615
SVR sigmoidale	0.5988	0.7769	58614	0.5452

**Tabella 2:** Risultati per il test su query R1 con datasize 250

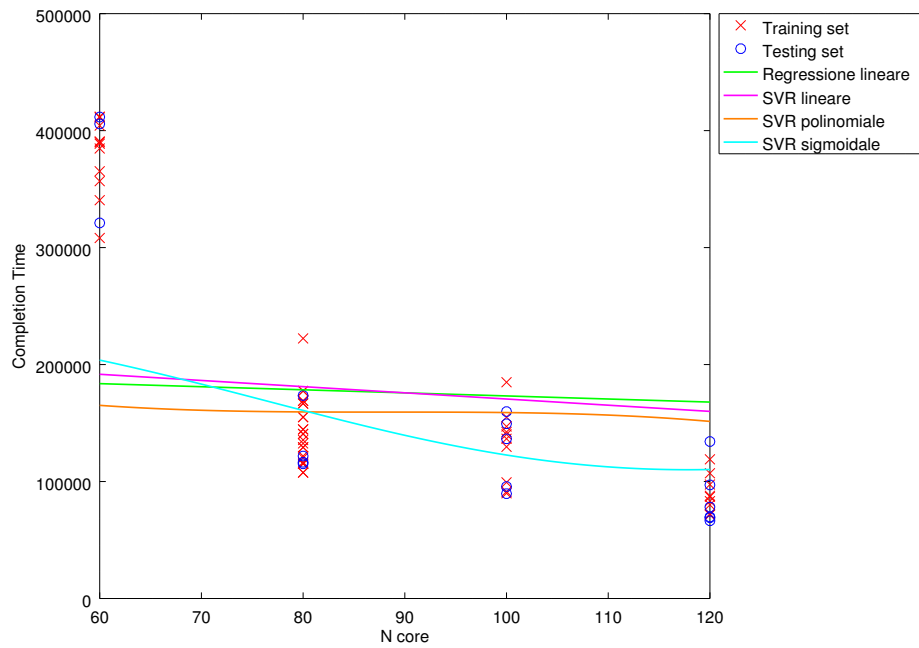


**Figura 2:** Plot per il test su query R1 con datasize 250

### 2.1.2 R1 – Datasize 500

Modello	RMSE	$R^2$	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.5279	0.7681	100565	1.8233
SVR lineare	0.2053	0.9737	96467	0.8926
SVR polinomiale	0.8232	0.9846	105694	4.2244
SVR sigmoideale	0.6578	0.8483	101982	1.9743

**Tabella 3:** Risultati per il test su query R1 con datasize 500

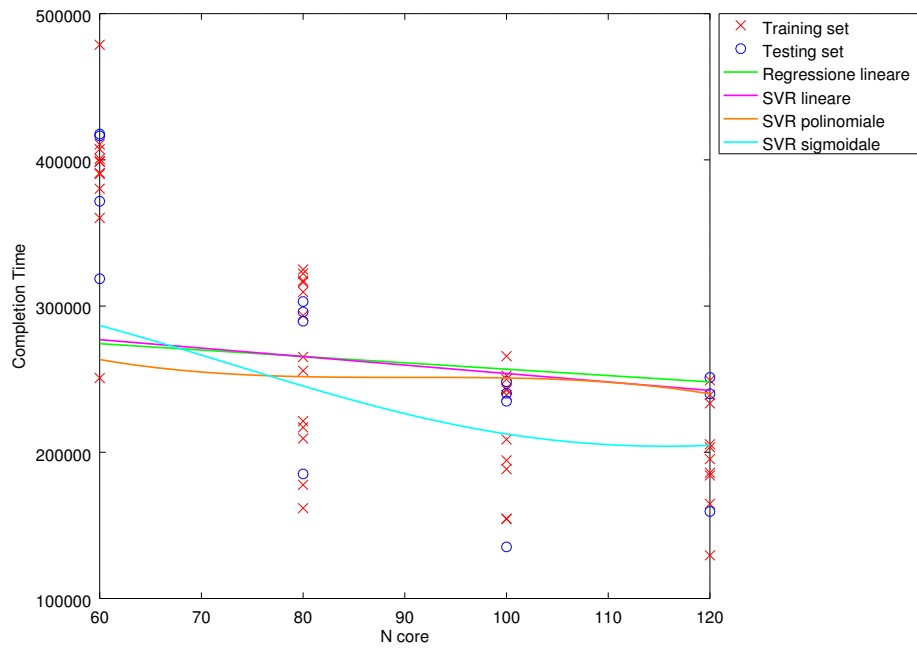


**Figura 3:** Plot per il test su query R1 con datasize 500

### 2.1.3 R1 – Datasize 750

Modello	RMSE	$R^2$	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.1606	0.9676	272212	1.2198
SVR lineare	0.1728	0.9644	273318	2.5279
SVR polinomiale	0.3743	0.8626	285239	1.1986
SVR sigmoideale	0.1082	0.9870	269280	0.3210

**Tabella 4:** Risultati per il test su query R1 con datasize 750

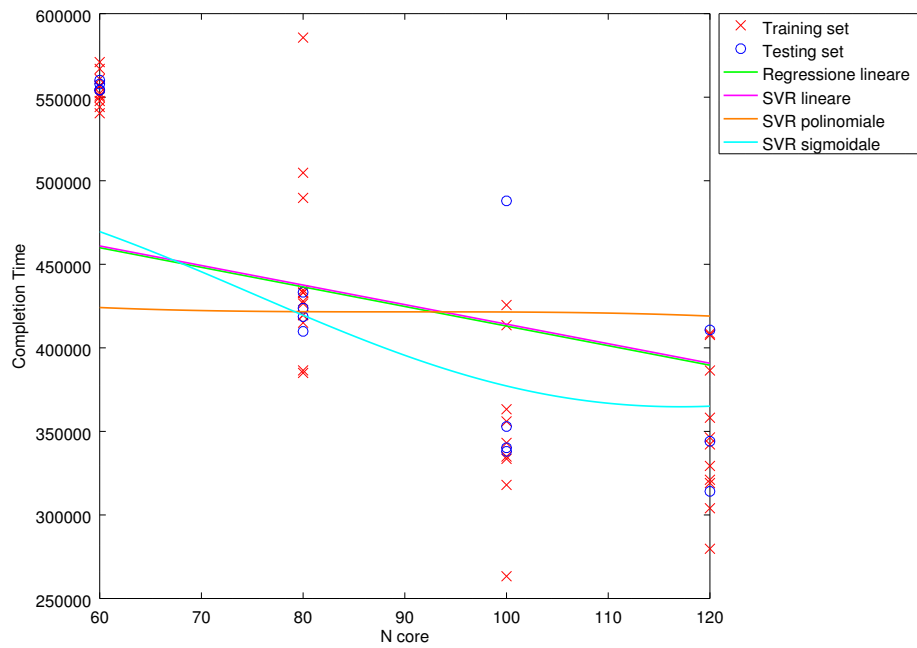


**Figura 4:** Plot per il test su query R1 con datasize 750

### 2.1.4 R1 – Datasize 1000

Modello	RMSE	$R^2$	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.1019	0.9875	431805	0.3050
SVR lineare	0.0943	0.9897	431858	0.3831
SVR polinomiale	0.6198	0.9260	473497	3.3538
SVR sigmoideale	0.0989	0.9920	432081	0.4642

**Tabella 5:** Risultati per il test su query R1 con datasize 1000



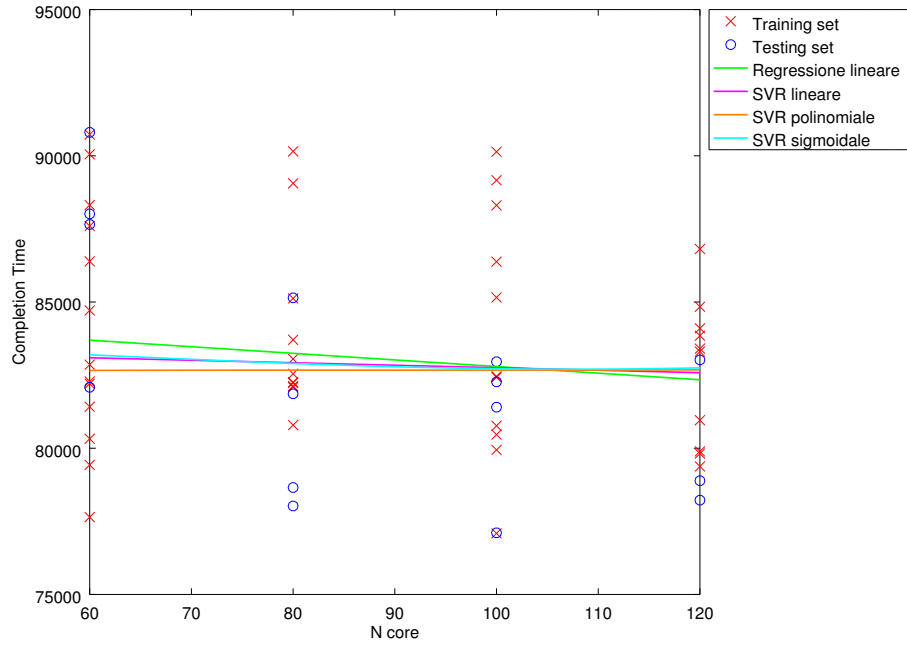
**Figura 5:** Plot per il test su query R1 con datasize 1000

## 2.2 Query R2

### 2.2.1 R2 – Datasize 250

Modello	RMSE	R <sup>2</sup>	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.2945	0.9300	83930	1.2256
SVR lineare	0.3331	0.9107	83944	1.3726
SVR polinomiale	0.5539	0.8927	84628	1.6366
SVR sigmoidale	0.4829	0.8611	84329	1.6959

**Tabella 6:** Risultati per il test su query R2 con datasize 250



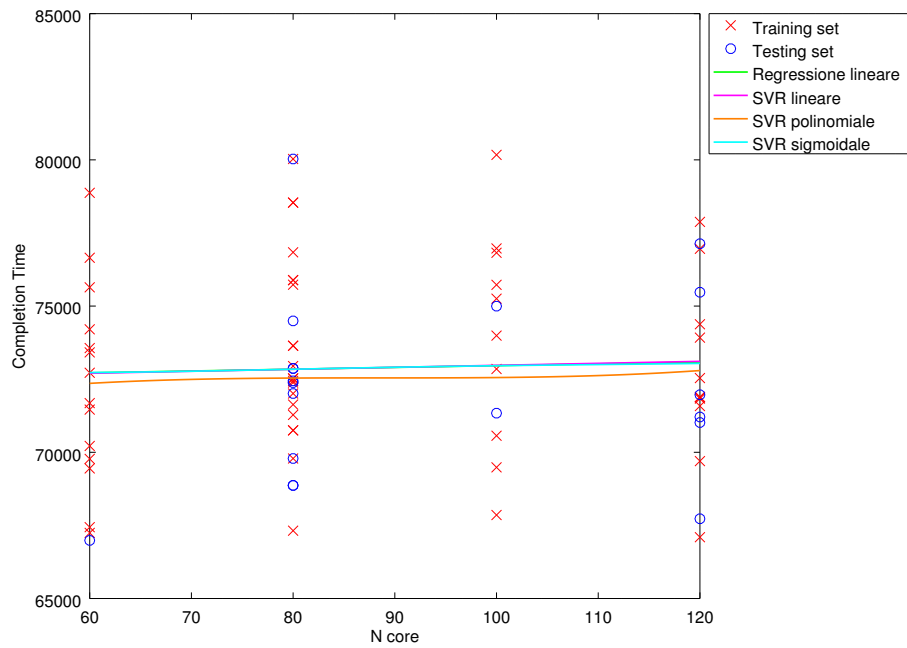
**Figura 6:** Plot per il test su query R2 con datasize 250



## 2.2.2 R2 – Datasize 500

Modello	RMSE	R <sup>2</sup>	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.1810	0.9688	73280	0.4963
SVR lineare	0.1800	0.9698	73280	0.4679
SVR polinomiale	0.4380	0.8193	73907	2.5618
SVR sigmoideale	0.2172	0.9578	73375	0.4690

**Tabella 7:** Risultati per il test su query R2 con datasize 500

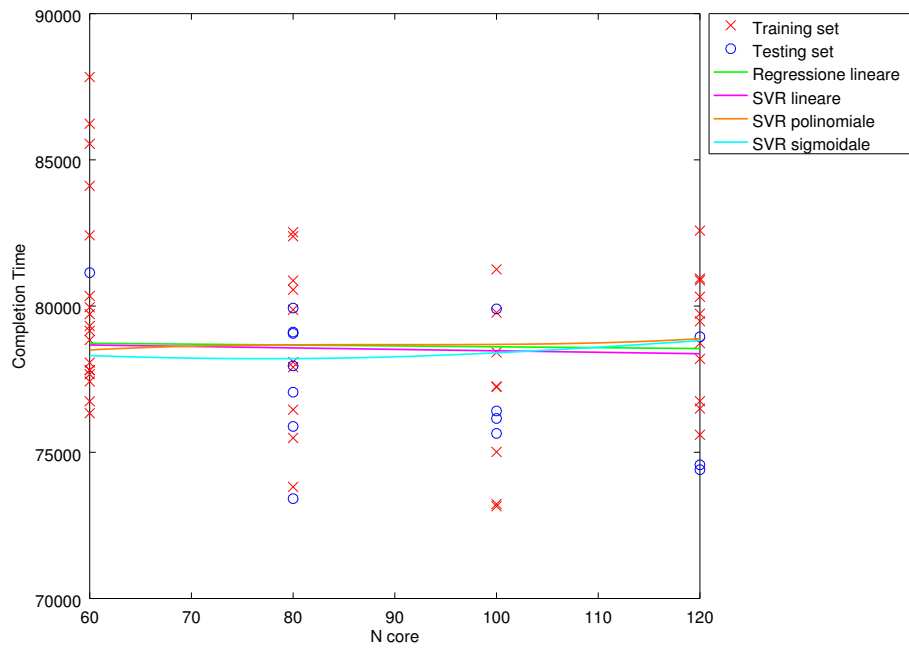


**Figura 7:** Plot per il test su query R2 con datasize 500

### 2.2.3 R2 – Datasize 750

Modello	RMSE	$R^2$	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.2172	0.9198	79129	1.6851
SVR lineare	0.2177	0.9219	79103	0.5003
SVR polinomiale	0.6016	0.7222	80166	9.7460
SVR sigmoideale	0.2593	0.8958	79206	0.4017

**Tabella 8:** Risultati per il test su query R2 con datasize 750

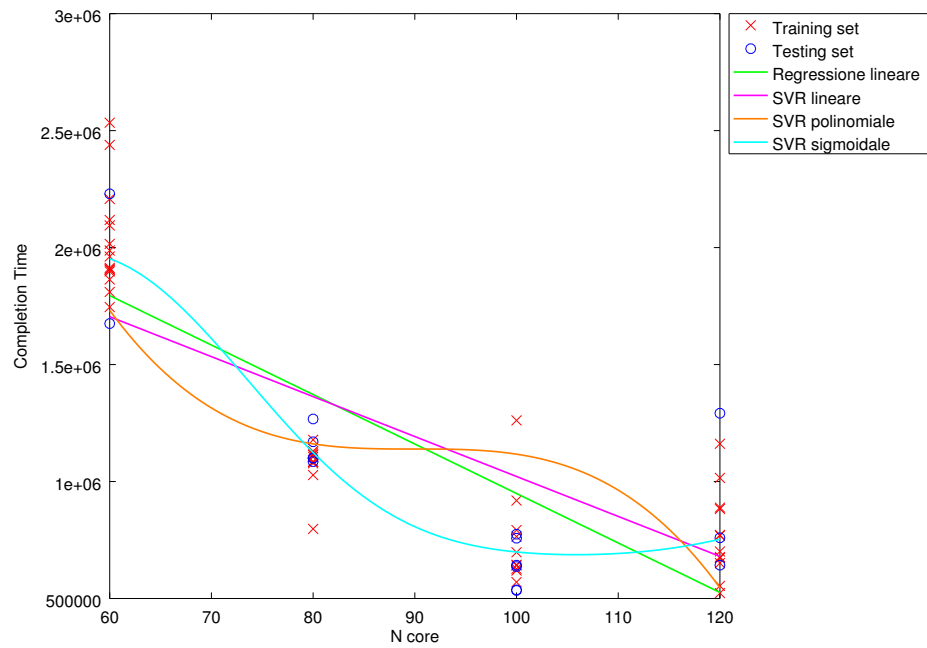


**Figura 8:** Plot per il test su query R2 con datasize 750

## 2.2.4 R2 – Datasize 1000

Modello	RMSE	R <sup>2</sup>	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.6206	0.4222	1461868	1.1535
SVR lineare	0.6184	0.5211	1449291	1.5072
SVR polinomiale	0.6906	0.3466	1456778	40.6253
SVR sigmoidale	0.3406	0.8269	1289489	0.6985

**Tabella 9:** Risultati per il test su query R2 con datasize 1000



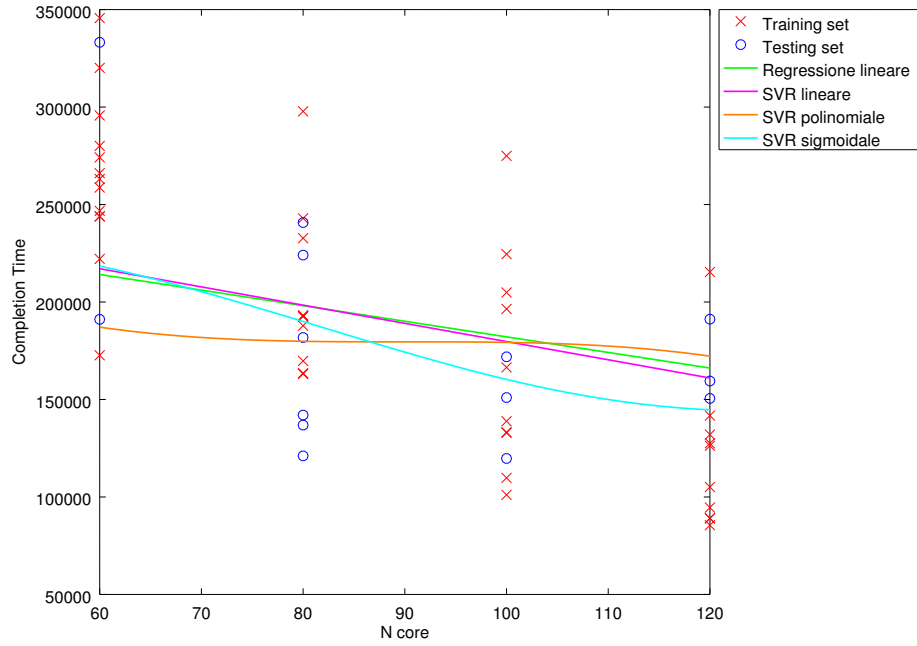
**Figura 9:** Plot per il test su query R2 con datasize 1000

## 2.3 Query R3

### 2.3.1 R3 – Datasize 250

Modello	RMSE	R <sup>2</sup>	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.1367	0.9728	196466	1.2421
SVR lineare	0.1449	0.9716	197333	1.8465
SVR polinomiale	0.2522	0.9379	203404	0.7703
SVR sigmoidale	0.3566	0.8594	208542	0.5891

**Tabella 10:** Risultati per il test su query R3 con datasize 250

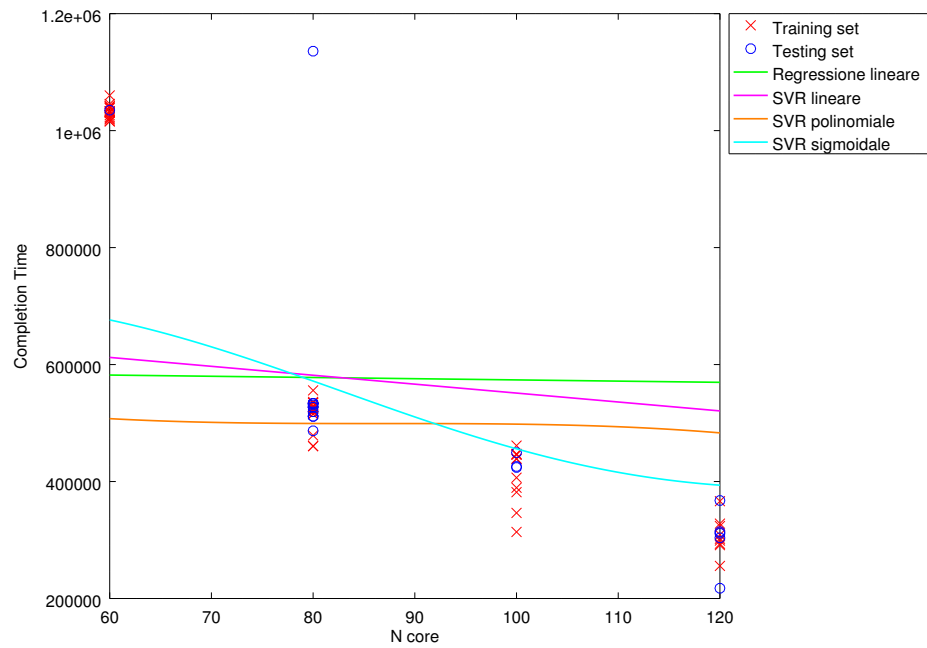


**Figura 10:** Plot per il test su query R3 con datasize 250

### 2.3.2 R3 – Datasize 500

Modello	RMSE	R <sup>2</sup>	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.4296	0.7341	616021	1.6412
SVR lineare	0.4369	0.7383	624688	1.1396
SVR polinomiale	0.5711	0.5604	639247	0.6813
SVR sigmoideale	0.5229	0.6132	631858	159.9640

**Tabella 11:** Risultati per il test su query R3 con datasize 500

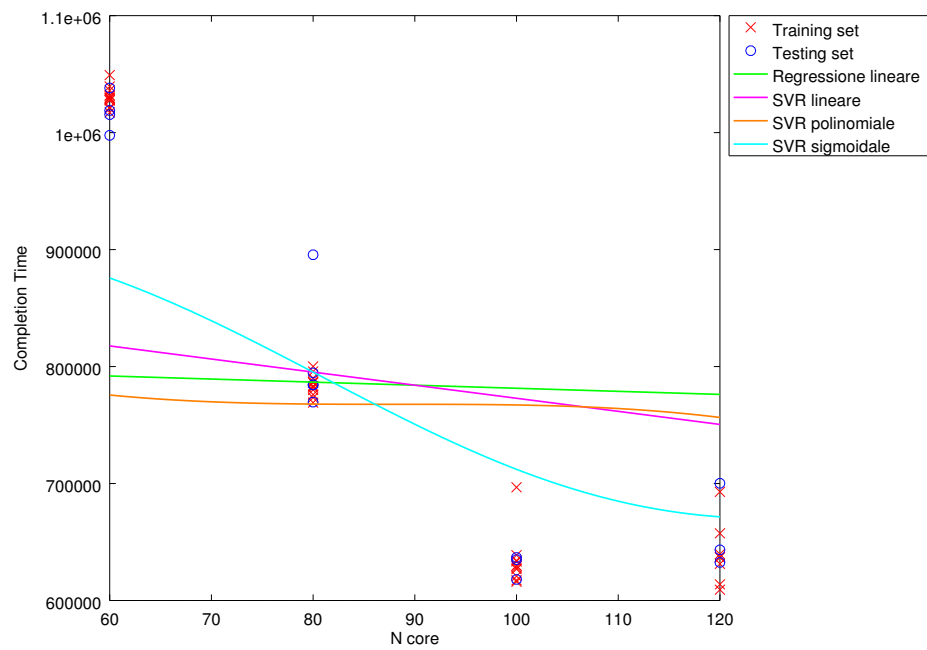


**Figura 11:** Plot per il test su query R3 con datasize 500

### 2.3.3 R3 – Datasize 750

Modello	RMSE	$R^2$	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.0336	0.9987	788712	0.1637
SVR lineare	0.0701	0.9953	794619	0.2754
SVR polinomiale	0.2331	0.9436	814941	0.8557
SVR sigmoideale	0.2472	0.9388	812792	1.2858

**Tabella 12:** Risultati per il test su query R3 con datasize 750

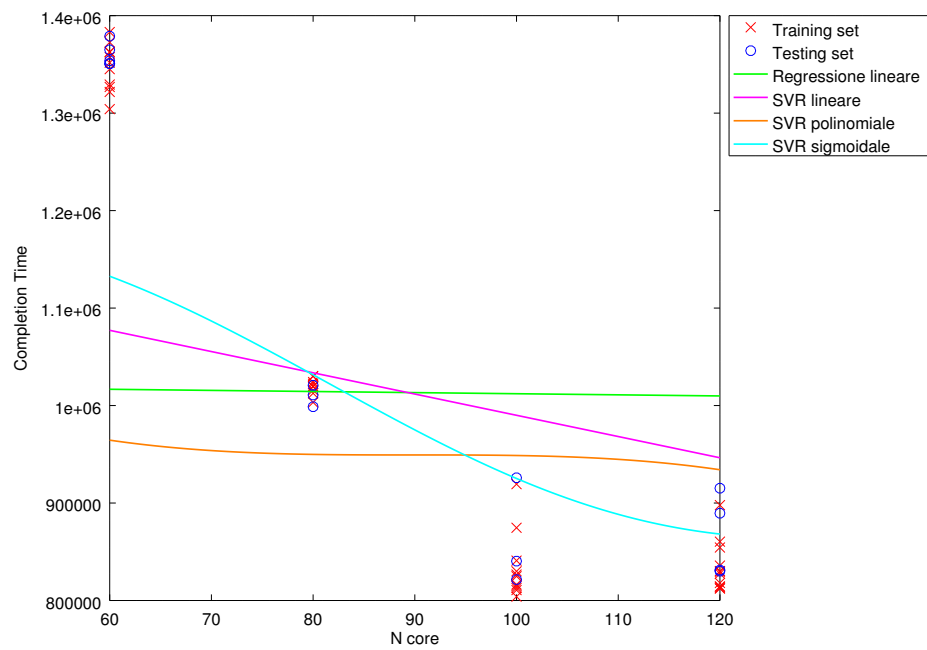


**Figura 12:** Plot per il test su query R3 con datasize 750

### 2.3.4 R3 – Datasize 1000

Modello	RMSE	$R^2$	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.0225	0.9995	1017476	0.2029
SVR lineare	0.1039	0.9916	1035277	0.3553
SVR polinomiale	0.3761	0.8810	1075249	0.4485
SVR sigmoidale	0.3269	0.9232	1063963	0.5882

**Tabella 13:** Risultati per il test su query R3 con datasize 1000



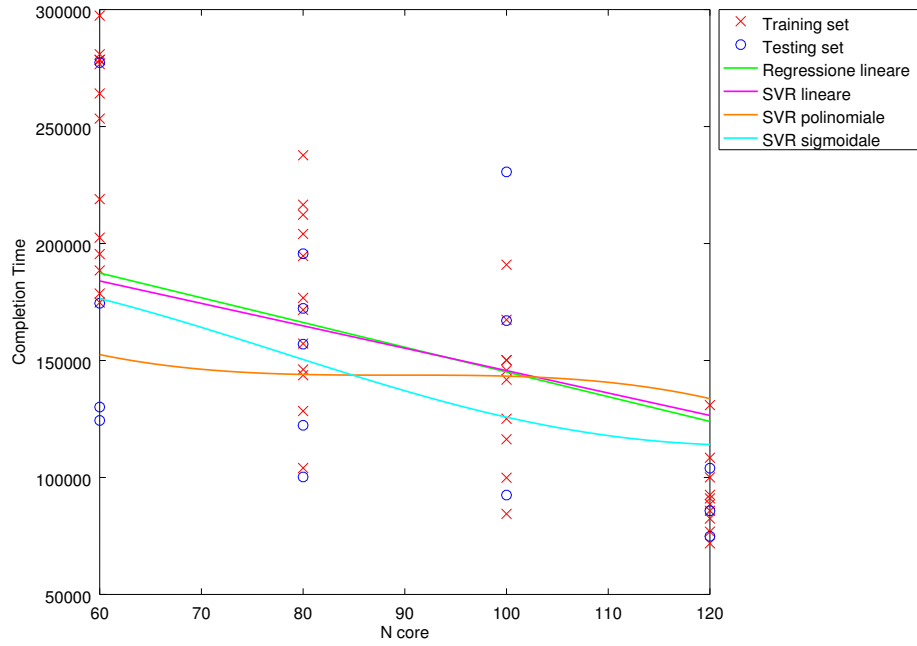
**Figura 13:** Plot per il test su query R3 con datasize 1000

## 2.4 Query R4

### 2.4.1 R4 – Datasize 250

Modello	RMSE	$R^2$	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.1366	0.9770	163760	0.2719
SVR lineare	0.1426	0.9758	164053	0.2698
SVR polinomiale	0.3301	0.8880	172301	5.7912
SVR sigmoidale	0.2201	0.9539	164805	0.3189

**Tabella 14:** Risultati per il test su query R4 con datasize 250



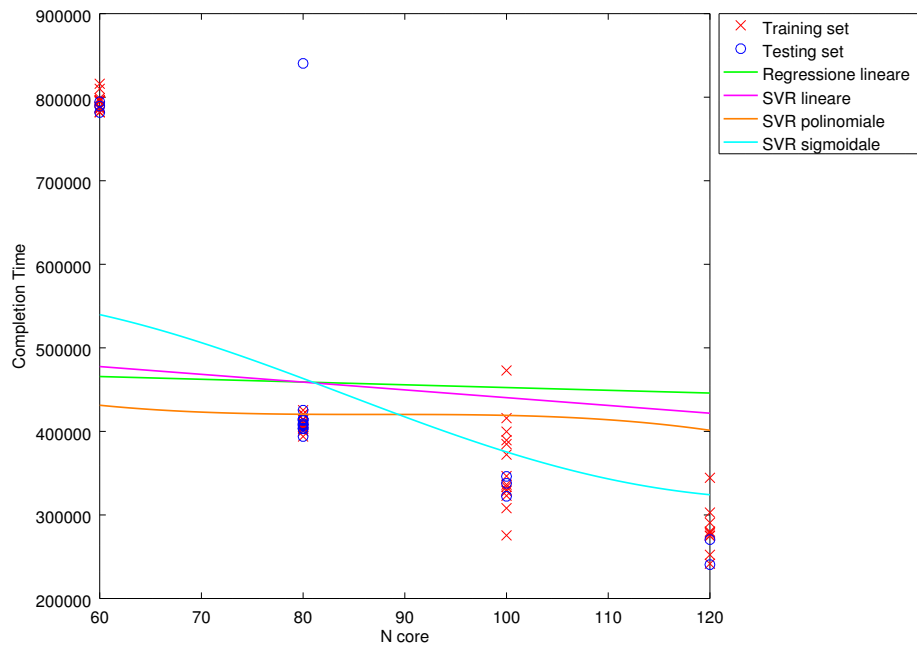
**Figura 14:** Plot per il test su query R4 con datasize 250



## 2.4.2 R4 – Datasize 500

Modello	RMSE	$R^2$	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.4873	0.7455	481376	1.6211
SVR lineare	0.4836	0.7583	488122	3.2325
SVR polinomiale	0.5518	0.6898	511493	2.2149
SVR sigmoideale	0.5141	0.7193	494295	1.1828

**Tabella 15:** Risultati per il test su query R4 con datasize 500

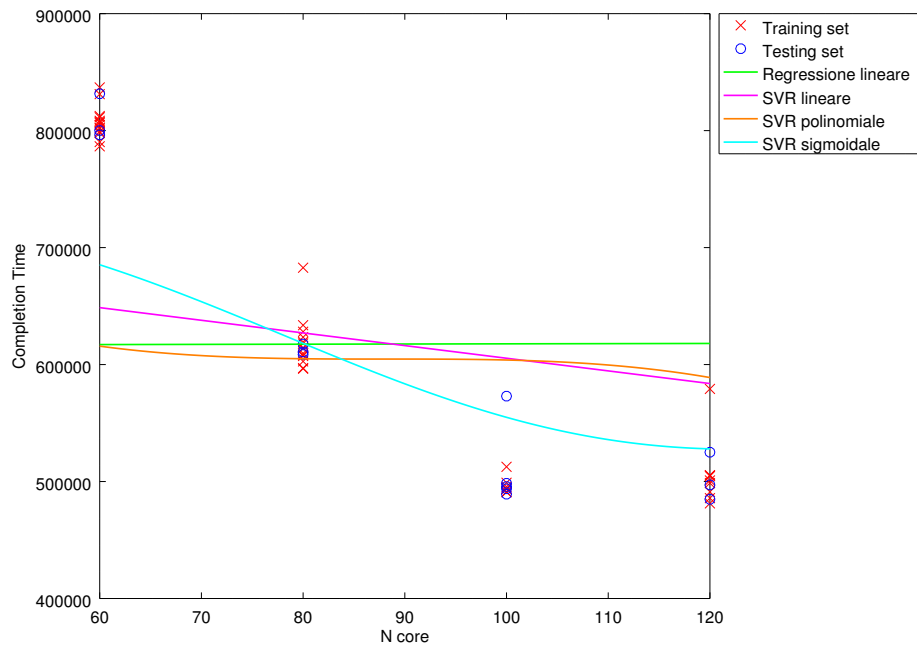


**Figura 15:** Plot per il test su query R4 con datasize 500

### 2.4.3 R4 – Datasize 750

Modello	RMSE	$R^2$	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.0248	0.9993	620160	0.1281
SVR lineare	0.0816	0.9936	627111	0.7623
SVR polinomiale	0.3360	0.9021	649491	0.7968
SVR sigmoidale	0.1712	0.9722	632970	0.4241

**Tabella 16:** Risultati per il test su query R4 con datasize 750

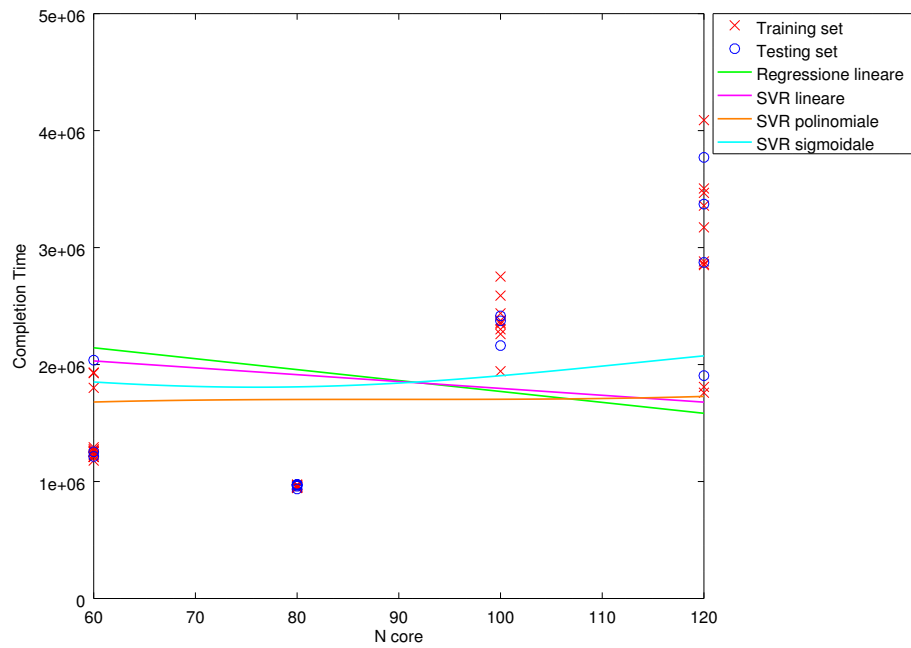


**Figura 16:** Plot per il test su query R4 con datasize 750

#### 2.4.4 R4 – Datasize 1000

Modello	RMSE	$R^2$	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.1125	0.9886	1937660	0.3304
SVR lineare	0.1180	0.9883	1933919	0.8125
SVR polinomiale	0.6788	0.8269	2236602	4.5318
SVR sigmoideale	0.2119	0.9649	2001964	0.2980

**Tabella 17:** Risultati per il test su query R4 con datasize 1000



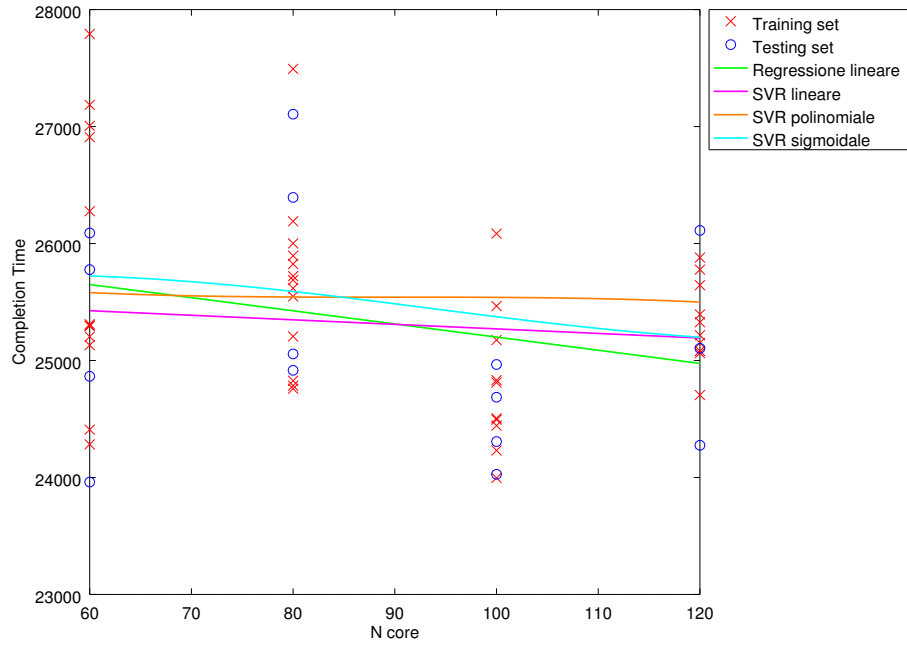
**Figura 17:** Plot per il test su query R4 con datasize 1000

## 2.5 Query R5

### 2.5.1 R5 – Datasize 250

Modello	RMSE	$R^2$	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.7396	0.5197	25756	1.4684
SVR lineare	0.7232	0.7663	25801	2.5330
SVR polinomiale	1.2789	0.1267	26188	5.1048
SVR sigmoideale	0.9120	0.5705	25976	7.3267

**Tabella 18:** Risultati per il test su query R5 con datasize 250

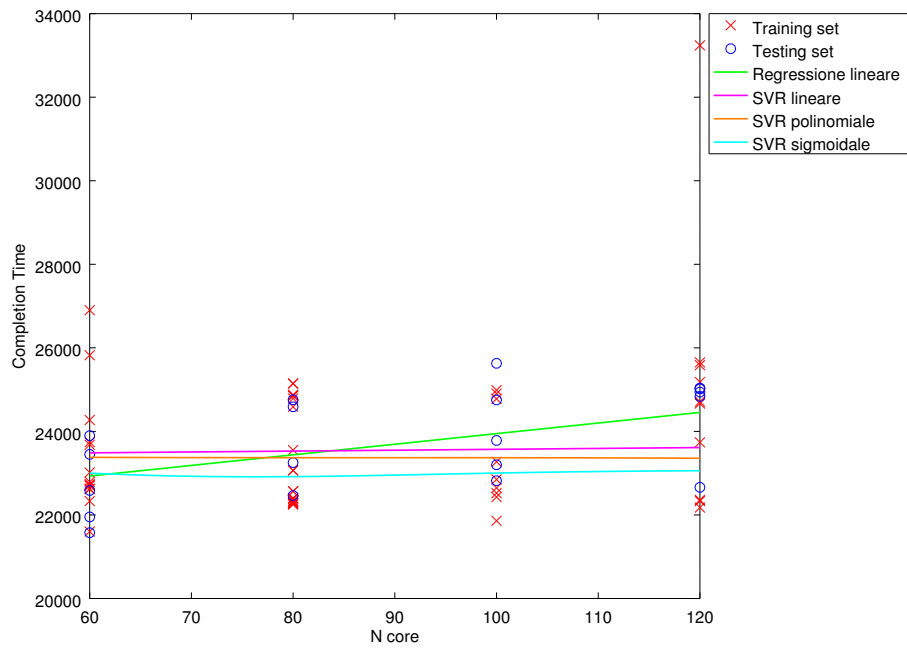


**Figura 18:** Plot per il test su query R5 con datasize 250

### 2.5.2 R5 – Datasize 500

Modello	RMSE	$R^2$	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.5126	0.5127	24308	1.2420
SVR lineare	0.1699	0.9472	23876	1.5843
SVR polinomiale	1.0731	0.6518	24596	1.3832
SVR sigmoideale	0.5384	0.4989	24179	0.9291

**Tabella 19:** Risultati per il test su query R5 con datasize 500

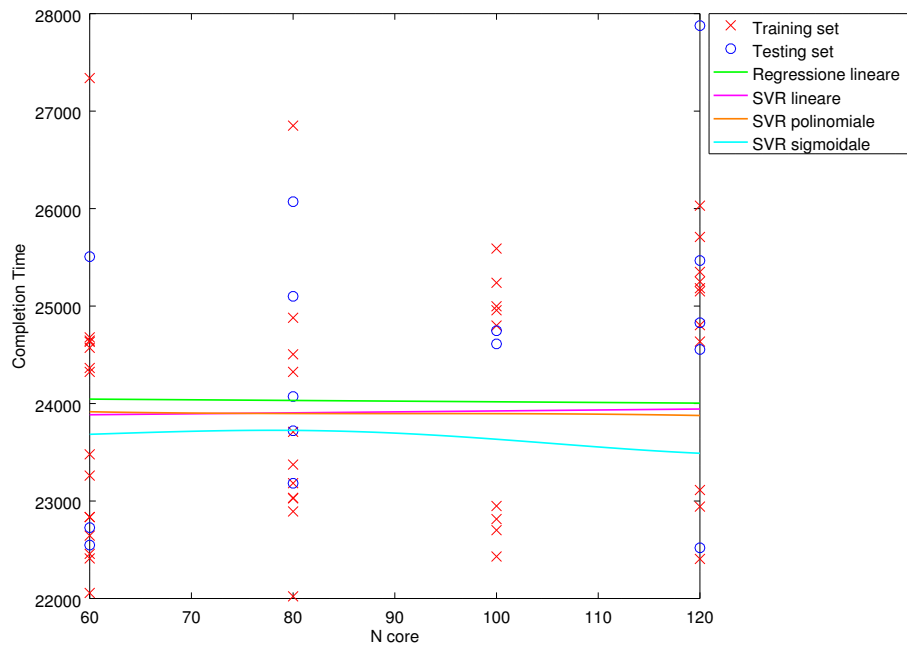


**Figura 19:** Plot per il test su query R5 con datasize 500

### 2.5.3 R5 – Datasize 750

Modello	RMSE	$R^2$	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.9679	0.1647	24739	1.1966
SVR lineare	0.9723	0.2832	24696	1.3507
SVR polinomiale	1.1618	0.0687	24899	2.0287
SVR sigmoideale	1.0636	0.1689	24810	1.5607

**Tabella 20:** Risultati per il test su query R5 con datasize 750

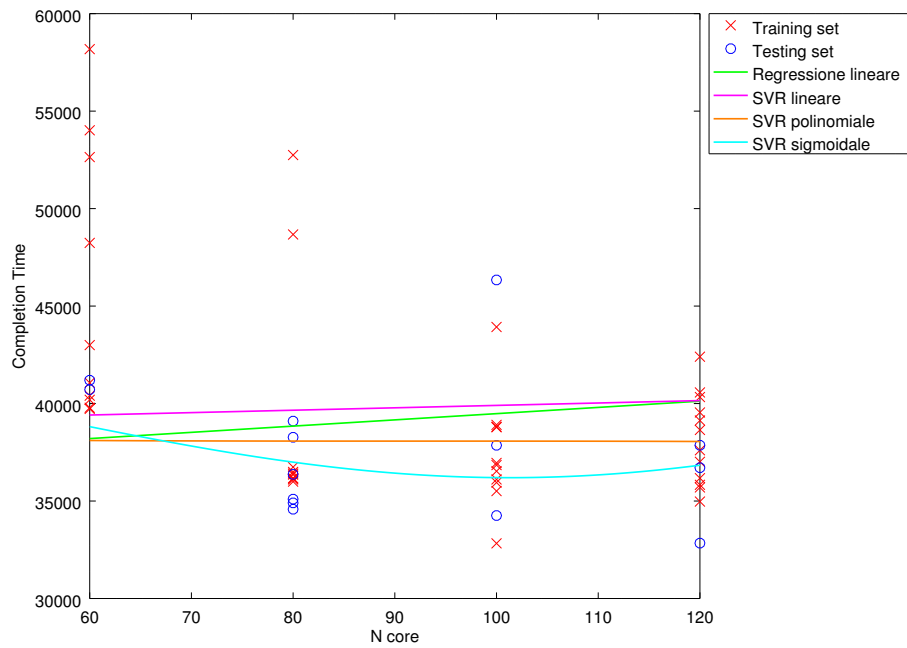


**Figura 20:** Plot per il test su query R5 con datasize 750

#### 2.5.4 R5 – Datasize 1000

Modello	RMSE	$R^2$	Errore assoluto medio	Errore relativo medio
Regressione lineare	0.6269	0.1119	41327	0.9856
SVR lineare	0.5173	0.7606	41106	0.8531
SVR polinomiale	0.4304	0.7565	40984	1.1478
SVR sigmoideale	0.3310	0.8067	40590	0.5291

**Tabella 21:** Risultati per il test su query R5 con datasize 1000



**Figura 21:** Plot per il test su query R5 con datasize 1000