

Advanced Machine Learning

Project 1 Logistic Regression



Albert Roathel, Marcel Affi

0100101,316056

MSc Data Science

Faculty of Mathematics and Information Systems

Warsaw University of Technology

2021

Datasets used

For this project we have opted to choose the following datasets, each one consisting of a binary class variable denoted as *Target* :

Dataset	Observations	Variables	Target
Titanic	1045	sex*, pclass*, parch*, fare	Survived
Diabetics	768	ALL	Diabetic
Titanic	1045	sex, pclass, parch, fare	Survived
Titanic	1045	sex, pclass, parch, fare	Survived
Titanic	1045	sex, pclass, parch, fare	Survived

Representing each class for each dataset in a radar plot clearly shows us the weight of each variable on the target outcome. Observations with missing variables were omitted.

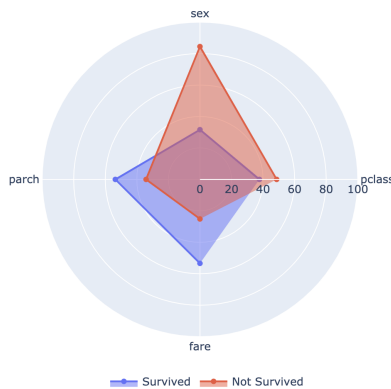


Figure 1: Titanic

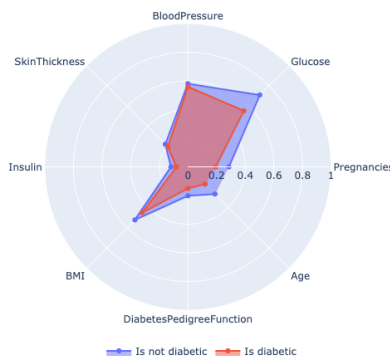


Figure 2: Diabetics

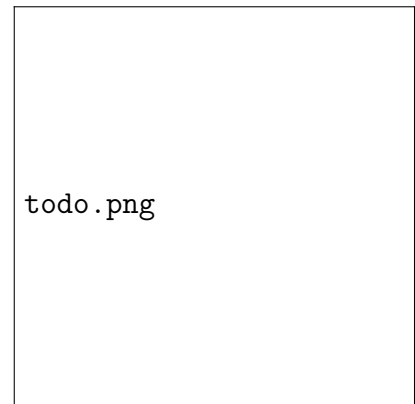


Figure 3: A really Awesome dataset

*undergone one hot encoding

Results

We compared the scores of our logistic regression model with the scores of models imported from the sickit learn (LDA, QDA, KNN, Logistic regression)

Scores							
Dataset	our scores			sickit learn scores			
	SGD	GD	IWLS	LDA	QDA	KNN	Logistic regression
Diabbetics	72.2%	70.1%	45%	75.3%	76.1%	72.2%	75.3%
Titanic	70	70	70	70	70	70	70
Titanic	70	70	70	70	70	70	70
Titanic	70	70	70	70	70	70	70
Titanic	70	70	70	70	70	70	70