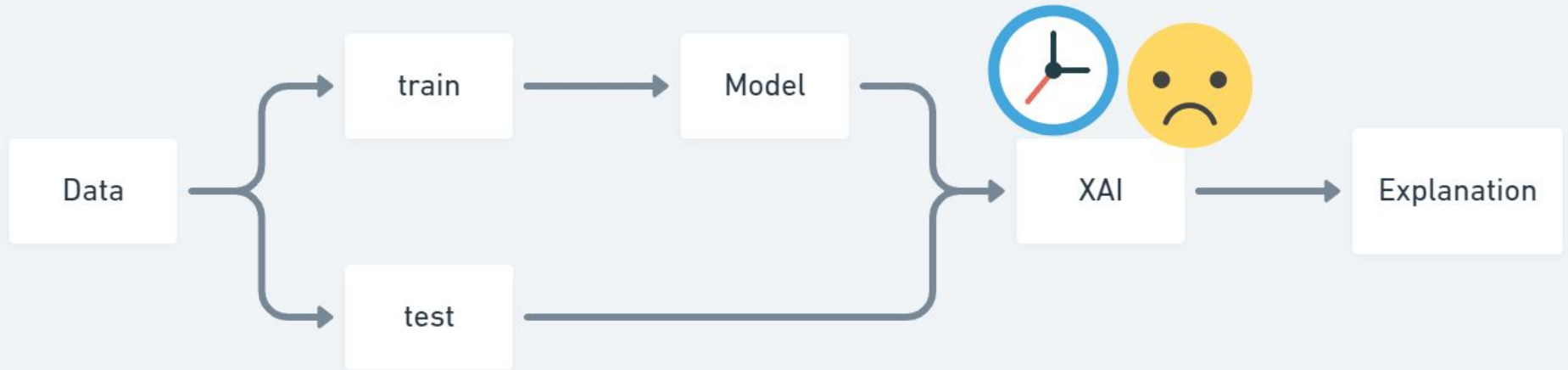


# Data compression for improved explanation estimation

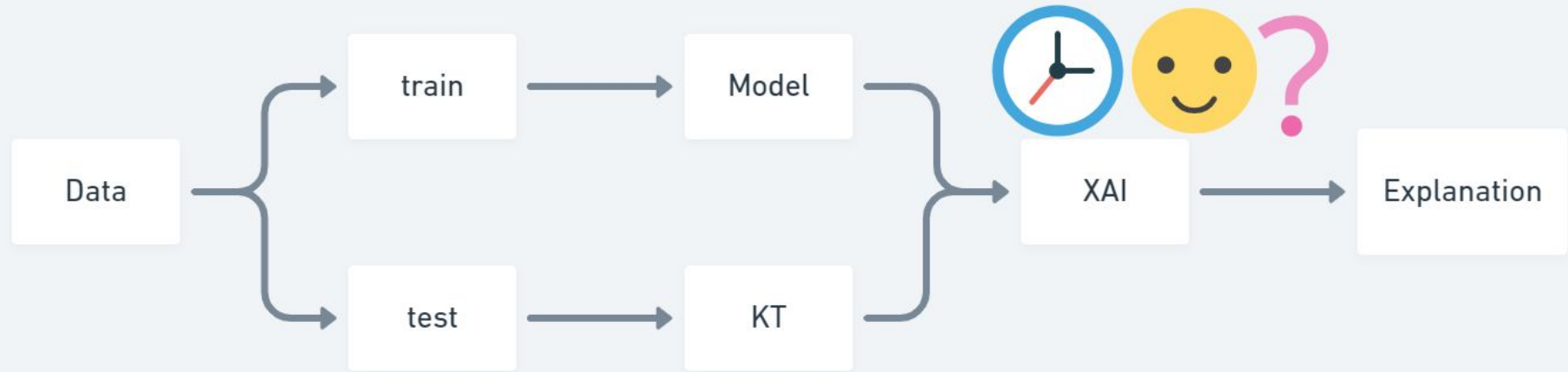
Checkpoint 3

Mateusz Biesiadowski, Paulina Kaczyńska, Ania Semik

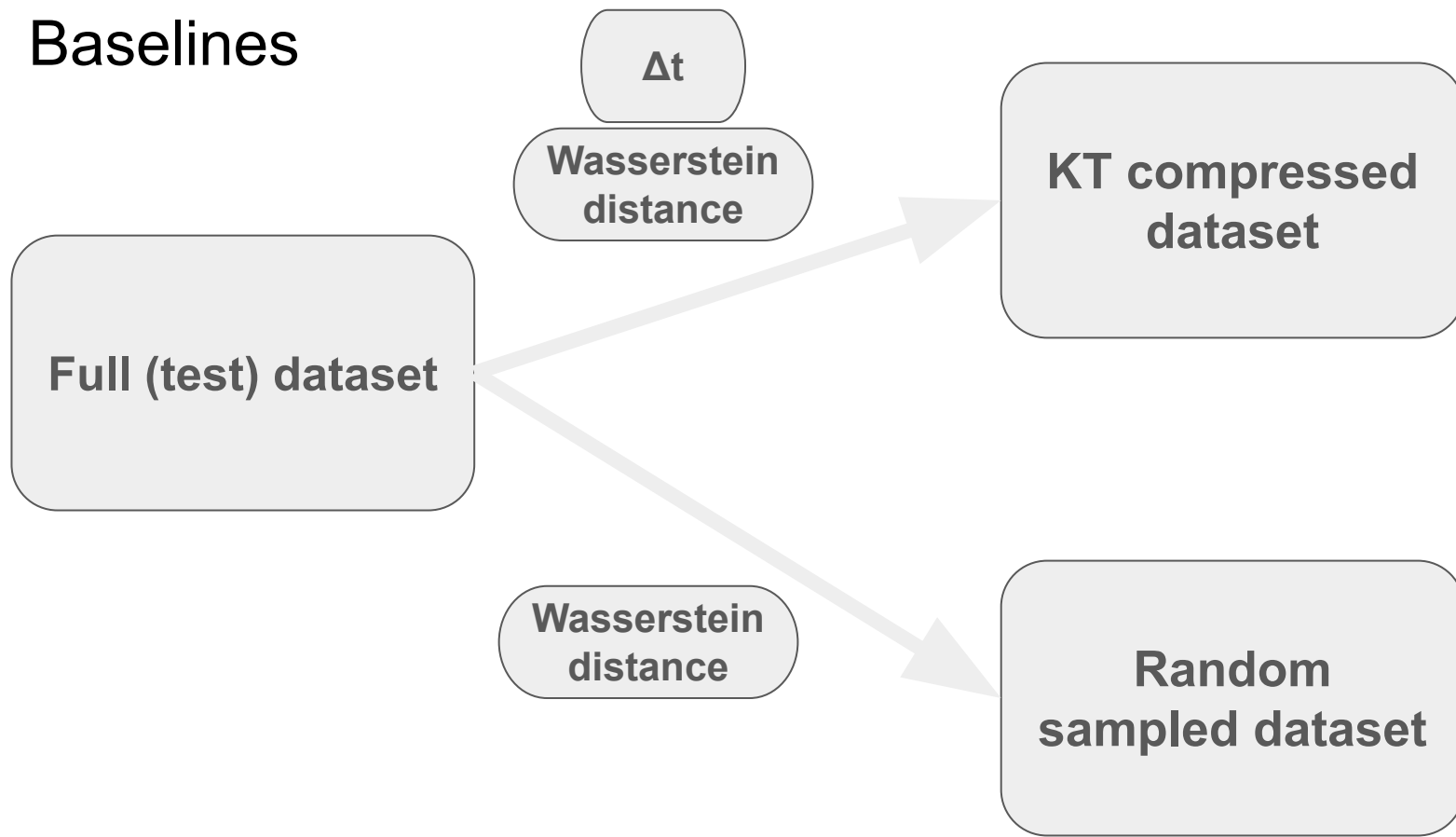
# Problem



# Solution - Kernel Thinning compression?

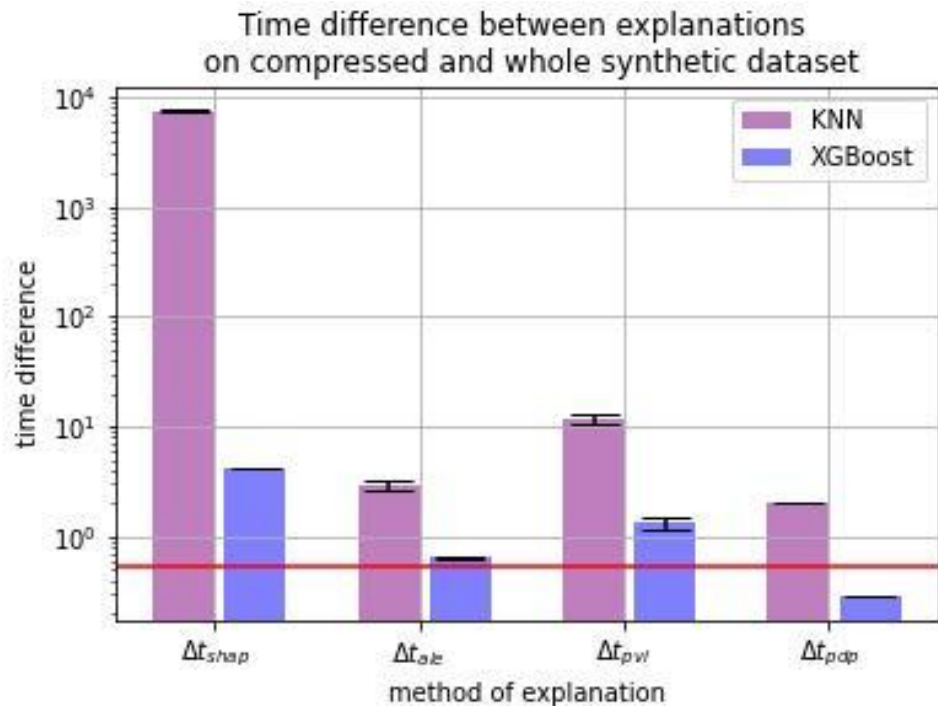


# Baselines



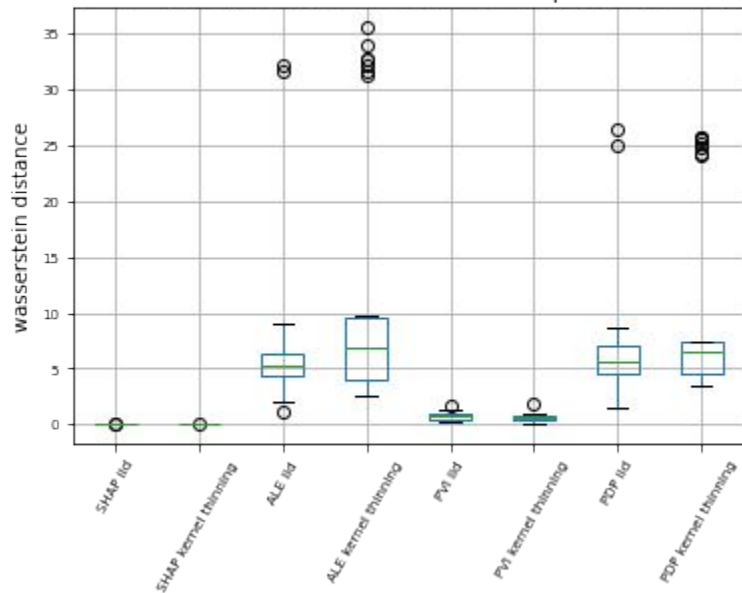
# Synthetic Dataset - Is time shorter?

Size: 5 000 (test size: 1024)  
make\_classification from sklearn

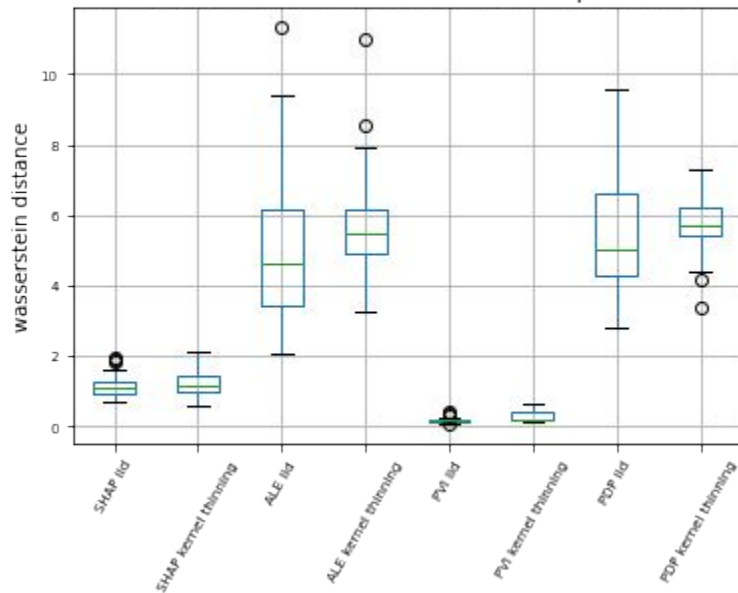


# Synthetic Dataset - Is explanation as good as on i.i.d.?

KNN - Wasserstein distance vs different explanation methods

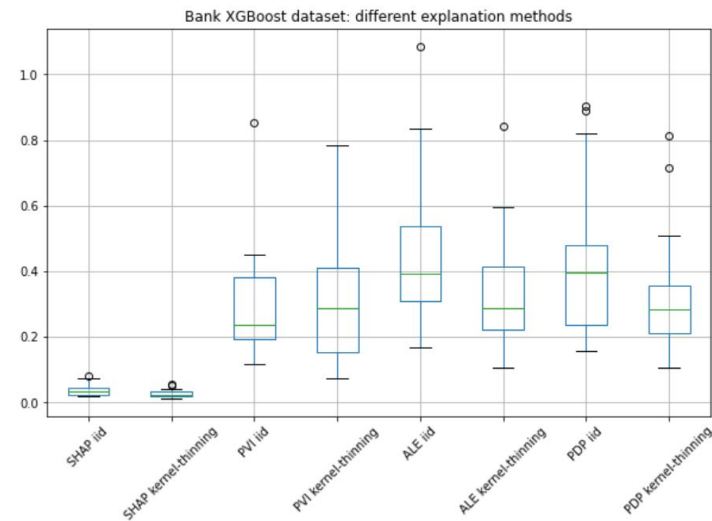
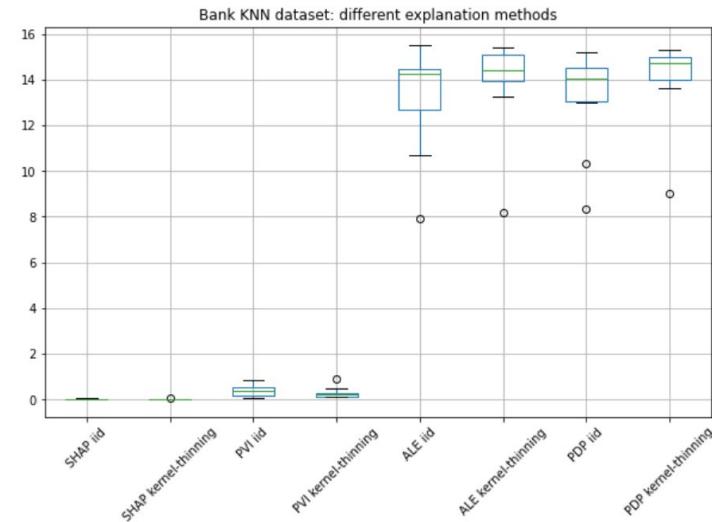
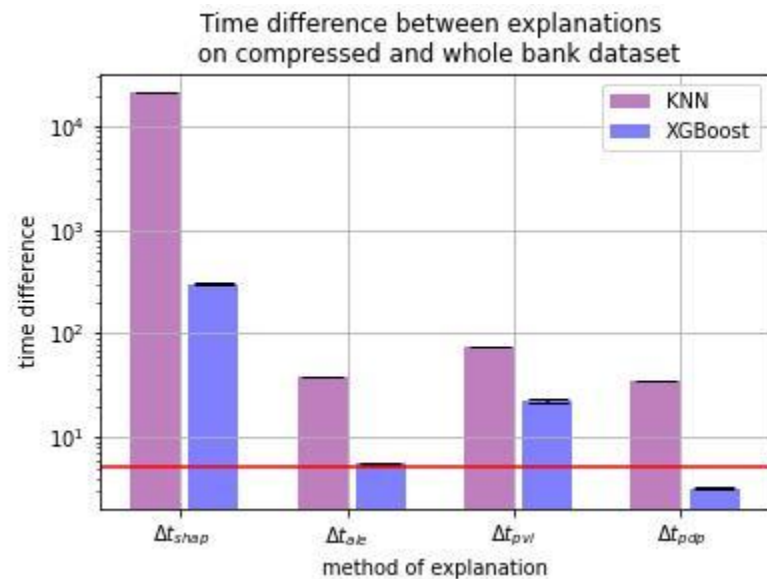


XGBoost - Wasserstein distance vs different explanation methods



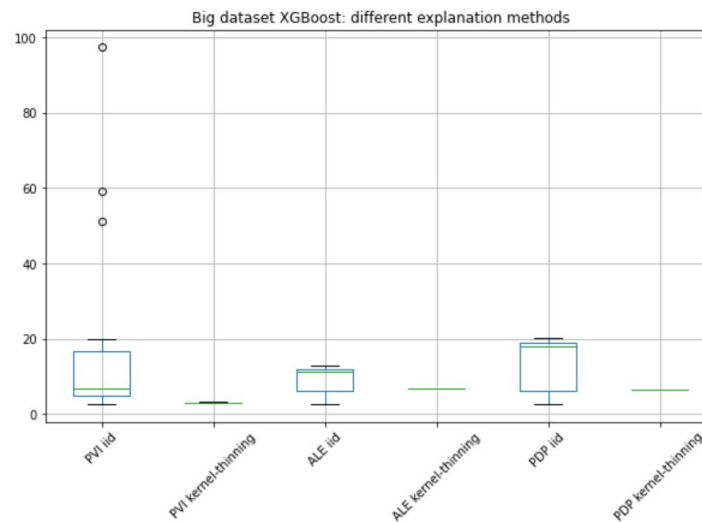
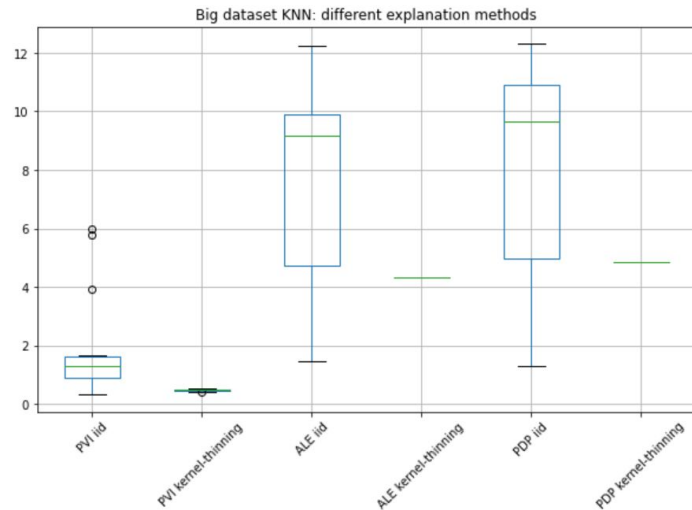
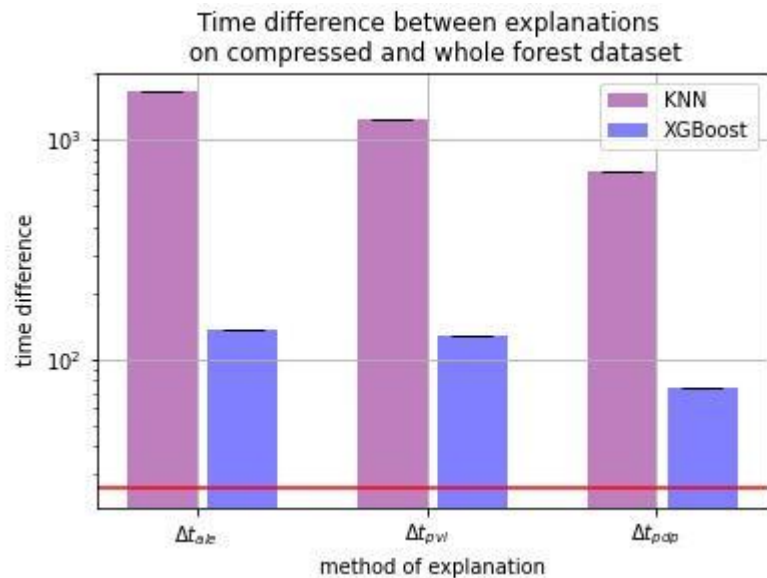
# Bank Marketing Dataset

Size:~45 000 (test size:~16000)



# Covertypes Dataset

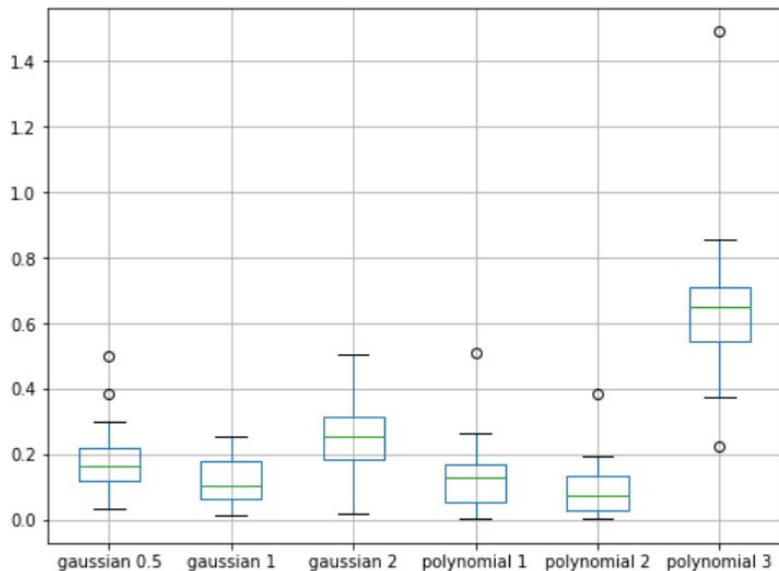
Size: 581 012



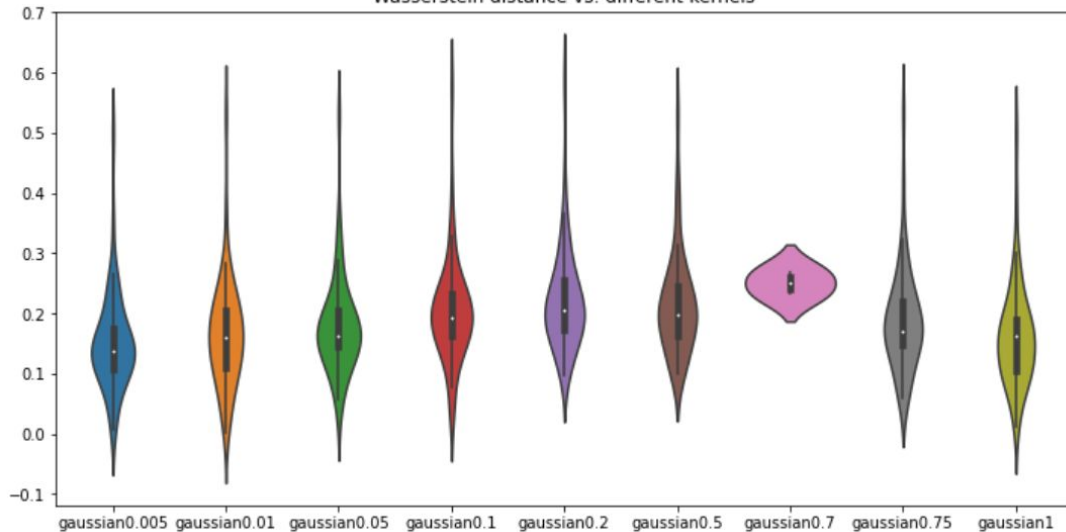


# Kernel function parameters on bank dataset

Wasserstein distance vs. different kernels



Wasserstein distance vs. different kernels



Thank you