

- Simple Linear Regression
- Multiple Linear Regression
- Support Vector Machine
- Decision tree
- Random Forest

Dataset: 50_Startups.csv (Profit prediction)

Independent variables (R&D spend, Administration, Marketing Spend, State)

Dependent variable/ Output : Profit

Simple Linear Regression:

Sklearn.linear_model.LinearRegression

Simple linear regression cannot be performed on multiple input variables

Multiple Linear Regression:

Sklearn.linear_model.LinearRegression

R2 Score	0.93
----------	------

Support Vector Machine:

Sklearn.svm.SVR

Kernel	Gamma	C	Epsilon	R2 Score
linear	scale(default)	1	0.1 (default)	0.90
linear	scale(default)	1000	0.1 (default)	-
linear	scale(default)	2000	0.1 (default)	-
linear	scale(default)	3000	0.1 (default)	-
linear	auto	1	0.1 (default)	0.90
linear	auto	1000	0.1 (default)	-
linear	auto	2000	0.1 (default)	-
linear	auto	3000	0.1 (default)	-
poly	scale(default)	1	0.1 (default)	-49195.75
poly	scale(default)	1000	0.1 (default)	-0.119
poly	scale(default)	2000	0.1 (default)	-0.043

poly	scale(default)	3000	0.1 (default)	0.005
poly	auto	1	0.1 (default)	-
poly	auto	1000	0.1 (default)	-
poly	auto	2000	0.1 (default)	-
poly	auto	3000	0.1 (default)	-
rbf(default)	scale(default)	1	0.1 (default)	-38108109.1
rbf(default)	scale(default)	1000	0.1 (default)	-46.22
rbf(default)	scale(default)	2000	0.1 (default)	-10
rbf(default)	scale(default)	3000	0.1 (default)	-4.6
rbf(default)	auto	1	0.1 (default)	-1
rbf(default)	auto	1000	0.1 (default)	-1
rbf(default)	auto	2000	0.1 (default)	-5
rbf(default)	auto	3000	0.1 (default)	-1
sigmoid	scale(default)	1	0.1 (default)	-3567106284.1
sigmoid	scale(default)	1000	0.1 (default)	-3610.6
sigmoid	scale(default)	2000	0.1 (default)	-913
sigmoid	scale(default)	3000	0.1 (default)	-410
sigmoid	auto	1	0.1 (default)	-1.2
sigmoid	auto	1000	0.1 (default)	-1.2
sigmoid	auto	2000	0.1 (default)	-1.2
sigmoid	auto	3000	0.1 (default)	-1.2
precomputed	scale(default)	1	0.1 (default)	-
precomputed	scale(default)	1000	0.1 (default)	-
precomputed	scale(default)	2000	0.1 (default)	-
precomputed	scale(default)	3000	0.1 (default)	-
precomputed	auto	1	0.1 (default)	-
precomputed	auto	1000	0.1 (default)	-

precomputed	auto	2000	0.1 (default)	-
precomputed	auto	3000	0.1 (default)	-

Kernel {'linear', 'poly', 'rbf', 'sigmoid', 'precomputed'} default : rbf

Degree {integer value} default : 3 [should be non-neg for poly kernel] ignored by other kernels

Gamma {'scale', 'auto'} default : scale

C

Epsilon

DecisionTree:

Sklearn.tree.DecisionTreeRegressor

Criterion	Splitter	max_features	R2 Score
squared_error(default)	best(default)	sqrt	0.80
squared_error(default)	best(default)	log2	0.41
squared_error(default)	random	sqrt	0.88
squared_error(default)	random	log2	-0.56
friedman_mse	best(default)	sqrt	0.65
friedman_mse	best(default)	log2	0.53
friedman_mse	random	sqrt	0.64
friedman_mse	random	log2	0.67
absolute_error	best(default)	sqrt	-0.027
absolute_error	best(default)	log2	0.65
absolute_error	random	sqrt	0.22
absolute_error	random	log2	-0.54
poisson	best(default)	sqrt	0.29
poisson	best(default)	log2	0.23
poisson	random	sqrt	0.62
poisson	random	log2	-0.78

Among all the algorithms that we have verified, Multiple Linear regression looks to give maximum accurate results for this data set.

Model	Max R2 Score calculated
Simple Linear Regression	-
Multiple Linear Regression	0.93
Support Vector Machine	0.90
Decision Tree	0.88