Mohammad Al-Mamun, Department of Physics, University of Tennessee, Knoxville.



Abstract

I used the survey data from CDC and present quantifiable features on different demographics to measure self-harm vulnerability. I used linear regression models to test significance of different features. Accuracy tests signify, linear regression is not adequate to model suicide rates.

Motivations

 Mental health issues and suicide rates are growing concerns among grad students.

- Intend to parameterize suicide rate from survey data. • Test regression models and verify if that can used to measure vulnerability in different demographics.
- Use features (eg. "Year", "Sex", "Age", "Race") to train these models.
- List significant independent variables from these models.
- Test whether these features are sufficient to make good predictions.
- Compare different models.
- Test the variations of coefficients on parameters.

Methods

I've selected injury based mortality survey data from CDC.

Current survey data relates categorical data (features)

continuous dependent variable (Death rate).

Regression models are appropriate to explore this dataset.

Made dummy columns for categorical variables using "Pandas" .

Log normalized the target column.

Made 80-20 split for train and test data.

Used "scikit-learn" to select regression models.

Trained these models with train data set.

Used predicted and test data to calculate mean squared error and R² score.

Made plots for predicted and test data for each model.

HE UNIVERSITY OF

ENNESSEE





Used different regression algorithms to generate models; Linear Regression, Lasso, and Ridge. Different models predicts different set of coefficients. Predicted data are similar in different models.

Results

Parameter	LR	Lasso	Ridge
Year	-7.99E-04	-8.67E-04	-8.64E-04
Female	5.88E+10	-3.41E-02	-6.81E-02
Male	5.88E+10	3.41E-02	9.86E-16
Age(<15)	-1.92E+11	-8.53E-02	-8.34E-02
Age(15-24)	-1.92E+11	-1.53E-02	-1.35E-02
Age(25-44)	-1.92E+11	-9.73E-03	-7.94E-03
Age(45-64)	-1.92E+11	7.97E-03	9.39E-03
Age(65-74)	-1.92E+11	3.27E-02	3.41E-01
Age(75+)	-1.92E+11	6.97E-02	7.11E-02
Hispanic	-2.25E+11	3.29E-04	0.00E+00
Non- hispanic black	-2.25E+11	2.85E-02	2.81E-02
Non- hispanic white	-2.25E+11	-2.89E-02	-2.91E-02

Accuracy tests: Mean squared error (MSE) and R² score.

MSE	R2 Score
0.0243	0.325
0.0244	0.324
0.0243	0.326

Summary

• Exploratory results show "Race" has significant impact on the suicide

• Linear regression models did not provide necessary significance for

Models have good MSE values but R² score need to be improved.