XGBoost Learning Rate Experiment

15th June 2022

1 Aim

To examine the effect of changing XGBoost learning rates on accuracy, model calibration, and variation in predicted 10K cohort thrombolysis rates between hospitals.

Notes:

A low learning rate modifies model predictions by a smaller amount with each iteration; this leads to greater regularisation of the model to avoid/reduce any potential over-fitting). XGBoost default learning rate is 0.3.

2 Methods

Adjust learning rate of XGBoost model during accuracy measurement and measurement of predicted 10K cohort thrombolysis rates.

3 Results

3.1 Accuracy

LR: 0.01, Accuracy: 0.8324

LR: 0.1, Accuracy: 0.8499

LR: 0.3, Accuracy: 0.852

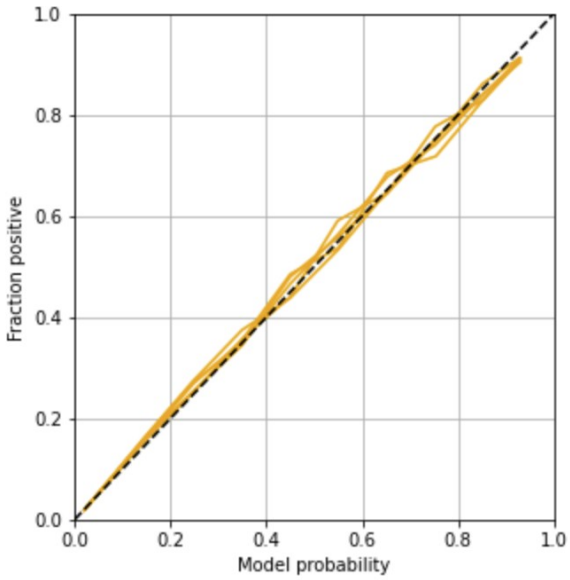
LR: 0.5, Accuracy: 0.8497

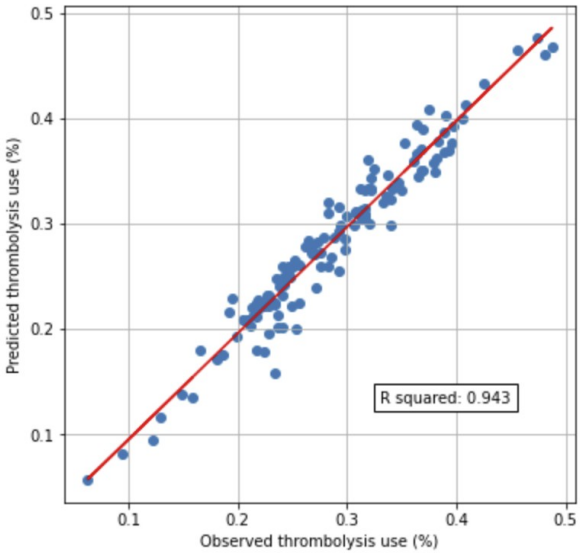
LR: 0.7, Accuracy: 0.8455

LR: 1, Accuracy: 0.8341

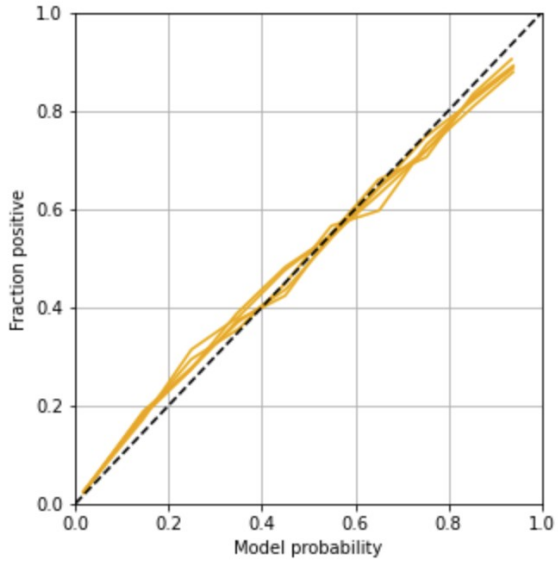
3.2 Model calibration and predicted vs. observed thrombolysis rates

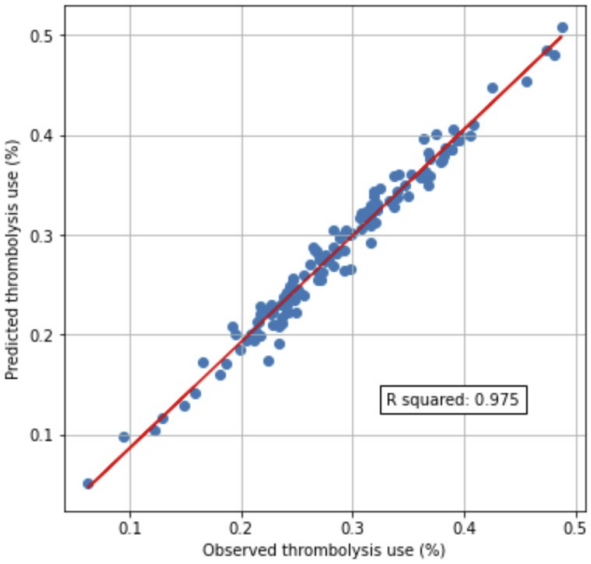
3.2.1 Learning rate 0.3



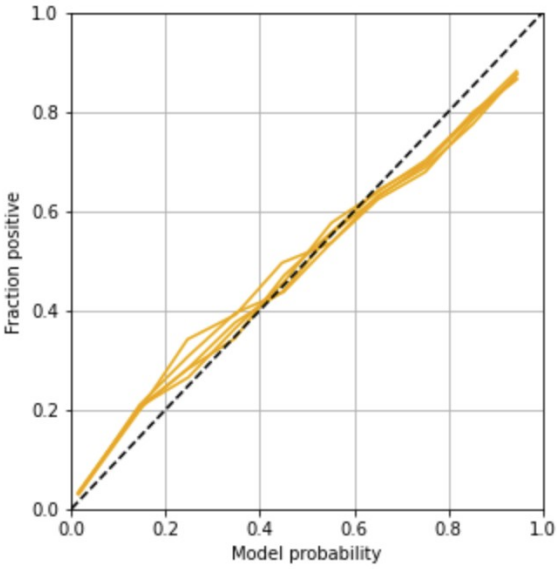


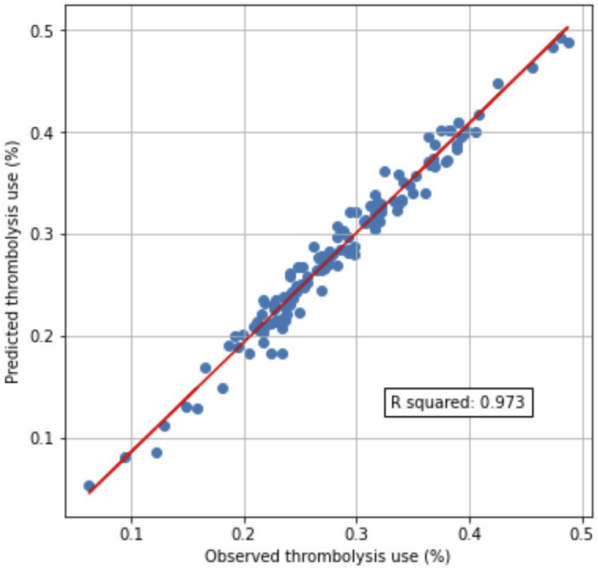
3.2.2 Learning rate 0.5



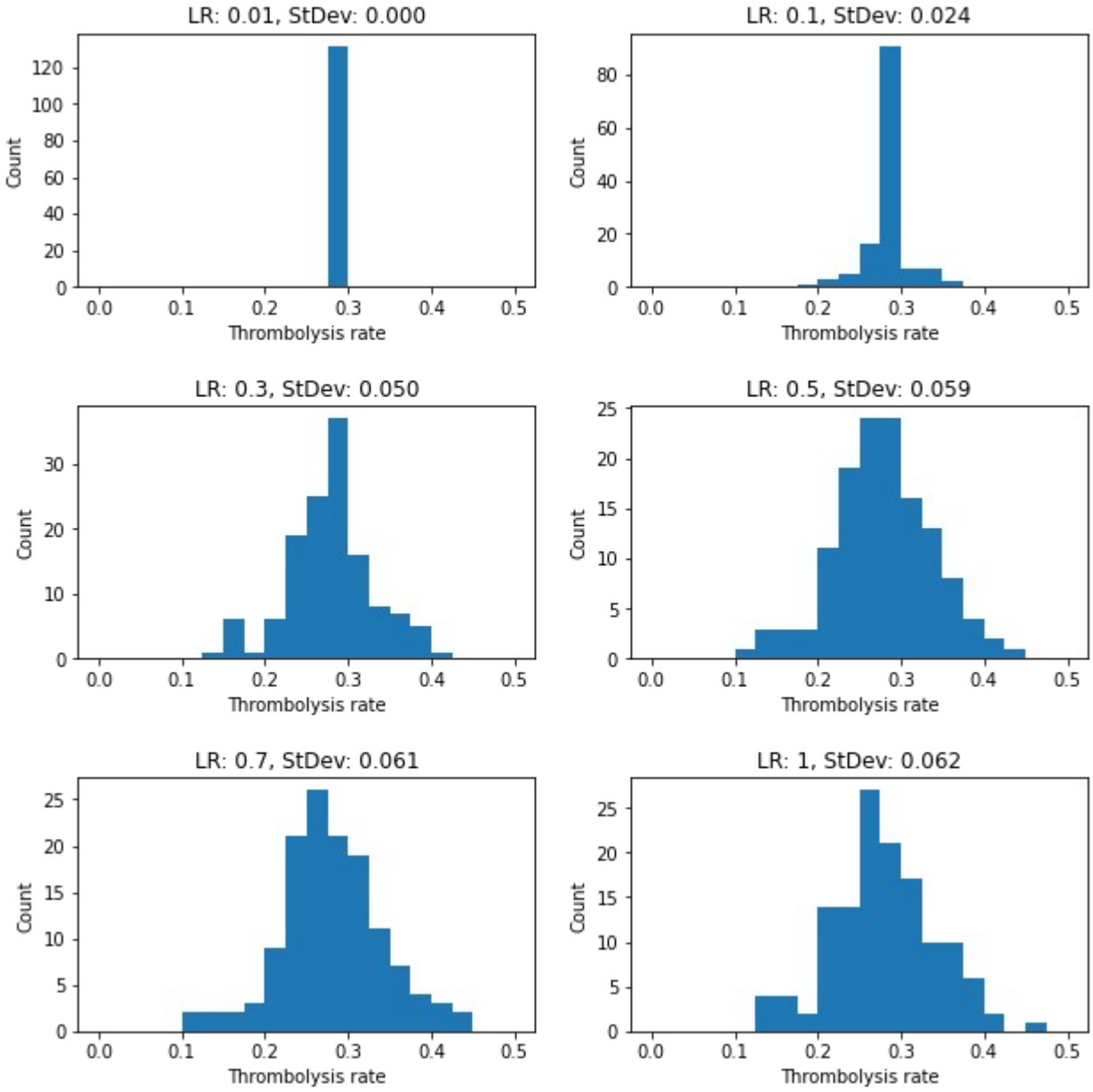


3.2.3 Learning rate 0.7





3.2.4 10K thrombolysis rate by hospital



4 Observations

- Reducing learning rate (greater regularisation) leads to loss of the effect of hospital ID when comparing 10K cohorts (effect plateaus at ~LR=0.5).
- Increasing learning rate leads to slight distortion of calibration; but this is not significant enough to affect predicted thrombolysis rates significantly.