

Machine Learning techniques to Model Data Intensive Application Performance

A. Battistello P. Ferretti

4 maggio 2016

Indice

1 Prime analisi	1
1.1 SVR vs Linear Regression	1

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 1.1 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 ²	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

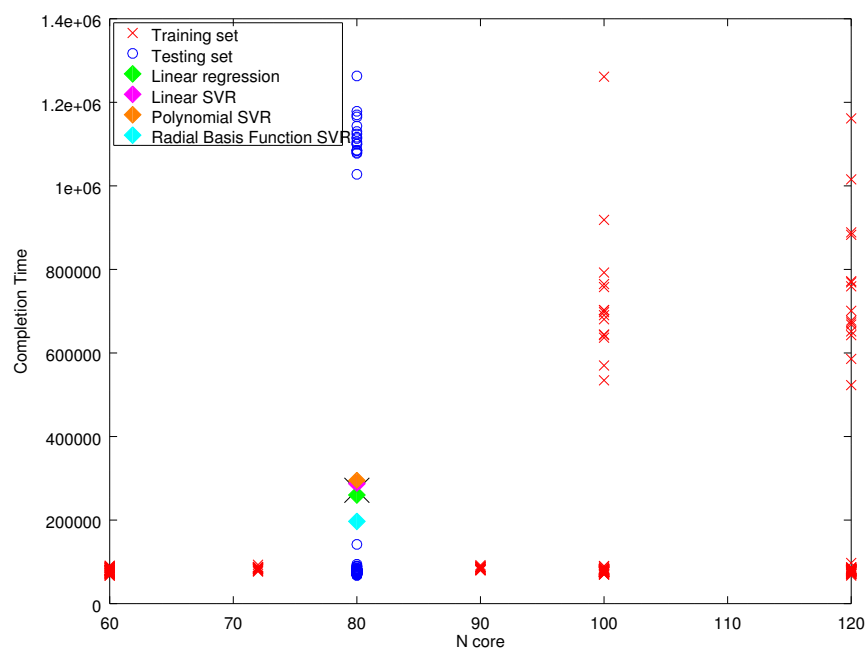


Figura 1: Test su numero di cores. La croce nera indica la media originale dei valori di test.