MACHINE LEARNING I FATALITY PREDICTION OF ROAD ACCIDENTS

BUĞRA DUMAN

KAMIL MATUSZELAŃSKİ

SCOPE

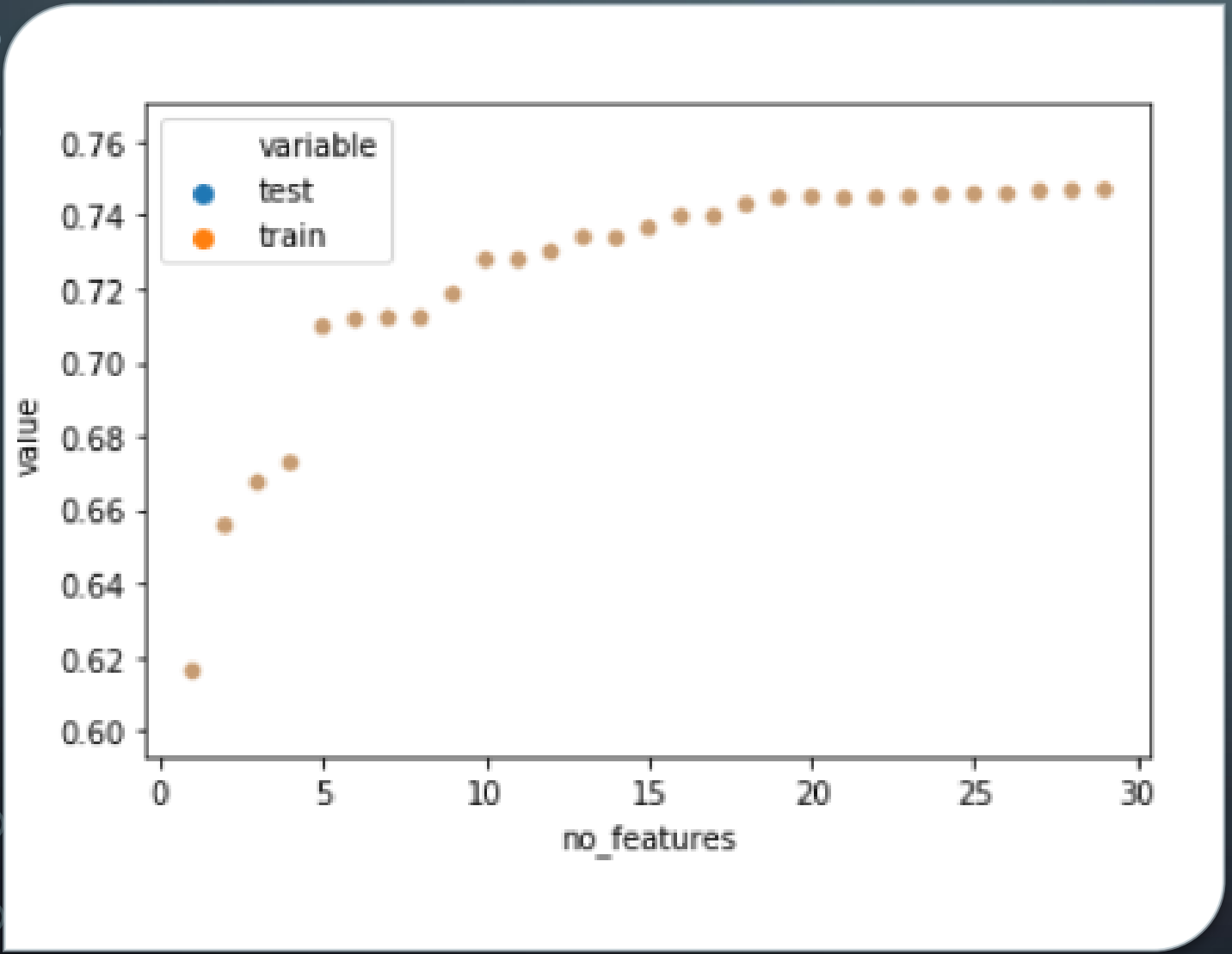
- ✓ General information
- ✓ EDA
- ✓ Modeling

GENERAL INFORMATION

- Car accidents in France from years 2005-2016, in total 839 985
- In each accident multiple people took part, totaling 1 876 005 observations
- Target variable: if injury was severe or fatal (20% of all users)
- Huge dataset, some methods infeasible

FEATURES

- ACCIDENT INFORMATION
- USER INFORMATION
- LOCATION INFORMATION

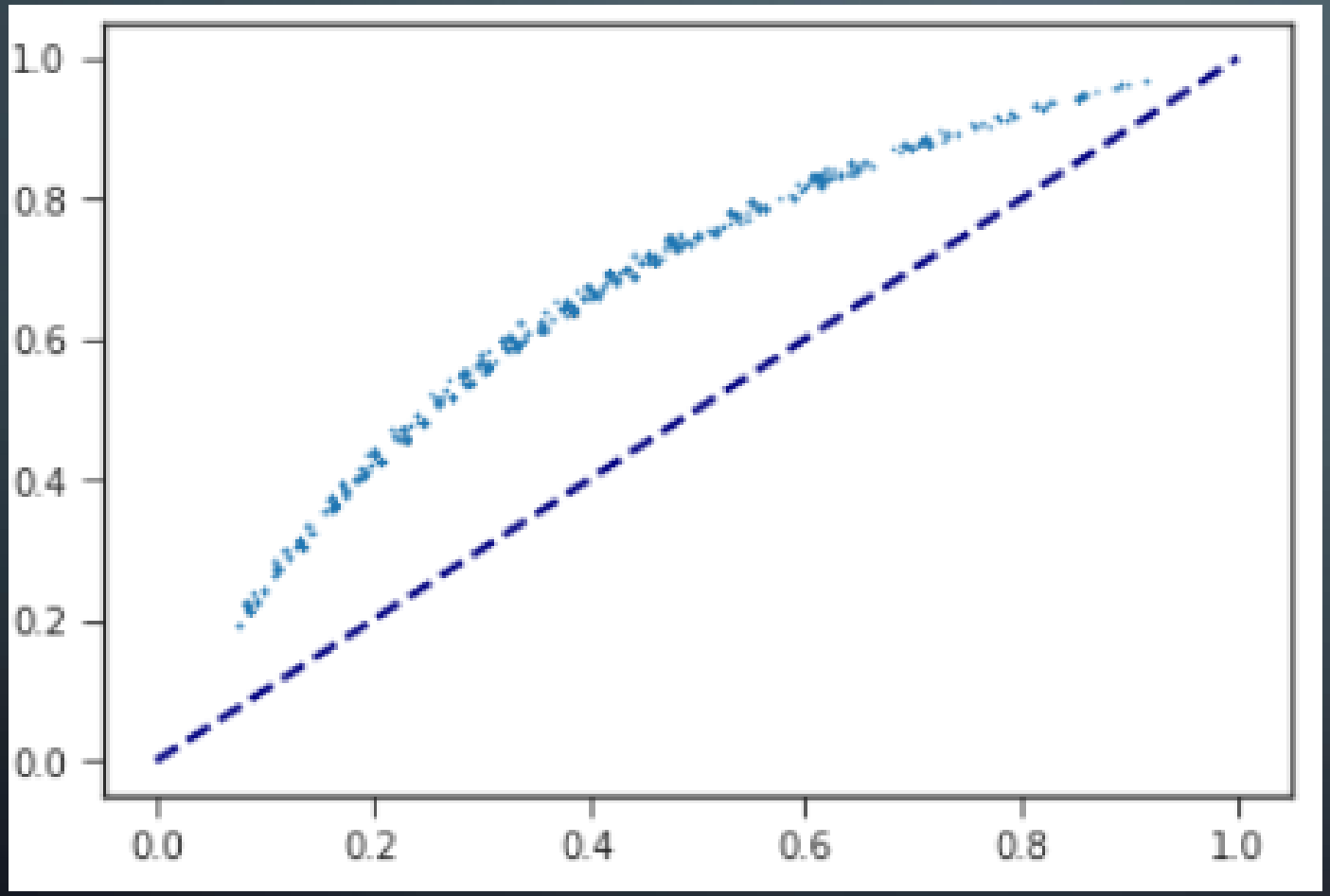


SELECTING CORRECT NUMER OF VARIABLES WITH LR

AUC TRAIN: 0.744

AUC TEST: 0.715

DOWNSAMPLING

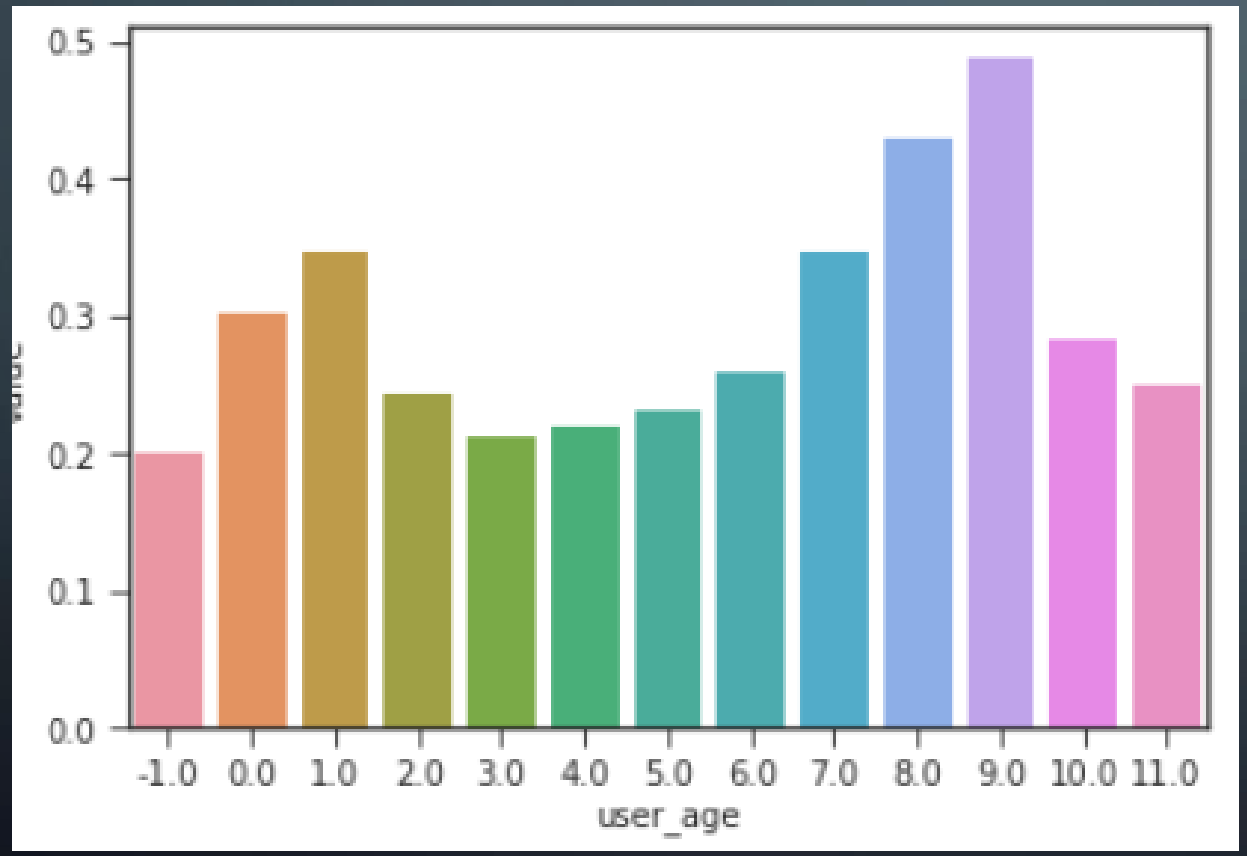


DROPPING OBSERVATIONS AT RANDOM

AUC TEST: 0.7153

MORE ADVANCED METHODS IMPOSSIBLE

FEATURE GENERATION



BINNING AGE VARIABLE

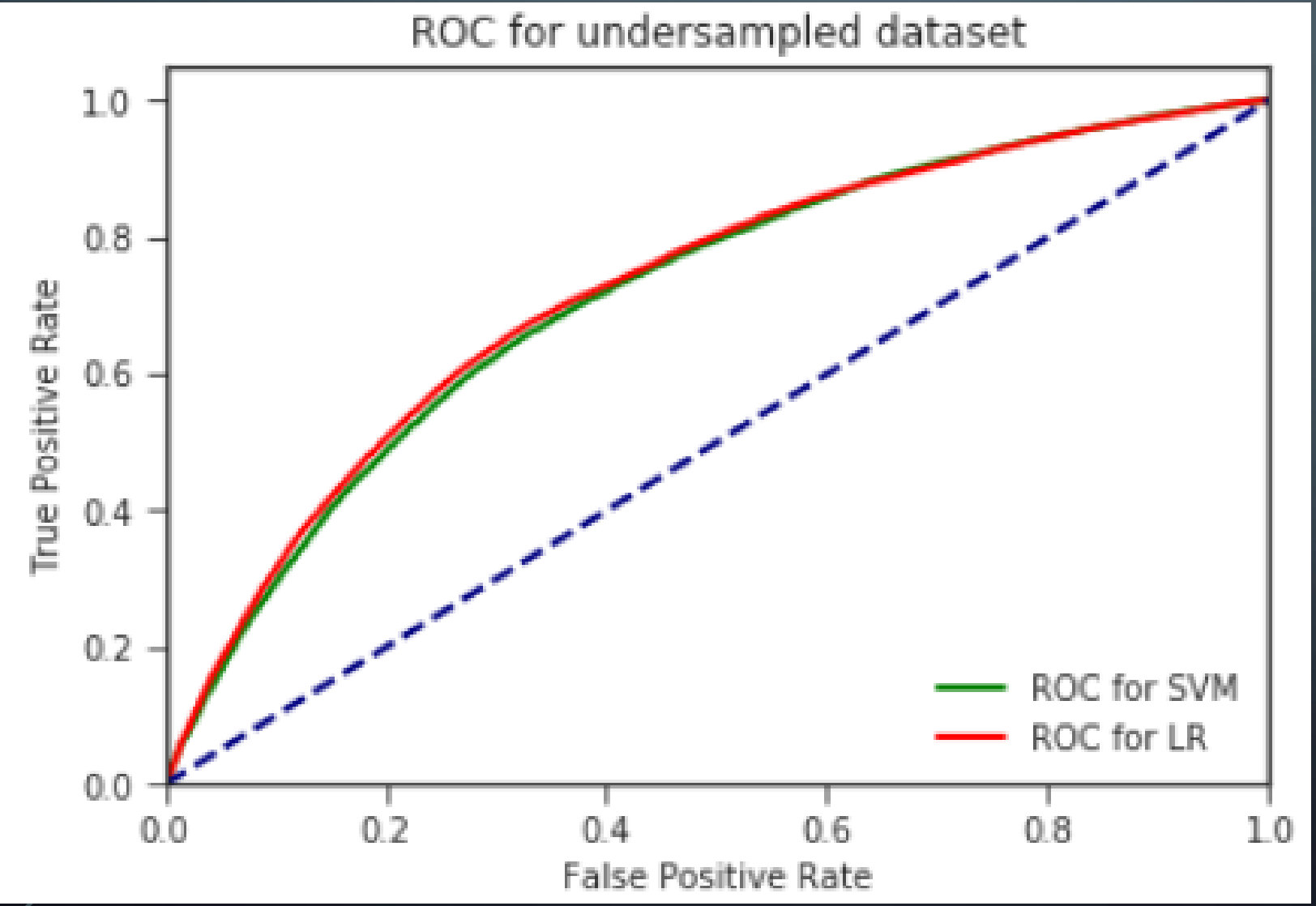
AUC TRAIN: 0.751

AUC TEST: 0.721

SVM

- Using stochastic gradient descent for paralel computation
- AUC train: 0.741
- AUC test: 0.714

FINAL MODEL



- LR with regularization
- Undersampled
- 20 features
- + 10 age dummies
- AUC train: 0.751
- AUC test: 0.721

THANK YOU