Box Cox transformation

& PCA Report

By Uddesh Soman

UECU Ujjan

* The given assignment includes Crowd at gym data with the target of prediction of number of people at gym for given input attributes.
* Apart from given attributes, minute (of hour) and day (of month) were extracted from date-time object as additional features
* Linear Regression was taken as base line model. Decision tree Regressor , Random Forest Regressor and XGBRegressor were also trained and used to predict the same data.
* Hyper parameters of Decision tree Regressor, Random Forest Regressor and XGBRegressor were set to obtain better results. The models mentioned in result are hyper parameter tuned.
* Timestamp and hour column were same so timestamp column was dropped out of model.
* The same models used to make predictions on original data were used to make predictions on data normalized using Box-Cox transformation.
* The component dimensionality was reduced to explain 90% variance and the resultant features were 7 PCA features.
* The New features were used to train the same models again and predictions were made using them.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Data set type | original data | Normalized data | Score gain After normalization (percentage) | PCA Reduced data | accuracy trade-off after PCA |
| Linear Regression | Training | 0.508 | 0.5474 | 3.94 | 0.5446 | 0.0028 |
| Linear Regression | Testing | 0.5153 | 0.5529 | 3.76 | 0.5509 | 0.002 |
| Decision tree Regressor | Training | 0.8671 | 0.9213 | 5.42 | 0.9235 | -0.0022 |
| Decision tree Regressor | Testing | 0.8238 | 0.8855 | 6.17 | 0.8398 | 0.0457 |
| Random forest Regressor | Training | 0.9136 | 0.95 | 3.64 | 0.9537 | -0.0037 |
| Random forest Regressor | Testing | 0.8784 | 0.9244 | 4.6 | 0.9099 | 0.0145 |
| XGBRegressor | Training | 0.9182 | 0.9422 | 2.4 | 0.9406 | 0.0016 |
| XGBRegressor | Testing | 0.9033 | 0.9341 | 3.08 | 0.9058 | 0.0283 |

Best Recommended model : XGBRegressor