

Assignment-Regression

1.Problem statement –

stage1-Supervised learning

stage2-Machine learning

stage3-Regression

2.Total number of rows-1338

Total number of columns-6

3.changed String to number – Smoker,sex

4.R2_score- best model-Random forest

To find Machine learning regression method using R2 value

1.Multiple Linear Regression – R2 value –0.78

2.SVM –

	R_score	C=0.01	C=1	C=10	C=100
linear	-0.113	-0.084	-0.1135	-0.0016	0.543
rbf	-0.088	-0.089	-0.0882	-0.0827	-0.126
poly	-0.0674	-0.088	-0.0674	-0.0982	-0.0987
sigmoid	-0.0899	-0.089	-0.0899	-0.0911	-0.1276

3.Decision Tree

criterion	splitter	R_score
Squared error	Best	0.68076
Friedman_mse	Best	0.7013
Absolute error	Best	0.6689
poisson	Best	0.6640
Friedman_mse	Random	0.6956
Absolute error	Random	0.07479
poisson	Random	0.6990
Squared error	Random	0.6624

Random Forest

N_estimators	R_score
10	0.8373
20	0.849
30	0.85
40	0.856
50	0.850
60	0.850
70	0.8521
80	0.854
90	0.853
100	0.853

Final machine learning best method of regression

Random forest R2 value=0.856

As a result of comparison between multiple algorithm ,We selected random forest algorithm as best model , algorithm R_score value=0.856
