

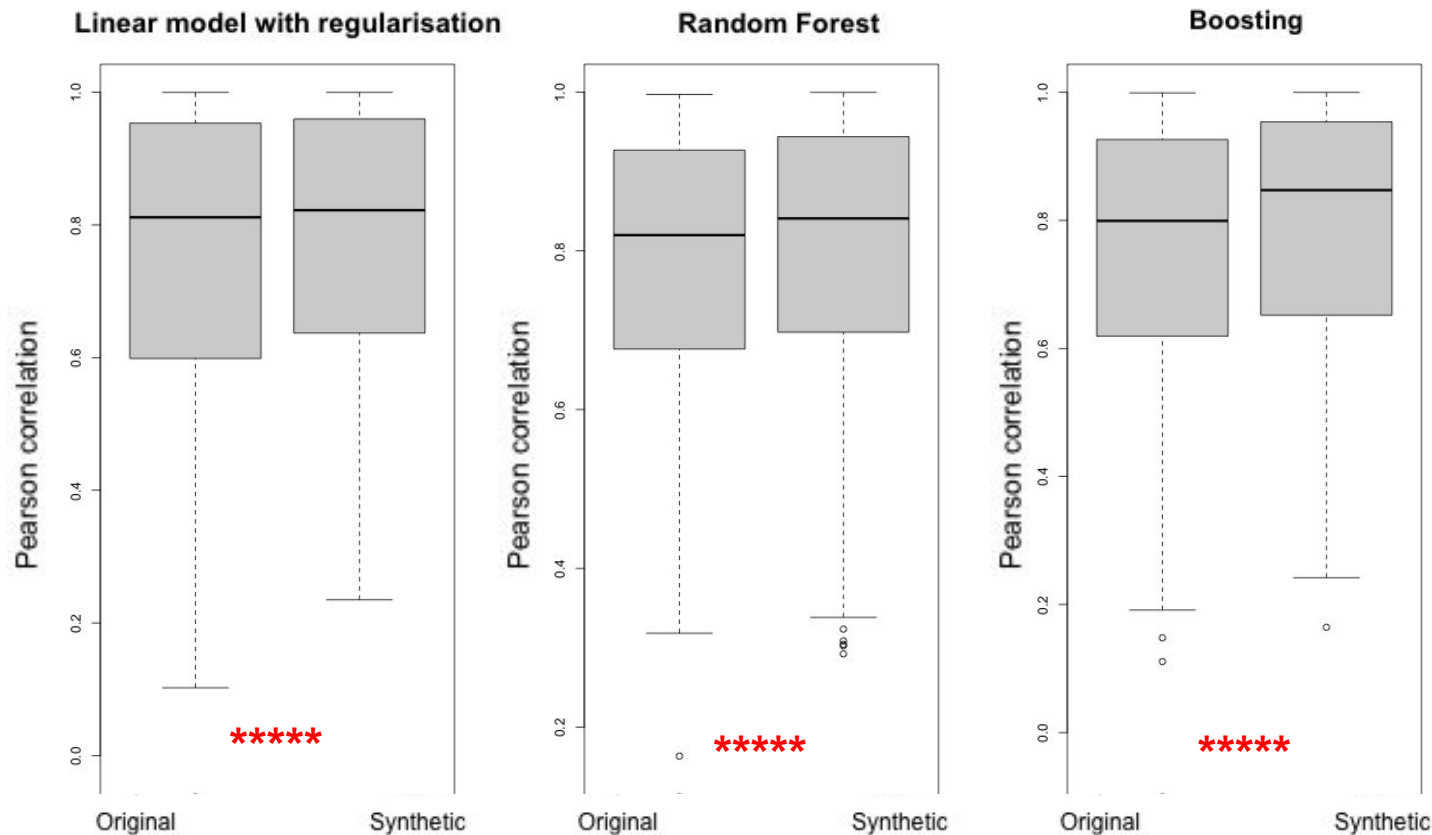
# Synthetic data outperforms original

( \*\*\*\*\* ) : statistically significant according to paired t-test

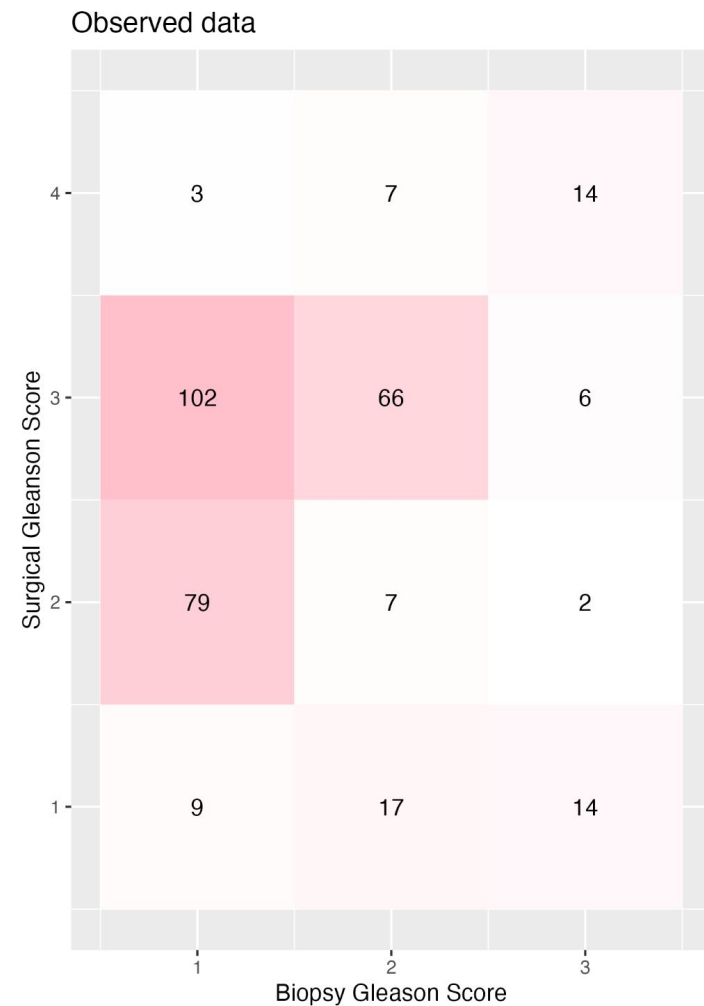
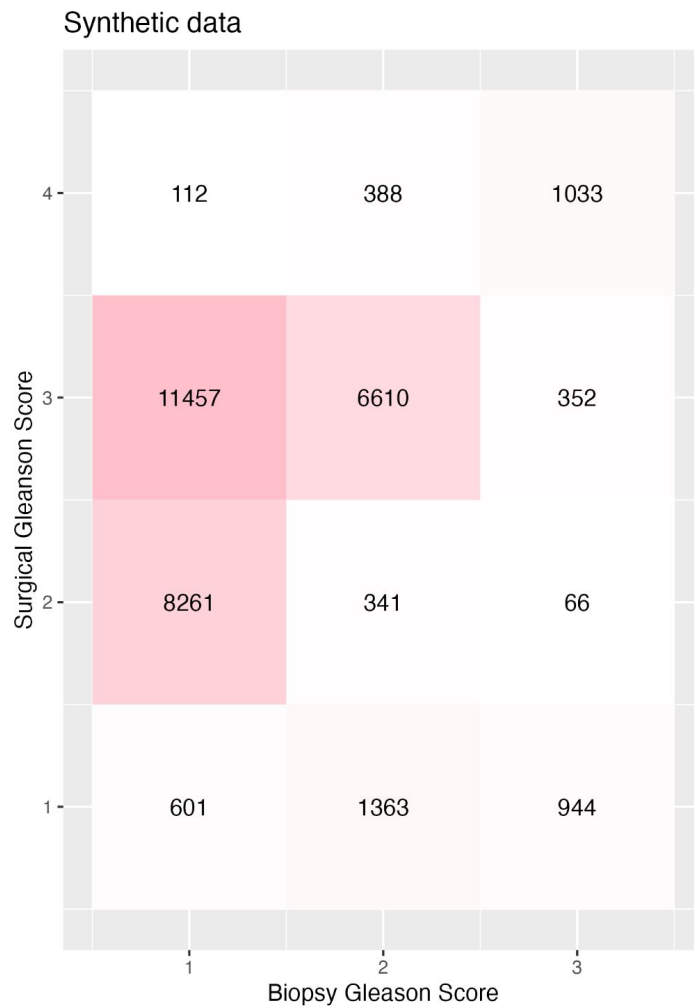
8 datasets with correlation from regression tests:

a randomly chosen feature **[set to y]**,

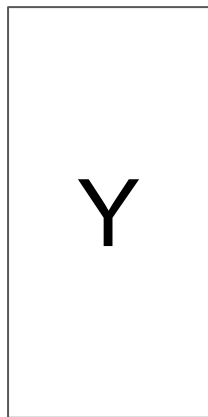
the other features **[set to x]**



Co-occurrence  
as a  
performance  
measure of  
intrinsic data  
topology  
captured and  
reflected in  
synthetic  
dataset  
in a prostate  
cancer dataset  
with gleason  
scores



Original data



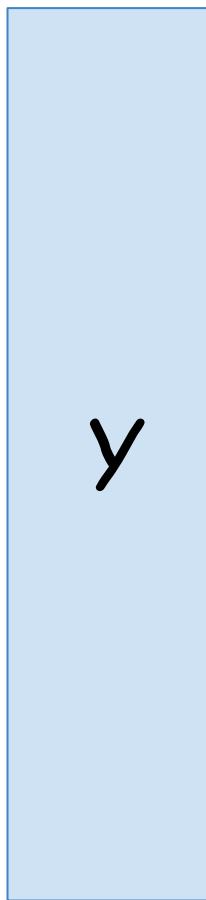
N samples

VS

N+M  
samples

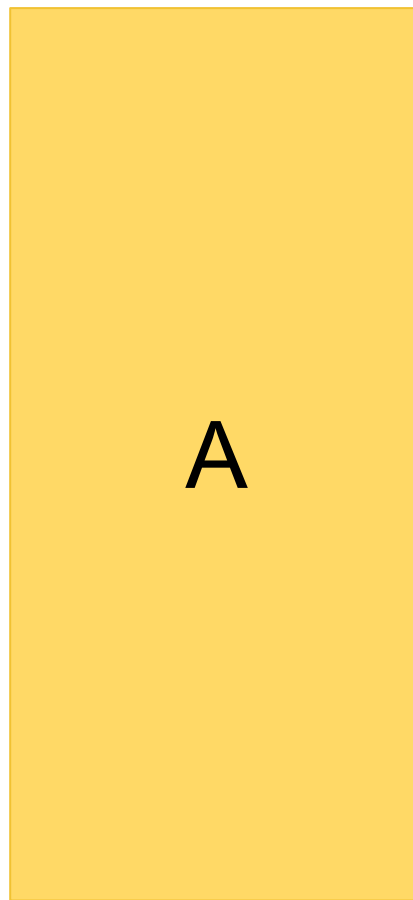


Synthetic  
**Amplified** Data



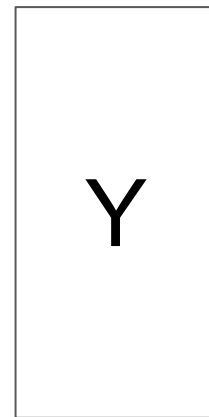
=

Synthetic  
Sample **Amplifier**



×

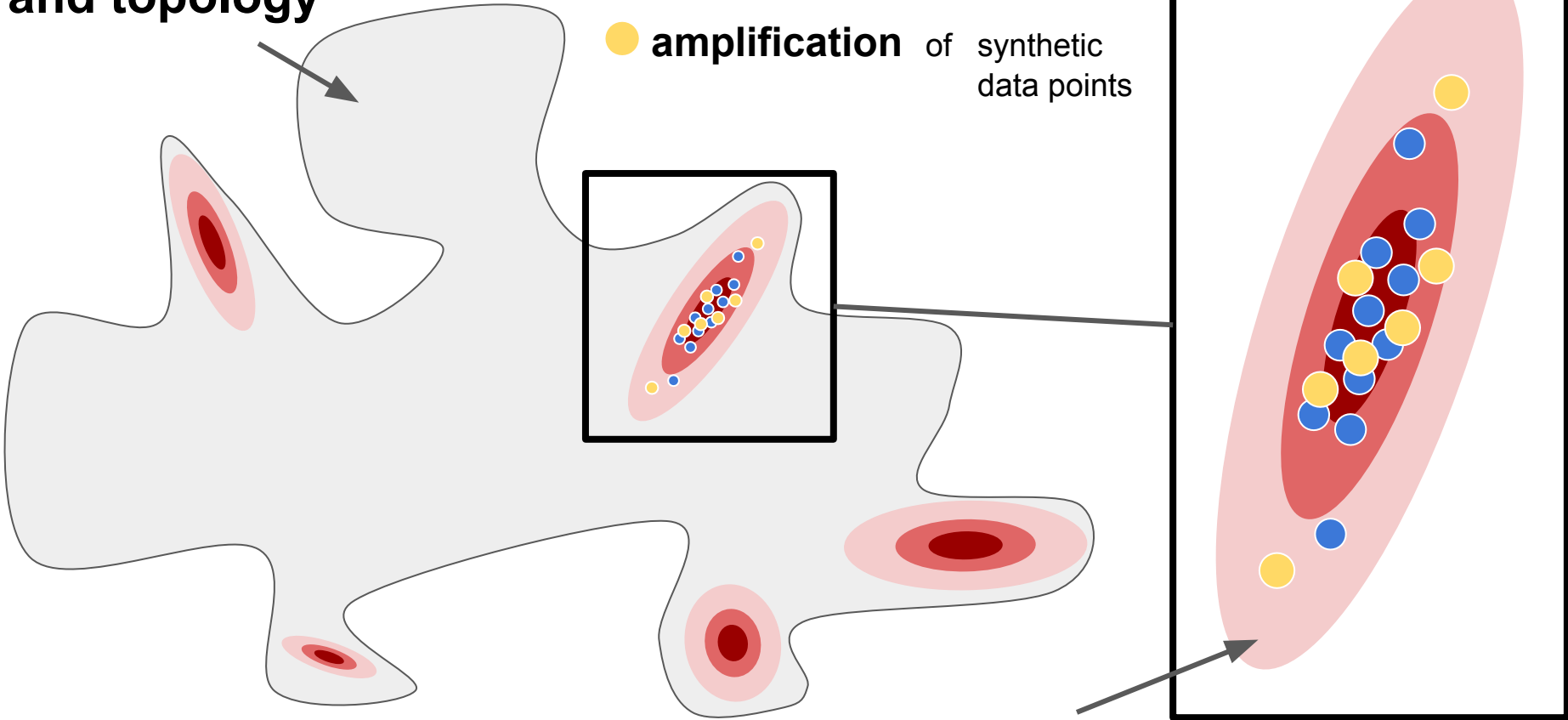
Original data



**original data shape  
and topology**

● original data points

● **amplification** of synthetic  
data points



**machine learnable manifold and topology**

- original data points
- **amplification** of synthetic data points

