Project Machine Learning

•••

Predict Final Grade of Students

Data

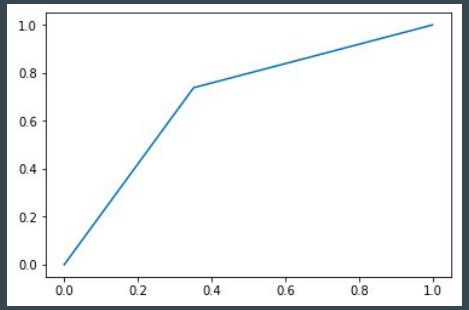
- ★ Students data set
  - 30 features
  - 3 grades (first period, second period, final)
- ★ Algorithms to solve Regression Problems
  - Final grade (0-20) vs. all features
- ★ Algorithms to solve Classification Problems
  - Final grade (failure/success) vs. all features

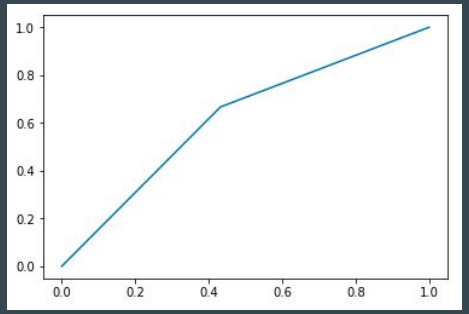
Regression Models

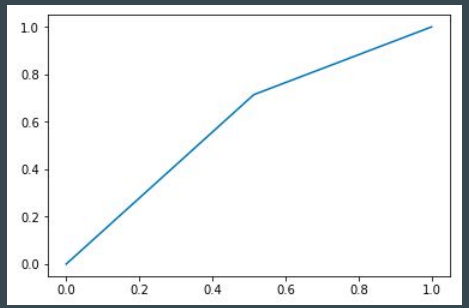
- ★ Multiple Linear Regression Algorithm
  - Performance training set: 0.314
  - Performance testing set: 0.044
- ★ K-Nearest Neighbors Algorithm
  - Performance training set: 0.375
  - Performance testing set: 0.101
- ★ Decision Tree Algorithm
  - Performance training set: 0.993
  - Performance testing set: 0.189
- ★ Conclusion
  - Results not so good >> try classification models

Classification Models

- ★ Logistic Regression Algorithm
  - AUC = 0.69
  - Performance training set: 0.75
  - Performance testing set: 0.69
- ★ Random Forest Classifier Algorithm
  - AUC = 0.62
  - Performance training set: 0.97
  - Performance testing set: 0.62
- ★ Support Vector Machine Algorithm
  - AUC = 0.62
  - Performance training set: 0.82
  - Performance testing set: 0.61
- ★ Conclusion:
  - Better, but still not great







The End

...

Noortje Taal | Data Analist