

Task 3: fitting simple models

- For reference, Git Repository for the hometask is available at the following link : <https://github.com/abdul0214/TransferwiseDataScience>
- In order to see the repository, kindly send me an e-mail for making it public or send me GitHub username so that I add you the relevant person in it.
- Task3 has the corresponding Jupyter Notebook which contains code.

We fit the following linear models on white wine dataset from the Wine Quality Data Set and the performance of each model is as follows:

Linear Models

Performance Reports of Linear Models:

Model: LinearRegression

Train r2_score : 0.2879431225091068
Test r2_score : 0.2506416786605612

Train mae : 0.572938287720041
Test mae : 0.6314120523577407

Model: Ridge

Train r2_score : 0.28794088486195923
Test r2_score : 0.2509695974252484

Train mae : 0.5729256779607267
Test mae : 0.6314134327036903

Model: RidgeCV

Train r2_score : 0.2879408848619589
Test r2_score : 0.25096959742516267

Train mae : 0.5729256779607212
Test mae : 0.6314134327037204

Model: SGDRegressor

Train r2_score : 0.28426268029063595
Test r2_score : 0.2552993717776122

Train mae : 0.5740127580524553
Test mae : 0.6319809628482271

Model: Lasso

Train r2_score : 0.0
Test r2_score : -0.005656503687872849

Train mae : 0.6535309995802152
Test mae : 0.7164796699690595

Model: ElasticNet

Train r2_score : 0.0

Test r2_score : -0.005656503687872849

Train mae : 0.6535309995802152
Test mae : 0.7164796699690595

Model: HuberRegressor

Train r2_score : 0.2855640385989474
Test r2_score : 0.25403100779974286

Train mae : 0.5722018489719723
Test mae : 0.6292486396404636

Model: LinearSVR

Train r2_score : 0.2811940002859512
Test r2_score : 0.25197573789623406

Train mae : 0.570006438691438
Test mae : 0.6281717953306518

Model: ExtraTreeRegressor

Train r2_score : 1.0
Test r2_score : 0.00665701881331382

Train mae : 0.0
Test mae : 0.5530612244897959

Model: DecisionTreeRegressor

Train r2_score : 1.0
Test r2_score : 0.029811866859623448

Train mae : 0.0
Test mae : 0.5816326530612245

An Overview of the Linear Modeling results:

Best r2_train : ExtraTreeRegressor 1.0

Best r2_test : SGDRegressor 0.2552993717776122

Best Train MAE : ExtraTreeRegressor 0.0

Best Test MAE : ExtraTreeRegressor 0.5530612244897959

Among linear models, Over-fitting as observed on the following models:

Models with Over-fittings ['ExtraTreeRegressor', 'DecisionTreeRegressor']

Among linear models, Under-fitting as observed on the following models

Models with Under-fittings ['LinearRegression', 'Ridge', 'RidgeCV', 'SGDRegressor', 'Lasso', 'ElasticNet', 'HuberRegressor', 'LinearSVR']

Non - Linear Models

Performance Reports of Non-Linear Models:

Model: RandomForestRegressor

Train R2 : 0.9348320572542389
Test R2 : 0.9348320572542389

Train mae : 0.15634762633996932
Test mae : 0.5247448979591837

Model: XGBRegressor

Train R2 : 0.9405719496284562
Test R2 : 0.9405719496284562

Train mae : 0.1479305915530693
Test mae : 0.557083260039894

Model: NuSVR

Train R2 : 0.4844599653533134
Test R2 : 0.4844599653533134

Train mae : 0.49715574570701
Test mae : 0.5905932715061741

Model: SVR

Train R2 : 0.498578529479269
Test R2 : 0.498578529479269

Train mae : 0.4463006214440417
Test mae : 0.5613065611105376

Model: AdaBoostRegressor

Train R2 : 0.33915274794057126
Test R2 : 0.33915274794057126

Train mae : 0.5699032857170218
Test mae : 0.6362242938253484

An Overview of the Non-Linear Modeling results:

Best r2_train : XGBRegressor 0.9405719496284562

Best r2_test : RandomForestRegressor 0.430739565846599

Best Train MAE : XGBRegressor 0.1479305915530693

Best Test MAE : RandomForestRegressor 0.5246020408163266

Among non-linear models, Over-fitting as observed on the following models:

Models with Over-fittings ['RandomForestRegressor', 'XGBRegressor']

Among linear models, Under-fitting as observed on the following models

Models with Under-fittings ['NuSVR', 'SVR', 'AdaBoostRegressor']